

Ergebnisbericht

Ergebnisbericht Projekt *Data Go - Open Data zu GovData*

Version: 1.0

HINWEIS: Bei dieser Version des Berichtes handelt es sich um eine gekürzte Version. Der vollständige Bericht enthält Inhalte, die aus wettbewerbsrechtlichen Gründen nicht für eine uneingeschränkte Veröffentlichung geeignet sind.

Metadaten des Dokumentes

Eigenschaft	Wert
Dokumentdatum	05.04.2023
Status	veröffentlicht
Lizenz	cc-by/4.0, Namensnennungstext: "Freie und Hansestadt Hamburg, Senatskanzlei, Amt ITD"
An diesem Dokument haben mitgewirkt:	Geschäfts- und Koordinierungsstelle GovData, Senatskanzlei Hamburg: Christian Wittig, Antje Göldner, Christian Horn Capgemini Deutschland GmbH: Tim Disselhoff, Ewgenia Borodkow, Olga Knoll, Kimberly Mlitz]init[AG: Adam Kirstein, Daniel Brenn, Ludger Rinsche, Dr. Lukas Cladders SEITENBAU GmbH: Ralph Nörenberg, Sebastian Gebhart, Younes Oualad

Abbildungsverzeichnis

Abbildung 1: Entwicklung der Veröffentlichungen pro Organisationsart auf GovData 2013 - 2022	11
Abbildung 2: Anteil Datenbereitstellende je Kategorie	12
Abbildung 3: Anteil veröffentlichte Datensätze je Kategorie.....	12
Abbildung 4: Anzahl an durchschnittlichen Veröffentlichungen pro Organisationsart im ersten Jahr und ab dem zweiten Jahr der initialen Veröffentlichung.....	13
Abbildung 5: Anzahl der Teilnehmenden nach Bundesland.....	17
Abbildung 6: Verteilung der an GovData angebotenen Institutionsarten	18
Abbildung 7: Schätzung zukünftig zu veröffentlichender Metadatenätze über GovData 2023 – 2025 pro Institution	19
Abbildung 8: Verteilung der an ein Landes- oder Fachportal angebotenen Institutionsarten	20
Abbildung 9: Verteilung der Landes- und Fachportale, an die eine Anbindung angegeben wurde	20
Abbildung 10: Schätzung zukünftig zu veröffentlichender Metadatenätze über Landes- oder Fachportale 2023 – 2025.....	21
Abbildung 11: Verteilung der Institutionsarten, die anderweitig Daten veröffentlichen oder dies planen ..	22
Abbildung 12: Schätzung zukünftig anderweitig im Sinne von Open Data zu veröffentlichender Datensätze 2023 – 2025 pro Institution.....	23
Abbildung 13: Anzahl geplanter Anbindungen an GovData oder ein Landes- oder Fachportal	24
Abbildung 14: Häufigste genannte Metadatenstandards	25
Abbildung 15: Häufigste genannte Schnittstellen.....	26
Abbildung 16: Häufigste genannte Open Data Portal-Technologien	27
Abbildung 17: Schätzung neuer Metadatenätze pro Jahr – 3 Szenarien.....	35
Abbildung 18: Häufigste genannte Metadatenstandards	39
Abbildung 19: Häufigste genannte Schnittstellen.....	39
Abbildung 20: Häufigste genannte Open Data Portal-Technologien	40
Abbildung 21: Exemplarische Struktur des Aufgabenkatalogs	45
Abbildung 22: Ist-Erhebung – Standardisierter Erhebungsbogen (Beispiel).....	54
Abbildung 23: Ist-Erhebung – Personalaufwand für die Anbindung von Institutionen nach Anbindungstyp (in VZÄ).....	58
Abbildung 24: Ergänzter Ist-Erhebung – Personalaufwand für die Anbindung von Institutionen nach Anbindungstyp (in VZÄ).....	59
Abbildung 25: Ansätze der Aufgabenkritik gemäß Organisationshandbuch	60
Abbildung 26: Prognostizierter Bearbeitungsaufwand für die automatische Anbindung pro Institution nach Aufgaben (in Stunden).....	62
Abbildung 27: Prognostizierter Bearbeitungsaufwand für die automatische Anbindung, inkl. Mapping, pro Institution nach Aufgaben (in Stunden)	63
Abbildung 28: Prognostizierter Bearbeitungsaufwand für die manuelle Anbindung pro Institution nach	

Aufgaben (in Stunden).....	64
Abbildung 29: Prognostizierter Bearbeitungsaufwand pro Institution nach Anbindungstyp (in Stunden)	64
Abbildung 30: Prognostizierter jährlicher Personalbedarf (2023-2025) nach Art der Anbindung und Szenario (in VZÄ)	68
Abbildung 31: Prognostizierter jährlicher Personalbedarf (2023-2025) nach internen und externen Bedarfen (in VZÄ).....	68
Abbildung 32: Aktuelle Systemarchitektur von GovData.....	70
Abbildung 33: Ablaufplan der Schwellenwertanalyse.....	73
Abbildung 34: CKAN Backend mit 2,3 Millionen Datensätzen	74
Abbildung 35: Erzeugungsdauer für neue Test-Datensätze (je 100.000).....	75
Abbildung 36: Festplattenbedarf von CKAN.....	75
Abbildung 37: Durchschnittliche CPU Auslastung beim Ausführen der JMeter Tests	76
Abbildung 38: Dauer der Ausführung von JMeter Tests	76
Abbildung 39: Fehlerquote der JMeter Tests	77
Abbildung 40: Harvestingablauf bei GovData	79
Abbildung 41: Hochrechnung der Harvestingdauer (in Stunden).....	80
Abbildung 47: Erstellung eines Datenbereitsteller-Accounts und einer ContributorID	87
Abbildung 48: Aktueller Prozess für die Erstellung des Datenbereitsteller-Accounts	88
Abbildung 49: Prozessschritte zur automatischen Datenanlieferung	89
Abbildung 50: Prozessschritte zur manuellen Datenbereitstellung	91
Abbildung 51: Formular von GovData (mehnteilig)	94
Abbildung 52: Einbindung des Mapping-Microservice in den Prozess	99
Abbildung 53: Datenbereitsteller-Account Soll	104
Abbildung 54: Automatische Anbindung Soll.....	105
Abbildung 55: Automatische Anbindung, inkl. Mapping Soll	106
Abbildung 56: Prognostizierter zusätzlicher jährlicher Personalbedarf (2023-2025) nach Art der Anbindung und Szenario (in VZÄ).....	108
Abbildung 57: Footer mit Verlinkung zu Social-Media Kanälen	139
Abbildung 58: Interaktionselemente zum Weiterverbreiten.....	140
Abbildung 59: Allgemeines Rollen- und Rechte-Modell	141
Abbildung 60: SHACL-Validierungsprozess beim Harvesting.....	151
Abbildung 61: Prozess der Metrikchecks bei der SHACL-Validierung.....	153
Abbildung 62: Harvesting Prozess	160
Abbildung 63: Dublettenprüfung bei Datasets.....	162
Abbildung 64: Supportprozess	170
Abbildung 65: Übersicht Anwendungsarchitektur.....	172
Abbildung 66: Gesamtsystemarchitektur.....	174

Abbildung 67: Übersicht der Schnittstellen des GovData-Portals 176

Inhaltsverzeichnis

1	Einführung	1
1.1	Die Rolle von GovData im Kontext Open Data und Open Government	2
1.2	Hintergrund und Zielstellung	3
2	Teilprojekt 1: Ermittlung Mengengerüste	6
2.1	Ergebnisse des Teilprojekts 1	6
2.2	Datenanalyse	7
2.2.1	Ziel der Datenanalyse	8
2.2.2	Methode und Vorgehensweise	8
2.2.3	Ergebnisse	11
2.2.4	Fazit	13
2.3	Online-Umfrage (Quantitative Erhebung)	13
2.3.1	Ziel der Online-Umfrage	14
2.3.2	Methode und Vorgehensweise	14
2.3.3	Ablauf	15
2.3.4	Ergebnisse	16
2.3.5	Fazit	27
2.4	Experteninterviews (Qualitative Erhebung).....	27
2.4.1	Ziel der Experteninterviews	28
2.4.2	Vorgehensweise.....	28
2.4.3	Ergebnisse	29
2.4.4	Fazit	31
2.5	Berechnung Mengengerüste.....	31
2.5.1	Schätzung der zu erwartenden Metadatenätze	32
2.5.2	Schätzung der zu erwartenden anzubindenden Stellen	35
2.5.3	Darstellung genutzter Metadatenstandards, Schnittstellen und Open Data Portal- Technologien.....	38
2.6	Zusammenfassung TP1	40
3	Teilprojekt 2: Ermittlung Personalbedarf.....	42
3.1	Ausgangssituation	42
3.1.1	Aktuelle Belieferung von GovData	42
3.1.2	Die Rolle von DCAT-AP.de (2.0)	43
3.2	Ist-Erhebung	44
3.2.1	Aufgabenerhebung.....	44
3.2.2	Auswahl der Erhebungsmethodik	45
3.2.3	Durchführung der Ist-Erhebung	46
3.2.4	Plausibilisierung der Ist-Erhebung	57

3.2.5	Ergebnisse der Ist-Erhebung	57
3.3	Ist-Analyse.....	59
3.3.1	Zweckkritik	60
3.3.2	Vollzugskritik	60
3.4	Soll-Konzeption	66
3.4.1	Ergänzung der Mengengerüste	66
3.4.2	Ergebnisse der Soll-Konzeption.....	67
4	Planung Systemarchitektur.....	69
4.1	Teilprojekt 3: Planung der technischen Systemarchitektur.....	69
4.1.1	Ziel von Teilprojekt 3.....	69
4.1.2	Aktuelle Systemarchitektur	69
4.1.3	Schwellenwertanalyse des aktuellen Systems	71
4.1.4	Untersuchung neuer Anwendungskandidaten	81
4.1.5	Neue Systemarchitektur.....	84
4.2	Prozessoptimierung für zukünftige Anbindungen	84
4.2.1	Analyse der Ist-Prozesse	85
4.2.2	Optimierungs- und Automatisierungspotentiale.....	95
4.2.3	Fazit	106
5	Teilprojekt 4: Erstellung Kostenmodell	108
5.1	Zusätzlicher Mehraufwand	108
5.2	Geschätzte Mehrkosten	109
5.3	Kostenmodell.....	110
5.3.1	Kostenmodell 1: Individuelle Abrechnung pro Anbindung.....	111
5.3.2	Kostenmodell 2: Umlage auf Bundesländer	112
5.4	Fazit.....	114
6	Zusammenfassung und Ausblick	115
7	Anhang	117
7.1	Entwicklung der Veröffentlichung von Metadatenätzen auf GovData.....	117
7.2	Fragenkatalog Online-Umfrage	118
7.3	Glossar Online-Umfrage	124
7.4	Mustererfassungsbogen PBE	126
7.5	SPARQL Abfrage für alle verfügbaren Metadaten auf GovData	127
7.6	Zuordnung ContributorID	129
7.7	Kostenmodell 2: Umlage auf Bundesländer – Variationen	133
7.8	Gesamtspezifikation GovData.....	136
7.9	Vergleichsmatrix.....	185
7.10	User Storys – Mindestanforderungen an Funktionsumfang der neuen Systemarchitektur	189
7.11	Lastenheft GovData – Formular.....	193

8	Literaturverzeichnis	206
----------	-----------------------------------	------------

1 Einführung

Der Deutsche Bundestag hat zum Ende der abgelaufenen Legislaturperiode das **Gesetzespaket "Änderung des E-Government-Gesetzes und Einführung des Gesetzes für die Nutzung von Daten des öffentlichen Sektors"** (DNG) beschlossen. Bereits die Vorgängerversion des *Gesetzes zur Förderung der elektronischen Verwaltung* (E-Government-Gesetz – EGovG) aus dem Jahr 2017¹ sah gem. §12a Abs. 1 EGovG a.F. die Bereitstellung durch Behörden des Bundes von „Daten, die sie zur Erfüllung ihrer öffentlich-rechtlichen Aufgaben erhoben haben oder durch Dritte in ihrem Auftrag haben erheben lassen, zum Datenabruf über öffentlich zugängliche Netze“ vor, ohne jedoch einen „Anspruch auf Bereitstellung dieser Daten“ zu begründen. Damit erweiterte das Gesetz den Zugang zu Daten der Bundesverwaltung für alle Bürger:innen, den das *Gesetz zur Regelung des Zugangs zu Informationen des Bundes* (Informationsfreiheitsgesetz – IFG) bereits im September 2005 formuliert hatte, auf die digitale Sphäre. Durch die Verabschiedung des Gesetzespakets rund um das Datennutzungsgesetz (DNG) im Juli 2021 wurde der Wert offener Verwaltungsdaten weiter hervorgehoben und Prozessschritte zur Umsetzung konkretisiert.

Zum einen wurde der Anwendungsbereich des §12a EGovG erweitert. Umfasste er bis dahin nur die unmittelbare Bundesverwaltung, so sind seit 2021 auch Behörden der mittelbaren Bundesverwaltung explizit zur Veröffentlichung von Daten verpflichtet. Zum anderen ist die Einführung des DNG, als Ablösung des bis dahin geltenden Informationsweiterverwendungsgesetzes (IWG), in Umsetzung der Public Sector Information (PSI) Richtlinie als Reaktion auf die stetige Zunahme von Daten sowie deren Bedeutung, auch in der Verwaltung, zu verstehen. Im Gesetzentwurf² wird festgehalten, dass der Zweck des DNG darin besteht, „die Nutzung von Daten zu steigern und damit insbesondere die Entwicklung innovativer Produkte und Dienstleistungen zu fördern“. Gleichzeitig legt der Gesetzgeber gem. §1 Abs. 2 DNG weiterhin Wert darauf, dass auch durch das DNG kein Anspruch auf Bereitstellung von oder Zugang zu Daten begründet wird. Vielmehr regelt das DNG die Anforderungen an die Veröffentlichung von Daten, so diese gem. §2 Abs. 1 DNG im Rahmen einer Bereitstellungspflicht, eines Anspruchs auf Zugang oder auf sonstige Art und Weise (z.B. freiwillig) erfolgt. In diesem Kontext ist für GovData insbesondere der §7 IV DNG relevant, der vorsieht, dass Metadaten zu maschinenlesbaren Daten – soweit möglich und sinnvoll – über das Portal zur Verfügung gestellt werden sollen. Auch wenn dies keine Veröffentlichungspflicht im engeren Sinne zu begründen scheint, so hält es öffentliche Bereitsteller von Daten dazu an, diese entsprechend über GovData verfügbar zu machen.

¹ Das EGovG in seiner ursprünglichen Form wurde am 23. Juli 2013 im Bundesgesetzblatt veröffentlicht. Der §12a EGovG wurde im Rahmen der Novellierung des Gesetzes im Jahr 2017 hinzugefügt.

² Deutscher Bundestag, Drucksache 19/27442, 09. Mär. 2021, Deutscher Bundestag.

1.1 Die Rolle von GovData im Kontext Open Data und Open Government

Das nationale Datenportal GovData ist ein ebenenübergreifendes Metadatenportal, über das Verwaltungsdaten von Bund, Ländern und Kommunen einfach auffindbar und für eine Weiterverwendung nutzbar gemacht werden können. Auf GovData selbst werden keine Daten bereitgestellt, d.h. die Daten verbleiben dezentral in den Infrastrukturen der datenbereitstellenden Institutionen. Auf GovData finden sich lediglich beschreibende Metadaten zu den Daten sowie eine entsprechende Verlinkung. Zur Veröffentlichung auf GovData sind alle Verwaltungsinstitutionen des Bundes sowie der vierzehn Länder berechtigt, die sich an der Finanzierung von GovData beteiligen und der entsprechenden Verwaltungsvereinbarung beigetreten sind³.

Auf Ebene der Bundesverwaltung wurde bereits im August 2010 das Regierungsprogramm „Vernetzte und transparente Verwaltung“⁴ verabschiedet. Hierin war „Open Government“ als eines der zentralen Modernisierungsprojekte aufgenommen. Darüber hinaus wurde auf Bund-Länder-übergreifender Ebene im September 2010 die Nationale E-Government-Strategie (NEGS)⁵ beschlossen. Etwa ein Jahr später folgte durch den IT-Planungsrat das Projekt „Förderung des Open Government“⁶ als eines von sieben Steuerungsprojekten aus dem Schwerpunktprogramm zur Umsetzung der NEGS. Beide Projekte arbeiteten eng zusammen; ihr Schwerpunkt lag vorerst vor allem im Bereich der offenen Verwaltungsdaten. Hierzu wurde u.a. der Wettbewerb „Apps für Deutschland“⁷, der 2011/2012 stattfand und von den zivilgesellschaftlichen Netzwerken „Government 2.0 Netzwerk“, „Open Data Network“ und „Open Knowledge Foundation Deutschland“ organisiert wurde, durch die Schirmherrschaft des Bundesinnenministers unterstützt. Darüber hinaus veröffentlichte das Bundesministerium des Innern Mitte 2012 die von Fraunhofer FOKUS federführend erstellte Studie „Open Government Data Deutschland“⁸. Eine der zentralen Empfehlungen der Studie deckte sich dabei mit dem NEGS-Projektziel, ein ebenenübergreifendes Datenportal aufzubauen. Dabei wurde ferner auch empfohlen,

³ GovData, „Das Portal“, GovData, zuletzt abgerufen am 20.09.2022, <https://www.govdata.de/web/guest/hilfe>.

⁴ Bundesministerium des Innern, „Regierungsprogramm. Vernetzte und transparente Verwaltung“, Sep. 2010, Bundesministerium des Innern.

⁵ IT-Planungsrat, „Die Nationale E-Government-Strategie“, IT-Planungsrat, zuletzt abgerufen am 08.12.2022, <https://www.it-planungsrat.de/der-it-planungsrat/nationale-e-government-strategie>.

⁶ IT-Planungsrat, „Steuerungsprojekt „Open Government““, IT-Planungsrat Beschluss 2012/43, 25. Okt. 2012, zuletzt abgerufen am 08.12.2022, <https://www.it-planungsrat.de/beschluss/beschluss-2012-43>.

⁷ Deutscher Städte- und Gemeindebund, „Wettbewerb „Apps für Deutschland““, 18. Okt. 2011, DStGB, zuletzt abgerufen am 08.12.2022, <https://www.dstgb.de/aktuelles/archiv/archiv-2011/wettbewerb-apps-fuer-deutschland/>.

⁸ J. Klessmann, P. Denker, I. Schieferdecker, S. E. Schulz et al., „Open Government Data Deutschland. Eine Studie zu Open Government in Deutschland im Auftrag des Bundesministerium des Innern“, 2012, Fraunhoferinstitut für offene Kommunikationssysteme FOKUS.

zwar den Schwerpunkt auf offenen Daten zu legen, zusätzlich und unter Kennzeichnung aber auch eingeschränkt nutzbare Daten aufzunehmen. GovData ist das Ergebnis dieser Arbeiten.

GovData hat am 19. Februar 2013 seinen Pilotbetrieb aufgenommen. Mit Übergang der Verantwortung für das Portal GovData vom Bund an die Geschäfts- und Koordinierungsstelle der Freien und Hansestadt Hamburg ging das Portal Anfang 2015 in den Regelbetrieb und wurde seither agil weiterentwickelt. Im April 2016 wurde ein vollständiger Relaunch vorgenommen. Seitdem wächst der Metadatenbestand durch eine zunehmende Anzahl an angebotenen datenbereitstellenden Institutionen kontinuierlich an. Eine weitere umfassende Überarbeitung von GovData erfolgte Ende 2021. In diesem Rahmen wurde die Suchfunktionalität erweitert und ein „Open Data Showroom“ zur Veranschaulichung von Einsatzmöglichkeiten offener Daten eingeführt.

In Konsequenz des im Juli 2021 beschlossenen Datennutzungsgesetzes, das vorsieht, dass „Metadaten [...], soweit möglich und sinnvoll, über das nationale Metadatenportal GovData zur Verfügung zu stellen“ sind (§7 IV DNG), sind zukünftig deutlich ansteigende Zahlen sowohl bei den datenbereitstellenden Institutionen wie auch bei den Datensätzen zu erwarten. Die Geschäfts- und Koordinierungsstelle GovData hat hierauf im IT-Planungsrat (IT-PLR) hingewiesen^{9,10}. Mit Beschluss 2021/48 wurde die Geschäftsstelle GovData in Konsequenz aufgefordert, "ein verursachungsgerechtes Kostenmodell für die Länder ohne zentralen Datenübergabepunkt, auf der Basis der Daten des kommenden Jahres, zur Verfügung" zu stellen. Dieser Bericht ist als Umsetzung des Beschlusses des IT-Planungsrates 2021/48 anzusehen und erarbeitet systematisch ein Kostenmodell, das die Grundlage für die Bereitstellung von Daten im Sinne des IT-PLR-Beschlusses für jene Länder bildet, die offene Daten nicht über eine zentrale Schnittstelle bereitstellen.

1.2 Hintergrund und Zielstellung

Die auf GovData bereitgestellten Daten stehen insbesondere zur Nutzung durch Verwaltungsmitarbeitende, Bürger:innen, Unternehmen und Wissenschaftler:innen zur Verfügung. Das übergreifende Ziel ist es, die „Datenschätze“ aus der Verwaltung besser zu nutzen und weiterzuverwenden. Die Weiterverwendung beinhaltet unter anderem die Analyse der Daten, um neue Erkenntnisse zu gewinnen und Applikationen zu entwickeln, und Anwendungsfelder zu erschließen.

Die Grundlage für den Betrieb von GovData ist in einer Verwaltungsvereinbarung¹¹ festgelegt, der bislang der

⁹ IT-Planungsrat, „Änderung des E-Government-Gesetzes und Datennutzung auf GovData“, IT-Planungsrat Beschluss 2021/18, zuletzt abgerufen am 21.09.2022, <https://www.it-planungsrat.de/beschluss/beschluss-2021-18>.

¹⁰ IT-Planungsrat, „Änderung des E-Government-Gesetzes und Datennutzung auf GovData“, IT-Planungsrat Beschluss 2021/48, zuletzt abgerufen am 21.09.2022, <https://www.it-planungsrat.de/beschluss/beschluss-2021-48>.

¹¹ Redaktions-Team GovData.de, „Verwaltungsvereinbarung GovData tritt in Kraft“, GovData, zuletzt abgerufen am 21.09.2022, <https://www.govdata.de/neues/-/blogs/verwaltungsvereinbarung-govdata-tritt-in-kraft>.

Bund und folgenden Länder¹² beigetreten sind:

- Baden-Württemberg
- Bayern
- Berlin
- Brandenburg
- Bremen
- Hamburg
- Hessen
- Mecklenburg-Vorpommern
- Niedersachsen
- Nordrhein-Westfalen
- Rheinland-Pfalz
- Sachsen
- Schleswig-Holstein
- Thüringen

Das Projekt „Data Go“ wurde Anfang März 2022 initiiert, mit dem Ziel, den Beschluss des IT-Planungsrates umzusetzen und bis Dezember 2022 ein verursachergerechtes Kostenmodell zu entwickeln. An der Umsetzung des Projekts sind vier Projektpartner beteiligt, die in vier Teilprojekten unter Leitung der Geschäfts- und Koordinierungsstelle (GKSt) GovData an der Umsetzung mitgewirkt haben. Die Unterteilung erfolgte, um den Projektverlauf besser steuern zu können. Obwohl das Gesamtprojekt nach Inhalten aufgeteilt wurde, um eine möglichst autarke Durchführung der einzelnen Teilprojekte zu erreichen, bestehen dennoch Abhängigkeiten, die es bei der Umsetzung zu berücksichtigen galt.

Im Nachfolgenden wird auf den Inhalt der jeweiligen Teilprojekte eingegangen:

Teilprojekt 1 umfasst die Ermittlung der Mengengerüste für Schnittstellen und Datensätze auf Bundes- und Landesebene sowie den nachgeordneten Kommunen, wissenschaftlichen Institutionen und öffentlichen Unternehmen. Das Ziel ist die Schaffung einer belastbaren Grundlage, anhand der die zukünftigen Anforderungen an die Systemarchitektur von GovData sowie Personalbedarfe abgeleitet werden können.

Teilprojekt 2 umfasst die Pilotierung von dreizehn Institutionen in Form eines begleiteten Durchlaufens des Anbindungsprozesses, sowie eine darauf bezogene Analyse. So soll für die Zeit ab 2023 ermittelt werden, welcher durchschnittliche Aufwand für die Anbindung neuer Datenbereitstellender und den Support bei GovData entsteht und mit wie viel Personalbedarf zu rechnen ist. Weiterhin war ein Ziel dieses Teilprojektes den Prozess zur Anbindung neuer datenbereitstellender Stellen zu analysieren und die Abläufe für zukünftige Anbindungen zu verbessern.

¹² GovData, „Das Portal“, Stand: 12/2022

Das Ziel von **Teilprojekt 3** ist es, eine optimierte Systemarchitektur für GovData zu entwerfen, welche den Anforderungen der kommenden Jahre gewachsen ist. In diesem Zusammenhang wurde geprüft, ob die Komponenten der bestehenden Infrastruktur den zu erwartenden neuen Rahmenbedingungen entsprechen oder aber durch besser geeignete Module ersetzt werden müssen. Zu diesem Zweck wurden sowohl eine Schwellenwertanalyse in Bezug auf die Performance des aktuellen Systems vorgenommen als auch alternative Anwendungen untersucht, um mögliche Kandidaten für eine neue Systemarchitektur zu ermitteln.

Teilprojekt 4 umfasst die Erstellung zweier alternativer Kostenmodelle, um die erwarteten Mehrkosten durch zusätzliches Personal zu decken. Das eine Kostenmodell stellt eine verursachergerechte Option dar, während das andere die Umlage der Mehrkosten auf die der Verwaltungsvereinbarung beigetretenen Mitglieder vorsieht.

2 Ermittlung Mengengerüste

Das Ziel der Ermittlung der Mengengerüste ist die Schaffung einer belastbaren Grundlage, anhand der die zukünftigen Anforderungen an die Systemarchitektur und Personalkapazitäten von GovData abgeleitet werden können. Dies erfordert insbesondere eine Schätzung der Anzahl der neuer Datenbereitsteller sowie des Zuwachses an bereitgestellten Datensätzen über GovData. Zur Annäherung an die benötigten Zahlen wurden die auf GovData verfügbaren Datenquellen analysiert und mit den Ergebnissen einer Online-Umfrage kombiniert. Des Weiteren wurden eine Reihe von Experteninterviews durchgeführt, um die gewonnenen Erkenntnisse mit qualifizierten Aussagen und Perspektiven anzureichern.

2.1 Ergebnisse des Teilprojekts 1

Die Ergebnisse des Teilprojekts 1 umfassen die Beantwortung folgender drei Fragen:

1. Wie viele Metadatenätze werden in den nächsten drei Jahren über GovData schätzungsweise bereitgestellt?
2. Wie viele anzubindende Stellen (manuell und automatisch) sind in den nächsten drei Jahren schätzungsweise zu erwarten?
3. Welche Metadatenstandards, Schnittstellen und Open Data Portal-Technologien werden zur Bereitstellung genutzt?

Um die Validität der Schätzung von zu erwartenden zukünftigen Metadatenätzen sowie anzubindenden Stellen unter den gegebenen Umständen zu erhöhen, wurden durch Anwenden statistischer Methoden Spannen anstelle von einzelnen Werten erarbeitet und im Folgenden drei Szenarien abgeleitet: ein minimales Szenario (am unteren Ende der errechneten Spanne), ein mittleres Szenario (das den Durchschnitt darstellt), und ein maximales Szenario (am oberen Ende der errechneten Spanne, genaue Erläuterung siehe Kapitel 2.5).

Für die zu erwartenden Metadatenätzen von 2023 bis 2025 wurden die Datenanalyse und die Online-Umfrage herangezogen. Es erfolgte wegen unterschiedlichen Veröffentlichungsverhaltens eine Unterteilung in bestehende datenveröffentlichende Stellen und neue datenveröffentlichende Stellen. In der Summe ergaben sich folgende Schätzungen hinsichtlich zusätzliche Metadatenätze für 2023 – 2025:

- **Minimales Szenario:** 38.180
- **Mittleres Szenario:** 53.202
- **Maximales Szenario:** 73.041

Für die zu erwartenden anzubindenden Stellen wurden vornehmlich die Ergebnisse der Online-Umfrage genutzt. Geschätzte Anzahlen entlang der drei Szenarien wurden auf Basis des Verteilers der Online-Umfrage berechnet. Dies basiert auf der Annahme, die im Verteiler enthaltenen Adressaten (ca. 1850) würden den Großteil der Institutionen abdecken, die über notwendige Ressourcen und Möglichkeiten zur Anbindung an GovData verfügen. Die Experteninterviews dienen hierbei der Validierung dieser Annahme. Aus ihnen lässt sich

erschließen, dass im Verhältnis zur Anzahl an Kommunen und Landesbehörden (über 10.000 deutschlandweit) nur mit geringer Partizipation seitens entsprechender Stellen zu rechnen sei. Die Berechnung geschätzten an GovData anzubindenden Stellen von 2023 bis 2025 basierend auf dem Verteiler führt zu folgenden Ergebnissen:

- Im minimalen Szenario 69, d.h. pro Jahr 23 Stellen
- Im mittleren Szenario 153, d.h. pro Jahr 51 Stellen
- Im maximalen Szenario 257, d.h. pro Jahr (gerundet) 62 Stellen

Hinsichtlich der Art der Anbindung (automatisch mit und ohne Mapping, manuell) werden Erfahrungswerte der GKSt GovData zugrunde gelegt, und damit 90% automatische und 10% manuelle Anbindungen geschätzt. Automatische Anbindungen, die das Mapping eines neuen Metadatenstandards inkludieren, sind begrenzt auf 7 noch nicht gemappte, aber durch die Teilnehmenden der Online-Umfrage genutzte Metadatenstandards.

Zur Erstellung einer Übersicht genutzter Metadatenstandards, Schnittstellen und Technologien für Open Data Portale wurden die entsprechenden Daten der Online-Umfrage ausgewertet. In den Ergebnissen wurden Institutionsarten in zwei Gruppen geteilt: Die Gruppe der Verwaltungsebenen umfasst Teilnehmende aus der Bundes-, Landes- und Kommunalverwaltung, während die Gruppe Forschungseinrichtungen und Sonstige Teilnehmende aus Forschungseinrichtungen, öffentlichen Unternehmen und sonstigen Institutionen umfasst. Zwischen den Gruppen bestehen deutliche Unterschiede hinsichtlich der genutzten Metadatenstandards, Schnittstellen und Open Data Portal-Technologien.

38,7% der Teilnehmenden geben die (geplante) Nutzung von Metadatenstandards an. Beide Gruppen nennen MODS und METS. In der Gruppe Verwaltungsebenen werden zusätzlich ISO 19115 (Inspire), DCAT-AP.de, und Encoded Archival Description genannt, in der Gruppe Forschungseinrichtungen und Sonstige DataCite Metadata Schema, Dublin Core, MARC und OAI-ORE.

Die (geplanten) Nutzung von Schnittstellen beantworten 58,5% der Teilnehmenden positiv. Beide Gruppen nutzen CKAN API bzw. planen dies. Die Gruppe Verwaltungsebenen nennt zudem CSW und WMS, die Gruppe der Forschungseinrichtungen und Sonstige OAI-PMH und REST API.

Ein Anteil von 41,5% der Teilnehmenden nutzt oder plant die Nutzung von Open Data Portal-Technologien. Nur ArcGIS wird hierbei von beiden Gruppen angegeben. Die Gruppe Verwaltungsebenen nennt noch CKAN und Liferay, die Gruppe der Forschungseinrichtungen und Sonstige Zenodo, dSPACE, DKAN, Invenio und Geoserver.

Kapitel 2.5 beschreibt die Ergebnisse in einem höheren Detailgrad, sowie die jeweils zugrundeliegende Logik der Berechnung bzw. Erarbeitung der einzelnen Aspekte der Mengengerüste.

2.2 Datenanalyse

Dieser Abschnitt behandelt die Methode und Vorgehensweise der im Rahmen von Teilprojekt 1 durchgeführten Datenanalyse. Dabei werden in den folgenden Abschnitten zunächst die Ziele der Datenanalyse sowie die

Methode und Vorgehensweise erläutert und im Anschluss die Ergebnisse beschrieben.

2.2.1 Ziel der Datenanalyse

Die Datenanalyse hat zum Ziel, die momentan über GovData bereitgestellten Metadatensätze zu analysieren und die Entwicklung hinsichtlich der Veröffentlichung von Open Data über die letzten zehn Jahre nachzuvollziehen. Darauf basierend erfolgt eine Ableitung von Annahmen, die zur Schätzung der zukünftigen Anzahl an Datenbereitstellern und Metadatensätzen beitragen sollen.

2.2.2 Methode und Vorgehensweise

Ausgangspunkt für die Datenanalyse bildet die auf der GovData Website bestehende SPARQL Maske¹³, über die alle momentan verfügbaren Informationen zu den Metadatensätzen abgefragt¹⁴ und als CSV-Datei heruntergeladen wurden. Zu erwähnen ist an dieser Stelle, dass die Metadatensätze auf GovData immer den aktuell via Harvesting erfassten Stand widerspiegeln. Der Analyse liegt damit der Datenstand zum Zeitpunkt der Abfrage (9. Juni 2022) zugrunde. Sollte eine datenbereitstellende Stelle in der Zwischenzeit neue Metadatensätze veröffentlicht haben oder die Veröffentlichung zurückgezogen haben, so ergibt sich daraus ggf. eine neue Datengrundlage, die zu Abweichungen oder Differenzen bei den im Folgenden dargestellten Zahlen führen kann.

Das Ergebnis der Abfrage umfasst 59.238 Metadatensätze mit den in Tabelle 1 dargestellten Eigenschaften. Da die Rohdaten nach dem Download in keiner eine Auswertung zulassende Form vorlagen, wurden die Daten in einem ersten Schritt in eine Tabellenform überführt mit den jeweiligen Eigenschaften als Spaltennamen.

In dem resultierenden Datensatz stellt jeder Metadatensatz somit eine Zeile dar. In diesem Zuge wurde auch als erstes die ContributorID für jeden Metadatensatz anhand der in Anhang 7.6 dargestellten Zuordnung in einen Klarnamen überführt, um eine einheitliche Analyse zu ermöglichen. So wird beispielsweise aus <http://dcat-ap.de/def/contributors/destatis> das Statistische Bundesamt.

¹³ GovData, „Schnittstellen. SPARQL Assistent“, GovData, zuletzt abgerufen am 16.09.2022, <https://www.govdata.de/web/guest/sparql-assistent>.

¹⁴ Die entsprechende SPARQL Query ist in Anhang 7.5 dargestellt

Eigenschaft	Beschreibung	Spaltenname im Datensatz
http://purl.org/dc/terms/title	Titel des Datensatzes	Titel
http://dcat-ap.de/def/dcatde/contributorID	Die ContributorID des Datenbereitstellers	Bereitsteller
http://purl.org/dc/terms/issued	Das fachliche Erstellungsdatum des Datensatzes	Veröffentlichungsdatum
http://purl.org/dc/terms/modified	Das fachliche Änderungsdatum des Datensatzes	Letzte Änderung
http://xmlns.com/foaf/0.1/name	Die veröffentlichende Stelle	VeröffentlichendeStelle
http://www.w3.org/ns/dcat#theme	Die Kategorie, unter der der Datensatz auf GovData eingestellt wird (kann mehrfach vorkommen)	- ¹⁵
http://purl.org/dc/terms/license	Die Lizenz zur Distribution (kann mehrfach vorkommen, mind. einmal pro Distribution)	- ¹⁶
http://purl.org/dc/terms/format	Das Format zur Distribution (kann mehrfach vorkommen, einmal pro Distribution)	Format
DistributionCount	Anzahl der Distributionen	DistributionCount

Tabelle 1: Eigenschaften der SPARQL-Abfrage vom 09.06.2022 sowie Spaltennamen in der gesäuberten Tabelle

Die erste Spalte der Tabelle zeigt die Eigenschaften der Datensätze aus der SPARQL-Abfrage. Die zweite Spalte erläutert ebendiese Eigenschaften. Die dritte Spalte nennt die korrespondierenden Spaltennamen im gesäuberten Datensatz.

Im zweiten Schritt erfolgte die Vereinheitlichung der veröffentlichenden Stellen. Bei der initialen Analyse der Rohdaten wurde deutlich, dass z.B. dieselbe veröffentlichende Stelle teils unterschiedlich in den Daten geschrieben wurde oder aber Stellen, die derselben Behörde zuzuordnen sind (bspw. unterschiedliche Referate desselben Ressorts) einzeln aufgeführt waren. Die Vereinheitlichung musste aufgrund der vielen Einzelfälle händisch vorgenommen werden.

In diesem Zusammenhang wurde zudem eine Zuordnung der jeweiligen veröffentlichenden Stelle zu den Organisationsarten

¹⁵ Eigenschaft wurde nicht in Datensatz überführt, da nicht für die Datenanalyse relevant.

¹⁶ Eigenschaft wurde nicht in Datensatz überführt, da nicht für die Datenanalyse relevant.

- Bundesverwaltung
- Landesverwaltung
- Kommunalverwaltung
- Öffentliche Unternehmen
- Forschungseinrichtung
- Sonstiges¹⁷

vorgenommen, um eine Vergleichbarkeit mit den in der Online-Umfrage (siehe Abschnitt 2.3) erfassten Organisationsarten zu ermöglichen. Die Zuordnung erfolgte ebenfalls händisch und in Abstimmung mit den Projektbeteiligten.

Im dritten Schritt wurden doppelt vorhandene Datensätze (Dubletten) entfernt, um eine Verzerrung der Analyseergebnisse zu vermeiden. Ein Datensatz wurde als Dublette definiert, wenn derselbe Metadatensatz über verschiedene Bereitsteller veröffentlicht wurde und somit im gezogenen Datensatz mehrfach vorlag. Ein Metadatensatz kann zum Beispiel sowohl über mCLOUD als auch die GDI-DE bereitgestellt und durch GovData von beiden Plattformen via Harvesting erfasst werden. Die Identifizierung der Dubletten erfolgte in zwei Schritten:

- Zunächst wurden basierend auf den Feldern *Bereitsteller*, *Titel*, *Veröffentlichende Stelle* und *Organisationsform* eindeutige Dubletten entfernt.
- Um auch die Dubletten zu erkennen, die über verschiedene Quellen auf GovData bereitgestellt werden, erfolgte im nächsten Schritt die Identifikation ebendieser basierend auf den Feldern *Titel*, *Veröffentlichende Stelle* und *Organisationsform*.

Bei der Entfernung der Dubletten wurde nach einer definierten Reihenfolge der zu entfernenden Bereitsteller vorgegangen. So wurden zunächst die Einträge entfernt, die über die mCLOUD bereitgestellt wurden, wenn derselbe Datensatz über ein Landesportal oder GDI-DE verfügbar war. Analog wurden dann die noch verbleibenden und über die GDI-DE bereitgestellten Einträge entfernt, wenn der Datensatz ebenfalls über ein Landesportal bereitgestellt wurde. Darüber hinaus wurden keine weiteren Dubletten identifiziert. Insgesamt wurden in dem Schritt der Dublettenbereinigung ca. 12.000 Dubletten entfernt.

Im letzten Schritt erfolgte noch eine Filterung basierend auf dem Veröffentlichungsdatum: Für die Datenanalyse wurden nur solche Datensätze betrachtet, deren Veröffentlichungsdatum innerhalb der letzten 10 Jahre liegt, um historische Verzerrungen zu vermeiden. Ebenso wurden für die Ermittlung von temporalen Trends Datensätze ohne Veröffentlichungsdatum entfernt.

Der nach den beschriebenen Schritten vorliegende, gesäuberte Datensatz umfasste schlussendlich 41.962 Metadatensätzen. Diese Daten wurden zunächst rein deskriptiv analysiert, um danach eine Annäherung an die zu erwartende Anzahl zukünftig auf GovData bereitgestellter Datensätze treffen zu können.

¹⁷ Umfasst z.B. wirtschaftliche Unternehmen, die teils über Fachportale Metadatensätze auf GovData veröffentlichen und veröffentlichende Stellen, die keiner der anderen Kategorien zuordenbar waren.

2.2.3 Ergebnisse

Die Zahl der über GovData bereitgestellten Metadatensätze folgt insgesamt einem positiven Trend (siehe Abbildung 1). Im Betrachtungszeitraum der letzten 10 Jahre stellen 650 veröffentlichende Stellen insgesamt 41.962 Metadatensätze über GovData bereit¹⁸. Auffällig sind die Bundesverwaltungen, welche ab 2018 beginnen regelmäßig Daten bereitzustellen. Landesverwaltungen stellen seit 2013 jedes Jahr relativ konstant Daten bereit, wobei auch hier ab 2019 ein deutlicher Anstieg zu vermerken ist. Kommunalverwaltungen weisen erst seit 2021 ein deutliches höheres Bereitstellungsvolumen auf.

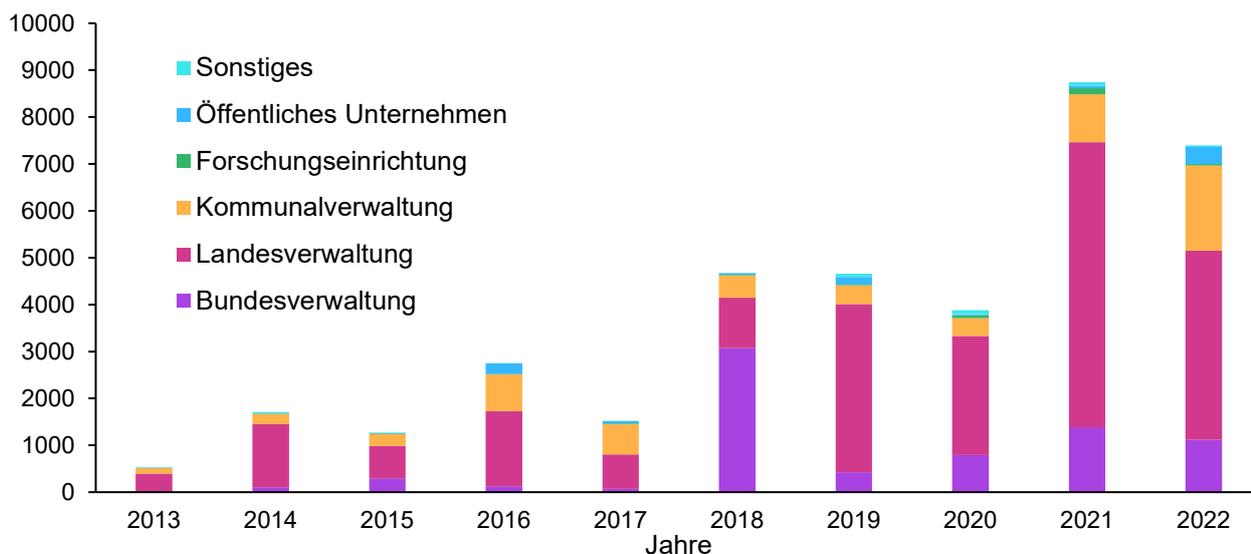


Abbildung 1: Entwicklung der Veröffentlichungen pro Organisationsart auf GovData 2013 - 2022

Seit 2013 ist insgesamt ein positiver Trend bei der Bereitstellung von Metadatensätzen zu erkennen. Für eine detaillierte Übersicht zu der Entwicklung der Veröffentlichungen für jede Organisationsart siehe Anhang 7.1.

Mit Blick auf das Veröffentlichungsverhalten veröffentlicht jede Stelle im Durchschnitt rund 65 Metadatensätze. Von den veröffentlichenden Stellen sind allerdings nur 91 (14%) als „aktiv“ einzuordnen. Als aktiv werden in diesem Kontext alle veröffentlichenden Stellen bezeichnet, die entweder in den Jahren 2020 und 2021 mindestens einen Datensatz veröffentlicht haben oder aber 2021 und zusätzlich bereits 2022 eine Veröffentlichung vorweisen können¹⁹. Mit Blick auf die Organisationsform aller veröffentlichender Institutionen lässt sich festhalten, dass mit 61,4% der Großteil der Landes- und Kommunalverwaltung zuzuordnen ist (siehe Abbildung 2).

¹⁸ Bereinigter Wert, siehe Abschnitt 1.2.2. Zu beachten ist außerdem, dass dies den Stand vom 09.06.2022 widerspiegelt und das Jahr 2022 somit noch nicht vollständig in die Analyse eingeflossen ist.

¹⁹ Da das Jahr 2022 zum Zeitpunkt der Datenanalyse noch nicht abgeschlossen war, wurden auch solche Stellen als aktiv eingeordnet, die zwar 2020 und 2021 veröffentlicht haben, jedoch nicht in 2022, da eine Veröffentlichung im Jahr 2022 nicht ausgeschlossen werden kann.

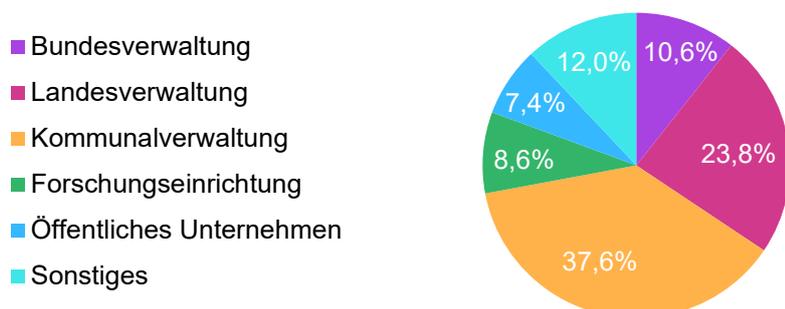


Abbildung 2: Anteil Datenbereitstellende je Kategorie

In der Verteilung aller veröffentlichender Institutionen nach Art der Organisation (n = 650) machen die Landes- und Kommunalverwaltung zusammen die Mehrheit aus.

Die Stellen der Landes- und Kommunalverwaltung sind es auch, die aktuell mit 78,7% den Großteil der Metadatenätze veröffentlichen. Forschungseinrichtungen, öffentliche Unternehmen und sonstige veröffentlichende Stellen sind hingegen lediglich für 3,7% der Metadatenätze auf GovData verantwortlich (siehe Abbildung 3).

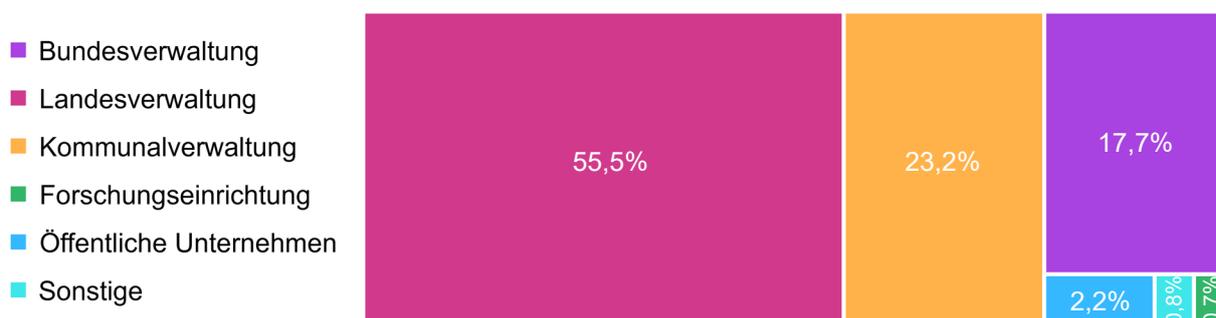


Abbildung 3: Anteil veröffentlichte Datensätze je Kategorie

Von allen 41.962 Metadatenätzen werden von Stellen der Landesverwaltung die meisten bereitgestellt, gefolgt von Stellen der Kommunalverwaltung und Stellen der Bundesverwaltung.

Für die Ermittlung von Trends ist es wichtig, das Veröffentlichungsverhalten der veröffentlichenden Stellen zu betrachten. So wurde zunächst angenommen, dass eine Institution, die sich entschließt, Daten zu veröffentlichen, im ersten Jahr eine höhere Anzahl an Datensätzen veröffentlicht als in den Folgejahren. Der Blick in die Daten zeigt jedoch das Gegenteil: Im ersten Jahr der Veröffentlichung werden durchschnittlich weniger Daten veröffentlicht als in den Folgejahren. Dies ist unabhängig von der Organisationsart zu beobachten, auch wenn dieser Faktor bei der Bundes- und Landesverwaltung deutlich höher ist (siehe Abbildung 4).

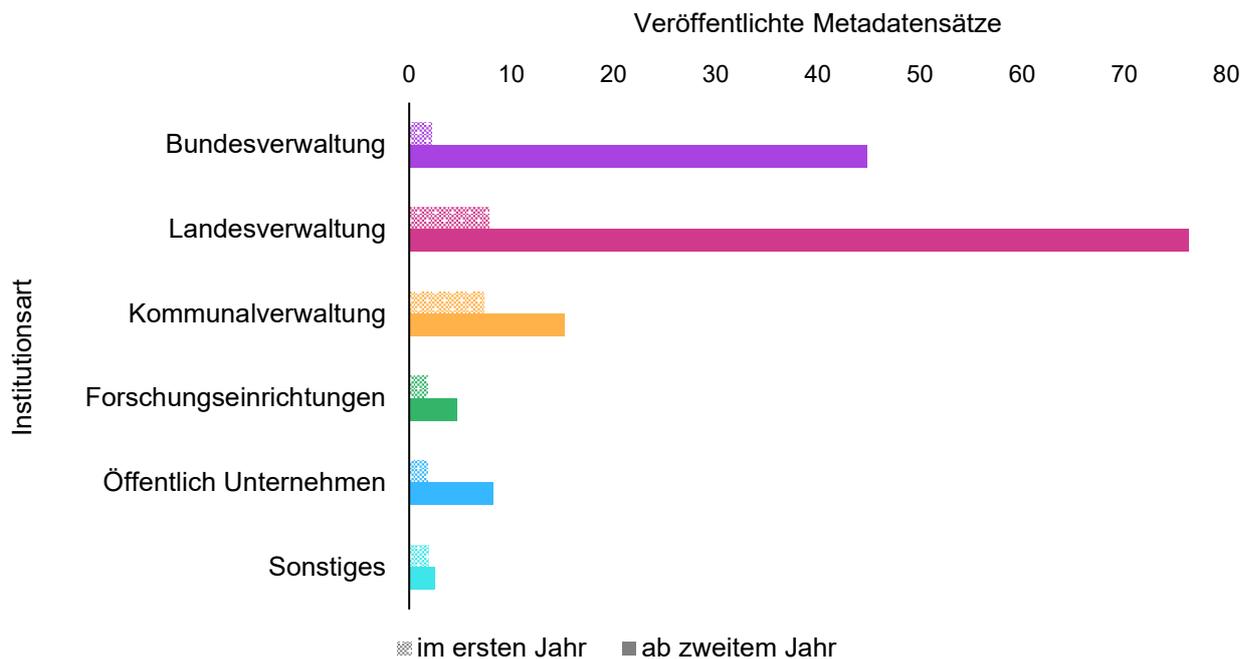


Abbildung 4: Anzahl an durchschnittlichen Veröffentlichungen pro Organisationsart im ersten Jahr und ab dem zweiten Jahr der initialen Veröffentlichung

Institutionen veröffentlichen im ersten Jahr nach ihrer Anbindung weniger Metadatensätze als in Folgejahren. Dieses Verhalten ist je nach Institutionsart unterschiedlich stark ausgeprägt.

2.2.4 Fazit

Die Datenanalyse liefert einen Einblick in den Bestand an Metadatensätzen auf GovData und die Entwicklung von Veröffentlichungen seit 2013. Teilweise konnten Annahmen bestätigt werden, z.B. der positive Trend bei den Mengen an veröffentlichten Metadatensätzen. Auch zeigt die Datenanalyse gegensätzliche Evidenz, z.B. bei neu angebotenen Datenveröffentlichenden, die, entgegen der Annahme eines stärkeren Veröffentlichungsverhaltens am Anfang, zunächst weniger Metadatensätze veröffentlichen und mit der Zeit mehr.

Die Datenanalyse ist entscheidend zur späteren Berechnung von Schätzungen der zu erwartenden Metadatensätze (siehe Kapitel 2.5.1) durch die Informationen, die sie zum Veröffentlichungsverhalten von datenbereitstellenden Institutionen liefert.

2.3 Online-Umfrage (Quantitative Erhebung)

Dieser Abschnitt behandelt die im Rahmen von Teilprojekt 1 durchgeführte Online-Umfrage. Dabei werden in den folgenden Abschnitten zunächst die Ziele der Online-Umfrage sowie die Methode und Vorgehensweise erläutert, bevor die Ergebnisse beschrieben werden.

2.3.1 Ziel der Online-Umfrage

Das Ziel der Befragung ist eine verlässliche Ermittlung der antizipierten Anforderungen an die Systemarchitektur von GovData. Sie soll ebenfalls wie die in Kapitel 2.2 beschriebene Analyse die dafür benötigte Datengrundlage schaffen, um die Schätzung der zu erwartenden Anzahl an Datenbereitstellenden und Datensätzen zu ermöglichen. Zudem soll ein Überblick der momentan genutzten Metadatenformate, Schnittstellen und Technologien für Open Data Portale zu generiert werden.

Die Zielgruppe dieser Umfrage entspricht

- der von §12 EGovG erfassten Behörden der mittelbaren und unmittelbaren Bundesverwaltung (ausgeschlossen Bundesunterbehörden),
- Behörden der Landesverwaltung (oberste und obere Landesbehörden),
- Behörden der Kommunalverwaltung (Landkreise und kreisfreie Städte),
- öffentliche Unternehmen (Stadt- und Gemeindewerke),
- und Forschungseinrichtungen (außeruniversitäre und staatliche Forschungseinrichtungen, Universitäten, (Fach-)Hochschulen, FuE-unterstützende Organisationen).

2.3.2 Methode und Vorgehensweise

Vor dem oben genannten Hintergrund und basierend auf der beschriebenen Problemstellung wurde die Online-Umfrage an insgesamt 1855 Behörden und Institutionen in Deutschland versendet. Die Adressaten wurden über eine umfassende Recherche²⁰ ermittelt. Der daraus resultierende Verteiler für die Umfrage setzt sich wie folgt zusammen:

- 218 Behörden der mittelbaren und unmittelbaren Bundesverwaltung (ausgeschlossen Bundesunterbehörden)
- 333 Behörden der Landesverwaltung (oberste und obere Landesbehörden)
- 346 Behörden der Kommunalverwaltung (Landkreise und kreisfreie Städte)
- 394 öffentliche Unternehmen (Stadt- und Gemeindewerke)
- 564 Forschungseinrichtungen (außeruniversitäre und staatliche Forschungseinrichtungen, Universitäten, (Fach-)Hochschulen, FuE-unterstützende Organisationen)

Die identifizierten Institutionen wurden per E-Mail zu der Umfrage eingeladen. Der Erhebungszeitraum erstreckte sich über mehr als fünf Wochen, vom 22.06.2022 bis 01.08.2022. Über diesen Zeitraum wurden in regelmäßigen Abständen insgesamt drei Erinnerungen an die Adressaten, ebenfalls per E-Mail, geschickt, um die Rücklaufquote zu erhöhen. Nach Abschluss des Erhebungszeitraums haben 332 Institutionen die Teilnahme an

²⁰ Die Behörden der Bundesverwaltung wurden über ein Adressverzeichnis der GKSt GovData ermittelt. Kreisfreie Städte und Landkreise als Behörden der Kommunalverwaltung wurden über das Statistische Bundesamt identifiziert. Für die Forschungseinrichtungen wurde die bei dem Hochschulkompass (hs-kompass.de) verfügbare Auflistung von Hochschulen zugrundegelegt. Alle weiteren Adressaten wurden über individuelle Online-Recherche ermittelt.

der Umfrage begonnen, somit lag die Rücklaufquote bei insgesamt 17,9%. Vollständig abgeschlossen wurde die Umfrage von 239 Teilnehmenden, was 12,9% der Institutionen des Gesamtverteilers entspricht.

Es bestehen bei der Nutzung einer Online-Befragung zumeist Herausforderungen, die sich in der persönlichen Distanz begründen. So ist die Rücklaufquote oft gering, da die Teilnehmenden selbst mehrere Hürden überwinden müssen (Aufrufen der Umfrage, aktive Entscheidung zur vollständigen Teilnahme, etc.), ohne dabei z.B. von einem Interviewer persönlich begleitet zu werden. Dies resultiert regelmäßig in einer hohen Anzahl an *Nonrespondents* bzw. *Dropouts*,²¹ was wiederum die Aussagekraft der ermittelten Mengengerüste gefährdet. Um dem entgegenzuwirken, wurden die Hürden für die Teilnehmenden so gering wie möglich gestaltet. Zum einen wurde der Umfrage ein offizielles Begleitschreiben beigelegt, welches den Hintergrund und die Zielsetzung der Umfrage erläuterte, und durch die Erinnerungen mehrfach um Teilnahme gebeten. Zum anderen war der Fragebogen auf schnelle Durchführbarkeit (maximal 10 Minuten) und ein nutzerfreundliches Design ausgelegt. Mit diesen Methoden sollte die Zahl der Rückläufer erhöht werden.

Die Umfrage wurde mit dem Online-Umfrage-Tool LimeSurvey durchgeführt und umfasste einen Fragenkatalog von insgesamt 37 Fragen (siehe Anhang 7.2). Die Anzahl an individuell zu beantwortenden Fragen für die Teilnehmenden war jedoch geringer und variierte, da es abhängig von verschiedenen Antworten mehrere mögliche Pfade gab (siehe Kapitel 2.3.3). Somit konnte ein Fokus auf zentrale Punkte zur Informationsgewinnung gelegt werden, was gleichzeitig die Nutzerfreundlichkeit erhöhte. Die Umfrage bestand sowohl aus Single Choice wie auch aus Multiple Choice Fragen. Der Großteil der Fragen waren Pflichtfragen, lediglich bei weiterführenden Fragen wie z.B. nach erlebten Hindernissen bei der Bereitstellung von Metadatensätzen war eine Antwort optional möglich.

2.3.3 Ablauf

Nachdem die Teilnehmenden über den Hintergrund und die Zielsetzung der Umfrage informiert wurden, sowie die Datenschutzerklärung akzeptiert hatten, gaben sie zu Beginn Auskunft über die Art ihrer Institution²². Bei Institutionen der Landes- und Kommunalverwaltung, und öffentlichen Unternehmen wurde zudem noch das jeweilige Bundesland abgefragt. Danach wurden ihnen Fragen zu einer bestehenden Anbindung an GovData, und wenn das nicht zutraf, an ein Landes- oder Fachportal gestellt.

Für Teilnehmende, die eine bestehende Anbindung ihrer Institution angaben, folgten Fragen nach der Art der Anbindung (manuell vs. automatisch), und nach bisher bereitgestellten sowie von 2023 bis 2025 geplanten Bereitstellungen von (Meta-)Datensätzen. Diejenigen, die eine Anbindung ihrer Institution an GovData angaben,

²¹Vgl. P. Wagner-Schelewsky und L. Hering (2019): *Online-Befragung*, in: N. Baur und J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung*, https://doi.org/10.1007/978-3-658-21308-4_49, S. 788 ff.

²² Die Antwortoptionen waren: Bundesverwaltung, Landesverwaltung, Kommunalverwaltung, Forschungseinrichtung, öffentliches Unternehmen und Sonstige

wurden zusätzlich nach möglichen Hindernissen befragt, die bei der Bereitstellung von Metadaten über GovData aufkommen. Teilnehmende, die keine Anbindung ihrer Institution angaben, wurden gefragt, ob sie anderweitig Daten veröffentlichten oder dies planten. Wo anwendbar, wurde um Angabe zu bisher veröffentlichten Datensätzen und geplanten Veröffentlichungen in den Jahren 2023 bis 2025 gebeten. Im Falle von Teilnehmenden, die gar keine bestehenden oder geplanten Datenveröffentlichungen im Sinne von Open Data ihrer Institution anführten, wurden diese gebeten, zu nennen, was ihre Institution daran hinderte.

Anschließend wurden die Teilnehmenden, die eine (geplante) anderweitige Datenveröffentlichung ihrer Institution oder eine Anbindung an ein „sonstiges“ Landes- oder Fachportal bejahten, zu bestehenden oder geplanten Metadatenstandards, Schnittstellen und Open Data Portal-Technologien befragt.

Diejenigen, die keine bestehende Anbindung ihrer Institution an GovData oder ein Landes- oder Fachportal angaben, wurden in einem separaten Pfad der Umfrage dahingehend befragt, ob eine Anbindung an GovData, und wenn nein, an ein Landes- oder Fachportal geplant sei. Bei bejahender Antwort wurde nach der Motivation zur Anbindung weitergefragt, bei verneinender nach Hindernissen.

Die Umfrage wurde nach Bestätigung der Teilnehmenden beendet und abgesendet.

2.3.4 Ergebnisse

Die Ergebnisse der Umfrage werden folgend aufgeteilt in:

- Einen allgemeinen Überblick über die Teilnehmenden (Kapitel 2.3.4.1)
- Eine Darstellung der Institutionen, die laut ihren Teilnehmenden bereits an GovData (Kapitel 2.3.4.2) oder ein Landes- oder Fachportal angebunden sind (Kapitel 2.3.4.3)
- Ein Einblick in die sonstige Landschaft mit anderweitig oder gar nicht an Open Data beteiligten Institutionen (Kapitel 2.3.4.4)
- Informationen zu geplanten Anbindungen an GovData oder ein Landes- oder Fachportal (Kapitel 2.3.4.5)
- Einen Überblick über die Nutzung von Metadatenstandards, Schnittstellen und Open Data Portal-Technologien (Kapitel 2.3.4.6)

2.3.4.1 Überblick Teilnehmende

Mit 239 abgeschlossenen Umfragen von 1855 Adressaten im Verteiler erreicht die Online-Umfrage eine Antwortquote von 12,9%. Im Verhältnis zu der Zahl an Adressaten nach Institutionsart haben Institutionen aus der Landesverwaltung die höchste Antwortquote mit 28,1% bzw. 94 Teilnehmenden, gefolgt von der Kommunalverwaltung mit 18,5% bzw. 64 Teilnehmenden und der Bundesverwaltung mit 15,1% bzw. 33 Teilnehmenden. Innerhalb der Bundesverwaltung waren Oberste Bundesbehörden und Bundesoberbehörden mit jeweils 14 und 10 Teilnehmenden am stärksten vertreten. Geringe Antwortquoten verzeichnen Forschungseinrichtungen mit einer Antwortquote von 5,1% bzw. 29 Teilnehmenden und öffentliche Unternehmen mit 2,3% bzw. 9 Teilnehmenden. Da es sich bei dieser Kategorisierung um eine subjektive Einschätzung der Teilnehmenden bei der Beantwortung der Umfrage handelt, existiert in den Ergebnissen der Umfrage im Gegensatz zum Verteiler auch eine Kategorie „Sonstige“ mit 10 Teilnehmenden.

Teilnehmende aus der Landesverwaltung, Kommunalverwaltung und öffentlichen Unternehmen (n = 167) wurden zudem auch nach Ihrem Bundesland befragt. Das am stärksten vertretene Bundesland ist Niedersachsen mit 25 Teilnehmenden, gefolgt von Bayern und Nordrhein-Westfalen mit jeweils 18. Bremen verzeichnet als einziges Bundesland keine Teilnehmenden (siehe Abbildung 5).

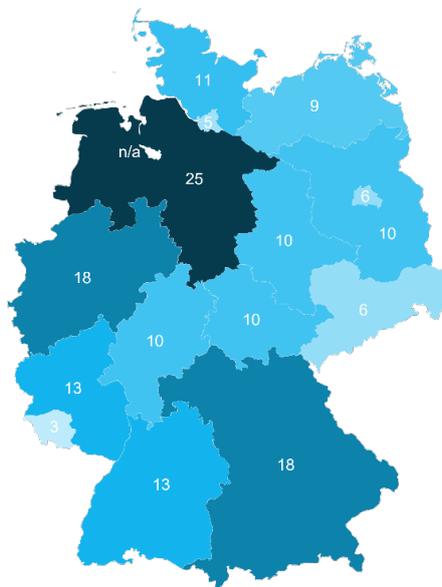


Abbildung 5: Anzahl der Teilnehmenden nach Bundesland

Teilnehmende, die ihre Institution als zur Landesverwaltung oder Kommunalverwaltung zugehörig, oder ein öffentliches Unternehmen klassifizierten (n = 167), wurden gefragt, welchem Bundesland ihre Institution zugehörig sei.

Hinsichtlich der Antworten auf Fragen nach bestehenden Anbindungen an GovData oder ein Landes- oder Fachportal und (geplanten) Datenveröffentlichungen²³ kann festgestellt werden, dass laut Aussagen der Teilnehmenden 74,9% der Institutionen in sich mit dem Thema Datenveröffentlichungen im Sinne von Open Data auseinandersetzen.

Nur insgesamt 5,9% der Teilnehmenden haben auf diese Fragen durchgängig mit „Weiß ich nicht“ geantwortet, was als positive Evidenz für den mehrwertstiftenden Informationsgehalt der Antworten und damit die Validität und Relevanz der Umfrage gewertet werden kann.

2.3.4.2 Direkt an GovData angebundene Institutionen

Von den 239 Teilnehmenden geben 28 (11,7%) eine direkte Anbindung ihrer Institution an GovData an. Stellen der Landesverwaltung nehmen dabei mit 42,9% den größten Anteil ein. Die Stellen der Bundesverwaltung nehmen folgen mit 39,3% der Institutionen, die laut den Teilnehmenden an GovData angebinden sind. 10,7% der Stellen sind aus der Kommunalverwaltung. Forschungseinrichtungen und sonstige Institutionen sind je mit

²³ Siehe Fragen 2.1 – 2.5 & 6.1 – 6.2 im Fragenkatalog der Online-Umfrage, Anhang 7.2

3,6% vertreten. Keine den öffentlichen Unternehmen zugeordneten Institutionen geben an, an GovData angebunden zu sein (siehe Abbildung 6).

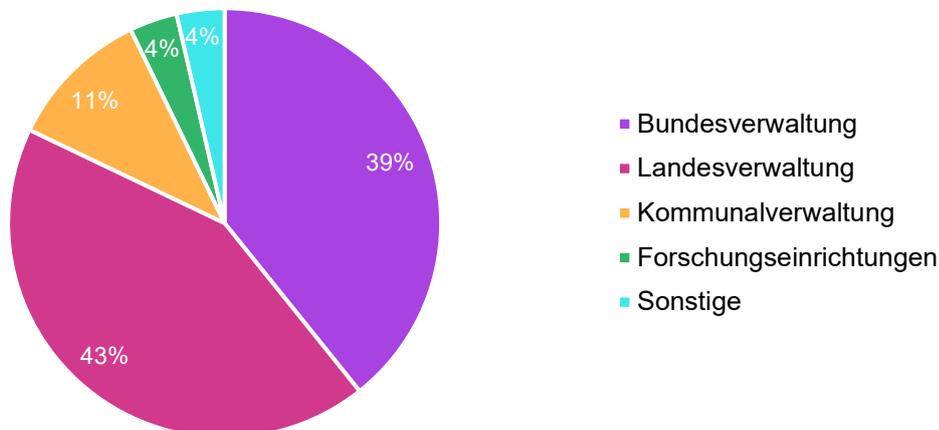


Abbildung 6: Verteilung der an GovData angebundenen Institutionsarten

Von 28 Teilnehmenden, die eine Anbindung ihrer Institution an GovData angeben, sind die meisten (82,1%) der Bundes- und Landesverwaltung zuzuordnen.

64,3% der an GovData angebundenen Institutionen geben eine automatische Anbindung an, und 32,1% eine manuelle Dateneingabe als Art der Anbindung. Auf Ebene der Institutionsarten ist zu vermerken, dass alle Teilnehmenden aus der Landes- und Kommunalverwaltung eine automatische Anbindung nennen, während die Mehrheit aus der Bundesverwaltung (63,6%) eine manuelle Dateneingabe angeben. Ein:e Teilnehmende:r gibt an, nicht zu wissen, um welche Art der Anbindung es sich handelt.

Zur zukünftigen Verbesserung der Nutzbarkeit wurde den Teilnehmenden, die eine bestehende Anbindung an GovData angaben, auch die Frage nach Hindernissen bei der Metadatenbereitstellung gestellt. Dabei geben 14,3% der Teilnehmenden an, regelmäßig auf Schwierigkeiten bei der Datenbereitstellung zu stoßen. Diese seien vor allem auf fehlende Personalkapazitäten und unklare Standardisierungsprozesse zurückzuführen.

Hinsichtlich der bisher über GovData bereitgestellten Metadatensätze geben 80,0% der Teilnehmenden an, bislang weniger als 100 Metadatensätze veröffentlicht zu haben. 14,3% geben an, die Zahl nicht zu wissen. Für die Bundes- und Landesverwaltung zeigt sich ein klares Bild hinsichtlich Trends bei der Bereitstellung von Metadatensätzen über GovData: Die Landesverwaltung gibt tendenziell höhere Werte an als die Bundesverwaltung. 66,7% der Teilnehmenden aus Stellen der Landesverwaltung nennen hohe Mengen (>250 veröffentlichte Metadatensätze), während Teilnehmende aus Stellen der Bundesverwaltung sich zu 81,8% im unteren Bereich positionieren (1-100).

Bei der Schätzung der Anzahl an Metadatensätze, die zukünftig (in den Jahren 2023-2025) über GovData für bereitgestellt werden sollen, enthalten sich je nach Jahr zwischen 25,0% und 28,6% der Teilnehmenden einer Angabe, indem sie „Weiß ich nicht“ wählen. Der Großteil schätzt vornehmlich geringe Mengen (1-50), nämlich je nach Jahr 42,9% bis 46,4% der Teilnehmenden. Zwischen 7,9% und 14,3% der Teilnehmenden geben hohe

Bereitstellungsmengen von über 500 Metadatensätzen an (siehe Abbildung 7).

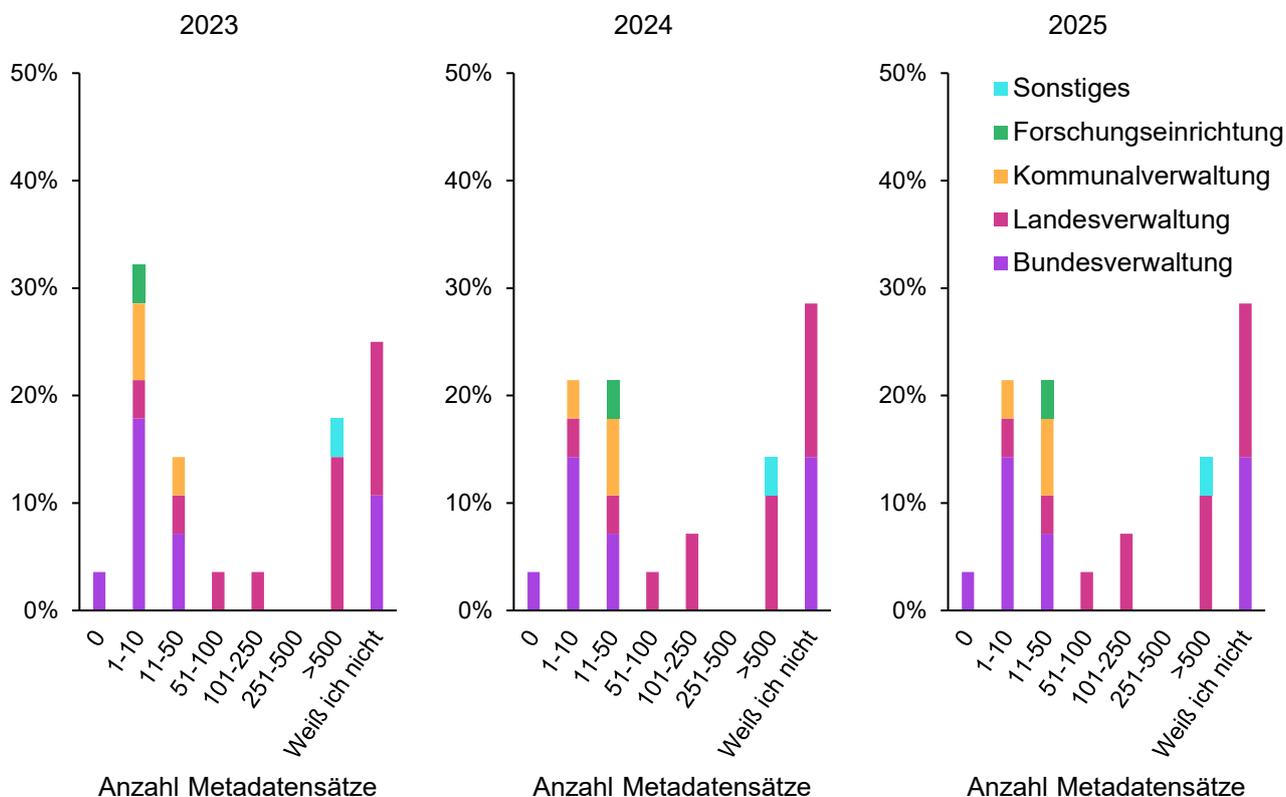


Abbildung 7: Schätzung zukünftig zu veröffentlichender Metadatensätze über GovData 2023 – 2025 pro Institution

Teilnehmende, die eine bestehende Anbindung ihrer Institution an GovData angeben ($n = 28$), schätzen die Anzahl an jährlich zu veröffentlichenden Metadatensätzen ihrer Institution in den Jahren 2023 bis 2025, getrennt nach Institutionsart.

2.3.4.3 An ein Landes- oder Fachportal angebundene Institutionen

Von 211 Institutionen, die angeben, nicht direkt an GovData angebunden zu sein oder es nicht zu wissen, beschreiben 62 (29,4%) eine bestehende Anbindung an ein Landes- oder Fachportal. Die höchste Anbindungsquote unter den Teilnehmenden weisen hierbei Stellen der Landesverwaltung mit 62,9% auf, gefolgt von Stellen der Kommunalverwaltung mit einem Anteil von 17,7% und Stellen der Bundesverwaltung mit 9,7%. Forschungseinrichtungen, öffentliche Unternehmen und sonstige Institutionen verzeichnen jeweils 3,2% (siehe Abbildung 8).

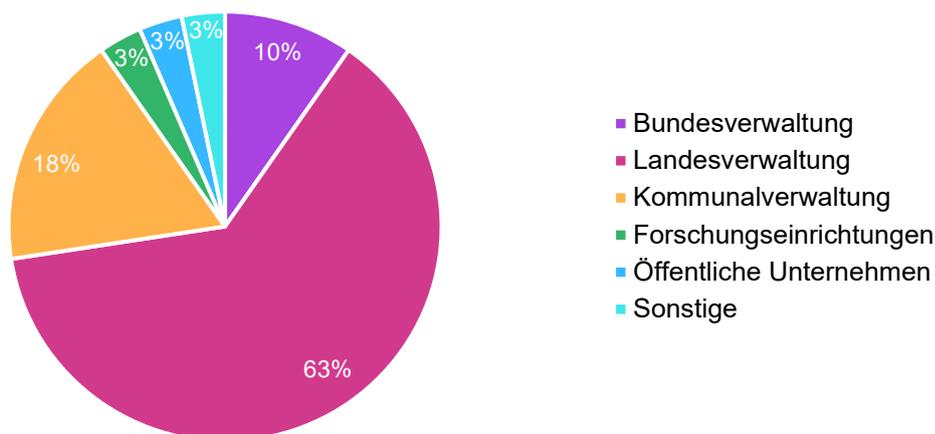


Abbildung 8: Verteilung der an ein Landes- oder Fachportal angebotenen Institutionsarten

Von 62 Teilnehmenden, die eine Anbindung ihrer Institution an ein Landes- oder Fachportal angeben, sind die meisten (62,9%) der Landesverwaltung zuzuordnen.

Die von 64,5% der Teilnehmenden genannten Landes- und Fachportale sind: Open.NRW, GDI-DE, mCLOUD und das Landesportal Brandenburg („Datenadler“) (siehe Abbildung 9)²⁴.

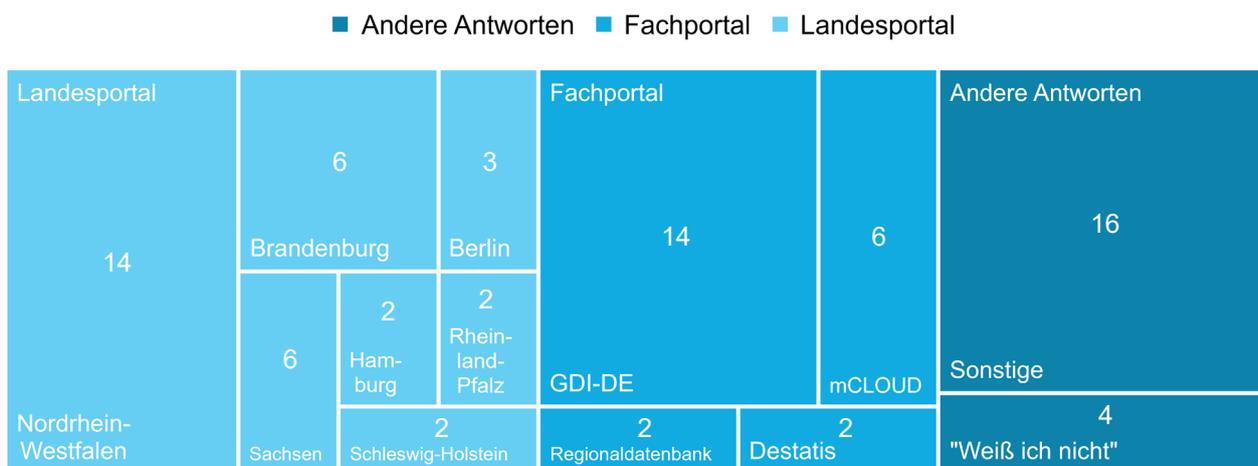


Abbildung 9: Verteilung der Landes- und Fachportale, an die eine Anbindung angegeben wurde

Die Teilnehmenden, die eine Anbindung ihrer Institution an ein Landes- oder Fachportal angeben (n = 62), nennen das oder die entsprechenden Landes- oder Fachportal(e). Mehrfachnennung möglich.

Hinsichtlich der Arten der Anbindungen nennen 45,2% eine automatisierte Anbindung gegenüber 37,1%, die eine manuelle Dateneingabe angeben. 17,7% können keine Aussage über die Art der Anbindung machen. In Bezug auf die bisher über Landes- oder Fachportale bereitgestellten Metadatensätze machen die Teilnehmenden zu 72,6% Angaben im niedrigen (1-50) bis mittleren (51-250) Bereich. Nur 9,7% der

²⁴ Frage aus der Gruppe „Anbindung an ein Open Data Portal“, siehe Anhang 7.2. Mehrfachnennungen möglich.

Teilnehmenden geben >250 Metadatenätze an. 17,7% können keine Aussage machen.

Hinsichtlich der Anzahl zukünftig über Landes- oder Fachportale bereitgestellte Metadatenätze von 2023 bis 2025 kann ein Drittel der Teilnehmenden keine Aussagen machen (je nach Jahr zwischen 30,6% und 37,1%). Unter den Teilnehmenden, die eine Einschätzung abgeben, wählen je nach Jahr zwischen 33,9% und 45,2% Werte im unteren Bereich (1-50). Zwischen 9,7% und 12,9% der Teilnehmenden geben Mengen im mittleren Bereich (51-250), und zwischen 8,1% und 9,7% Mengen im hohen Bereich (>250) (siehe Abbildung 10).

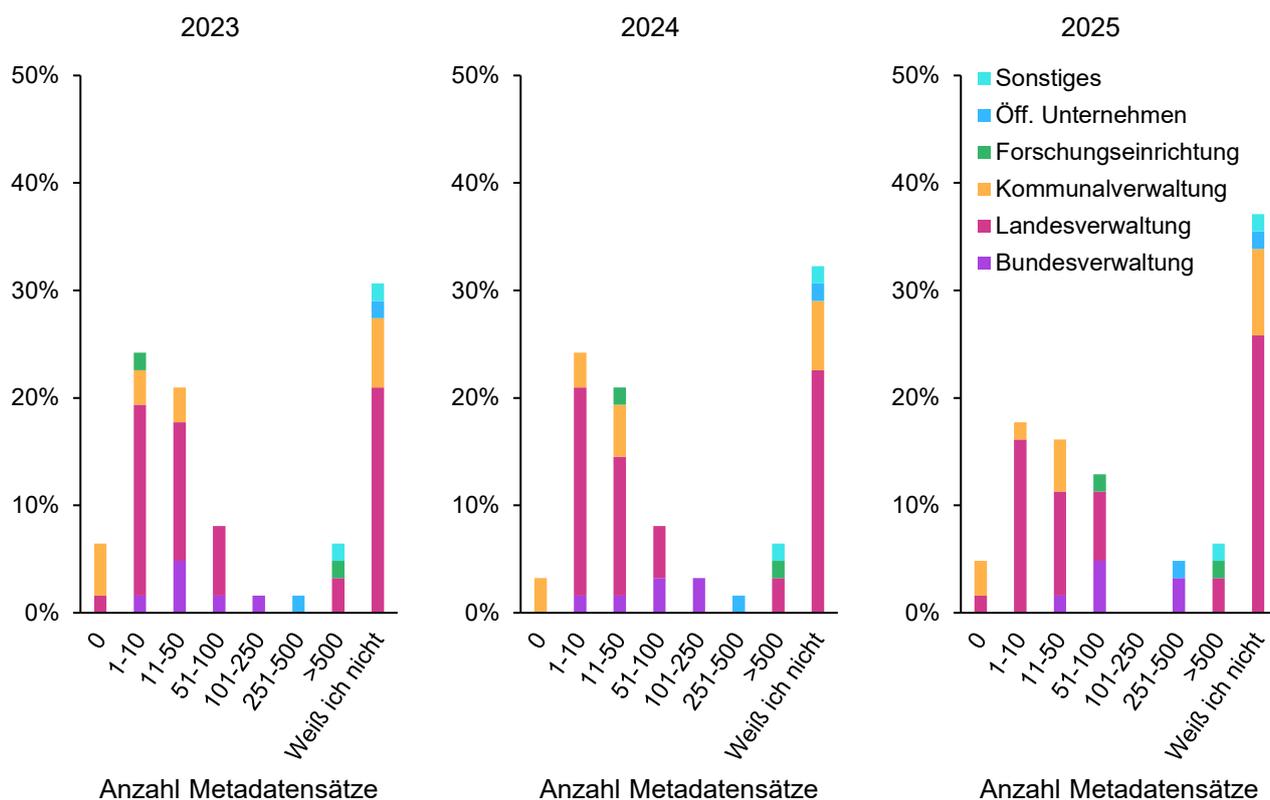


Abbildung 10: Schätzung zukünftig zu veröffentlichender Metadatenätze über Landes- oder Fachportale 2023 – 2025

Teilnehmende, die eine bestehende Anbindung ihrer Institution an ein Landes- oder Fachportal angeben (n = 62), schätzen die Anzahl an jährlich zu veröffentlichenden Metadatenätzen ihrer Institution in den Jahren 2023 bis 2025, getrennt nach Institutionsart.

2.3.4.4 Weitere Institutionen

Von den 149 Teilnehmenden, die keine Anbindung an GovData oder ein Landes- oder Fachportal angeben oder es nicht wissen, beschreiben 70 (47,0%), dass ihre Institution anderweitig Daten im Sinne von Open Data veröffentlicht. Weitere 19 (12,8%) planen eine Veröffentlichung innerhalb der nächsten drei Jahre. In dieser Gruppe haben Teilnehmende aus der Kommunalverwaltung mit 30,3% den größten Anteil, gefolgt von der

Landesverwaltung und den Forschungseinrichtungen mit jeweils 25,8%. Teilnehmende aus der Bundesverwaltung geben zu 10,1% eine anderweitige (geplante) Datenveröffentlichung an, und die aus öffentlichen Unternehmen zu 4,5%. Den kleinsten Anteil nehmen mit 3,4% sonstige Institutionen ein (siehe Abbildung 11).

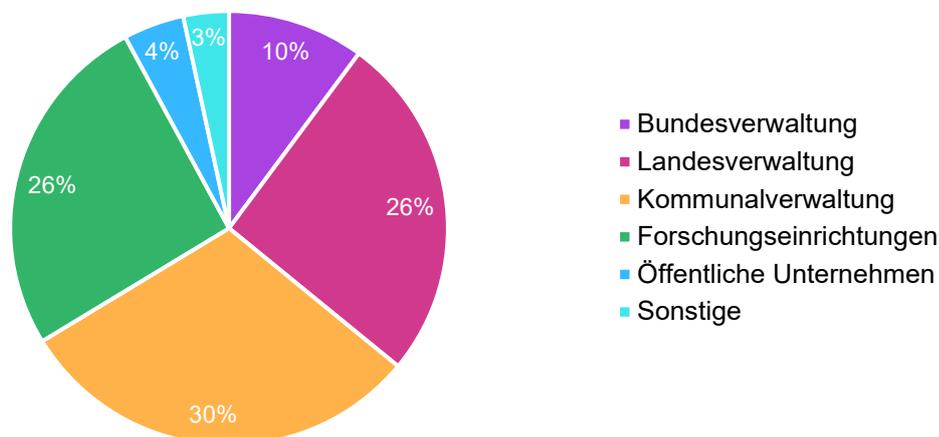


Abbildung 11: Verteilung der Institutionsarten, die anderweitig Daten veröffentlichen oder dies planen

Von 89 Teilnehmenden, die eine bestehende oder geplante anderweitige Datenveröffentlichung im Sinne von Open Data angeben (also nicht an GovData oder ein Landes-/Fachportal angebunden sind, aber anderweitig Daten veröffentlichen), sind knapp vier Fünftel (82,0%) der Landesverwaltung, Kommunalverwaltung oder Forschungseinrichtungen zuzuordnen.

71,4% der 70 Teilnehmenden, die eine bestehende Datenveröffentlichung ihrer Institution im Sinne von Open Data beschreiben, machen Angaben zu bisher veröffentlichten Datensätzen. Diese sind vor allem im niedrigen (1-50) und hohen Bereich (>250) angesiedelt, mit jeweils 31,4% und 28,6% der Angaben. Nur 11,4% der Teilnehmenden nennen Werte im mittleren Bereich (51-250).

Bei der Einschätzung der zukünftigen Mengen von zu veröffentlichenden Datensätzen wählen die relative Mehrheit der 89 Teilnehmenden, die eine bestehende oder geplante Datenveröffentlichung ihrer Institution im Sinne von Open Data nennen, die Antwortoption „Weiß ich nicht“ (je nach Jahr 40,4% bis 47,2%). Für 2023 schätzt die ein wesentlicher Teil noch eine Veröffentlichung von 1-10 Datensätzen (22,5%), während für 2024 und 2025 die Schätzungen oft im Bereich 11-100 liegen (23,6% für 2024, 25,8% in 2025). Anzumerken ist, dass für 2023 noch 11,1% der Teilnehmenden aus der Kommunalverwaltung (n = 27) mit 0 bereitzustellenden Datensätzen rechnen (siehe Abbildung 12).

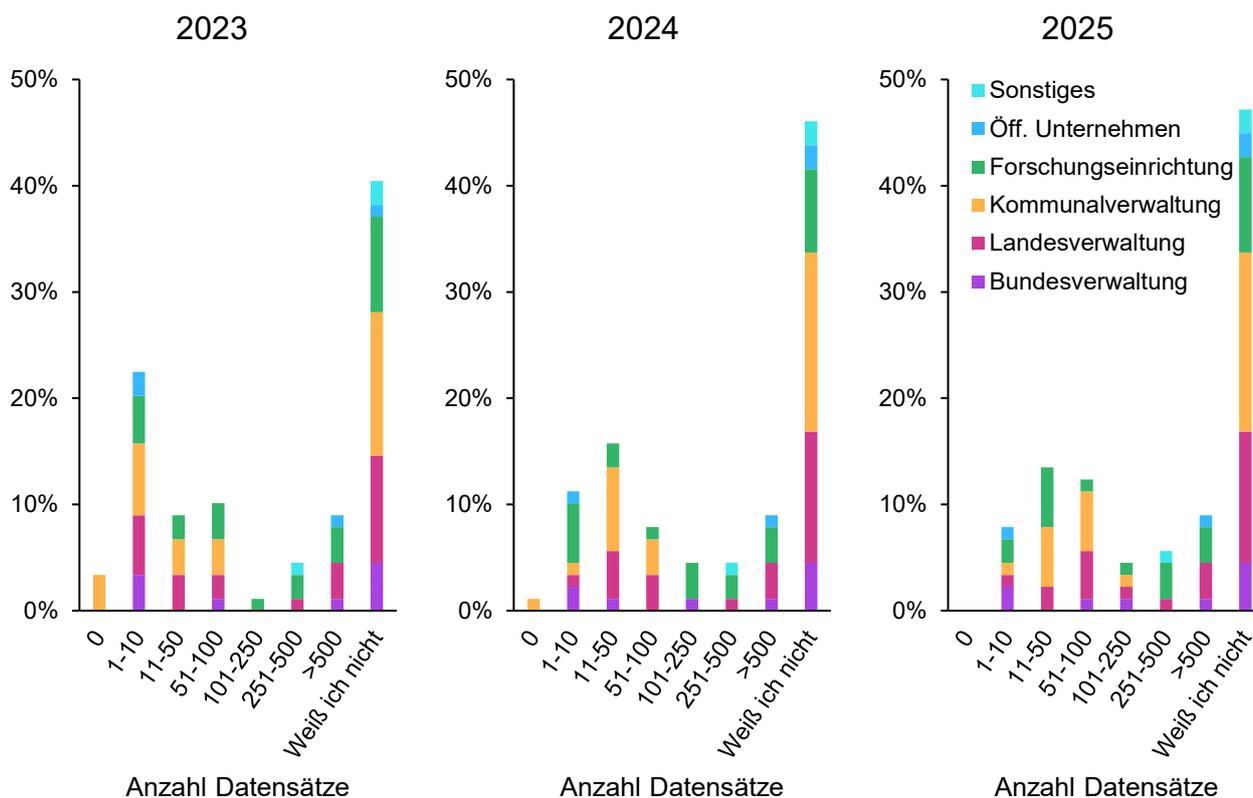


Abbildung 12: Schätzung zukünftig anderweitig im Sinne von Open Data zu veröffentlichender Datensätze 2023 – 2025 pro Institution

Teilnehmende, die eine bestehende oder geplante Datenveröffentlichung im Sinne von Open Data angeben (n = 89), schätzen die Anzahl an jährlich im Sinne von Open Data zu veröffentlichenden Datensätzen ihrer Institution in den Jahren 2023 bis 2025, getrennt nach Institutionsart.

Insgesamt 10,9% der gesamten Teilnehmenden geben weder eine Anbindung an ein Open Data-Portal, noch eine bestehende oder geplante anderweitige Datenveröffentlichung an. Als zentrale Gründe für die Ablehnung (Mehrfachnennung möglich) nennen 69,2% dieser Gruppe eine fehlende Priorität des Themas und je 34,6% fehlende Personalkapazitäten und einen geringen wahrgenommenen Nutzen.

2.3.4.5 Geplante Anbindungen

Von den 89 Teilnehmenden, die eine anderweitige bestehende oder geplante Datenveröffentlichung im Sinne von Open Data angeben, sagen 31 (34,8%) aus, dass ihre Institution eine Anbindung an ein Open Data Portal plane. Von diesen 31 planen etwa 45,2% die Anbindung an GovData²⁵. Den größten Anteil geplanter

²⁵Die Frage nach einer geplanten Anbindung an GovData war der Frage nach einer geplanten Anbindung an ein Landes- oder Fachportal vorangestellt. Teilnehmenden, die die Frage nach einer geplanten Anbindung an GovData mit „Ja“ beantworteten, wurde die Frage nach einer geplanten Anbindung an ein Landes- oder Fachportal nicht gestellt.

Anbindungen an GovData oder ein Landes- oder Fachportal machten unter den Institutionen die Stellen der Landesverwaltung (n = 23) aus, deren Teilnehmende dies in 60,9% der Fälle bejahten. Von diesen bezogen sich etwa ein Drittel (35,7%) auf eine Anbindung an GovData. Es folgen Stellen der Kommunalverwaltung (n = 27) mit 44,4% positiven Antworten, davon die Hälfte in Bezug auf GovData. Teilnehmende aus Stellen der Bundesverwaltung (n = 9) und öffentlichen Unternehmen (n = 4) befanden sich im mittleren Bereich mit jeweils 22,2% und 25,0% Bejahungen, alle in Bezug auf eine Anbindung an GovData. Forschungseinrichtungen (n = 23) planen laut ihren Teilnehmenden nur zu 8,7% eine Anbindung an ein Open Data Portal, davon keine an GovData. Keine als „Sonstige“ klassifizierte Institution (n = 3) plant laut den Teilnehmenden eine Anbindung (siehe Abbildung 13).

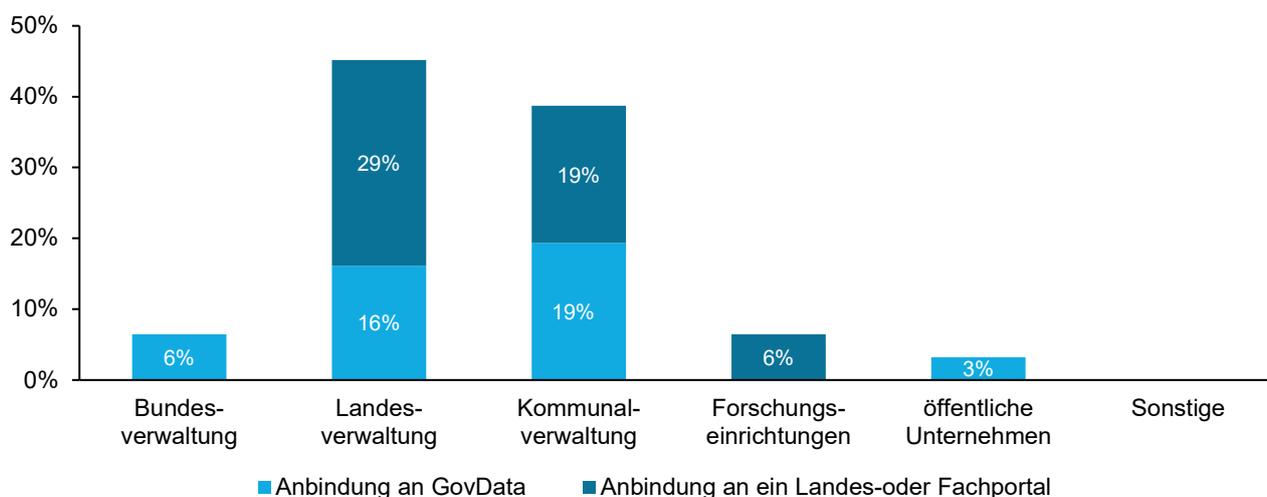


Abbildung 13: Anzahl geplanter Anbindungen an GovData oder ein Landes- oder Fachportal

31 Teilnehmende geben eine geplante Anbindung ihrer Institution an ein Open Data Portal an. 45,2% dieser Gruppe nennen GovData als das geplante Portal.

Die Motivationen für eine geplante Anbindung an GovData oder ein Landes- oder Fachportal (Mehrfachnennungen möglich) waren sowohl auf eine landes- oder bundesrechtliche Verpflichtung als auch auf Freiwilligkeit zurückzuführen. Von den 14 Teilnehmenden, die eine geplante Anbindung ihrer Institution an GovData angeben, nennen 50,0% Freiwilligkeit, 42,9% eine bundesrechtliche Verpflichtung und 35,7% eine landesrechtliche Verpflichtung²⁶. Von den 17 Teilnehmenden, die eine geplante Anbindung ihrer Institution an ein Landes- oder Fachportal angeben, nennen 52,9% Freiwilligkeit, 41,2% eine landesrechtliche Verpflichtung und 23,5% eine bundesrechtliche Verpflichtung.

Nennungen der Landes- oder Fachportale, an die eine Anbindung geplant ist, variieren. Nur zwei Portale, nämlich Berlin Open Data und GDI-DE, werden mehr als einmal genannt.

²⁶ Mehrfachnennungen waren möglich.

Die 53 Teilnehmenden, die eine geplante Anbindung an GovData oder ein Landes- oder Fachportal ihrer Institution mit „Nein“ beantworten, geben unterschiedliche Gründe dafür an (Mehrfachnennungen möglich). 41,5% nennen fehlende Personalkapazitäten, 32,1% eine fehlende Priorität, und 18,9% fehlende Expertise. Weitere 30,2% nennen nicht gelistete Gründe, unter anderem fehlende Vorgaben oder eine fehlende Passung der Daten zu GovData. Knapp ein Viertel (24,5%) der Teilnehmenden haben die Frage nicht beantwortet.

2.3.4.6 Metadatenstandards, Schnittstellen und Open Data Portal-Technologien

Für einen Überblick der technischen Landschaft über existierende Informationen hinaus beantworteten Teilnehmende, die eine bestehende oder geplante anderweitige Veröffentlichung ihrer Institution oder keine Anbindung an ein in der Umfrage gelistetes Landes- oder Fachportal angeben (n = 106), Fragen zur Verwendung von Metadatenstandards, Schnittstellen und Datenportalen.

Die Frage, ob Metadatenstandards verwendet werden, beantworten 41 (38,7%) der Teilnehmenden mit „Ja“, 49 (46,2%) mit „Nein“ und 16 (15,1%) mit „Weiß ich nicht“. Die am häufigsten genannten Metadatenstandards sind in Abbildung 14) dargestellt.

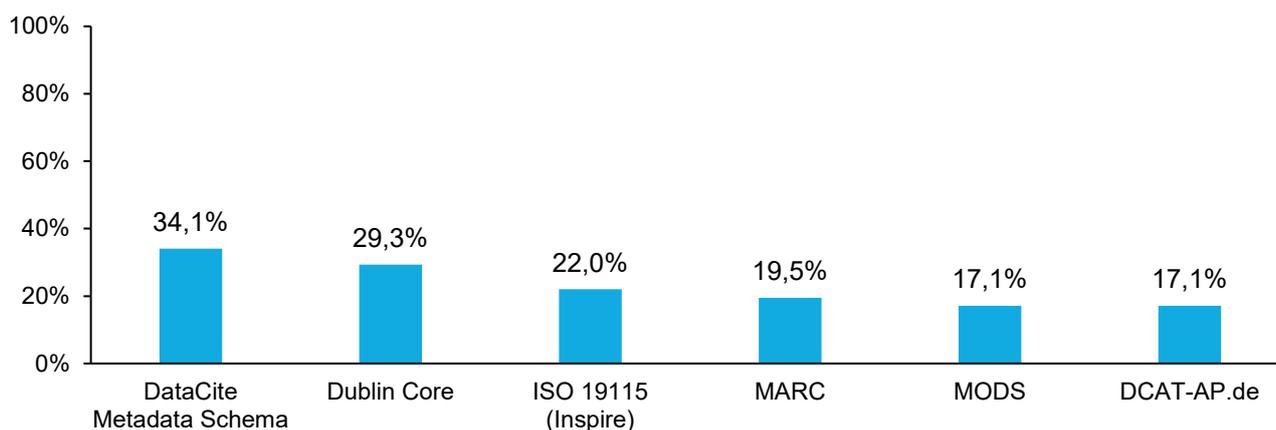


Abbildung 14: Häufigste genannte Metadatenstandards

Die 41 Teilnehmenden, die die Frage nach der Nutzung von Metadatenstandards in ihrer Institution bejahen, geben vor allem die angezeigten sechs Metadatenstandards an. Mehrfachnennungen waren möglich.

Zwischen den Institutionsarten sind dabei Unterschiede zu beobachten: DataCite Metadata Schema und Dublin Core werden vor allem von Teilnehmenden aus Forschungseinrichtungen genannt, und ISO 19115 von Teilnehmenden aus Stellen der Landesverwaltung.

Teilnehmende, die eine geplante Anbindung ihrer Institution an GovData angeben (n = 14), beantworten die Frage nach der Nutzung von Metadatenstandards zu 42,9% mit „Ja“ und 57,1% mit „Nein“. Diese Institutionen nennen ISO 19115 (Inspire) (50,0%) und DCAT-AP.de (33,3%) als genutzte Metadatenstandards.

Die Frage nach der Nutzung von Schnittstellen wurde von 62 (58,5%) der Teilnehmenden mit „Ja“ beantwortet, von 18 (17,0%) mit „Nein“ und von 26 (24,5%) mit „Weiß ich nicht“. Von den Teilnehmenden, die „Ja“ antworten, können knapp ein Drittel (32,3%) nicht sagen, welche Schnittstelle genutzt wird. Die am häufigsten genannten

Schnittstellen sind in Abbildung 15 dargestellt.

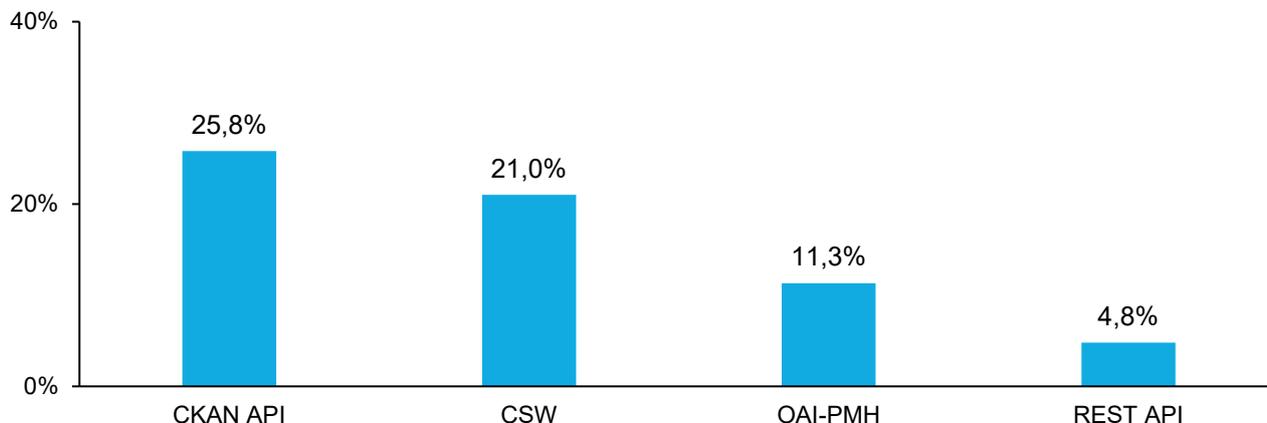


Abbildung 15: Häufigste genannte Schnittstellen

Die 62 Teilnehmenden, die die Frage nach der (geplanten) Nutzung von Schnittstellen in ihrer Institution bejahen, geben vor allem die angezeigten vier Schnittstellen an. Mehrfachnennungen waren möglich.

92,9% der Teilnehmenden, die eine geplante Anbindung ihrer Institution an GovData angeben, beantworten die Frage nach der Nutzung von Schnittstellen mit „Ja“ und 7,1% mit „Weiß ich nicht“. Von den positiven Antworten geben 42,9% die Nutzung von entweder CSW oder CKAN API an, und 14,3% die Nutzung beider Schnittstellen. Etwa ein Drittel (30,7%) der Teilnehmenden aus dieser Gruppe geben an, nicht zu wissen, welche Schnittstelle genutzt wird.

Zuletzt wurde nach der Existenz oder Planung von Open Data Portalen gefragt. Von den 106 Befragten antworteten 44 (41,5%) mit „Ja“, 47 (44,3%) mit „Nein“ und 15 (14,2%) mit „Weiß ich nicht“. Die Gruppe derjenigen, die ein bestehendes oder geplantes Datenportal ihrer Institution beschreiben, wurden nach der zugrundeliegenden Technologie gefragt. 34,1% konnten die Frage nicht beantworten. Die restlichen Teilnehmer gaben unterschiedliche Technologien an (siehe Abbildung 16).

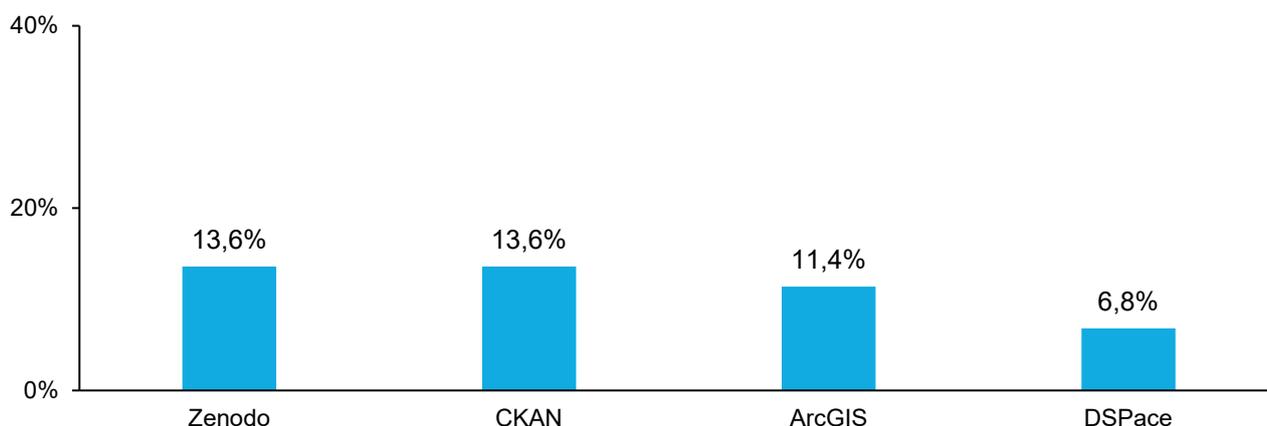


Abbildung 16: Häufigste genannte Open Data Portal-Technologien

Die 44 Teilnehmenden, die die Frage nach der Existenz eines (geplanten) eigenen Open Data Portal in ihrer Institution bejahen, geben vor allem die angezeigten vier zugrundeliegenden Technologien an. Mehrfachnennungen waren möglich.

Von den Teilnehmenden, die eine geplante Anbindung ihrer Institution an GovData angeben, beantworten 57,1% die Frage nach Datenportalen mit „Ja“, 35,7% mit „Nein“ und 7,1% mit „Weiß ich nicht“. Bei den Technologien, nach denen nach einer positiven Antwort gefragt wurde, ist nur CKAN hervorzuheben, das von 50,0% genannt wird. 6 weitere Technologien werden nur je einmal gewählt. 14,3% der Befragten konnten keine Antwort geben.

2.3.5 Fazit

Zusammengefasst zählt die Online-Umfrage wesentlich auf das Gesamtergebnis des Projekts ein. Durch die Erhebung repräsentativer Daten wurden:

- Schätzungen zu den zu erwartenden datenbereitstellenden Institutionen
- und den zu erwartenden Datensätze neu dazukommender Datenbereitstellender berechnet
- sowie ein Überblick über in der Landschaft genutzte Metadatenstandards, Schnittstellen und Technologien für Portale erstellt

Darüber hinaus liefert die Umfrage einen wertvollen Einblick in die bestehende Open Data Landschaft in Deutschland aus dem Blick der Teilnehmenden, durch Aussagen zu Anbindungen an GovData und Landes- und Fachportale, anderweitige Datenveröffentlichungen im Sinne von Open Data, und Werten zu bisher gelieferten Datensätzen pro Institution.

2.4 Experteninterviews (Qualitative Erhebung)

Als dritte Methode wurden im Zuge von TP 1 Experteninterviews eingesetzt. Insgesamt gaben im Zeitraum vom 13. bis zum 22. Juli 2022 vier Personen, die durch ihre berufliche Auseinandersetzung mit dem Thema Open Data als Experten in diesem Bereich identifiziert worden waren, je ein Interview.

Im Folgenden werden die Ergebnisse der Experteninterviews detailliert hinsichtlich ihrer Ziele, der

Vorgehensweise und der Ergebnisse erläutert.

2.4.1 Ziel der Experteninterviews

Das übergeordnete Ziel der Experteninterviews war die Einschätzung der Anzahl neuer Anbindungen an GovData in den nächsten drei Jahren durch Stellen der Kommunal- und/oder Landesverwaltung. Die Eingrenzung erfolgte, um eine identifizierbare Grundgesamtheit als Basis der Schätzung benennen zu können. Die Spezifikation „Kommunalverwaltung“ oder „Landesverwaltung“ war abhängig von der Expertise der jeweiligen Interviewten.

Um die Ergebnisse aus der Umfrage und Datenanalyse kontextualisieren zu können, war die Einschätzung der Entwicklung Open Data Kultur ein weiteres Ziel der Experteninterviews. Zuletzt wurden auch Informationen zu Hindernissen, die bei der Bereitstellung von Open Data oder Anbindung an GovData auftauchen, gesammelt, um langfristig Hinweise für weitere Handlungsbedarfe zu liefern.

2.4.2 Vorgehensweise

2.4.2.1 Fragebogenerstellung

Der Leitfaden für die Experteninterviews wurde in Abstimmung mit der Geschäfts- und Koordinierungsstelle (GKSt) GovData entwickelt. Die Interviews wurden semi-strukturiert konzipiert: die zentralen Fragen wurden festgelegt, die Durchführung war aber insofern frei, als dass die Reihenfolge verändert werden konnte und den Interviewten der Raum gegeben wurde, davon losgelöst von verwandten Themen zu berichten, wo sinnvoll.

Die wesentliche Struktur des Leitfadens war:

1. Information über das Projekt, die Zielstellung des Interviews und Einholen der Einwilligung zur Protokollierung
2. Präsentation bisheriger Erkenntnisse aus dem Projekt
3. Interviewfragen
 - a. Block 1: Allgemeine Open Data Landschaft
 - i. Einstellung der Institutionen gegenüber Open Data
 - ii. Zur Veröffentlichung verpflichtende Gesetze auf Landes- und Kommunalebene
 - iii. Motivation der Institutionen zur Datenveröffentlichung
 - iv. Anzahl datenveröffentlichender Institutionen aktuell und bis 2025
 - b. Block 2: Anbindungen an GovData
 - i. Anzahl sich anbindender Institutionen bis 2025
 - ii. Art der Anbindung (manuell vs. automatisch)
 - c. Block 3: Hindernisse
 - i. Bei der Veröffentlichung von Open Data
 - ii. Bei der Anbindung an GovData
4. Abschluss und Information zum weiteren Vorgehen

2.4.2.2 Expertenauswahl

Für die Auswahl der Interviewpartnerinnen und -partner wurde zunächst eine Liste geeigneter Personen recherchiert, und mit der GKSt GovData abgestimmt und priorisiert. Die finale Auswahl, mit denen Termine vereinbart werden konnten, bestand aus vier Teilnehmenden: zwei, die sich mit Open Data auch auf Kommunalebene in Deutschland befassten, und je einem mit Wissen auf Landesebene in Niedersachsen und auf Kommunalebene in Nordrhein-Westfalen.

2.4.2.3 Interviewvorbereitung

Im Vorfeld jedes Gesprächs wurden Dokumente zur Mitschrift sowie Präsentationsmaterial vorbereitet. Das Präsentationsmaterial mit insgesamt 16 Folien diente der Visualisierung der dargestellten Informationen und als Anker für die Gesprächspartnerinnen und Gesprächspartner während der Beantwortung der Fragen.

2.4.2.4 Interviewdurchführung

Die Interviews fanden als 60-75-minütiger Online-Videoanruf statt. Von Seiten Capgemini Invents waren immer zwei Teilnehmende anwesend – eine/r in der Rolle als Interviewer:in und der/die andere als Beobachter:in. Abhängig von der Expertise der/des Interviewten wurden die Fragen auf Gesamtdeutschland, Niedersachsen oder Nordrhein-Westfalen sowie auf Kommunal- oder Landesebene bezogen.

Das Gespräch wurde jeweils mit einer Tonaufzeichnung dokumentiert, die nach der Protokollerstellung gelöscht wurde.

2.4.2.5 Interviewnachbereitung

Im Nachgang an jedes Interview wurde basierend auf der Mitschrift und der Tonaufzeichnung ein Verlaufsprotokoll erstellt. Alle Interviewten erhielten das Protokoll zur Überprüfung auf die richtige Darstellung aller Aussagen. Inhaltlich gab es seitens der Expertinnen und Experten keine wesentlichen Beanstandungen.

2.4.2.6 Auswertung

Die Protokolle der Interviews wurden auf unterschiedlichen Ebenen miteinander verglichen und ausgewertet. Zunächst wurden die zentralen Aussagen aus den Interviews im Gesamten erarbeitet, indem die Aussagen der einzelnen Interviewten miteinander verglichen und zusammengefasst wurden. Anschließend wurde das gleiche Verfahren auf die einzelnen Themengebiete des Fragebogens angewandt.

2.4.3 Ergebnisse

2.4.3.1 Kernaussagen

Über alle Interviews hinweg konnten vier Kernaussagen identifiziert werden, bei denen weitestgehend Einigkeit herrschte:

1. Direkten Anbindungen an GovData durch Stellen der Landes- oder Kommunalverwaltung sind in nur geringem Maße zu erwarten.
2. Open Data wird von Institutionen langsam als Notwendigkeit erkannt. Das Verständnis dafür ist aber

noch segmentiert und stark von Individuen in den einzelnen Stellen abhängig.

3. Die weitere Entwicklung von Open Data hängt stark von politischem Einsatz auf Landesebene und der Bereitstellung entsprechender Ressourcen für Stellen der Landes- und Kommunalverwaltung ab.
4. Eine Einschätzung zukünftiger Mengengerüste für GovData oder der allgemeinen Open Data Landschaft ist auch für die Interviewten nicht vollumfänglich möglich.

2.4.3.2 Detaillierte Ergebnisse

Im Rahmen der Einschätzung der allgemeinen Open Data Landschaft in Deutschland beschreiben die Experten und Expertinnen durchgängig eine positive Entwicklung des Verständnisses in den letzten Jahren. Diese Entwicklung wird als tendenziell eher von Einzelpersonen abhängig beschrieben. Positiv ist anzumerken, dass Open Data mittlerweile mit weniger Vorbehalten behaftet ist als vor einigen Jahren. Das ist vor allem in Umfeldern der Fall, in denen Open Data stärker betrieben wird. Auf Kommunalebene bedeutet das, dass benachbarte Kommunen durch das Betreiben von Open Data einen Einfluss auf umliegende Kommunen haben können. Die Expertinnen und Experten sehen hier weiterhin die Notwendigkeit eines stärkeren politischen Engagements, inklusive der Bereitstellung entsprechender personeller und finanzieller Ressourcen. Zum Zeitpunkt der Interviews benennen die sie keine Gesetze, die ein entsprechendes politisches Engagement durch Verpflichtung zur flächendeckenden Datenveröffentlichung für Kommunen aufzeigen. Zwar schaffen einige wenige Kommunen Regelungen in Verwaltungsvorschriften, die zur Datenveröffentlichung motivieren und Anreize schaffen sollen – diese seien aber eher Einzelfälle. Durch das fragmentierte Bild von Open Data in Deutschland ist es den Interviewten nicht möglich, eine konkrete Anzahl zur Datenveröffentlichung motivierter Institutionen zu benennen. Sie beschreiben aber eine solche Motivation als Ergebnis aus Gesetzgebung und Kultur: durch Gesetze soll Handlungssicherheit und durch einen Kulturwandel auf allen Ebenen ein Open Data-freundliches Umfeld geschaffen werden.

Vor dem Hintergrund dieser allgemeinen Einschätzungen geben die Expertinnen und Experten an, die Benennung einer konkreten Anzahl aktuell und zukünftig veröffentlichender Institutionen auf Landes- und / oder Kommunalebene sei nur bedingt möglich. Für Kommunen in Nordrhein-Westfalen wurde die Spanne 30 bis 40 genannt, zusätzlich zu den Kommunen, die an Open.NRW angebunden sind. Auf Bundesebene belaufen sich die Schätzungen bereits datenveröffentlichender Kommunen auf einen mittleren dreistelligen Bereich mit 250 bis 700. Mit einer langsameren zu erwartenden Entwicklung gibt ein:e Interviewte:r an, die Zahl veröffentlichender Kommunen könne auf bis zu 1000 ansteigen, wenn der Einstieg in die Veröffentlichung von Daten über Geoplattformen stattfinden würde. Ein:e andere:r dagegen schätzt ca. 130 zusätzliche Kommunen in den nächsten Jahren. Die geschätzten Mengen sind auch durch Hindernisse bedingt, die Institutionen das Betreiben von Open Data erschweren. Als Hindernisse nennen die Expertinnen und Experten vor allem sechs Felder:

- Kultur: Der Wandel hin zu einer Fehlerkultur, in der große Bedenken hinsichtlich z.B. möglicher Probleme mit Datenqualität akzeptiert werden, ohne das Unterfangen zu beenden, ist langsam.
- Ressourcen: Vor allem personelle wie auch finanzielle Mittel sind für Open Data oft nicht vorhanden.

- technisches Verständnis: Personen können oft nicht mit den technischen Aspekten von Open Data umgehen.
- rechtliche Sicherheit: Es besteht Unsicherheit hinsichtlich der rechtlichen Lage, sollte es zu Problemen kommen.
- Systeme und Technik: Bei der Konzeption neuer IT-Systeme wird Open Data oft nicht mitgedacht.
- Datenzugehörigkeit: Daten liegen oft bei unterschiedlichen Akteuren.

Auch bei der Schätzung von sich in Zukunft anbindenden Stellen der Landes- und Kommunalverwaltung sind die Expertinnen und Experten vorsichtig. Dass sich Institutionen hier direkt anbinden werden, halten alle für möglich. Die konkreten Zahlen bewegen sich allerdings auf einem niedrigen Niveau: Es wird entweder ein niedriger zweistelliger Prozentsatz erwartet (auf einer deutlich längeren Zeitleiste als die besprochenen drei Jahre von 2023 bis 2025) oder eine absolute Zahl im mittleren zweistelligen Bereich (50 bis 100 Kommunen). Dabei sehen die Interviewten vor allem die automatisierte Anbindung als zentral an. Eine:r gibt an, eine manuelle Anbindung könne die Eintrittsschwelle senken, sei aber aufgrund der höheren benötigten Ressourcen nicht als langfristige Lösung zu betrachten.

Hindernisse, die im Kontext mit GovData gesehen werden, werden nur wenige genannt:

- Datenverlust: Daten, die grundsätzlich erkenntnisreich oder notwendig sind (z.B. PDF-Dateien mit Erklärungstexten), könnten potenziell verloren gehen, wenn Datensätze entfernt werden, um die Datenqualität der Gesamtheit an (Meta-)Datensätzen zu steigern.
- Verständlichkeit: Es fehlt eine einheitliche Taxonomie zur Kategorisierung und Beschreibung von Datensätzen.
- Nutzbarkeit: Die Nutzbarkeit wird unter der steigenden Zahl an Datensätzen leiden, weswegen erwägt werden muss, wie diese Datensätze am besten zu repräsentieren sind, und das Portal für APIs geeignet gemacht werden muss.

2.4.4 Fazit

Die Experteninterviews liefern über die quantitativen Methoden hinaus eine Kontextualisierung der Daten. Die geringe Schätzung zukünftiger Anbindungen aus der Kommunalverwaltung (im Vergleich zu der hohen Zahl an Kommunen in Deutschland) validiert die Eingrenzung des Verteilers als Grundgesamtheit für die Berechnung der Mengengerüste hinsichtlich der zu erwartenden anzubindenden Stellen (siehe Kapitel 2.5).

Zusätzlich bieten die Experteninterviews Einsichten zu Hindernissen im Kontext von Open Data oder der Anbindung an GovData, die zukünftig als Handlungsfelder genutzt werden können, um das Konzept weiter voranzutreiben.

2.5 Berechnung Mengengerüste

Im Folgenden erfolgt die Herleitung der zu schätzenden Mengengerüste unter Einbeziehung der Ergebnisse der beschriebenen Methoden (Datenanalyse, Online-Umfrage, Experteninterviews). Die Zahlen bilden den Kern von

Arbeitspaket 1 und sollen vor allem als Antworten auf folgende Fragen dienen:

1. Wie viele Metadatenätze werden in den nächsten drei Jahren über GovData schätzungsweise bereitgestellt?
2. Wie viele anzubindende Stellen (manuell und automatisch) sind in den nächsten drei Jahren schätzungsweise zu erwarten?
3. Welche Metadatenstandards, Schnittstellen und Open Data Portal-Technologien werden genutzt?

2.5.1 Schätzung der zu erwartenden Metadatenätze

2.5.1.1 Vorgehen

Die Gesamtzahl zukünftig über GovData bereitgestellte Metadatenätze für 2023 bis 2025 lässt sich formal definieren als Summe

1. der Anzahl zukünftiger Metadatenätze bestehender datenveröffentlichender Stellen und
2. der Anzahl zukünftiger Metadatenätze neuer datenveröffentlichender Stellen.

Während das Veröffentlichungsverhalten für beide Teile modelliert werden muss, ist zumindest die für den ersten Teil notwendige Anzahl der bestehenden datenveröffentlichenden Stellen bekannt. Die Anzahl neuer datenveröffentlichender Stellen wiederum muss ebenfalls geschätzt werden (siehe Kapitel 2.5.2).

2.5.1.2 Schätzung Anzahl zukünftiger Metadatenätze bestehender datenveröffentlichender Stellen

Bei der Schätzung der Metadatenätze bestehender veröffentlichender Institutionen wurde zwischen den jeweiligen Institutionsarten unterschieden, da die Datenanalyse auf voneinander abweichende Veröffentlichungsmuster hinweist und dadurch eine entsprechende Schätzung abgegeben werden kann.

Zunächst wurde für jedes Jahr im Beobachtungszeitraum 2013 - 2021 pro Organisationsart die durchschnittliche Anzahl der neu veröffentlichten Metadatenätze sowie die Anzahl der veröffentlichenden Stellen ermittelt²⁷. Damit eventuelle Ausreißer die Schätzung nicht verzerren, wurde ein Akzeptanzintervall von 1,5 Standardabweichungen gewählt und ggf. eine Substitution mit dem Mittelwert vorgenommen: Liegt also für eine Institutionsart in einem Jahr die Anzahl neu veröffentlichter Datensätze außerhalb dieses Intervalls, so wurde der Wert durch das arithmetische Mittel der jeweiligen Institutionsart ersetzt.

Die berechnete Anzahl an veröffentlichten Datensätzen innerhalb eines Jahres ergibt sich aus drei Faktoren:

- (a) Die insgesamt Anzahl an veröffentlichenden Stellen pro Institutionsart im Beobachtungszeitraum
- (b) Der durchschnittliche Anteil an Veröffentlichungen pro Institutionsart²⁸

²⁷ Für die Erstellung der Schätzung wurden Metadatenätze mit Veröffentlichungsdatum im Jahr 2022 nicht berücksichtigt, da das Jahr zum Zeitpunkt der Analyse noch nicht abgeschlossen war.

²⁸ Nicht alle veröffentlichenden Stellen veröffentlichen jedes Jahr Daten. Entsprechend ist der durchschnittliche Anteil an Veröffentlichungen pro Jahr der Quotient aus der durchschnittlichen Anzahl von veröffentlichenden Stellen, welche im Jahr mindestens einen Datensatz veröffentlicht haben, und der gesamten Anzahl an veröffentlichenden Stellen pro Institutionsart im Beobachtungszeitraum.

- (c) Die durchschnittliche Anzahl an Veröffentlichungen pro Institution und Institutionsart innerhalb der letzten 3 Jahre²⁹

Das Produkt der zwei Faktoren (a) und (b) ergibt die Anzahl an veröffentlichten Datensätzen innerhalb eines Jahres pro Institutionsart.

Um ein sich änderndes Veröffentlichungsverhalten zu berücksichtigen, wurde weiterhin das Wachstum der durchschnittlichen Anzahl an Veröffentlichungen der letzten 3 Jahre als Prozentwert berechnet. Zur Berechnung der Schätzung für die Jahre 2023 – 2025 wurde die durchschnittliche Anzahl an Veröffentlichungen pro Jahr (c) jedes Jahr mit dem zuvor berechneten Wachstumswert multipliziert und die Anzahl an veröffentlichten Metadatenätzen analog berechnet.

Die folgende Beispielrechnung soll die Berechnung anhand des Beispiels der geschätzten Veröffentlichungen für die derzeit angebotenen Institutionen der Bundesverwaltung im Jahr 2023 verdeutlichen:

- (a) Insgesamt haben 69 Institutionen der Bundesverwaltung im Beobachtungszeitraum mindestens ein Mal, d.h. mindestens einen Datensatz über GovData zur Verfügung gestellt (siehe auch Abbildung X).
- (b) Nicht jede Institution hat jedes Jahr auch Datensätze veröffentlicht bzw. über GovData bereitgestellt. Dies haben im Durchschnitt nur 16,9% (ca. 12) der Institutionen der Bundesverwaltung getan.
- (c) Unter Betrachtung der Jahre 2019 bis 2021 wurden durchschnittlich 44,77 Datensätze von Institutionen der Bundesbehörden pro Jahr über GovData bereitgestellt. Da dieses nicht als konstant angenommen werden kann, wurde zusätzlich das Wachstum der durchschnittlichen Anzahl an Veröffentlichungen berechnet (27% pro Jahr) und mit dem ursprünglichen Wert aus dem Vorjahr multipliziert. Entsprechend wurden zur Berechnung der Zahlen für das Jahr 2023 im Durchschnitt 57 Veröffentlichungen angelegt.

Für die Ermittlung der Veröffentlichungen der Bundesverwaltungen im Jahr 2023 wurden die drei Werte aus (a), (b) und (c) multipliziert ($69 * 0,169 * 57$), sodass im Ergebnis dort 665 Metadatenätze geschätzt werden. Die Schätzung der Metadatenätze für alle weiteren Institutionsarten erfolgte entsprechend analog.

2.5.1.3 Schätzung Anzahl zukünftiger Metadatenätze neuer datenveröffentlichender Stellen

Für die Schätzung der Anzahl an Metadatenätzen von neuen datenveröffentlichenden Stellen innerhalb eines Jahres wurden die Ergebnisse aus der Umfrage herangezogen (siehe Kapitel 2.3.4). Die Gesamtsumme an zukünftigen veröffentlichenden Stellen ergibt sich aus den direkten Anbindungen an GovData und den Anbindungen an ein Landes- oder Fachportal. Die untere und obere Grenze des berechneten Konfidenzintervalls auf 95%-igem Konfidenzniveau³⁰, das für eine höhere Validität der Ergebnisse der Umfrage zu den zukünftigen Anbindungen berechnet wurde, stellen jeweils die Mindest- und Maximalzahl an zukünftig

²⁹ Es wird angenommen, dass das grundsätzliche Veröffentlichungsverhalten sich entsprechend der Entwicklung rund um das Thema Open Data in den letzten Jahren dynamisch verändert hat. Um die Berechnung nicht aufgrund zu langer Betrachtungszeiträume zu beeinflussen wurden nur die letzten drei Jahre herangezogen.

³⁰ Definition Konfidenzintervall: Intervall, das unter wiederholter zufälliger Stichprobenprüfung unter identischen Bedingungen den wahren Wert in den dem Konfidenzniveau entsprechenden Prozent der Fälle enthalten würde. (vgl. M. Smitshon (2003): Confidence Intervals (No. 140). Sage, ISBN: 9780761924999, S. 1 f.)

zu erwartbaren veröffentlichenden Stellen dar. Zusammen mit dem Mittelwert aus der Umfrage ergeben sich so drei Werte, anhand derer drei verschiedene Szenarien für die neu angebotenen Stellen erstellt wurden. Es wurde unter der Annahme gearbeitet, dass sich die Zahl der neuen veröffentlichenden Stellen gleichmäßig über die drei zu schätzenden Jahre 2023 bis 2025 verteilt, sich also in jedem Jahr ein Drittel der erwarteten Gesamtanzahl anbinden wird.

Analog zu der Berechnung von bereits angebotenen Stellen (siehe oben) wurde mit angepassten Werten für den Anteil an Veröffentlichungen (b) und der durchschnittlichen Anzahl an Veröffentlichungen pro Jahr (c) die Anzahl an Metadatenätzen geschätzt. Auch hier ergibt sich Anzahl an veröffentlichenden Datensätzen innerhalb eines Jahres aus dem Produkt der Faktoren (a), (b) und (c), wobei die Gesamtsumme der zukünftigen veröffentlichenden Stellen (a) wie erläutert aus den Antworten der Umfrage abgeleitet wurde. Auch die Berücksichtigung des jährlichen Wachstums für die durchschnittliche Anzahl an Veröffentlichungen pro Jahr wurde berücksichtigt. Weiterhin wurde das Ergebnis der Datenanalyse miteinbezogen, wonach eine neu angebotene Stelle im ersten Jahr ein anderes Veröffentlichungsmuster aufweist als in den Folgejahren (siehe Kapitel 2.2.3). Deshalb flossen diese in den Folgejahren an die Anbindung mit eigenen angepassten Werten für den Anteil an Veröffentlichungen und Wachstum der durchschnittlichen Veröffentlichungen pro Jahr in die Schätzung mit ein.

2.5.1.4 Ergebnis

Die Ergebnisse der dargelegten Schätzung sehen vor, dass in den Jahren 2023–2025 geschätzt 25.525 Metadatenätze von bereits bestehenden Stellen veröffentlicht werden, unabhängig vom Szenario. Von den neuen veröffentlichenden Stellen werden im Zeitraum 2023–2025 geschätzt zwischen 12.654 und 47.515 Metadatenätze veröffentlicht werden. Für die finale Schätzung der Mengengerüste wurden die beiden Teilschätzungen der Metadatenätzen von bereits bestehenden Anbindungen und neuen Anbindungen addiert. Dies bedeutet hinsichtlich der Schätzung von zusätzlichen Metadatenätzen für 2023–2025 insgesamt zwischen 38.180 und 73.041 Metadatenätze (siehe Tabelle 2).

Szenario	Zusätzliche Metadatenätze 2023–2025			
	Bestehende Datenbereitstellende		Neue Datenbereitstellende	Summe
Minimales Szenario	25.525		12.654	38.180
Mittleres Szenario	25.525	+	27.676	= 53.202
Maximales Szenario	25.525		47.515	73.041

Tabelle 2: Berechnete Schätzung zusätzlicher Metadatenätze auf GovData von 2023 bis 2025

Die Gesamtschätzung der von 2023 bis 2025 insgesamt dazukommenden Metadatenätzen auf GovData basiert auf den Teilschätzungen „Bestehende Datenbereitstellende“ und „Neue Datenbereitstellende“, die unterschiedliches Veröffentlichungsverhalten aufweisen.

Basierend auf den Ergebnissen der Umfrage sowie der Datenanalyse ist somit damit zu rechnen, dass bis zu 73.000 neue Metadatensätze auf GovData bereitgestellt werden (siehe Abbildung 17). Dabei muss die grundsätzliche Annahme beachtet werden, dass in dieser Zeit keine wesentlichen regulatorischen Änderungen, z.B. eine Erweiterung der grundsätzlichen Veröffentlichungspflicht von Daten durch Landes- oder Kommunalbehörden, vorgenommen werden. An dieser Stelle ist darauf zu verweisen, dass derzeit eine Diskussion um die Verschärfung der Verpflichtung aus §12 a EGovG geführt wird. Potenzielle Änderungen würden die zu erwartenden Zahlen vermutlich deutlich erhöhen. Allerdings kann eine solche Änderung der Gesetzeslage zu diesem Zeitpunkt weder antizipiert noch zielführend in die Schätzung miteinbezogen werden.

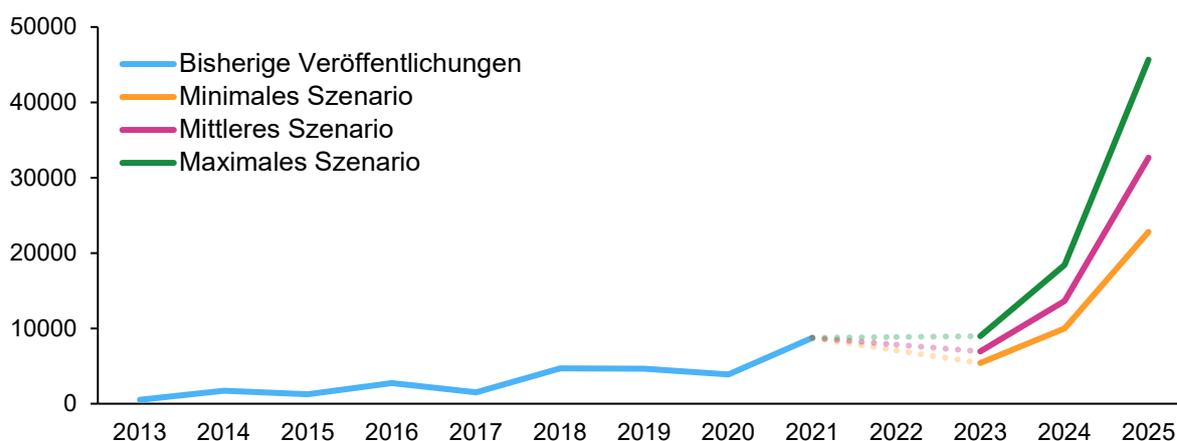


Abbildung 17: Schätzung neuer Metadatensätze pro Jahr – 3 Szenarien

Unter der Annahme, dass die aktuellen regulatorischen Umstände bestehen bleiben, ist je nach Szenario in Summe für die Jahre 2023 bis 2025 mit bis zu etwa 73.000 neuen Metadatensätzen zu rechnen.

2.5.2 Schätzung der zu erwartenden anzubindenden Stellen

2.5.2.1 Vorgehen

Die Schätzung der zu erwartenden anzubindenden Stellen basiert vorrangig auf den Ergebnissen der Online-Umfrage. Dabei wird von der Annahme ausgegangen, dass der für den Versand genutzte Verteiler den Großteil der Institutionen abdeckt, die über die personellen und finanziellen Ressourcen und Möglichkeiten zur Anbindung an GovData bzw. Veröffentlichung von Open Data verfügen. Diese Annahme wurde im Rahmen der Experteninterviews validiert. Aus den Informationen der Expertinnen und Experten lässt sich schließen, dass im Verhältnis zur Anzahl an Kommunen und Landesbehörden (über 10.000 deutschlandweit) nur mit geringer Partizipation seitens entsprechender Stellen zu rechnen sei.

Die Berechnung der Institutionen, die sich in den Jahren 2023-2025 an GovData anbinden möchten und werden, basiert auf dem identifizierten Verteiler und erfolgte über mehrere Schritte:

1. Bereits an GovData angeschlossene Institutionen (direkt oder über ein Landes- oder Fachportal angebunden) wurden von der Grundgesamtheit des Verteilers ausgeschlossen. Diese Institutionen

wurden durch einen manuellen Abgleich der Institutionen im Verteiler mit den Daten von GovData identifiziert. Die resultierende Gesamtzahl stellt die Basis für die weitere Berechnung dar.

2. Der Anteil der Institutionen, die bereits Daten veröffentlichen oder dies zu planen, wurde anhand der Umfrage berechnet. Weiterhin wurde hier mit dem Ziel, Ergebnisse mit erhöhter Validität zu erarbeiten, Konfidenzintervalle auf 95%-igem Konfidenzniveau berechnet³¹. Dadurch entstand zusätzlich um den Einzelwert herum eine Spanne, an deren Ober- und Untergrenze die Berechnung des nächsten Schrittes zusätzlich zum Mittelwert erfolgte.
3. Zur Berechnung der Anzahl der Institutionen, die sich zukünftig an GovData anbinden werden, wurde der Anteil der Institutionen in der Umfrage herangezogen, deren Teilnehmende eine geplante Anbindung an GovData angeben. Dieser Anteil, inklusive Konfidenzintervall auf 95%-igem Konfidenzniveau, wurde auf die zuvor errechneten Werte der Spanne angewandt.
4. Zuletzt wurde die Anzahl der Institutionen geschätzt, die sich zukünftig an ein Landes- oder Fachportal anbinden werden. Dafür wurde die Spanne des Anteils der Institutionen, die bereits Daten veröffentlichen oder dies planen (siehe Punkt 2) durch Subtraktion der entsprechenden Werte aus Punkt 3 auf die notwendige Zielgesamtheit reduziert. Anschließend wurde analog zu Punkt 3 der Anteil der Institutionen aus der Umfrage, deren Teilnehmende angeben, sich zukünftig an ein Landes- oder Fachportal anbinden zu wollen, inklusive Konfidenzintervall auf 95%-igem Konfidenzniveau auf die entsprechende Spanne angerechnet.

Durch die Anwendung von Konfidenzintervallen entstand auf jeder Berechnungsebene eine geschätzte Spanne, die entsprechend der Wahrscheinlichkeit des gewählten Konfidenzniveaus (in diesem Fall 95%) statistisch gesehen den erwartbaren Wert der Gesamtpopulation enthält. Basierend auf der finalen Spanne wurden Szenarien für die Anzahl zukünftig anzubindender Stellen abgeleitet: die Untergrenze für das minimale Szenario, der Mittelwert für das mittlere Szenario, und die Obergrenze für das maximale Szenario.

Zur Berechnung des Anteils an automatischen und manuellen Anbindungen unter den Institutionen, die sich an GovData anbinden werden, die auf bisherigen Erfahrungen der GKSt GovData beruhende Annahme getroffen, dass sich ca. 90% der Institutionen automatisch und 10% manuell anbinden werden.

Der Aufwand, der durch Mapping von neuen Metadatenstandards entstehen wird, ist durch die Zahl der Metadatenstandards ohne bestehendes Mapping begrenzt. Bestehende Mappings waren durch Abstimmung mit der GKSt GovData bekannt. Die Zahl der neu mit einem Mapping zu versehenen Metadatenstandards wurde durch die Umfrage ermittelt und auf diejenigen Standards begrenzt, die mehrfach ausgewählt bzw. genannt wurden.

³¹ Definition Konfidenzintervall: Intervall, das unter wiederholter zufälliger Stichprobenprüfung unter identischen Bedingungen den wahren Wert in den dem Konfidenzniveau entsprechenden Prozent der Fälle enthalten würde. (vgl. M. Smithson (2003): *Confidence Intervals* (No. 140). Sage, ISBN: 9780761924999, S. 1f.)

2.5.2.2 Ergebnis

Die Berechnung der Spanne erwartbarer Datenbereitsteller über GovData wurde konzeptuell umgesetzt wie beschrieben. Von den erstellten Szenarien, welche den minimalen, mittleren und maximalen Wert repräsentieren, wird das mittlere Szenario als vorrangig betrachtet.

Durch Abgleich der Institutionen im Verteiler mit den bei GovData als Datenbereitstellende³² gelisteten Institutionen wurde die Grundgesamtheit von 1855 auf 1631 Institutionen reduziert. Von diesen 1631 Institutionen werden sich in den nächsten 3 Jahren im mittleren Szenario schätzungsweise 153 (bei einer Spanne von 69 bis 257) Institutionen direkt an GovData anbinden, und schätzungsweise 186 (bei einer Spanne von 86 bis 325) an ein Landes- oder Fachportal. Es wurde angenommen, dass sich die Zahl der neu veröffentlichenden Stellen gleichmäßig über die drei zu schätzenden Jahre 2023 bis 2025 verteilt. Damit wird ein jährlicher Zuwachs um ein Drittel der Gesamtanzahl angegeben. Die geschätzten Werte neu anzubindender Stellen an GovData und an Landes- und Fachportale sind in Tabelle 3 dargestellt.

Die Anteile der Art der Anbindung werden entsprechend bisheriger Erfahrungswerte seitens der GKSt GovData wie folgt geschätzt: 90% automatisch und 10% manuell. Weitere Aufwände für das Mapping neuer Metadatenstandards bei automatischen Anbindungen sind begrenzt auf die Zahl noch nicht gemappter Metadatenstandards, die von Teilnehmenden in der Umfrage angegeben wurden³³. Dies sind die folgenden 7: Dublin Core, DataCite Metadata Schema, Encoded Archival Description, MARC, METS, MODS, OAI-ORE. Die Anteile der Institutionen nach Art der Anbindung (manuell, automatisch, automatisch mit Mapping) pro Jahr sind in Tabelle 3 dargestellt.

³² Sowohl direkt als auch indirekt über Landes- oder Fachportale.

³³ Teilnehmende, die nach genutzten Metadatenstandards gefragt wurden: n = 41

		Szenario					
		Minimal		Mittel		Maximal	
		Gesamt	Pro Jahr	Gesamt	Pro Jahr	Gesamt	Pro Jahr
Daten-bereitstellende	<i>GovData</i>	69	23	153	51	257	86
	<i>Landes-/ Fachportal</i>	86	29	186	62	325	108
Anbindungen	<i>Automatisch</i>	55	18	131	44	224	75
	<i>Automatisch inkl. Mapping</i>	7	2	7	2	7	2
	<i>Manuell</i>	7	2	15	5	26	9

Tabelle 3: Zu erwartende anzubindende Stellen inkl. Art der Anbindung von 2023 bis 2025

Die Werte der Szenarien basieren auf Berechnungen von Konfidenzintervallen der Umfrageergebnisse zu geplanten Anbindungen an GovData oder ein Landes- oder Fachportal und dem erwarteten Verhältnis der Anbindungsarten basierend auf Erfahrungswerten der GKSt GovData. Die Summe der Gesamtwerte für automatische Anbindung und automatisch Anbindung inkl. Mapping ergibt die erwarteten 90% automatischer Anbindungen. Das mittlere Szenario, von welchem zentral ausgegangen wird, ist zur Hervorhebung grau hinterlegt. Es wurde auf ganze Zahlen gerundet, weswegen die Summe der Anbindungen pro Jahr nach Art der Anbindung von den Werten zu gesamten Anbindungen an GovData abweichen kann.

2.5.3 Darstellung genutzter Metadatenstandards, Schnittstellen und Open Data Portal-Technologien

2.5.3.1 Vorgehen

Aus den Ergebnissen der Online-Umfrage wurde eine Übersicht erstellt, welcher Anteil an Institutionen insgesamt angegeben hat, Metadatenstandards, Schnittstellen und Open Data Portale zu verwenden oder dies zu planen. Hierfür wurden ebenfalls Konfidenzintervalle herangezogen.

Darauf basierend wurde eine Darstellung der genutzten / geplanten Metadatenstandards, Schnittstellen und Open Data Portal-Technologien erstellt. Wo sinnvoll, wurden Informationen nach Institutionsart getrennt, um unterschiedliche Nutzungsintensitäten zu verdeutlichen.

2.5.3.2 Ergebnis

Aus der Umfrage konnte abgeleitet werden, dass 38,7% der Teilnehmenden angeben, ihre Institution verwende einen Metadatenstandard, mit dem 95%-igen Konfidenzintervall [29,4% - 48,0%]. Es bestehen Unterschiede in den genutzten Metadatenstandards zwischen den Ebenen der Verwaltung (Bundesverwaltung, Landesverwaltung, Kommunalverwaltung) und Forschungseinrichtungen und Sonstigen (öffentliche Unternehmen, sonstige Institutionen). Die genutzten Metadatenstandards für beide Gruppen sind in Abbildung 18 dargestellt.

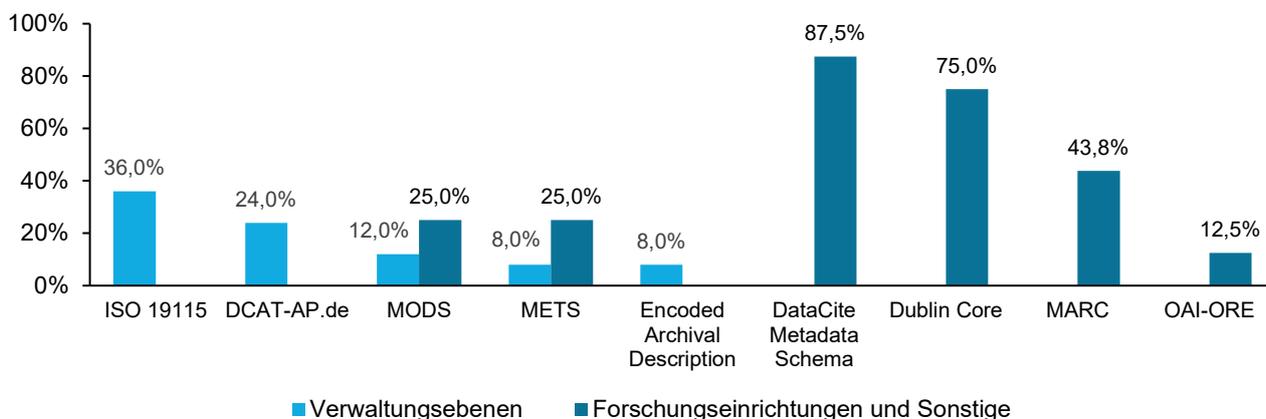


Abbildung 18: Häufigste genannte Metadatenstandards

Die Gruppen der Ebenen der Verwaltung (Bundesverwaltung, Landesverwaltung, Kommunalverwaltung) und der Forschungseinrichtungen und Sonstige geben unterschiedliche genutzte Metadatenstandards an. Mehrfachnennungen waren möglich.

58,5% der Teilnehmenden geben an, ihre Institution verwende Schnittstellen oder plane dies, mit dem 95%-igen Konfidenzintervall [49,1% - 67,9%]. Es bestehen Unterschiede zwischen den Ebenen der Verwaltung und Forschungseinrichtungen sowie sonstigen Institutionen. Abbildung 19 stellt die überwiegend angegebenen Schnittstellen dar.

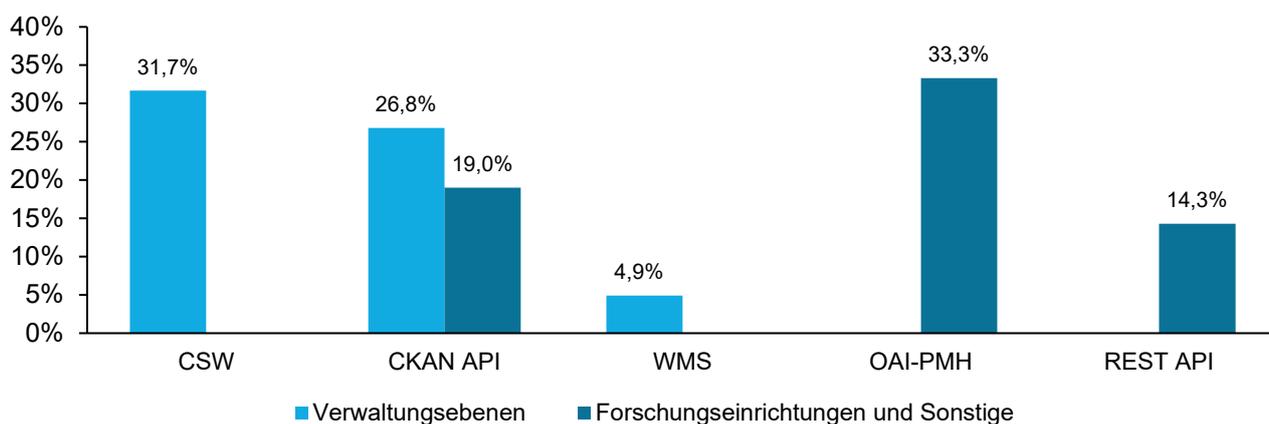


Abbildung 19: Häufigste genannte Schnittstellen

Die Gruppen der Ebenen der Verwaltung (Bundesverwaltung, Landesverwaltung, Kommunalverwaltung) und der Forschungseinrichtungen und Sonstige geben unterschiedliche genutzte Schnittstellen an. Mehrfachnennungen waren möglich.

Bei der Nutzung oder Planung Open Data Portale konnte basierend auf den Angaben der Teilnehmenden ein Anteil von 41,5% identifiziert werden, mit dem Konfidenzintervall [32,1% - 50,9%]. Wie zuvor, besteht auch hier ein Unterschied hinsichtlich der genutzten Technologien zwischen den Ebenen der Verwaltung und Forschungseinrichtungen und sonstigen Institutionen (siehe Abbildung 20).

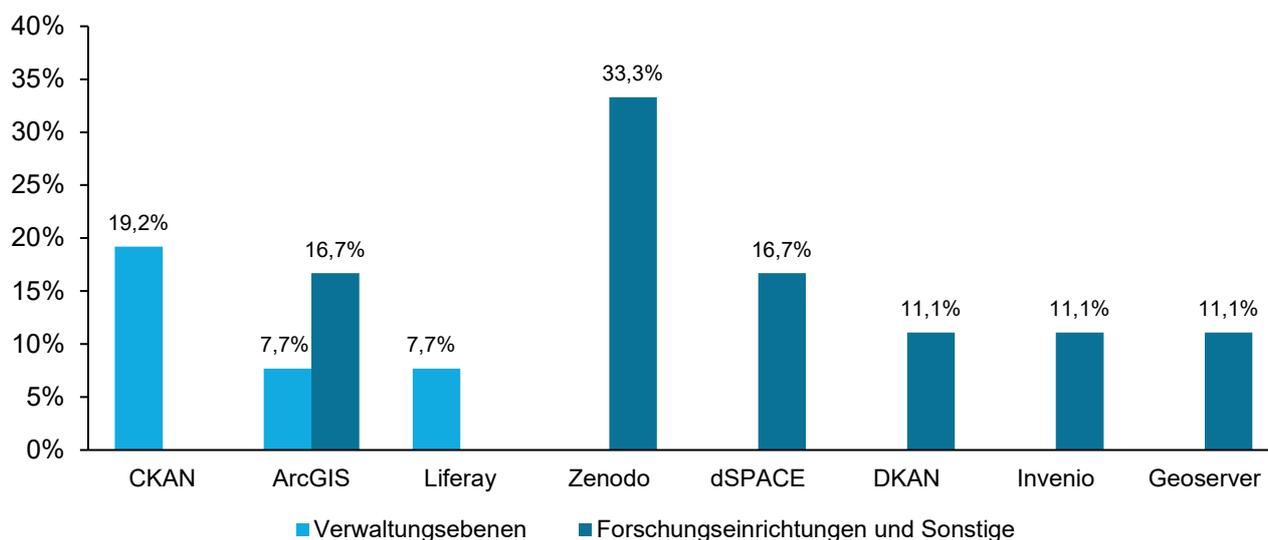


Abbildung 20: Häufigste genannte Open Data Portal-Technologien

Die Gruppen der Ebenen der Verwaltung (Bundesverwaltung, Landesverwaltung, Kommunalverwaltung) und der Forschungseinrichtungen und Sonstige geben größtenteils unterschiedliche genutzte Technologien für ihre Open Data Portale an. Mehrfachnennungen waren möglich.

2.6 Zusammenfassung

Ziel des Teilprojekts 1 war die Generierung einer Datengrundlage für sowie die Berechnung der für das Gesamtprojekt benötigten Mengengerüste. Diese dienen im Gesamtprojekt als belastbare Grundlage, anhand der die zukünftigen Anforderungen an die Systemarchitektur von GovData sowie Personalbedarfe abgeleitet werden können. Zur Erarbeitung der Mengengerüste wurden drei Methoden eingesetzt:

- Eine Datenanalyse zu den auf GovData veröffentlichten Metadatenansätzen;
- eine Online-Umfrage zur Schaffung der benötigten Datengrundlage und im Fall der Online-Umfrage zur Generierung eines Überblicks der momentan genutzte Metadatenformate, Schnittstellen und Technologien für Open Data Portale;
- Experteninterviews, um die gewonnenen Daten zu kontextualisieren, zu validieren, und tiefere Einblicke in die Open Data Landschaft zu erhalten

Die Ergebnisse umfassen die zu erwartenden zusätzlichen Metadatenansätze von 2023 bis 2025, sowie die zu erwartenden Anbindungen an GovData pro Jahr im Zeitraum 2023 bis 2025 mit zusätzlichem Fokus auf die Art der Anbindung. Im Zuge der Beantwortung der entsprechenden Fragen wurden drei Szenarien entwickelt, die

einen minimalen, einen mittleren, und einen maximalen Zuwachs prognostizieren. Das mittlere Szenario wird dabei als vorrangig betrachtet. Berechnungen werden dennoch für alle Szenarien durchgeführt. Teilprojekt 2 nutzt die Daten der zu erwartenden Anbindungen zur Berechnung des Personalbedarfs entlang der Szenarien. Diese Ergebnisse zahlen dann im Weiteren auf das in Teilprojekt 4 erstellte Kostenmodell ein.

3 Ermittlung Personalbedarf

Übergreifendes Ziel von Teilprojekt 2 ist es, den künftigen Aufwand und – daraus abgeleitet – den künftigen Personalbedarf für die Anbindung von Institutionen an das GovData-Portal mittels anerkannter und belastbarer Methoden annäherungsweise zu ermitteln.

Hierfür wurden im Rahmen einer Ist-Erhebung in unterschiedlichen Iterationsstufen die Pilotinstitutionen an das GovData-Portal angebunden und bei der Anbindung durch Beratung und technische Hilfsmittel (z.B. einen Validator zur Prüfung der Metadaten) unterstützt. Die dabei entstehenden Aufwände wurden anhand eines standardisierten Erhebungsbogens dokumentiert. Die Grundlage des methodischen Vorgehens zur Ermittlung des Personalbedarfs ist angelehnt an das Handbuch für Organisationsuntersuchungen und Personalbedarfsermittlung des Bundesministeriums des Innern und Für Heimat (nachfolgend „Organisationshandbuch“)³⁴. Das Vorgehen nach Organisationshandbuch gewährleistet den Einsatz anerkannter Verfahren und entsprechende Akzeptanz vor Kontrollinstanzen sowie eine Vergleichbarkeit der Vorgehensweisen in der gesamten Bundesverwaltung.

Das Vorgehen unterteilt sich in drei Phasen, die im Nachfolgenden näher erläutert werden: Ist-Erhebung, Ist-Analyse, und Soll-Konzeption.

3.1 Ausgangssituation

3.1.1 Aktuelle Belieferung von GovData

Datenbereitstellende Institutionen haben grundsätzlich zwei Möglichkeiten, die Metadaten zu ihren Daten auf GovData bereitzustellen: manuell per Eingabeformular oder automatisch mittels Maschine-zu-Maschine Kommunikation. Aus Gründen der Skalierbarkeit und Qualitätssicherung wird nach Möglichkeit die automatisierte Variante präferiert. Im Falle einer manuellen Bereitstellung hilft die Handreichung des Kompetenzzentrum Open Data (CCOD) des Bundesverwaltungsamtes beim Befüllen der einzelnen Datenfelder des GovData-Formulars³⁵. Aufgrund des für GovData genutzten Metadatenstandards (siehe Kapitel 3.1.2) ist für die automatische Bereitstellung Fall zu differenzieren zwischen automatischen Anbindungen, in denen auf Seiten der datenbereitstellenden Institution die Metadaten bereits im notwendigen Format vorgelegt werden können, und Anbindungen, in denen die Anforderungen erst per Mapping-Verfahren abgebildet werden müssen. Gemäß des Beschlusses 2018/30 des IT-Planungsrates müssen alle Datenbereitstellenden die Metadaten im

³⁴ Organisationshandbuch, „2.4.3 Personalbedarfsermittlung (PBE) – PBE-Leitfaden“, Bundesministerium des Innern, für Bau und Heimat / Bundesverwaltungsamt, zuletzt abgerufen am 10.08.2022, https://www.orghandbuch.de/OHB/DE/OrganisationshandbuchNEU/2_Organisationsmanagement/2_4_Ressourcen/2_4_3_Leitfaden/leitfaden-node.html.

³⁵ Bundesverwaltungsamt, „Leitfaden Metadaten“, 24. Aug. 2020, zuletzt abgerufen am 21.11.2022, https://www.bva.bund.de/SharedDocs/Downloads/DE/Behoerden/Beratung/Methoden/open_data_leitfaden_metadaten.pdf

Format DCAT-AP.de bereitstellen³⁶. Einzige Ausnahme ist hier die Geodateninfrastruktur Deutschland (GDI-DE): Aufgrund eines bilateralen Abkommens wird hier durch GovData ein Mapping der ISO-Metadaten auf DCAT-AP.de erstellt.

Bis dato müssen die Metadaten von Datenbereitstellenden aus den beteiligten Bundesländern über einen zentralen Übergabepunkt für GovData bereitgestellt werden. Ausnahme hiervon sind bislang Datenbereitstellende aus Bundesländern, die der Verwaltungsvereinbarung GovData beigetreten sind, jedoch noch kein Open Data Portal haben (z.B. ist die Stadt Rostock direkt an GovData angebunden).

In allen Fällen ist vorab der Zugang zu GovData mittels Einrichtung eines Accounts und die Prüfung der Berechtigung zum Bereitstellen von Daten per Mail an GovData zu initiieren³⁷.

3.1.2 Die Rolle von DCAT-AP.de (2.0)

DCAT-AP.de³⁸ ist die deutsche Adaption des europäischen Metadatenstandards DCAT-AP (Data Catalogue – Application Profile) zur Beschreibung von Datenkatalogen und ihren Inhalten. Ende 2016 wurde beschlossen, dass eine solche Adaption als allgemeiner Austauschstandard für Verwaltungsdaten und als Grundlage für die Datenanlieferung an GovData entwickelt werden soll. Seit Ende 2018 ist DCAT-AP.de verpflichtender Standard³⁹ für die Anlieferung von Metadaten an GovData. GovData verantwortet auch die Pflege- und Weiterentwicklung des Standards. Seit Anfang 2022 liegt DCAT-AP.de in Version 2.0 vor.

DCAT-AP basiert auf dem Resource Description Framework (RDF) des World Wide Web Consortium (W3C), dem zentralen Gremium für die Standardisierung von Internet-Technologien. Das W3C definiert für RDF den Anspruch, dass Daten durch offene Lizenzen, maschinenlesbare Formate und Strukturierung sowie offene Technikstandards letztlich in eine semantische Verbindung im Sinne des „semantic web“ überführt werden können sollten.

Zentral für DCAT-AP.de ist die Definition von „verpflichtenden“, „empfohlenen“ und „optionalen“ Attributen. Diese definieren, welche Informationen in jedem Fall anzugeben sind, welche empfohlen sind und welche optional für zusätzliche Informationen genutzt werden können. Einen Überblick über den Aufbau von DCAT-AP.de 2.0 liefert die zugehörige Spezifikation⁴⁰. Das Konventionenhandbuch⁴¹ trifft zusätzliche Regelungen für

³⁶ IT-Planungsrat, „Standardisierungsbedarf ‚Metadatenstruktur für offene Verwaltungsdaten‘“, IT-Planungsrat Beschluss 2018/30, 28. Jun. 2018, zuletzt abgerufen am 28.11.2022, <https://www.it-planungsrat.de/beschluss/beschluss-2018-30>.

³⁷ Die Prozesse der Anbindung an GovData sind im Anhang 6.7.1 näher beschrieben.

³⁸ DCAT-AP.de, „Vokabulare und Dokumente für DCAT-AP.de“, zuletzt abgerufen am 21.11.2022, <https://www.dcat-ap.de/def/>

³⁹ IT-Planungsrat, „Standardisierungsbedarf zur Metadatenstruktur für offene Verwaltungsdaten“, IT-Planungsrat Beschluss 2018/11, 16. Apr. 2018, zuletzt abgerufen am 21.11.2022, <https://www.it-planungsrat.de/beschluss/beschluss-2018-11>

⁴⁰ DCAT-AP.de, „DCAT-AP.de Spezifikation 2.0“, 28. Feb. 2022, zuletzt abgerufen am 21.11.2022, <https://www.dcat-ap.de/def/dcatde/2.0/spec/>

⁴¹ DCAT-AP.de, „DCAT-AP.de Konventionenhandbuch 2.0“, 28. Feb. 2022, zuletzt abgerufen am 21.11.2022, <https://www.dcat-ap.de/def/dcatde/2.0/implRules/>

die Bereitstellung von DCAT-AP.de-Metadaten für GovData.

3.2 Ist-Erhebung

3.2.1 Aufgabenerhebung

Um die Ermittlung eines belastbaren Personalbedarfs für die dauerhafte Wahrnehmung der Aufgaben zur technischen Anbindung von Institutionen zu ermöglichen, war zunächst die Erstellung eines vollständigen Aufgabenkataloges erforderlich, welcher die im Untersuchungsbereich anfallenden Tätigkeiten systematisch aufführt. Die Erhebung der Aufgaben erfolgte durch die unmittelbaren Prozessbeteiligten und somit Fachexperten und Fachexpertinnen des Untersuchungsbereichs.

Die Struktur des finalen Aufgabenkatalogs gliedert sich in zwei Ebenen, welche durch eine eindeutige Nummerierung sowie eine visuelle Abgrenzung dargestellt werden (siehe Abbildung 21).

- **Aufgabenbereiche (grau und einstellige Gliederungsziffer):** Die Aufgabenbereiche beschreiben die unterschiedlichen Anbindungstypen, die unterschiedliche Aufwände verursachen:
 - o Manuelle Anbindung
 - o Automatische Anbindung
 - o Automatische Anbindung, inkl. eines Mappings

Die pilotierten Institutionen wurden diesen drei Typen initial zugeordnet, bei Bedarf wurde die Zuordnung im Laufe des Projektes angepasst.

- **Aufgaben (weiß und zweistellige Gliederungsziffer):** Die Aufgaben bilden die tiefste Gliederungsebene im Aufgabenkatalog. Diese Ebene wurde zur Erhebung und Konzeption von Aufwänden, Bearbeitungsmengen und Personalbedarfen verwendet. Bei der Erstellung der Aufgabenkataloge wurde besonders darauf geachtet, dass die Granularität der Aufgaben differenziert genug ist, damit die Prozessbeteiligten die einzelnen Aufgaben voneinander abgrenzen können, jedoch auch grob genug, um einen angemessenen Aufwand für die Erhebungen zu gewährleisten.

Weitere Bestandteile des Aufgabenkatalogs sind:

- **Detaillierte Beschreibung:** Bei Bedarf zur näheren Erläuterung bzw. Abgrenzung der Aufgaben untereinander wurde eine detaillierte Beschreibung ergänzt.
- **Aufgabencharakterisierung:** Die Zuordnung der Aufgabencharakterisierung erfolgte stets auf der Ebene der Aufgaben. Es wurde unterschieden zwischen:
 - o *Den dispositiv-kreativen Aufgaben:* Hierbei handelt es sich häufig um konzeptionelle und gestalterische Aufgaben, die in ihren Bearbeitungsschritten und inhaltlichen Ausgestaltungen sehr unterschiedlich und daher kaum vergleichbar sind.
 - o *Den quantifizierbaren Aufgaben:* Quantifizierbare Aufgaben sind wiederholt anfallende Aufgaben, die auch in der Zukunft mit ähnlichen oder gleichen Bearbeitungsschritten und ähnlichen Bearbeitungszeiten ausführbar sind.

Bei Tätigkeiten, die sowohl quantifizierbare als auch dispositiv-kreative Elemente enthalten (s. g. Mischaufgaben), wurde die von beiden zutreffendere Charakterisierung ausgewählt.

- **Bezugseinheit:** Zur Zählung von bearbeiteten (Ist-Erhebung) oder zukünftig zu bearbeitenden (Soll-

Konzeption) Fällen / Vorgängen wurde für quantifizierbare Aufgaben jeweils eine Bezugseinheit – die zu zählende Einheit – definiert.

Aufgabenkatalog				
Nr.	Bezeichnung (Aufgabenbereich, Aufgabe)	Detaillierte Beschreibung (bei Bedarf)	Aufgaben- charakterisierung (dispositiv-kreativ oder quantifizierbar)	Bezugseinheit (nur bei quantifizierbaren Aufgaben)
1				
Manuelle Anbindung				
1.1	Beratung der Pilotinstitution, inklusive Vor- und Nachbereitung	Beratung per Mail, Telefon, Videokonferenz zu allen Themen im Anlegen einer Organisation in CKAN.	dispositiv-kreativ	
1.2	Accountmanagement	Anlegen des Datenbereitstelleraccounts. Erstellen und Implementieren der contributorID.	dispositiv-kreativ	
1.3	Inhaltliche Abstimmung mit GovData und Dienstleistern inklusive Vor- und Nachbereitung	Austausch per Mail, Telefon, Videokonferenz zu Inhalten des Teilprojektes zwei. Austausch innerhalb und zwischen den Dienstleistern sowie mit der Erstellung von Präsentationen, Texten, Artikeln und weiteren Unterlagen im Rahmen des	dispositiv-kreativ	
1.4	Erstellung Präsentation und weitere Unterlagen	Erstellung von größeren Protokollen im Rahmen des Teilprojektes.	quantifizierbar	Anzahl der Präsentationen / Unterlagen
1.5	Erstellung von größeren Protokollen		quantifizierbar	Anzahl der Protokolle
2				
Automatisch				
2.1	Beratung der Pilotinstitution, inklusive Vor- und Nachbereitung	Beratung per Mail, Telefon, Videokonferenz zu allen Themen im Rahmen von DataGo.	dispositiv-kreativ	
2.2	Testharvesting	Durchführen des Testharvestings und Erstellen des Testprotokolls mit Hinweisen zur Problemlösung.	quantifizierbar	Anzahl der durchgeführten Testharvestings

Abbildung 21: Exemplarische Struktur des Aufgabenkatalogs

Der Aufgabenkatalog gliedert sich in drei Aufgabenbereiche nach Art der Anbindung (automatisch, automatisch inkl. Mapping, manuell) mit jeweils korrespondierenden Aufgaben und enthält zusätzliche Informationen, wie die Nummerierung, Bezeichnung, detaillierte Beschreibung der Aufgabe, Charakterisierung der Aufgaben nach dispositiv-kreativ oder quantifizierbar und Bezugseinheit der jeweiligen quantifizierbaren Aufgaben.

3.2.2 Auswahl der Erhebungsmethodik

Ablauf und Methodik der Personalbedarfsermittlung orientierten sich an den Vorgaben des Organisationshandbuchs.

Eine anerkannte Methode zur Ermittlung des Personalbedarfs ist die Selbstaufschreibung. Bei der **Selbstaufschreibung**⁴² wird der Personalbedarf auf Basis von aktuellen Ist-Daten ermittelt bzw. erhoben. Das Verfahren wird gemäß Organisationshandbuch zur systematischen Datengewinnung von Zeiten und Mengen einzelner Aufgaben / Tätigkeiten / Prozessschritte an einem oder mehreren Arbeitsplätzen verwendet. In der Regel werden die Selbstaufschreibungen *strukturiert* (d.h. mithilfe von Aufschreibungsformularen) und entweder mit dem Fokus auf die wahrgenommenen Aufgaben der Organisation (*arbeitsplatzbezogene tägliche Selbstaufschreibung*) oder Prozesse (*prozessbezogenen Selbstaufschreibung / Laufzettelverfahren*) durchgeführt. Bei der Selbstaufschreibung handelt es sich um eine Eigenerhebung. Dabei erfassen die Beschäftigten tagesaktuell in einem definierten Erhebungszeitraum und mithilfe eines Tools (z.B. Excel-Tabelle) die in der Untersuchung erforderlichen Informationen (in der Regel Bearbeitungszeiten und -mengen pro Aufgabe). Da die Selbstaufschreibung ohne Beobachtung erfolgt, ist eine Plausibilisierung der Daten notwendig. Grundsätzlich ist die Selbstaufschreibung primär für quantifizierbare Aufgaben für die Ermittlung der mittleren Bearbeitungszeiten geeignet. Dabei ist zu berücksichtigen, dass die Aufgaben nicht nur

⁴² Organisationshandbuch, „2.4.3 Personalbedarfsermittlung (PBE) – PBE-Leitfaden“, Kapitel 11.2, zuletzt abgerufen am 17.08.2022

zählbar sind, sondern auch im Erhebungszeitraum häufig genug anfallen, um eine Repräsentativität der Erhebung zu gewährleisten.

Zur Gewährleistung der maximalen Datenaktualität und Erreichung der höchsten Datengüte wurde zur Ermittlung der Arbeitsaufwände das Verfahren der **arbeitsplatzbezogenen täglichen Selbstaufschreibung** gewählt. Dabei wurden alle Prozessbeteiligten angehalten, tagesaktuell ihre Aufwände für die anfallenden definierten Aufgaben zu erfassen und in regelmäßigen Abständen an die Projektgruppe zur Plausibilisierung zu übermitteln.

3.2.3 Durchführung der Ist-Erhebung

3.2.3.1 Prozessbeteiligte

Die Anbindung der Pilotinstitutionen wurde federführend von drei Mitarbeitenden der Geschäfts- und Koordinierungsstelle GovData geleitet, koordiniert und begleitet. Maßgeblich unterstützt wurden diese durch sieben Kräfte der externen Dienstleister init und SEITENBAU. Die Aufwandsermittlung umfasste alle Prozessbeteiligten gleichermaßen.

3.2.3.2 Initiierung einer Pilotierung

Für die Erhebung der Aufwände für Anbindungen von Institutionen an das GovData-Portal wurde eine Pilotierung initiiert, die die Anbindung von 13 Institutionen mit unterschiedlichen Anbindungstypen, die in die drei Kategorien (1) Automatisch, (2) Manuell und (3) Automatisch, inkl. Mapping, geclustert wurden.

Die Anbindung der Pilotinstitutionen, und damit die Aufwandsermittlung, startete Mitte Dezember 2021 und endete Mitte September 2022. Durch den langen Erhebungszeitraum konnte gewährleistet werden, dass der Großteil der Arbeiten für die Anbindung der Institutionen abgebildet werden und somit eine umfassende und belastbare Aufwandsermittlung stattfinden konnte. Dabei konnte die Anbindung von acht Institutionen – über alle drei Anbindungstypen – vollständig abgeschlossen werden. Die verbliebenen Aufwände wurden anhand der vorliegenden Auswertungsergebnisse mittels einer Expertenschätzung der Prozessbeteiligten erhoben. Die jeweiligen Institutionen, deren Anbindungstypen und der Status der Anbindung zum Ende der Piloterhebung sind in Tabelle 4 dargestellt.

Institution	Art der Anbindung	Status der Anbindung zum Ende der Piloterhebung
Stadt Frankfurt am Main	Automatisch	1 Testharvesting ausstehend
Stadt Freiburg im Breisgau	Automatisch	1 Testharvesting ausstehend
Stadt Oldenburg	Automatisch	Anbindung abgeschlossen
Stadt München	Automatisch	2-3 Testharvestings ausstehend
Land Niedersachsen	Automatisch	Anbindung abgeschlossen
Metropolregion Rhein-Neckar	Automatisch	Anbindung abgeschlossen
Robert-Koch-Institut	Automatisch	2-3 Testharvestings ausstehend
Stadt Augsburg	Manuell	Anbindung abgeschlossen
Stadt Nürnberg	Manuell	Anbindung abgeschlossen
Stadtwerke Osnabrück	Manuell	Anbindung abgeschlossen
Bundesanstalt für Materialprüfung	Manuell	Anbindung abgeschlossen
Open Agrar	Automatisch, inkl. Mapping	Fertigstellung Mapping + Konzeption & Umsetzung der Implementierung ausstehend
Thünen Institut - TISDAR (Geonode)	Automatisch, inkl. Mapping	Anbindung abgeschlossen

Tabelle 4: Institutionen der Piloterhebung, inkl. Art und Status der Anbindung

Im Rahmen der Piloterhebung wurden Anbindungen für 13 Institutionen durchgeführt, von denen 8 bis zum Zeitpunkt der Berichtserstellung abgeschlossen waren. 53,8% der Institutionen wurden automatisch angeunden, 30,8% manuell und 15,4% automatisch inkl. Mapping.

3.2.3.3 Vorannahmen

Bei der Konzeption des Projektes wurden mehrere Annahmen auf Basis der Erfahrungen der GKSt GovData getroffen.

Allgemein wurde angestrebt Pilotinstitutionen aus unterschiedlichen Stellen des föderalen Systems, von der Kommunal- bis zur Bundesebene für DataGo zu akquirieren. Entsprechend der erwarteten hohen Komplexität wurde ein agiles Projektvorgehen vorgeschlagen, um sich an wandelnde Anforderungen und Zuschnitte anpassen zu können. Trotz der institutionellen Vielfalt wurde grundsätzlich davon ausgegangen, alle teilnehmenden Institutionen in einem gemeinsamen Prozess durch die Anbindung zu führen, auch mit Blick auf die genannte Unterschiedlichkeit in Bezug auf die institutionelle Verortung sowie im Hinblick auf den technischen Entwicklungsstand bei der Anbindbarkeit. Nach einer ersten Einschätzung war davon auszugehen, dass die meisten Pilotinstitutionen für die automatisierte Anbindung technische Anpassungsarbeiten zur Datenbereitstellung würden vornehmen müssen. Insgesamt wurde für alle Fälle mit einem Bedarf an umfangreichen Unterstützungsleistungen in Form von Workshops, Webinaren und individueller Betreuung gerechnet. Besonderer Bedarf an unterstützenden Materialien (z.B. Handreichungen) und technischen Hilfsmitteln (z.B. Validierungsartefakten) wurde ebenfalls angenommen.

Im Vorhinein wurden verschiedene Risiken im Projektverlauf identifiziert. Hierzu zählte das für die Initiierungsphase vermutete Problem, Pilotinstitutionen nicht in ausreichender Zahl für eine aussagekräftige Untersuchung zu finden. Ein mögliches Risiko für das Projekt bestand in der Abschätzung erforderlicher Kompetenzen und Aufwände für die einzelnen Einrichtungen. Um dem entgegenzuwirken, wurden erweiterte Unterstützungsleistungen vorbereitet. Weiterhin bestand das Risiko, dass die teilnehmenden Pilotinstitutionen aufgrund nicht vorhersehbarer Faktoren wie z.B. unüblicher technischer Lösungen nicht repräsentativ für die zu erwartende Gesamtheit der in Zukunft an GovData anzubindenden Institutionen stehen könnten, und die Ergebnisse der Pilotierung somit in ihrer Aussagekraft verringert wären. Um diesem Risiko gegenzusteuern, sollte die Auswahl in Abstimmung mit dem Lenkungsausschuss von GovData getroffen werden. Zudem wurde die Möglichkeit einer Nacherhebung von Beginn an eingeplant. Im Prozess selbst wurde das Risiko einer im Einzelfall langsamer als im Projektzeitplan vorgesehen verlaufenden Anbindung verortet. Um die Planungssicherheit zu erhöhen, wurden mit den Pilotinstitutionen bereits frühzeitig terminierte Zielvereinbarungen getroffen. Gleichzeitig sollten im Rahmen der agilen Projektumsetzung im Bedarfsfall mögliche Potenziale zur Verlängerung der Projektlaufzeit genutzt werden. Akquise der Pilotinstitutionen

Die Akquise von Pilotinstitutionen erfolgte bereits ab November 2021. Ziel war es, Pilotinstitutionen zu finden, die möglichst die gesamte Bandbreite der vom DNG genannten Institutionen abbilden. Mit Blick auf den Umfang und den zeitlichen Rahmen des Projektes konnte nur eine exemplarische Auswahl getroffen werden. Insbesondere der erwartete Bedarf an zusätzlichen Anpassungen an den Softwaresystemen der Pilotinstitutionen in Kombination mit der kurzen Projekt- und Vorlaufzeit schränkte die Auswahl ein.

GovData selbst konnte eine Reihe von Institutionen ausmachen, die für das Projekt in Frage kamen. Dies waren vornehmlich Portale und Organisationen mit Anbindungswunsch, so dass insbesondere Kommunen und Landesportale schnell in ausreichendem Maße gefunden werden konnten. Schwieriger gestaltete sich die Akquise der öffentlichen Unternehmen und Organisationen aus dem Bereich Wissenschaft und Forschung. Zum einen stellen öffentliche Unternehmen i.d.R. ihre Daten bereits über kommunale Datenportale bereit und lagen damit außerhalb des Projektfokus, zum anderen ist bei Forschungsinstitutionen die Bekanntheit von Open Government Data im Allgemeinen und des GovData-Portals sowie von DCAT-AP.de im Speziellen noch sehr gering. Der Anreiz, Metadaten bei GovData als zentralem Portal bereitzustellen, schien für die Institutionen aus dem Bereich Wissenschaft und Forschung viel geringer zu sein als bei Institutionen aus der Verwaltung. Damit einhergehend war auch die Bereitschaft geringer, Metadaten mittels DCAT-AP.de bereitzustellen, da dies bei Forschungsinstitutionen kein gängiger Standard ist.

Von insgesamt 22 kontaktierten Institutionen konnten 13 für eine Teilnahme am Projekt gewonnen werden. Es handelte sich um sieben Kommunen, einen Regionalverbund, ein Bundesland sowie vier Forschungseinrichtungen. Damit wurde die ursprünglich anvisierte Anzahl von maximal zehn anzubindenden Institutionen deutlich erweitert. Mit Blick auf das im Vorhinein schwer zu bemessende Risiko, dass Institutionen innerhalb der Projektlaufzeit die Anbindung nicht schaffen könnten, wurde bewusst entschieden mit einer größeren Zahl zu starten, um möglichst valide Ergebnisse zu generieren.

3.2.3.4 Einteilung der Pilotinstitutionen

Um die Pilotinstitutionen bedarfsgerecht bei der Anbindung zu unterstützen, wurden sie gemäß der Art der geplanten Anbindung an GovData in Gruppen gegliedert. Weiterhin gab es in wenigen Fällen im Projektverlauf Neuordnungen, weil sich beispielsweise die technischen Rahmenbedingungen bei Institutionen anders als angenommen entwickelten.

Letztlich kamen so drei Gruppen mit folgender Charakteristik und Zusammensetzung zustande:

3.2.3.4.1 Gruppe „Automatische Anbindung“

Die Institutionen dieser Gruppe sollten im Rahmen des Teilprojektes als Quelle für das produktive Harvesting an GovData angeschlossen werden. Hierfür müssen die Institutionen ihre Metadaten DCAT-AP.de-konform im RDF-Format über einen URL zur Verfügung stellen.

Diese Bereitstellung der Metadaten für GovData kam im Rahmen des engen Projektzeitraumes von wenigen Monaten nur für Pilotinstitutionen in Frage, die bereits eine technische Lösung zur Bereitstellung eines Metadatenkatalogs implementiert hatte oder diese im Projektzeitraum noch realisieren konnten.

Mit vier Kommunen, einem Bundesland, einem länderübergreifender Regionalverband sowie einem Forschungsinstitut war die Gruppe mit automatischer Anbindung die größte im Rahmen von „Data Go“ (siehe Tabelle 4). Mit Blick auf die Metadatenqualität und die Skalierbarkeit ist die automatische Bereitstellung zu favorisieren, insbesondere dann, wenn die Daten durch die Institution bereits anderweitig online veröffentlicht werden.

3.2.3.4.2 Gruppe „Automatische Anbindung, inkl. Mapping“

Die Forschungsinstitutionen aus dieser Gruppe stellten eine fachlich neue Richtung für GovData dar, sollten aber technisch die Voraussetzungen für eine automatisierte Anbindung erfüllen. Eine Besonderheit dieser Gruppe war die Nutzung anderer Metadatenstandards als DCAT-AP.de, so dass bei einer Harvestinglösung zusätzlicher Aufwand zu erwarten war.

Die Bereitstellung der Metadaten verzögerte sich daher bis außerhalb des Projektzeitraums, da die technischen Gegebenheiten zwar vorhanden waren, aber die Implementierung eines Mappings einen höheren Zeitaufwand in Anspruch nahm, der im Projektzeitraum nicht umgesetzt werden konnte. Durch interne Verschiebungen bei den Gruppen waren hier die wenigsten teilnehmenden Einrichtungen vertreten (siehe Tabelle 4).

3.2.3.4.3 Gruppe „Manuelle Anbindung“

Die Institutionen dieser Gruppe (siehe Tabelle 4) zeichneten sich dadurch aus, dass bei ihnen die technischen Voraussetzungen für eine automatisierte Anbindung zum Zeitpunkt der Projektdurchführung nicht vorlagen. Für eine Anbindung an GovData war deshalb das manuelle Anlegen von Metadatenätzen über das auf der GovData-Website bereitgestellte Formular notwendig.

In allen betreffenden Fällen wurden jedoch bereits Schritte unternommen, um eine automatische Anbindung zukünftig zu gewährleisten. Die betreffenden Kommunen waren als mittelgroße Städte alle in der Konzeptions- oder Aufbauphase für eigene Strukturen für offene Verwaltungsdaten, u.a. auch im Kontext von Smart City-Bestreben. Auch die in dieser Gruppe geführte Forschungsinstitution befand sich im Aufbau eigener Strukturen,

konnte die Bedingungen für eine automatisierte Anbindung während der Projektlaufzeit letztlich jedoch nicht gewährleisten und wurde deshalb zunächst manuell angebunden.

3.2.3.5 Onboarding

Ziel der Onboarding Phase war es, die Pilotinstitutionen in das Projekt einzuführen und Zielvereinbarungen für die einzelnen Institutionen zu treffen.

Bereits bei der Akquise der potenziellen Pilotinstitutionen wurden erste Informationen zum Inhalt und Ablauf des Projektes gegeben. Dem schloss sich ein zweistündigen Kick-off-Workshops am 06.04.2022 mit Teilnehmenden aller Pilotinstitutionen in einem Online-Format an. Hier wurden sie detaillierter über das Projektvorgehen und ihre Ansprechpersonen informiert und die im Vorfeld getätigten Annahmen über inhaltliche und technische Kenntnisse sowie technische Voraussetzungen im interaktiven Arbeitsformat konsolidiert.

In der Folge wurden die Gruppen separat betreut, um die sich aus den spezifischen Kontexten ergebenden Informationsbedarfe zielgenau abzudecken und die im weiteren Prozess einlaufenden Informationen sinnvoll zu bündeln.

3.2.3.6 Beratung der Pilotinstitutionen

Die Einzeltermine mit den Pilotinstitutionen verfolgten das zentrale Ziel, die Anbindung der Pilotinstitutionen innerhalb der Projektlaufzeit zu realisieren und dabei die spezifischen Unterstützungsbedarfe der Institutionen zu berücksichtigen⁴³. Weiterhin sollten in diesem Rahmen Erkenntnisse über die bestehenden Prozesse validiert und Möglichkeiten der Optimierung herausgearbeitet werden. Die Termine fanden im Kontext von Online-Videokonferenzen statt, in denen Protokoll geführt wurde. Darüber hinaus bestand für jede Pilotinstitutionen stets die Möglichkeit zur Kommunikation via E-Mail.

In den einzelnen Gruppen verliefen diese Einzeltermine folgendermaßen:

3.2.3.6.1 Gruppe „Automatische Anbindung“

Mit den dieser Gruppe zugeordneten Institutionen wurden in Folge des Kick-Off-Workshops Einzeltermine vereinbart. In diesen wurde das Informationsmaterial zur Anmeldung bei GovData und zum Erstellen eines Datenbereitsteller-Accounts weitergegeben und die Prozessschritte zur automatischen Datenanlieferung erläutert.

Mit den Institutionen gestaltete sich der Prozess der Anbindung als unterschiedlich langwierig. Teilweise bestand mehrfacher Besprechungsbedarf. Dies lag zum einen darin begründet, dass für die bestehenden Plattformen die Ausgabefähigkeit von DCAT-AP.de erst hergestellt werden musste. Hierbei wurde von einer Institution auch angemerkt, dass eine Ausweitung der akzeptierten Metadatenformate und Schnittstellen zu einer besseren Verbreitung bei der Nutzung von GovData führen könnte.

⁴³ Weiterhin wurden in diesem Rahmen Erkenntnisse über die bestehenden Prozesse validiert und Möglichkeiten der Optimierung herausgearbeitet (siehe Kapitel 4.2.2). Da die Umsetzung dieser Optimierungen jedoch nicht Teil des Projektes war und wenn überhaupt erst mittelfristig durch die FITKO erfolgen kann, werden diese weiterhin nicht bei der Berechnung der Personalbedarfe (Kapitel 3.4) berücksichtigt.

Drei Institutionen setzten die CKAN-Software ein, jeweils eine DKAN, DUVA oder InGrid. Eine Sonderrolle nimmt das RKI ein. Dieses plant, bestehende Metadaten mit einem eigenen Skript in RDF-Daten umzuwandeln und diese dann über ihre GitHub-Repositories zu veröffentlichen.

Für die Bereitstellung von DCAT-AP.de-konformen Metadaten bei CKAN-Portalen konnte eine bereits vorhandene Extension (ckanext-dcatde) genutzt werden. Die anderen Software-Produkte verfügten über verschiedenartige Export-Schnittstellen, die angepasst werden mussten.

Dafür waren in mehreren Fällen IT-Dienstleister der Verwaltungen eng in den Prozess eingebunden. Diese hatte teilweise über die reine Anbindung hinausgehenden Besprechungsbedarf, um Möglichkeiten der Vereinfachung des zukünftigen Prozesses auf Verwaltungsseite einzuschätzen.

Besondere Bedeutung im Hinblick auf den Gesamtprozess war bereits im Vorfeld dem DCAT-AP.de-Validator zugedacht worden. Dieser wurde von den Dienstleistern]init[und SEITENBAU im Rahmen ihrer Tätigkeiten für GovData parallel zum Projekt fertiggestellt und an die beteiligten Institutionen weitergeleitet. Damit konnten diese ihre Metadatensätze eigenständig auf die Erfüllung der Anforderungen prüfen. Von diesem Angebot wurde zumeist Gebrauch gemacht, in einem Fall wurde der Validator darüber hinaus auch in die Redaktionssoftware eines kommunalen Datenportals eingebunden.

Neben dem Validator und Beratungsgesprächen wurden zur Unterstützung der Institutionen auch Testharvestings angeboten. Bei Testharvestings importiert GovData mit Unterstützung seiner Dienstleister testweise die Metadaten der Institutionen im Portal und gibt dann in einem Protokoll Rückmeldung über etwaige Nachbesserungsbedarfe. Die Testharvestings verliefen in den meisten Fällen unproblematisch, sodass es nur wenig Klärungsbedarf zu einzelnen Fehlermeldungen gab. Ebenso komplikationslos verlief der Übergang in den Regelbetrieb des Harvestings.

Im Anschluss an die automatische Anbindung wurden mit allen beteiligten Institutionen individuelle Feedback-Termine durchgeführt, in denen Fragen zur Motivation zur Teilnahme, zu den Prozessschritten und zum zukünftigen Umgang mit der GovData-Anbindung beantwortet wurden.

3.2.3.6.2 Gruppe „Automatische Anbindung, inkl. Mapping“

Bereits im Verlauf des Kick-Off-Workshops wurde deutlich, dass die Positionen und Voraussetzungen der einzelnen Institutionen sehr unterschiedlich sind. Insbesondere war das Metadatenformat DCAT-AP.de im Kontext von Forschungsdaten noch weitestgehend unbekannt. Stattdessen gibt es mehrere andere Metadatenformate, die in diesem Feld und abhängig von den jeweiligen Forschungsgebieten den Standard bieten. Die Notwendigkeit zur Erstellung eines Mappings wurde in diesem ersten Termin deutlich, eine Einigung auf ein konkretes Format konnte hier jedoch noch nicht getroffen werden. Gleichzeitig wurde während des Kick-Off-Termins zum ersten Mal der Vorschlag für eine alternative Anlieferung an GovData in Form einer Push-Schnittstelle angesprochen.

Alle in dieser Gruppe gefassten Institutionen erhielten im Verlauf von Mai und Juni einstündige Einzeltermine, die als Videokonferenz durchgeführt wurden. Ziel dieser Termine war es, eine möglichst einheitliche Mappinglösung für Forschungsdaten zur automatisierten Anlieferung an das GovData-Portal zu finden. Dabei zeigte sich, dass insbesondere die XML-Standards DataCite, Dublin Core und MARC eine hohe Verbreitung

aufweisen. Diese Erkenntnisse wurden auch durch die entsprechenden Ergebnisse aus der Umfrage bestätigt (siehe Kapitel 2.5.3). Nach einer eingehenderen Prüfung dieser drei Standards und einer weiteren Abstimmung mit den Pilotinstitutionen fiel die Entscheidung letztlich auf DataCiteXML. MARC findet vor allem im bibliothekarischen Bereich Anwendung und hat damit bei nicht bibliothekarisch betreuten Forschungsrepositorien keine Verbreitung. Dublin Core liefert einen zu unpräzisen Metadatenstand für ein konzises Mapping aus. Letztlich wird das DataCite-Format von allen Forschungseinrichtungen für das Melden von DOIs verwendet, was für einen hohen Verbreitungsgrad dieses Formats bei Forschungsinstitutionen im Allgemeinen spricht.

Die durchgeführten Einzeltermine – in der Regel 2-3 je Einrichtung – betrafen die konkrete Gestaltung der einzelnen Anbindung und variierten daher stark zwischen den einzelnen Institutionen. Das Thünen-Institut wird indirekt durch das Harvesting von GDI-DE sowie OpenAgrar miterfasst. Das Portal OpenAgrar zeichnet sich aktuell als erste Anbindung über das einzurichtende Mapping ab, da über die OAI-PMH Schnittstelle der Plattform ein Harvesting in verschiedenen Formaten, u.a. DataCiteXML, sowie eine Eingrenzung auf die relevanten Inhalte, die Forschungsdaten, möglich ist. Dabei stellte sich heraus, dass über die Plattform OpenAgrar mehrere Einrichtungen ihre Daten bereitstellen.

Zum Abschluss wurden mit den einzelnen Pilotinstitutionen standardisierte Feedbackgespräche geführt, um einen Überblick über den Ablauf der Anbindung und die noch ausstehenden Bedarfe für eine zukünftige Anbindung weiterer Forschungseinrichtungen zu gewinnen

3.2.3.6.3 Gruppe „Manuelle Anbindung“

Aus dem Kick-off-Workshop wurden für die Institutionen aus dieser Gruppe folgende Erkenntnisse mitgenommen: Für viele Institutionen gibt es Informationsbedarf zur Sinnhaftigkeit von Open Data, verbunden mit der konkreten Frage, welche Datensätze aus den nachgeordneten Dienststellen überhaupt bzw. prioritär auf GovData bereitzustellen sind. Die Bedeutung und Funktionalität von DCAT-AP.de in diesem Kontext spielte dort bislang kaum eine Rolle. In den Mittelpunkt der Untersuchung im Rahmen der Folgetermine rückte neben den Beratungen zu den aufgeworfenen Fragen zum allgemeinen Vorgehen bei der Bereitstellung Open Data insbesondere die Funktionalität und Nutzerfreundlichkeit des GovData-Formulars als Mittel zum manuellen Einpflegen von Metadaten.

Alle in dieser Gruppe gefassten Institutionen erhielten im Verlauf von Mai und Juni einstündige Einzeltermine, die als Videokonferenz durchgeführt wurden. Ziel dieser Termine war das gemeinsame manuelle Anlegen eines Metadatensatzes auf GovData, bei dem sie Schritt für Schritt durch die Datenfelder des Eingabeformulars für Metadaten begleitet und aufkommende Fragen und Unklarheiten besprochen werden konnten. In der Folge wurden die teilnehmenden Institutionen angehalten, eigenständig weitere Metadatensätze anlegen.

Um die Bereitstellung von Open Data im Rahmen von Data Go und auch darüber hinaus nachhaltig zu fördern, wurden drei Handreichungen erarbeitet. Im Rahmen einer Präsentation „DCAT-AP.de in aller Kürze“ wurden die Funktionalität des Metadatenstandards, sein Nutzen im Open Data-Kontext sowie mögliche Anwendungsszenarien beschrieben. In einer zweiten Handreichung zu „Musterdatensätze für GovData“ wurde die Frage der Auswahl von Datensätzen speziell für Kommunen thematisiert und Auswahlkriterien sowie

Beispiele präsentiert.

Eine dritte Handreichung betraf im Rahmen einer Step-by-Step-Anleitung die Beschreibung des Prozesses der Anmeldung auf der GovData-Website und die Schritte zum Datenbereitstellen auf manuellem wie automatischem Wege. Diese Handreichung wurden allen beteiligten Institutionen im Rahmen der Einladung zu den Einzelterminen zur Verfügung gestellt.

Die Einzeltermine, bei denen von den Pilotinstitutionen Metadatensätze bei GovData angelegt wurden, wurden umfassend protokolliert und für jedes Datenfeld Verständnisschwierigkeiten und Probleme beim Ausfüllen festgehalten. Auf dieser Basis wurde eine tabellarische Anforderungsanalyse erstellt. In zwei weiteren Workshop-Terminen Anfang Juli wurden Vertreter:innen der bereits in Einzelterminen befragten Institutionen mit jeweils einer Vertreterin einer Institution zusammengebracht, die länger zurückreichende Erfahrungen mit der manuellen Bereitstellung von Daten auf GovData besitzt. Die Vermittlung der „Power User:innen“ erfolgte durch das Kompetenzzentrum Open Data des Bundesverwaltungsamtes. Durch das Zusammenbringen von bisher nur im Rahmen des Projektes mit dem Formular auf GovData in Kontakt gekommenen Vertreter:innen mit „Power User:innen“ sollten die in den Einzelterminen gefundenen Problemfelder zum einen konsolidiert, und zum anderen ein offener Austausch angeregt werden, um auch über die einzelnen Datenfelder hinausgehende Problematiken zu erfassen. Die Ergebnisse wurden mit denen der Einzeltermine in einer gemeinsamen Tabelle zusammengeführt⁴⁴.

3.2.3.7 Erhebungsbogen

Für die Durchführung der arbeitsplatzbezogenen täglichen Selbstaufschreibung wurde ein standardisierter Erhebungsbogen entwickelt (Ausschnitt aus diesem siehe Abbildung 22).

Der Erhebungsbogen folgte einer festen Struktur, um die Anwenderfreundlichkeit der Erfassung sowie die spätere einheitliche Auswertung der Daten gewährleisten zu können. Auswählbar bzw. editierbar waren die gelb hinterlegten Felder:

- Datum
- Pilotinstitution
- Aufgabe
- Aufwand in Minuten
- Ggf. Anzahl abschließend bearbeiteter Fälle bei quantifizierbaren Aufgaben
- Kommentar

Sobald in einer Zeile eines der Felder befüllt wurde, färbten sich die korrespondierenden Felder so lange rot ein, bis hier ebenfalls eine entsprechende Angabe erfolgte. Somit wurde die Vollständigkeit der Erfassung gewährleistet. Die grau hinterlegten Felder

- Typ der Anbindung
- Aufgabennummer
- Tätigkeitsbeschreibung

⁴⁴ Diese bildet die Grundlage für die Analyse der Ist-Prozesse wie der Optimierung- und Automatisierungspotentiale (siehe Kapitel 4.2).

- Definition der Mengeneinheit
- Anzahl abschließend bearbeiteter Fälle bei dispositiv-kreativen Aufgaben
- Berechnete Aufwand in Stunden

befüllten bzw. berechneten sich automatisch und waren für die Nutzer:innen gesperrt, um fehlerhafte und / oder widersprüchliche Angaben zu vermeiden.

Projekt DataGo
 Teilprojekt 2: Ermittlung des Personalbedarfes für die Anbindung aller Schnittstellen und deren dauerhaften Support
 Erfassende Person: ***

Bitte befüllen Sie die editierbaren Felder

Hinweise:
 - Die **editierbaren Felder** sind durch **gelb hinterlegt**. Die **grün hinterlegten Felder** erhalten wichtige Zusatzinformationen. Diese werden **automatisch befüllt** und sind für die Bearbeitung **gesperrt**.
 - Für die Erhebung ist es wesentlich, dass die benötigten **Informationen vollständig** vorliegen. Sobald Sie eine editierbare Zelle befüllt haben, läßt sich daher alle **korrespondierenden Felder** entsprechend rot ein, damit Sie die Bearbeitung keine Information vergessen wird.
 - Bei **vielen Aufgaben** sind neben den **Aufwänden Fallzahlen** zu erfassen. Bitte beachten Sie bei der Erfassung die **zuvor definierte Mengeneinheit**, also das zu zählende Element. In der **Spalte K** Bitte erfassen Sie **ausschließlich final bearbeitete Fälle**. Falls Sie bspw. mit der Erstellung eines großen Protokolls am Montag beginnen, diese am Dienstag fortsetzen und am Freitag finalisieren, so tragen Sie Bitte am Montag und Dienstag der Fallzahl "0" ein und Freitag der Fallzahl "1".
 - **Für die Erfassung mit mehreren Institutionen gleichzeitig** (Bspw. in einer gemeinsamen Videokonferenz), so fassen Sie bitte **pro Institution eine Zeile** aus und teilen Sie den tatsächlichen Aufwand der Beratung (Bspw. 2 Stunden für eine gemeinsame Videokonferenz) durch die Anzahl der beteiligten Institutionen und tragen den errechneten einzelnen Datenwert (Bspw. 60 Minuten) in die jeweilige Zeile ein.
 - Den Aufwand in Minuten bitte möglichst genau eintragen (ausgenommen doppelte Aufwände bei Sammelberatungen).



Datum der Erhebung	Pilotinstitution	Typ der Anbindung	Aufgabe	Aufgabennummer	Tätigkeitsbeschreibung	Definition Mengeneinheit	Anzahl abschließend bearbeiteter Fälle (Bspw. Mengeneinheit: Stunden)	Aufwand in Minuten	Aufwand in Stunden	Kommentar
06.04.2022	Metropole Rhein-Neckar	Automatisch	Beratung der Pilotinstitution, inklusive Vor- und Nachbereitung	2.1	Beratung per Mail, Telefon, Videokonferenz zu allen Themen im Rahmen von DataGo.	Bitte keine Fallzahl erfassen		60	1,00	
06.04.2022	Stadt München	Automatisch	Beratung der Pilotinstitution, inklusive Vor- und Nachbereitung	2.1	Beratung per Mail, Telefon, Videokonferenz zu allen Themen im Rahmen von DataGo.	Bitte keine Fallzahl erfassen		60	1,00	
06.04.2022	Stadt Flakfurt am Main	Automatisch	Beratung der Pilotinstitution, inklusive Vor- und Nachbereitung	2.1	Beratung per Mail, Telefon, Videokonferenz zu allen Themen im Rahmen von DataGo.	Bitte keine Fallzahl erfassen		60	1,00	
07.04.2022	PKI	Automatisch_ggf	Beratung der Pilotinstitution, inklusive Vor- und Nachbereitung	3.1	Beratung per Mail, Telefon, Videokonferenz zu allen Themen im Rahmen von DataGo.	Bitte keine Fallzahl erfassen		60	1,00	
07.04.2022	Freistaat Bayern	Automatisch	Beratung der Pilotinstitution, inklusive Vor- und Nachbereitung	2.1	Beratung per Mail, Telefon, Videokonferenz zu allen Themen im Rahmen von DataGo.	Bitte keine Fallzahl erfassen		60	1,00	
07.04.2022	Stadt Augsburg	Manuell	Beratung der Pilotinstitution, inklusive Vor- und Nachbereitung	1.1	Beratung per Mail, Telefon, Videokonferenz zu allen Themen im Rahmen von DataGo.	Bitte keine Fallzahl erfassen		60	1,00	
07.04.2022	Stadt Nürnberg	Manuell	Erstellung von größeren Protokollen	1.5	Erstellung von größeren Protokollen im Rahmen des Teilprojektes.	Anzahl der Protokolle	1	120	2,00	
07.04.2022	Stadt Oldenburg	Automatisch	Beratung der Pilotinstitution, inklusive Vor- und Nachbereitung	2.1	Beratung per Mail, Telefon, Videokonferenz zu allen Themen im Rahmen von DataGo.	Bitte keine Fallzahl erfassen		45	0,75	
07.04.2022	Bundesanstalt für Materialforschung	Manuell	Beratung der Pilotinstitution, inklusive Vor- und Nachbereitung	1.1	Beratung per Mail, Telefon, Videokonferenz zu allen Themen im Rahmen von DataGo.	Bitte keine Fallzahl erfassen		45	0,75	
08.04.2022	Land Niedersachsen	Automatisch	Beratung der Pilotinstitution, inklusive Vor- und Nachbereitung	2.1	Beratung per Mail, Telefon, Videokonferenz zu allen Themen im Rahmen von DataGo.	Bitte keine Fallzahl erfassen		45	0,75	

Abbildung 22: Ist-Erhebung – Standardisierter Erhebungsbogen (Beispiel)

Zur Durchführung der täglichen Selbstaufschreibung im Rahmen der Anbindung der Pilotinstitutionen wurde ein standardisierter Erhebungsbogen, dessen Ausschnitt hier dargestellt ist, entwickelt. Mit diesem konnten Prozessbeteiligte ihre Aufwände für jeweilige Pilotinstitutionen und Aufgaben dokumentieren.

3.2.3.8 Berücksichtigung von Verteilzeiten

Alle während der Arbeitszeit anfallenden Zeiten, die nicht unmittelbar zu den Arbeitsaufgaben im engeren Sinne gehören, werden als Verteilzeiten bezeichnet. Dabei wird zwischen sachlichen und persönlichen Verteilzeiten unterschieden.⁴⁵

Bei den sachlichen Verteilzeiten wird zudem zwischen planmäßigen und unplanmäßigen Verteilzeiten unterschieden. Die **planmäßigen sachlichen Verteilzeiten** beinhalten regelmäßig wiederkehrende Tätigkeiten, während die **unplanmäßigen sachlichen Verteilzeiten** Zeiten darstellen, die aus störungsbedingten Unterbrechungen des Arbeitsablaufs resultieren.

Bei **persönlichen Verteilzeiten** handelt es sich um Unterbrechung des Arbeitsablaufs aufgrund von persönlichen Angelegenheiten (z. B. persönliche Verrichtungen sowie kurze Erholungspausen). Im Fokus der Erhebung stand die Aufwandsermittlung für die Anbindung von Institutionen, daher wurden ausschließlich Aufgaben berücksichtigt, die unmittelbar der Anbindung dienen. Planmäßig sachliche Verteilzeiten, wie mögliche Fortbildungstage einzelner Prozessbeteiligter oder Abstimmungen außerhalb der Piloterhebung, wurden explizit nicht berücksichtigt. Unplanmäßige sachliche Verteilzeiten (wie z. B. kurze IT-Störungen) und persönliche Verteilzeiten (wie z. B. kurze Pausen für persönliche Verrichtungen) sind implizit in den Bearbeitungszeiten und somit den berechneten Aufwänden enthalten.

⁴⁵ Organisationshandbuch, „2.4.3 Personalbedarfsermittlung (PBE) – PBE-Leitfaden“, Kapitel 2.4.3.5.4.3, zuletzt abgerufen am 11.10.2022

Ermittlung der Jahresarbeitszeit

Eine maßgebliche Berechnungsgröße bei der Personalbedarfsermittlung ist die Jahresarbeitszeit. Hierbei ist die „Normalarbeitskraft“ als abstrakte Größe von Bedeutung. Die Jahresarbeitszeit einer Normalarbeitskraft charakterisiert die normale zeitliche Verfügbarkeit von Beschäftigten bezogen auf ein Arbeitsjahr. Die für die Aufgabenerledigung benötigte Arbeitszeit dividiert durch die errechneten Jahresarbeitszeitminuten (JAM) einer Normalarbeitskraft ergibt den Personalbedarf für ein Jahr in Vollzeitäquivalenten (VZÄ).⁴⁶

Die Berechnung Jahresarbeitszeit wurde folgendermaßen durchgeführt (siehe Tabelle 5):

- **Berechnung der Jahresarbeitstage je Laufbahngruppe**
 - o Abzug der Wochenenden von den Gesamttagen eines Jahres
 - o Abzug der Feiertage und weitere arbeitsfreier Tage von den verbleibenden Tagen
 - Feiertage, die nicht immer an einem Wochentag stattfinden, bei denen dies jedoch möglich ist, wurden mit dem Faktor 5/7 berücksichtigt.
 - Für die Berechnung der Feiertage wurde Hessen als Referenzland gewählt, da die Anbindungen künftig durch eine Geschäfts- und Koordinierungsstelle der Föderalen IT-Kooperation (FITKO) mit Sitz in Frankfurt am Main ausgeführt werden.
- **Abzug der Ausfalltage durch Erkrankungen, Kuren etc. von den verbleibenden Tagen**
 - o Hierzu wurde die Statistik für offizielle Krankmeldungen ab drei Tagen für alle Arbeitnehmerinnen und Arbeitnehmer in Deutschland herangezogen. Die aktuellste Datenlage bezog sich auf das Jahr 2019.⁴⁷
- **Abzug der Ausfalltage durch Urlaub**
 - o Hierzu wurden Durchschnittswerte für Gesamtdeutschland der Techniker Krankenkasse herangezogen.⁴⁸
- **Berechnung der Jahresarbeitsminuten (JAM)**
 - o Die verbleibenden Tage stellen die Jahresarbeitstage einer Normalarbeitskraft dar. Mit Hilfe der durchschnittlichen Wochenarbeitszeit und der durchschnittlichen Jahresarbeitstage wurden die Jahresarbeitsstunden und -minuten berechnet.

⁴⁶ Organisationshandbuch, „2.4.3 Personalbedarfsermittlung (PBE) – PBE-Leitfaden“, Kapitel 2.4.3.5.4.7, zuletzt abgerufen am 24.08.2022

⁴⁷ Statistisches Bundesamt, „Krankheitstage je Beschäftigten pro Jahr“, 2022, Statistisches Bundesamt, zuletzt abgerufen am 24. 08. 2022, https://www.destatis.de/DE/Themen/Arbeit/Arbeitsmarkt/Qualitaet-Arbeit/Dimension-2/_Interaktiv/2-4-krankenstand.html

⁴⁸ Die Techniker, „Wer hat die meisten Urlaubstage in Deutschland?“, zuletzt abgerufen am 24.08.2022, https://www.tk-lex.tk.de/web/guest/externalcontent?_leongshared_template=HAUFEDetail&_leongshared_externalcontentid=15_PORTLET_17979109

A	Jahrestage	365,0
1	Wochenenden	
1.1	Samstage	52,0
1.2	Sonntage	52,0
B	Verbleibend	261,0
2	Arbeitsfreie Tage	
	Feiertage am Wochentag (Hamburg)	
2.1	Karfreitag	1,0
	Ostermontag	1,0
	Christi Himmelfahrt	1,0
	Pfingstmontag	1,0
	Fronleichnam	1,0
C	Verbleibend	256,0
	Feiertage, die auf einen Wochentag fallen können	
2.2	Neujahr	0,7
	Tag der Arbeit	0,7
	Tag der deutschen Einheit	0,7
	1. Weihnachtsfeiertag	0,7
	2. Weihnachtsfeiertag	0,7
D	Verbleibend	252,4
	Sonstige arbeitsfreie Tage, die auf einen Wochentag fallen können	
2.3	Heiligabend (24.12.)	0,7
	Silvester (31.12.)	0,7
	Sonstige	-
E	Verbleibend	251,0
3	Ausfalltage durch Erkrankung	
3.1.	Durchschnitt in Deutschland (offizielle Krankmeldungen ab 3 Tagen)	10,9
F	Verbleibend	240,1
4	Ausfalltage durch Urlaub	
4.1	Durchschnitt in Deutschland	28,9
G	Verbleibend	211,2
5	Arbeitszeiten einer Normalarbeitskraft	
5.1	Berechnete Jahresarbeitstage (Normalarbeitskraft)	211,2
5.2	Tägliche Arbeitszeit (Normalarbeitskraft)	8,0
5.3	Wöchentliche Arbeitszeit (Normalarbeitskraft)	40,0
	Jahresarbeitsstunden (Normalarbeitskraft)	1689,6
	Jahresarbeitsminuten (Normalarbeitskraft)	101376,0

Tabelle 5: Berechnung der Jahresarbeitszeit

Die Tabelle stellt die schrittweise Berechnung der Jahresarbeitszeit in Stunden und Minuten dar. Durch die Subtraktion von Wochenenden, arbeitsfreien Tagen und Durchschnittswerten zu Krankheit und Urlaub verbleiben effektiv 211,2 Arbeitstage, die für die Erledigung von Aufgaben eingesetzt werden können.

3.2.4 Plausibilisierung der Ist-Erhebung

Die im Rahmen der Selbstaufschreibung eingegangenen Erhebungsbögen wurden zunächst während der gesamten Erhebungsphase durch das Projektteam einer formalen Plausibilisierung unterzogen und bei Bedarf Rücksprache mit den einzelnen Prozessbeteiligten gehalten. Prüf Aspekte der formalen Plausibilisierung waren:

- Vollständigkeit der Angaben (Datum, Pilotinstitution, Aufgabe, Aufwand und ggf. Fallzahl)
- Korrekte bzw. plausible Datumsangaben (z. B. Erfassung von mehr als acht Stunden an einem Tag)
- Vollständige Erfassung von Fallzahlen
- Einheitliche Erfassung von gleichen Aufgaben / Terminen (z. B. Kick-Off-Veranstaltung mit den Pilotinstitutionen)
- Auswertung der Kommentare auf weitere Hinweise zur Erhebung

Die formal plausibilisierten Daten wurden konsolidiert und aggregiert aufbereitet. Dabei wurden die Aufwände, differenziert nach den einzelnen Institutionen, Aufgabenbereichen (Anbindungstypen) und Aufgaben, ausgewertet, und die Gesamtaufwände sowie die durchschnittlichen Aufwände pro Institution berechnet.

Im Anschluss wurden die aggregierten Ergebnisse den verantwortlichen Prozessbeteiligten im Rahmen eines Workshops präsentiert und gemeinsam diskutiert. Dabei standen die folgenden Fragestellungen im Fokus der Betrachtung:

- **Sind die erfassten Fallzahlen und berechneten mittleren Bearbeitungszeiten plausibel?**

Bei einigen der als quantifizierbar charakterisierten Aufgaben waren mehrere der Prozessbeteiligten an der Bearbeitung beteiligt, so dass die Möglichkeit bestand, dass die Fallzahl mehrmals erfasst worden ist und somit die berechnete mittlere Bearbeitungszeit verzerrt dargestellt wurden. Bei diesen Aufgaben wurde der Gesamtaufwand in die weitere Betrachtung einbezogen.

- **Ist die Schwerpunktverteilung der Aufgaben plausibel?**

Hierbei stand die Verteilung der Aufwände auf die einzelnen Aufgaben im Fokus. Es wurde analysiert, ob die Schwerpunktverteilung plausibel ist oder hieraus Buchungsfehler abzuleiten waren, die einer Korrektur bedürften.

- **Sind in der Erfassung alle Aufwände enthalten, die im Rahmen der Anbindung der 13 Institutionen relevant sind?**

Hierbei ging es um die Fragestellung, welche Aufwände bis zum Ende der Piloterhebung nicht erfasst werden konnten, da sie zum Zeitpunkt der Berichtslegung noch ausstehend waren. Um ein möglichst vollständiges und repräsentatives Ergebnis zu erhalten, wurden diese verbliebenen Aufwände durch die Prozessbeteiligten identifiziert und mittels einer Expertenschätzung auf Basis der bereits ermittelten Ergebnisse und Erkenntnisse geschätzt.

3.2.5 Ergebnisse der Ist-Erhebung

Die im Rahmen der Piloterhebung angefallenen Aufgaben für die Anbindung von Institutionen umfassten insgesamt 732,6 Stunden bzw. 0,43 VZÄ. Die meisten Aufwände wurden dabei mit insgesamt 0,19 VZÄ auf die automatischen Anbindungen der insgesamt sieben Institutionen aufgewendet, gefolgt von manuellen Anbindungen mit einem Personaleinsatz von 0,17 VZÄ für vier Institutionen. Die beiden Institutionen mit automatischer Anbindung plus Mapping erforderten 0,07 VZÄ (siehe Abbildung 23).

Wie bereits beschrieben, konnte die Anbindung bei einigen der Institutionen zum Ende der Piloterhebung nicht vollständig abgeschlossen werden. Die verbliebenen Aufwände wurden im Rahmen eines Workshops geschätzt.

Unter Berücksichtigung dieser ergänzenden Schätzung gestaltet sich das Ergebnis wie folgt:

Die Gesamtaufwände belaufen sich auf insgesamt 986 Stunden bzw. 0,58 VZÄ (siehe Abbildung 24). Die höchsten Aufwände mit insgesamt 0,22 VZÄ entfallen hier weiterhin auf die automatischen Anbindungen, die jedoch mit sieben auch die höchste Anzahl an Institutionen umfassen. Im Durchschnitt wurden für die automatische Anbindung einer Institution 53,18 Stunden bzw. 0,03 VZÄ aufgewandt. Für die beiden Institutionen, die neben der automatischen Anbindung noch ein Mapping erforderten, wurden insgesamt 0,19 VZÄ aufgewandt. Pro Institution entspricht das einem Einsatz von 162,54 Stunden bzw. 0,10 VZÄ. Die manuellen Anbindungen, die vier Institutionen umfassen, wurden mit insgesamt 0,17 VZÄ bebucht. Pro Institution wurden hier folglich 72,16 Stunden bzw. 0,04 VZÄ eingesetzt. Gemäß der ergänzten Ist-Erhebung wurden im Rahmen der Piloterhebung somit in Relation zur Anzahl der Institutionen für die automatischen Anbindungen, inkl. Mapping, die meisten Aufwände verwendet, gefolgt von den manuellen Anbindungen und zuletzt den automatischen Anbindungen.

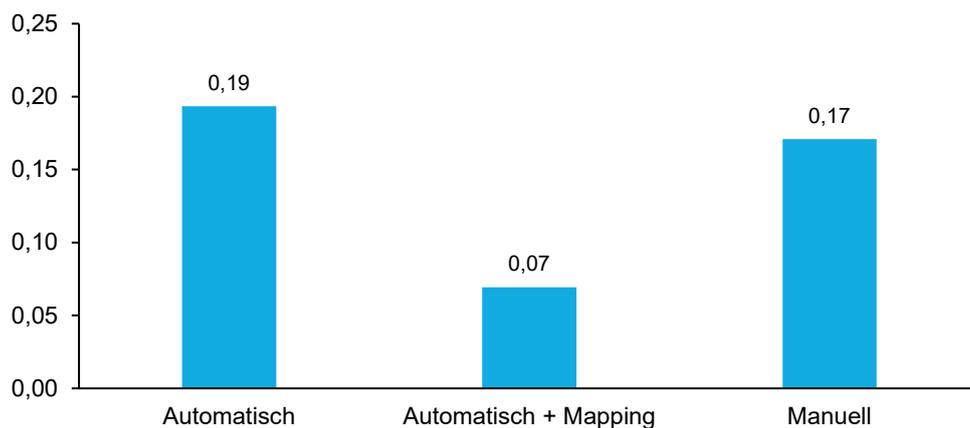


Abbildung 23: Ist-Erhebung – Personalaufwand für die Anbindung von Institutionen nach Anbindungstyp (in VZÄ)

Im Rahmen der Anbindung der Pilotinstitutionen wurde über alle Anbindungsarten ein Aufwand von insgesamt 0,43 VZÄ erhoben. Dies umfasst sowohl die abgeschlossenen als auch noch nicht abgeschlossenen Anbindungen.

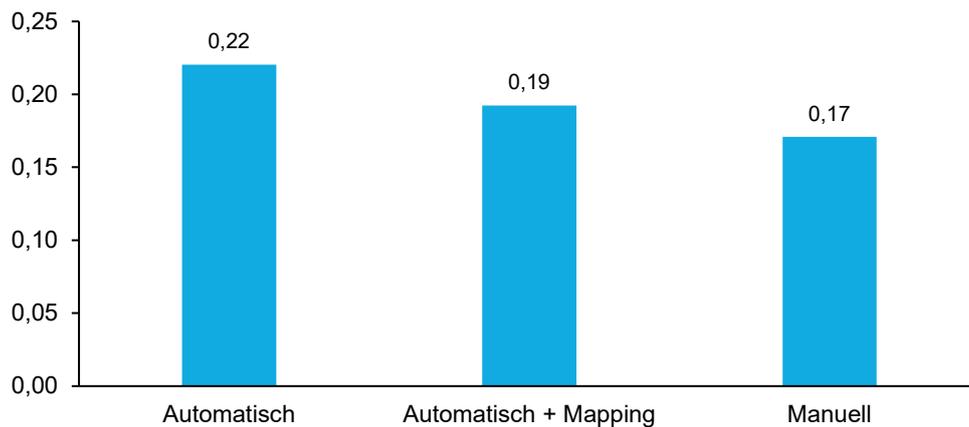


Abbildung 24: Ergänzte Ist-Erhebung – Personalaufwand für die Anbindung von Institutionen nach Anbindungstyp (in VZÄ)

Die verbleibenden Aufwände für die noch nicht abgeschlossenen Anbindungen wurden geschätzt und in der Ist-Erhebung ergänzt. Final wurden die Aufwände für die Anbindung von 13 Institutionen im Rahmen der Pilotierung auf insgesamt 0,58 VZÄ festgelegt.

3.3 Ist-Analyse

Auf Basis der Ist-Erhebung – und hier insbesondere der berechneten Aufwände pro Institution – folgte die aufgabenkritische Betrachtung. Die Aufgabenkritik dient der Identifikation von Optimierungspotenzialen hinsichtlich der Aufgabenwahrnehmung, der Ressourcenverteilung und der Ablauforganisation. Die Aufgabenkritik unterteilt sich grundsätzlich in zwei Ansätze (siehe Abbildung 25).

- **Zweckkritik:** Die Zweckkritik stellt die Frage nach der Erledigung der richtigen Aufgaben – ist es notwendig, bestimmte (Teil-)Aufgaben wahrzunehmen? Hieraus ergeben sich weitere Fragen: Erfüllen die (Teil-)Aufgaben das mit ihnen verbundene Ziel? Ist der aufgewendete personelle und finanzielle Ressourcenbedarf angemessen, um das mit der (Teil-)Aufgabe verfolgte Ziel zu erreichen? Die Zweckkritik setzt am Aufgabenkatalog an und klärt, welche Aufgaben auch weiterhin wahrgenommen werden und welche Leistungsbreite für die beizubehaltenden Aufgaben benötigt wird.
- **Vollzugskritik:** Die Vollzugskritik untersucht, ob Aufgaben, deren Notwendigkeit nicht angezweifelt wird, in Zukunft kostengünstiger, schneller und qualitativ besser wahrgenommen werden können. Die Vollzugskritik versucht folglich zu klären, ob die wahrgenommenen Aufgaben richtig erledigt werden. Diese Frage zielt damit nicht nur auf die Optimierung von Geschäftsprozessen, sondern ermittelt auch, wer für die Erbringung dieser (Teil-)Aufgabe grundsätzlich geeignet ist (z. B. unter dem Gesichtspunkt der Wirtschaftlichkeit). Im Rahmen der Vollzugskritik wird damit geklärt, ob die betrachtete (Teil-)Aufgabe zwangsläufig im Untersuchungsbereich weiterhin erbracht werden muss (Frage nach der Leistungstiefe). Bei der Frage nach der Leistungstiefe geht es nicht darum, wie intensiv eine Aufgabe wahrgenommen wird. Es wird vielmehr danach gefragt, welchen Teil einer Aufgabe der öffentliche

Sektor selbst erstellt, und in welchem Umfang Leistungen eingekauft werden.⁴⁹

Für die Durchführung der Aufgabenkritik wurde ein Expert:innen-Workshop mit den verantwortlichen Prozessbeteiligten – sowohl den internen Mitarbeitenden von GovData als auch den Mitarbeitenden der externen Dienstleister – durchgeführt.

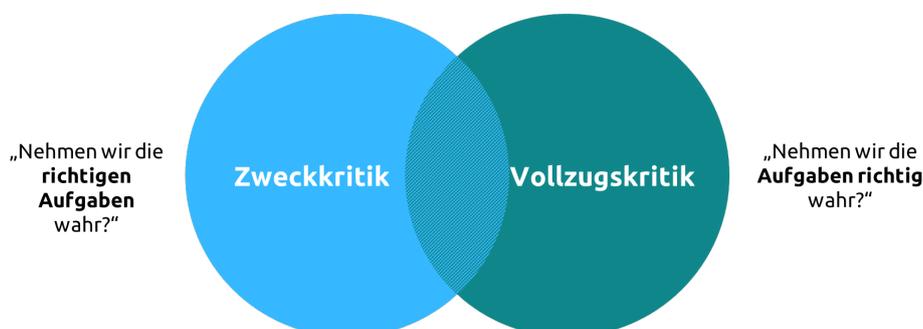


Abbildung 25: Ansätze der Aufgabenkritik gemäß Organisationshandbuch

Die Grafik zeigt die Aufteilung von Aufgabenkritik in Zweckkritik und Vollzugskritik auf. Während sich die Zweckkritik mit der Wahrnehmung der richtigen Aufgaben befasst, stellt die Vollzugskritik die Art und Weise der Wahrnehmung von Aufgaben in den Fokus.

3.3.1 Zweckkritik

Im Rahmen der Zweckkritik wurde der Aufgabenkatalog einer kritischen Betrachtung unterzogen. Dabei standen die folgenden Fragestellungen im Fokus:

- Werden die im Rahmen der Piloterhebung angefallenen Aufgaben auch in Zukunft anfallen?
- Sind außerhalb der Piloterhebung künftig neue bzw. zusätzliche Aufgaben zu erwarten, die bisher nicht Bestandteil des Aufgabenkatalogs waren?

Die wesentlichen Ergebnisse der gemeinsamen Diskussion waren, dass die Aufgabe „Inhaltliche Abstimmung mit GovData und Dienstleistern inklusive Vor- und Nachbereitung“ sich weitestgehend mit der Aufgabe „Beratung der Pilotinstitution, inklusive Vor- und Nachbereitung“ deckt und daher künftig nicht erforderlich sein wird. Der laufende Support hingegen, der erst nach der Anbindung der Institutionen anfällt, konnte im Rahmen der Piloterhebung nicht betrachtet werden und ist daher künftig als neue Aufgabe zu berücksichtigen.

3.3.2 Vollzugskritik

Im Rahmen der Vollzugskritik wurden die in der Pilotierung erhobenen Aufwände kritisch betrachtet und die Fragestellung diskutiert, ob diese künftig in der gleichen Intensität anfallen. Dabei lagen der projektbezogenen Pilotierung Rahmenbedingungen zugrunde, die sich voraussichtlich im laufenden Geschäft verändern werden:

- An der Piloterhebung nahmen insbesondere Institutionen teil, die zum einen eine hohe Motivation und zum anderen eine ausgeprägte fachliche Expertise bezüglich Open Data aufweisen. Künftig werden

⁴⁹ Organisationshandbuch, „2.4.3 Personalbedarfsermittlung (PBE) – PBE-Leitfaden“, Kapitel 3, zuletzt abgerufen am 13.10.2022

beide Aspekte vermutlich deutlich heterogener ausfallen. Dennoch wurden die einzelnen Institutionen im Rahmen des Projektes eng, individuell und persönlich begleitet (u. a. gemeinsames Ausfüllen von Formularen, Durchführung von Feedbackterminen, Klärung grundsätzlicher Fragen zum Thema Open Data ohne direkten GovData-Bezug). Im laufenden Geschäft ist diese intensive Betreuung voraussichtlich nicht erforderlich und größtenteils auch nicht leistbar. Zudem kann künftig auf den bereits entwickelten standardisierten Unterlagen und Formularen aufgebaut werden, so dass der Aufwand für die Beratung von Einzelinstitutionen tendenziell im Vergleich zu Pilotierung sinken wird.

- Innerhalb der Pilotierung wurden zahlreiche Arbeitsartefakte in Form von Präsentationen und Unterlagen (u. a. Handreichung DCAT-AP.de in a Nutshell, Step-by-Step-Anleitung für die Anmeldung bei GovData) erstellt, die für künftige Entwicklungen nachnutzbar und nur noch geringfügig an die Spezifika der Institutionen angepasst werden müssten, so dass auch hier der Aufwand voraussichtlich sinken wird.

Die Auswirkungen der sich perspektivisch verändernden Rahmenbedingungen auf den künftigen Personalbedarf wurden entsprechend der Annahmen der verantwortlichen Prozessbeteiligten quantifiziert.

Der sich ergebende prognostizierte Aufwand verteilt sich je nach Anbindung und Aufgaben unterschiedlich. Bei der automatischen Anbindung (siehe Abbildung 26) wird der Beratungsaufwand voraussichtlich auch weiterhin den höchsten Personalaufwand binden. Pro Institution werden hier 16 Stunden veranschlagt. Weiterhin sind bei den automatischen Anbindungen mit zwei bis drei Testharvestings (im Durchschnitt 2,50) pro Institution zu rechnen, die einen Aufwand von jeweils 5,50 Stunden und insgesamt 13,75 Stunden erfordern. Aufwände für den laufenden Support, der im Rahmen der Piloterhebung noch nicht angefallen ist, neu geschätzt. Die Basis für die Schätzung bildete dabei die im Vorfeld durchgeführte Personalbedarfsermittlung bei der FITKO zum Produktmanagement. Die für den laufenden Support relevanten Aufgaben wurden identifiziert und die entsprechenden Aufwände (anteilig) addiert. Die berechneten Werte wurden durch die Erfahrungswerte der GovData verifiziert. Pro Institution wurde hier dadurch ein Aufwand von 4,75 Stunden errechnet. Präsentationen und Unterlagen wurden, wie bereits beschrieben, im Rahmen der Anbindung der Pilotinstitutionen entwickelt und können künftig größtenteils wiederverwendet werden. Für die Anpassung der Dokumente an die Spezifika der jeweiligen Institutionen wurden vier Stunden veranschlagt. Ähnlich hohe Aufwände verursacht auch die Erstellung von größeren Protokollen mit 3,65 Stunden. Gemeint ist hier v. a. der Einmalaufwand für die Erweiterung der Software-Implementierung, der künftig voraussichtlich ebenfalls anfallen wird und mit der Anzahl der Institutionen verrechnet wurde. Für das Accountmanagement werden, unabhängig vom Anbindungstyp, durchschnittlich 1,30 Stunden pro Institution veranschlagt, wie auch im Rahmen der Piloterhebung ermittelt.

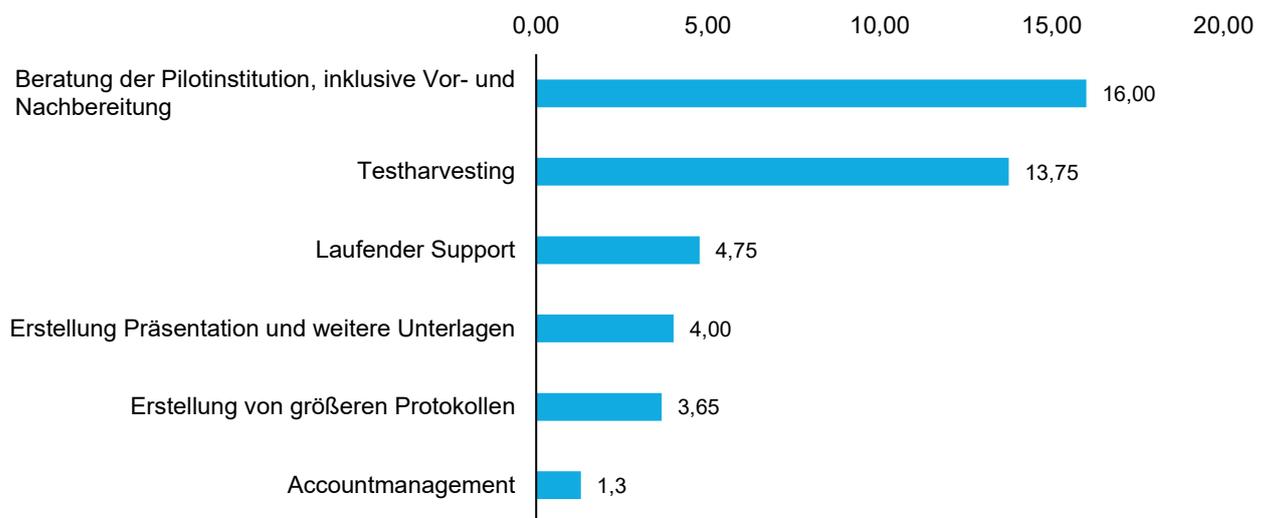


Abbildung 26: Prognostizierter Bearbeitungsaufwand für die automatische Anbindung pro Institution nach Aufgaben (in Stunden)

Pro automatische Anbindung nimmt die Beratung der Institution mit durchschnittlich 16 Stunden gemäß Prognose den größten Raum ein, gefolgt vom Testharvesting mit 13,75 Arbeitsstunden. Die restlichen anfallenden Aufgaben nehmen jeweils unter 5 Arbeitsstunden in Anspruch.

Bei den automatischen Anbindungen inkl. Mapping (siehe Abbildung 27) deckt sich der Aufwand bei dem Großteil der Aufgaben (Beratungsleistungen, Testharvesting, Erstellung von Präsentationen und weiteren Unterlagen und Accountmanagement) mit dem der automatischen Anbindungen. Für die Erstellung von größeren Protokollen wurde der Aufwand aus der Piloterhebung fortgeschrieben, da sich hier künftig keine inhaltlichen Änderungen an der Aufgabenstruktur oder -wahrnehmung erwarten lassen. Für den laufenden Support wurde auf Basis der Personalbedarfsermittlung bei der FITKO zum Produktmanagement ein Aufwand von 6,08 Stunden bzw. 365 Minuten ermittelt. Für die Erstellung von Mapping bei der Anbindung von Forschungseinrichtungen wurde ein anteiliger Zeiteinsatz von insgesamt zehn Stunden pro Institution prognostiziert. Davon entfallen durchschnittlich acht Stunden pro Institutionen auf das Aufsetzen eines Mappings (von DataCiteXML zu DCAT-AP.de) und weitere zwei Stunden auf die laufende Pflege des Mappings. Zu beachten ist dabei, dass für die Erstellung der Mapping-Software noch etwa 30 Personentage bzw. 240 Stunden anfallen werden, die zum Zeitpunkt der Berichtslegung noch nicht geleistet wurden. Dieser Zeitanteil ist in der Prognose nicht berücksichtigt, da es sich um einen einmaligen Aufwand handelt, der im Datensatz der ergänzten Ist-Erhebung bereits veranschlagt wurde.

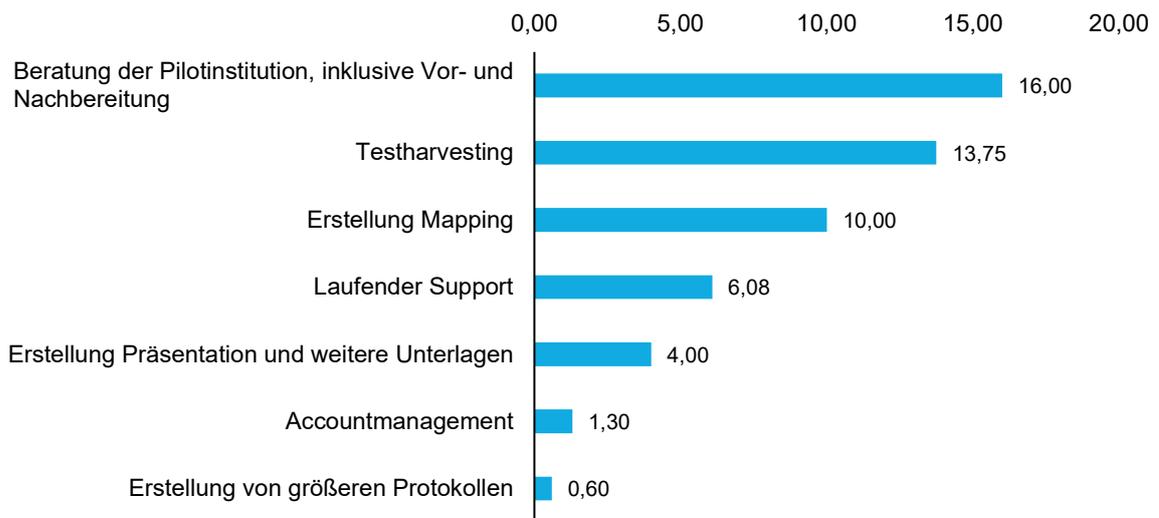


Abbildung 27: Prognostizierter Bearbeitungsaufwand für die automatische Anbindung, inkl. Mapping, pro Institution nach Aufgaben (in Stunden)

Pro automatische Anbindung nimmt die Beratung der Institution mit durchschnittlich 16 Stunden gemäß Prognose den größten Raum ein, gefolgt vom Testharvesting mit 13,75 Arbeitsstunden. Die Erstellung des Mappings bedarf im Schnitt 10,00 Arbeitsstunden und der laufende Support 6,08 Stunden. Die restlichen anfallenden Aufgaben nehmen jeweils unter 5 Arbeitsstunden in Anspruch.

Auch die manuellen Anbindungen (siehe Abbildung 28) weisen zum größten Teil die gleichen Aufwände auf wie die automatischen und die automatischen inkl. Mapping. Für den laufenden Support wurde hier auf Basis der Personalbedarfsermittlung bei der FITKO zum Produktmanagement ein Aufwand von 2,42 Stunden bzw. 145 Minuten ermittelt.

Die Aufwände pro anzubindende Institution im Gesamtergebnis sind in Abbildung 29 dargestellt. Für eine automatische Anbindung werden 43,45 Stunden benötigt, während die automatischen Anbindungen, inkl. Mapping, aufgrund der zusätzlichen Aufgabe und des aufwändigeren laufenden Supports 51,73 Stunden erfordern. Die manuelle Anbindung weist mit insgesamt 24,36 Stunden den geringsten Bearbeitungsaufwand auf.

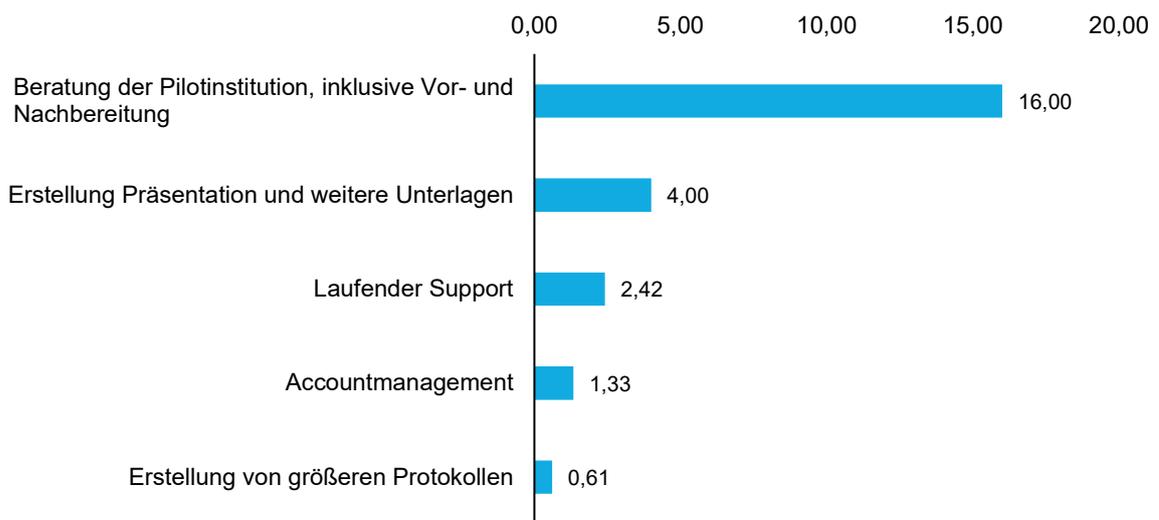


Abbildung 28: Prognostizierter Bearbeitungsaufwand für die manuelle Anbindung pro Institution nach Aufgaben (in Stunden)

Pro manuelle Anbindung nimmt die Beratung der Institution mit durchschnittlich 16 Stunden gemäß Prognose den größten Raum ein. Die restlichen anfallenden Aufgaben nehmen jeweils unter 5 Arbeitsstunden in Anspruch.

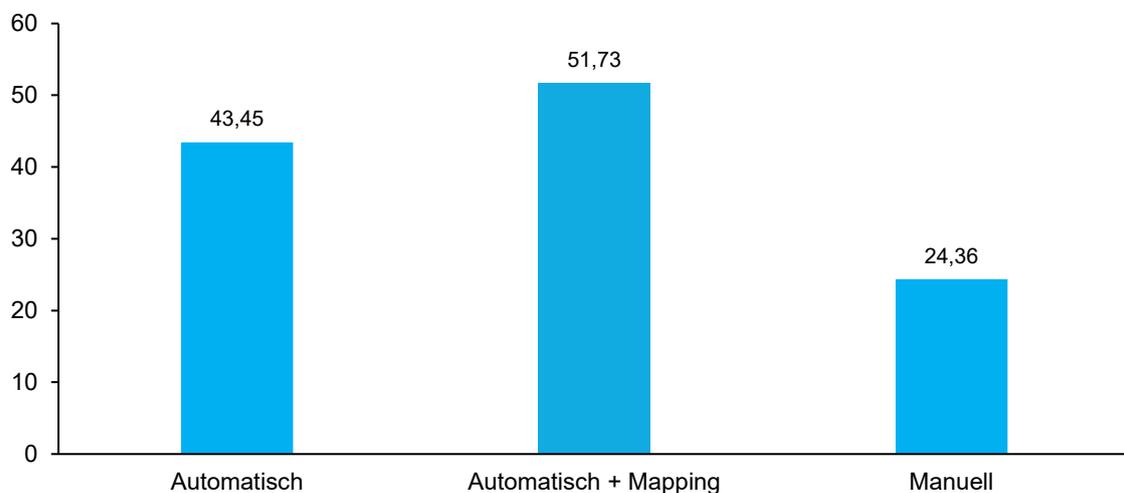


Abbildung 29: Prognostizierter Bearbeitungsaufwand pro Institution nach Anbindungstyp (in Stunden)

Insgesamt weist eine automatische Anbindung inkl. Mapping mit 51,73 Stunden gemäß Prognose den höchsten Bearbeitungsaufwand auf. Eine automatische Anbindung fällt mit 43,45 Arbeitsstunden ins Gewicht. Eine manuelle Anbindung bedarf eines Aufwands von 24,36 Arbeitsstunden.

Neben der Quantifizierung künftiger Bearbeitungsaufwände war ein weiterer Bestandteil der Vollzugskritik die Analyse, welche Aufgaben intern, durch Mitarbeitende der öffentlichen Verwaltung selbst, wahrgenommen werden können und in welchem Umfang Leistungen eingekauft werden müssen. Auf Basis der Piloterhebung

und der Erfahrungswerte der Prozessbeteiligten wurden Annahmen getroffen, die in Tabelle 6 aufgelistet sind. Der Großteil der Aufgaben kann nach Einschätzung der Prozessbeteiligten durch internes Personal der öffentlichen Verwaltung geleistet werden. Externes Personal wird insbesondere für die Durchführung von Testharvestings und die Erstellung von Mappings benötigt.

Aufgabe	Interner Anteil	Externer Anteil
<i>Automatische Anbindung</i>		
Beratung der Pilotinstitution, inklusive Vor- und Nachbereitung	80	20
Testharvesting	-	100
Accountmanagement	80	20
Laufender Support	80	20
Erstellung Präsentation und weitere Unterlagen	80	20
Erstellung von größeren Protokollen	80	20
<i>Automatische Anbindung inkl. Mapping</i>		
Beratung der Pilotinstitution, inklusive Vor- und Nachbereitung	80	20
Testharvesting	-	100
Accountmanagement	80	20
Laufender Support	80	20
Erstellung Präsentation und weitere Unterlagen	80	20
Erstellung von größeren Protokollen	80	20
Erstellung Mapping	20	80
<i>Manuelle Anbindung</i>		
Beratung der Pilotinstitution, inklusive Vor- und Nachbereitung	80	20
Accountmanagement	80	20
Laufender Support	80	20
Erstellung Präsentation und weitere Unterlagen	80	20
Erstellung von größeren Protokollen	80	20

Tabelle 6: Aufteilung der Bearbeitungsanteile für die jeweiligen Aufgaben je Anbindungstyp nach intern und extern (in %)

Die Aufgaben, die mit einer Anbindung anfallen, werden nicht alle intern durchgeführt werden können, sondern müssen zum Teil extern eingekauft werden. Basierend auf den Ergebnissen der Piloterhebung und Erfahrungswerten der Prozessbeteiligten wurde für jede Aufgabe der Anteil bestimmt, der intern bedient werden kann, und der, der voraussichtlich extern geleistet werden muss. Bei den meisten Anteilen beläuft sich das Verhältnis intern-extern auf 80-20, aber vor allem technische Aspekte können größtenteils nicht intern bedient werden, sodass bspw. das Testharvesting zu 100% von externen Dienstleistenden durchgeführt werden müsste.

3.4 Soll-Konzeption

Die Ergebnisse der aufgabenkritischen Betrachtungen bilden die Ausgangslage für die Soll-Konzeption. Auf Basis der Ergebnisse der Zweckkritik wurde zunächst der Aufgabenkatalog angepasst. Im Rahmen der vollzugskritischen Betrachtung wurden zudem die abgestimmten und plausibilisierten Bearbeitungszeiten pro Anbindung auf Ebene der einzelnen Aufgaben ergänzt. Für die Berechnung des Personalbedarfs bedarf es neben den Bearbeitungszeiten pro Anbindung bzw. Institution auch der korrespondierenden Mengengerüste, konkret der erwarteten Anzahl an anzubindenden Institutionen:

$$\frac{\text{Bearbeitungszeit in Stunden} \times \text{Anzahl der anzubindenden Institution} / \text{Soll-Jahresarbeitsstunden}}{=}$$

Personalbedarf in VZÄ

3.4.1 Ergänzung der Mengengerüste

Die Anzahl der erwarteten anzubindenden Institutionen pro Anbindungstyp wurde unmittelbar aus den Ergebnissen des TP 1 abgeleitet (vgl. Kapitel 2.5.2). Hier wurden Werte für drei Szenarien (minimal, mittel und maximal) berechnet, wobei das mittlere Szenario als maßgeblich gilt. Unter der Annahme, dass sich die Zahl der neu veröffentlichen Stellen gleichmäßig über die drei zu schätzenden Jahre 2023 bis 2025 verteilt, wurde die jeweilige Anzahl an Institutionen der drei Szenarien pro Jahr für jede Aufgabe des Aufgabenkatalogs angesetzt und mit den korrespondierenden Bearbeitungszeiten multipliziert, so dass sich pro Szenario ein gleichmäßig verteilter jährlicher Personalbedarf errechnet.

	Aufwand pro Anbindung in Stunden	Szenario		
		Minimal	Mittel	Maximal
Anbindungen		23	51	86
Automatisch	43,45	18	44	75
Automatisch inkl. Mapping	51,70	2	2	2
Manuell	24,36	2	5	9

Tabelle 7: Aufwand pro Anbindung und jährlich zu erwartende anzubindende Institutionen nach Anbindungsart und Szenario

Den Bearbeitungsaufwänden pro Anbindung, differenziert nach Anbindungsart, wird die jeweils korrespondierende geschätzte Anzahl jährlich anzubindender Institutionen gegenübergestellt. Das mittlere Szenario, von welchem zentral ausgegangen wird, ist zur Hervorhebung grau hinterlegt.

Die einzige Ausnahme bildet hier die Aufgabe zum laufenden Support. Die Anzahl der Institutionen wird mit jedem Jahr steigen, da die Institutionen, die im Vorjahr angebinden wurden, ebenfalls weiterhin der laufenden

Unterstützungsleistungen bedürfen. Daher wurde bei dieser Aufgabe für die Berechnung der Menge der Mittelwert über den gesamten Betrachtungszeitraum der drei Jahre 2023-2025 angesetzt. Die Aufwände pro Anbindung nach Anbindungstyp für jedes Szenario sind in Tabelle 7 dargestellt.

3.4.2 Ergebnisse der Soll-Konzeption

Im Ergebnis stellt sich der jährliche Personalbedarf in den jeweiligen Szenarien wie folgt dar:

Bei Betrachtung der Aufwände für jedes Szenario nach Art der Anbindung (siehe Abbildung 30) wird deutlich, dass die automatischen Anbindungen in allen drei Szenarien aufgrund der hohen erwartbaren Mengen (90% aller Anbindungen) am stärksten ins Gewicht fallen.

Differenziert nach Szenarien sind im minimalen Szenario insgesamt 0,61 VZÄ erforderlich, davon entfallen 0,51 VZÄ auf die automatischen Anbindungen, 0,07 VZÄ auf die automatischen Anbindungen inkl. Mapping, und 0,03 VZÄ auf die manuellen Anbindungen.

Für das mittlere Szenario wurde ein Personalbedarf von insgesamt 1,40 VZÄ errechnet. Die automatischen Anbindungen erfordern 1,26 VZÄ, die manuellen 0,08 VZÄ und die automatischen inkl. Mapping, 0,07 VZÄ. Der in Relation geringe Personalbedarf bei den automatischen Anbindungen inkl. Mapping, trotz des hohen Bearbeitungsaufwands pro Institution erklärt sich durch die geringe Anzahl an anzubindenden Institutionen in diesem Bereich, die pro Jahr auf zwei Anbindungen prognostiziert wurde.

Im maximalen Szenario fällt ein Personalbedarf von insgesamt 2,35 VZÄ an, davon entfallen 2,14 VZÄ auf die automatischen Anbindungen, 0,14 VZÄ auf die manuellen Anbindungen und 0,07 VZÄ auf die automatischen Anbindungen inkl. Mapping.

Ein Großteil der Personalbedarfe (60 %) kann durch internes – der öffentlichen Verwaltung zugehöriges – Personal gedeckt werden. Bei den übrigen 40 % der Personalbedarfe sind voraussichtlich externe Dienstleistern hinzuzuziehen (siehe Abbildung 31).

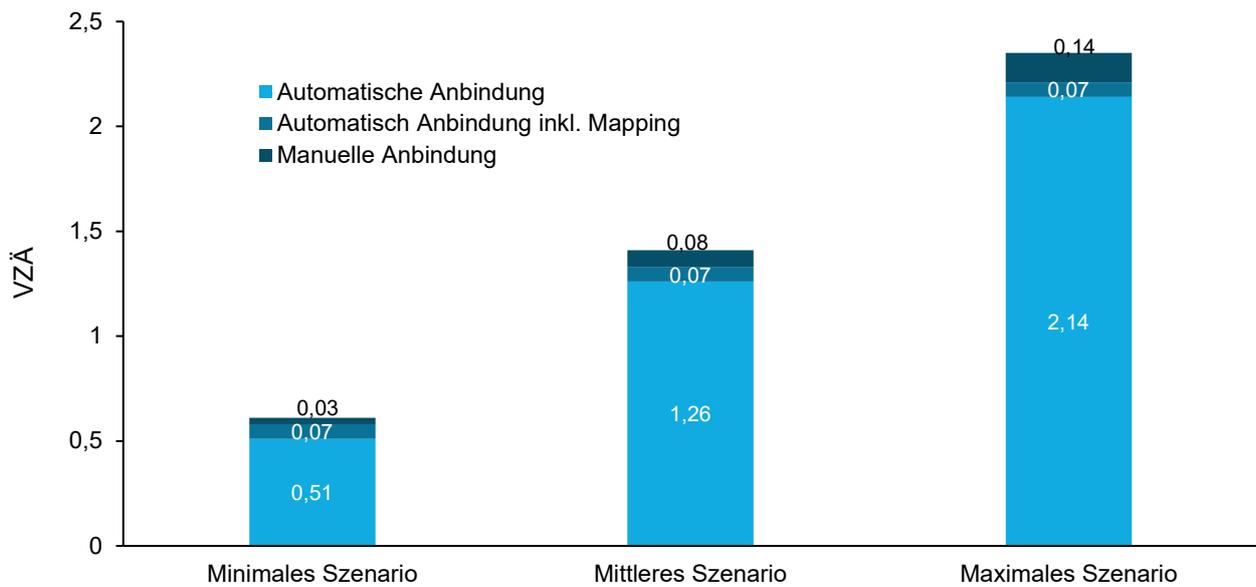


Abbildung 30: Prognostizierter jährlicher Personalbedarf (2023-2025) nach Art der Anbindung und Szenario (in VZÄ)

Der prognostizierte jährliche Personalbedarf variiert je nach Szenario, da hier auch jeweils unterschiedlich viele anzubindende Institutionen erwartet werden. Für die Begründung des Personalbedarfs wird das mittlere Szenario verwendet, während das minimale und maximale die Spanne darstellen. Unterschiede bei den Summen zwischen Text und Grafik (z. B. mittleres Szenario: 1,40 vs. 1,41) kommen durch Rundungen zustande.

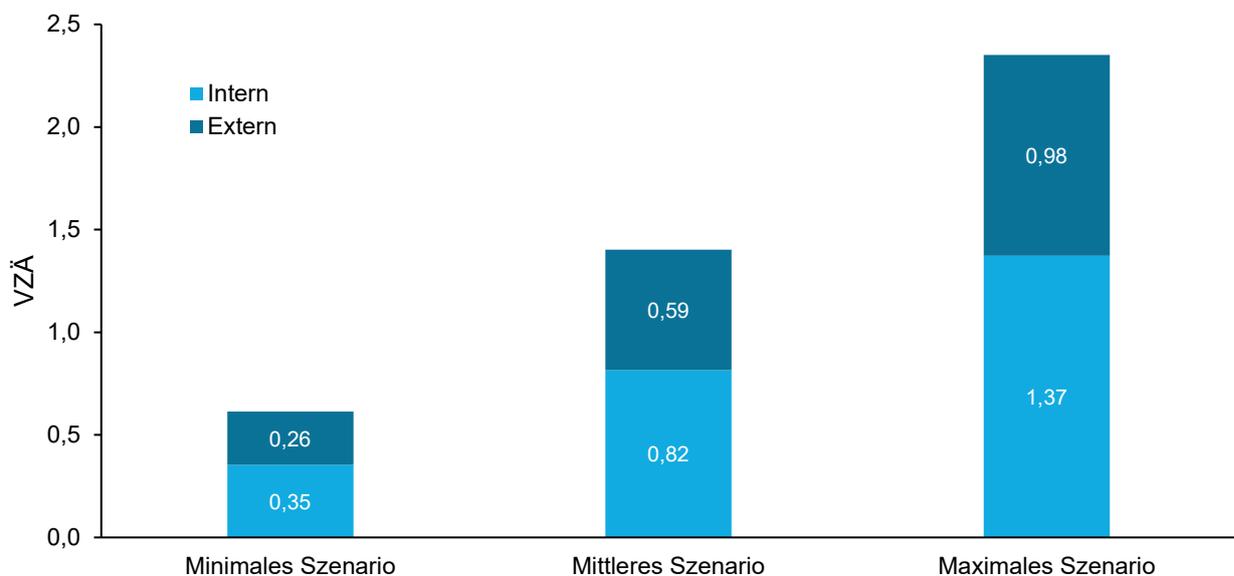


Abbildung 31: Prognostizierter jährlicher Personalbedarf (2023-2025) nach internen und externen Bedarfen (in VZÄ)

Insgesamt können ca. 60% der Aufgaben intern bedient werden. Unterschiede bei den Summen zwischen Text und Grafik (z. B. mittleres Szenario: 1,40 vs. 1,41) kommen durch Rundungen zustande.

4 Planung Systemarchitektur

Dieses Kapitel teilt sich auf in zwei Abschnitte in denen die Weiterentwicklung von GovData im Mittelpunkt steht. Der folgende Abschnitt 4.1 ist das Ergebnis intensiver Untersuchungen der technischen Infrastruktur im Rahmen des Teilprojektes 3, auf deren Grundlage eine optimierte Systemarchitektur für die Anforderungen der kommenden Jahre entwickelt wurde.

Im Abschnitt 4.2 werden dagegen auf Grundlage der Arbeiten im Teilprojekt 2 konkrete Ergebnisse dargestellt, die dazu beitragen, die Datenbereitstellung organisatorisch zu verbessern.

Die hier vorliegende Berichtsfassung enthält aus wettbewerbsrechtlichen Gründen eine gekürzte Form der Untersuchungs- und Bewertungsergebnisse.

4.1 Planung der technischen Systemarchitektur

4.1.1 Ziel von Teilprojekt 3

Die Ergebnisse der Untersuchungen aus Teilprojekt 1 zeigen, dass in den kommenden zwei Jahren die Zahl der datenbereitstellenden Institutionen und damit auch die Anzahl der zu verarbeitenden Datensätze bei GovData erheblich steigen werden. Vor diesem Hintergrund muss die aktuelle Systemarchitektur überdacht und damit sichergestellt werden, dass GovData auch zukünftig diesen Datenmengen gewachsen ist. Die Aufgabe des Teilprojektes 3 ist es zu prüfen, ob die Komponenten der bestehenden Infrastruktur den zu erwartenden neuen Rahmenbedingungen entsprechen oder aber durch besser geeignete Module ersetzt werden müssen. Zu diesem Zweck wurden sowohl eine Schwellenwertanalyse in Bezug auf die Performance des aktuellen Systems vorgenommen, als auch alternative Anwendungen untersucht, um mögliche Kandidaten für eine neue Systemarchitektur zu ermitteln. Das Ziel dieses Teilprojektes ist es, eine optimierte Systemarchitektur für GovData zu entwerfen, welche den Anforderungen der kommenden Jahre gewachsen ist.

4.1.2 Aktuelle Systemarchitektur

Als Basis für das Portal wird Liferay 7.4 als Portlet-Software und Content Management System (CMS) verwendet. In das Liferay Portal sind Suchfunktionen sowie Bearbeitungsformulare für Metadatenätze und Showcases integriert. Diese Funktionen kommunizieren über eine API mit dem Backend, womit Datensätze und Showcases im Frontend (im Web-Browser auf dem Client) geladen, editiert und wieder gespeichert werden können.

Für den Datenkatalog wird CKAN 2.9 eingesetzt. CKAN bietet neben einer Benutzeroberfläche und einer API auch ein Harvesting von Datenquellen. Über die CKAN-UI werden aktuell die Zuordnung von Datenbereitsteller-Konten zu Organisationen sowie die Aktualisierung von Datensatz-Kategorien vorgenommen. Über die CKAN-API können Datensätze abgefragt und geändert werden. Die Harvester können externe Datenendpunkte abfragen und deren Daten in CKAN abspeichern und aktualisieren. Die Daten aus dem Katalog werden in einem Elasticsearch-Suchindex gehalten, um die Zugriffszeiten aus dem Portal zu optimieren. Des Weiteren existiert

ein Apache Fuseki Triplestore, der die Daten parallel zum CKAN als Triples bereitstellt. Diese Triplestore-Daten sind für Nutzer über einen SPARQL-Endpunkt im Frontend abrufbar. Neben den Metadaten befinden sich auch die Validierungsergebnisse der Metadatensätze in einem eigenen Graphen im Triplestore. Diese werden mit dem ITB-SHACL-Validator generiert. Eine Validierung von Metadaten findet immer dann statt, wenn ein Datensatz sich ändert und im Backend gespeichert oder aktualisiert wird.

Der Ist-Zustand der Systemarchitektur von GovData ist in Abbildung 32 vereinfacht dargestellt. Für eine ausführliche Beschreibung der Architektur siehe Anhang 7.8.

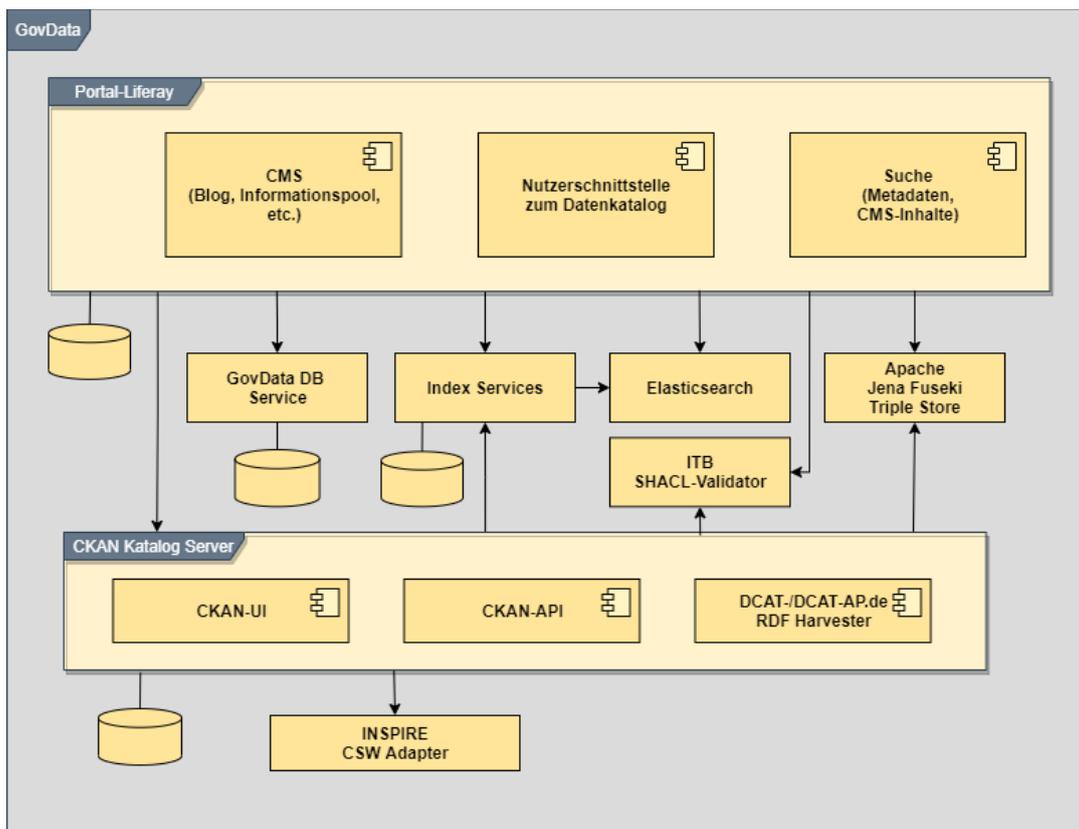


Abbildung 32: Aktuelle Systemarchitektur von GovData

4.1.2.1 Erfahrungen mit der aktuellen Systemarchitektur

Von den vorhandenen Komponenten kann im Rahmen der Neustrukturierung jede Komponente ausgetauscht werden, für die eine besser geeignete Option gefunden wird.

Die Erfahrungen im Betrieb von GovData haben gezeigt, dass das zur Zeit verwendete CMS-System nicht mehr zeitgemäß ist und zunehmend in der Entwicklung vernachlässigt wird.

Apache Jena Fuseki, der ITB-SHACL-Validator, der GovData-DB-Service sowie der CKAN sind den Erfahrungen von GovData nach unter den bisher geltenden Bedingungen performante, stabile und gut wartbare Anwendungen. Eine Einschränkung ergibt sich dabei in Bezug auf das CKAN Harvesting, welches bereits zum aktuellen Stand mit ~70.000 Datensätzen im GovData-Portal einen zu geringen Durchsatz für eine schnelle Aktualisierung des Gesamtdatenbestandes bietet.

4.1.3 Schwellenwertanalyse des aktuellen Systems

4.1.3.1 Ziel der Schwellenwertanalyse

Da das GovData-Portal in Zukunft deutlich mehr datenbereitstellende Institutionen und damit zu verarbeitende Datensätze erwartet, muss die aktuelle Infrastruktur untersucht und herausgefunden werden, ob diese den steigenden Anforderungen gewachsen ist. Nach den aktuellen Erkenntnissen aus dem Teilprojekt 1 muss die Infrastruktur für eine mittlere sechsstellige Anzahl an Datensätzen geeignet sein.

Ein Indikator dieser Eignung ist die performante Bedienbarkeit aller kritischen Funktionalitäten im Portal, wie beispielsweise das Suchen und Anzeigen der Metadatenätze.

Da etwa 97% der bisher bei GovData veröffentlichten Datensätze automatisiert über das sogenannte Harvesting importiert wird, ist zu erwarten, dass dies auch größtenteils bei der Anbindung neuer Datenbereitstellender der Fall sein wird. Es soll daher auch analysiert werden, wie sich die wachsende Anzahl von Endpunkten und Datensätzen auf die Gesamtdauer des Harvestings auswirkt. Da die Datenhaltung in CKAN erfolgt, werden auch die dafür standardmäßigen Erweiterungen `ckanext-harvest`⁵⁰ (Basis der Harvesting-Funktionalität) und `ckanext-dcat`⁵¹ (Import von RDF-Endpunkten) eingesetzt. Die Implementierung des Harvestings nutzt zwar eine Queue für die Aktualisierung der einzelnen Datensätze, diese wird jedoch ausschließlich sequentiell abgearbeitet. Auch das Einlesen von Daten aus den Harvestingquellen erfolgt seriell. Lediglich das Einlesen der Daten der Harvestingquellen und die Aktualisierung der einzelnen Datensätze findet parallel statt. Es ist daher anzunehmen, dass durch die steigende Anzahl an Harvestingquellen, und der damit verbundenen höheren Anzahl zu importierender Datensätze auch die Dauer für das Harvesting aller Endpunkte ansteigen wird. Dadurch könnten die Harvester möglicherweise nicht mehr so häufig ausgeführt werden wie gewünscht. Das Ziel ist bisher, dass die Aktualisierung aller angebotenen Endpunkte innerhalb von 48 Stunden möglich sein soll. Deshalb wurde geprüft, welche Optimierungen möglich sind, um die aktuelle Harvestingfrequenz von zwei Tagen einhalten zu können.

Wie im Folgenden beschrieben, wurden die ermittelten Indikatoren im Verlaufe einer sukzessiven Erhöhung der Anzahl an Datensätzen im Portal überprüft. Die ermittelten Grenzen stellen einen wesentlichen Aspekt für die Planung der zukünftigen Infrastruktur dar.

4.1.3.2 Beschreibung der Analyseumgebung

Für die Performancetests wurde eine Testumgebung genutzt, die sich zur Produktionsumgebung in einigen Punkten unterscheidet. Die Produktionsumgebung ist ein verteiltes System, bei dem die Anwendungen der verschiedenen Zonen (App, Service, Daten) auf eigene virtuelle Server (umgesetzt als virtuelle Maschine (VM)) aufgeteilt wurden. Darüber hinaus existiert eine VM für den Logging-Server.

Im Gegensatz zur verteilten Umgebung der Produktionsumgebung wurden die Messungen auf einer einzelnen

⁵⁰ CKAN, `ckanext-harvest`, GitHub, zuletzt abgerufen am 25.11.2022, <https://github.com/ckan/ckanext-harvest>

⁵¹ CKAN, `ckanext-dcat`, GitHub, zuletzt abgerufen am 25.11.2022, <https://github.com/ckan/ckanext-dcat>

VM ausgeführt, auf der alle Anwendungen installiert wurden. Die Testumgebung besitzt drei virtuelle CPUs sowie 24GB RAM. Die VM für die Ausführung der Tests liegt auf einem ESX-Server, auf dem sich noch andere VMs befinden, die Last erzeugen und die Leistung beeinflussen können. Da die Produktionsumgebung auch auf einer virtualisierten Umgebung betrieben wird, deren Hardware ebenfalls von anderen Systemen mitgenutzt wird, wurde die Annahme getroffen, dass der Overhead für den Netzwerkverkehr in der verteilten Umgebung zwischen den VMs, die stärkere Belastung der einzelnen VM des Testsystems ausgleicht.

4.1.3.3 Verwendete Tools

Apache JMeter

Apache JMeter ist eine Open Source Software zum Erstellen und Ausführen von Lasttests für Web Anwendungen. In der Software können Testabläufe definiert werden, wie auf eine Anwendung zugegriffen werden soll. Unser Testablauf besteht aus einfachen Suchanfragen im Portal, mit angemeldetem und nicht angemeldetem User (Gast), sowie Anfragen an die CKAN-API. Die Anfragen im Testablauf wurden in zwanzig parallelen Threads ausgeführt, welche die gleichzeitige Nutzung von zwanzig Anwendern simulieren sollten. Dieser Vorgang wurde je fünf Mal wiederholt, um temporäre Performanceschwankungen auf der Umgebung ausgleichen zu können.

Datenerzeugungs-Skript

Für die Generierung neuer Datensätze wurde ein Python-Skript gebaut, welches über die CKAN-API neue Datensätze anlegt. Dabei wird für jeden Datensatz ein randomisierter String erzeugt, welcher als Identifier verwendet wird. So kann sichergestellt werden, dass jeder Metadatenatz einzigartig ist und kein neuer Datensatz einen bestehenden überschreibt. Die Testdatensätze beinhalten Werte in den meisten Metadatenatz-Feldern, so dass sie eine Repräsentation eines durchschnittlichen Metadatenatz im Portal sind. Über die CKAN-API können zwei bis drei Datensätze pro Sekunde angelegt werden. Dadurch liegen die Metadatenätze in CKAN vor, aber noch nicht im Triplestore. Die Erzeugung der Daten im Triplestore und die Analyse der Auswirkungen auf die Performance wird in Abschnitt 784.1.3.5 behandelt.

4.1.3.4 Durchführung der Schwellenwertanalyse

Versuchsablauf

Nach dem Aufsetzen einer Testumgebung wurden über ein Python-Skript schrittweise je 100.000 Metadatenätze über die CKAN-API eingespielt. Nach jedem Schritt wurden auf der Umgebung verschiedene Tests ausgeführt und währenddessen die Systemauslastung notiert. Zunächst erfolgte eine manuelle Testung der Portal- und CKAN-Backend-Funktionen. Es wurden Metadatenätze erzeugt, editiert und gelöscht, um zu verifizieren, dass die Anwendung in angemessener Zeit reagiert. Anschließend fanden jeweils die automatisierten JMeter-Tests statt. Diese simulieren Zugriffe von 20 Benutzern zeitgleich auf das Portal und die CKAN-API. 20 parallele aktive Nutzer sind eine Anzahl, die in der Praxis im GovData-Portal für gewöhnlich nicht erreicht wird. Dadurch kann überprüft werden, ob die Umgebung auch bei einem hohen Nutzeraufkommen noch zuverlässig reagiert. Wenn die Umgebung nach einem erfolgreichen Durchlauf der beschriebenen Tests kein ungewöhnliches Verhalten gezeigt hat, wurden die nächsten 100.000 Datensätze eingespielt und der Ablauf wieder von vorne begonnen. Dieser Vorgang wurde so lange wiederholt, bis spürbare Performance-Einbußen in der Infrastruktur zu beobachten waren. Den soeben beschriebenen Ablauf zeigt Abbildung 33.

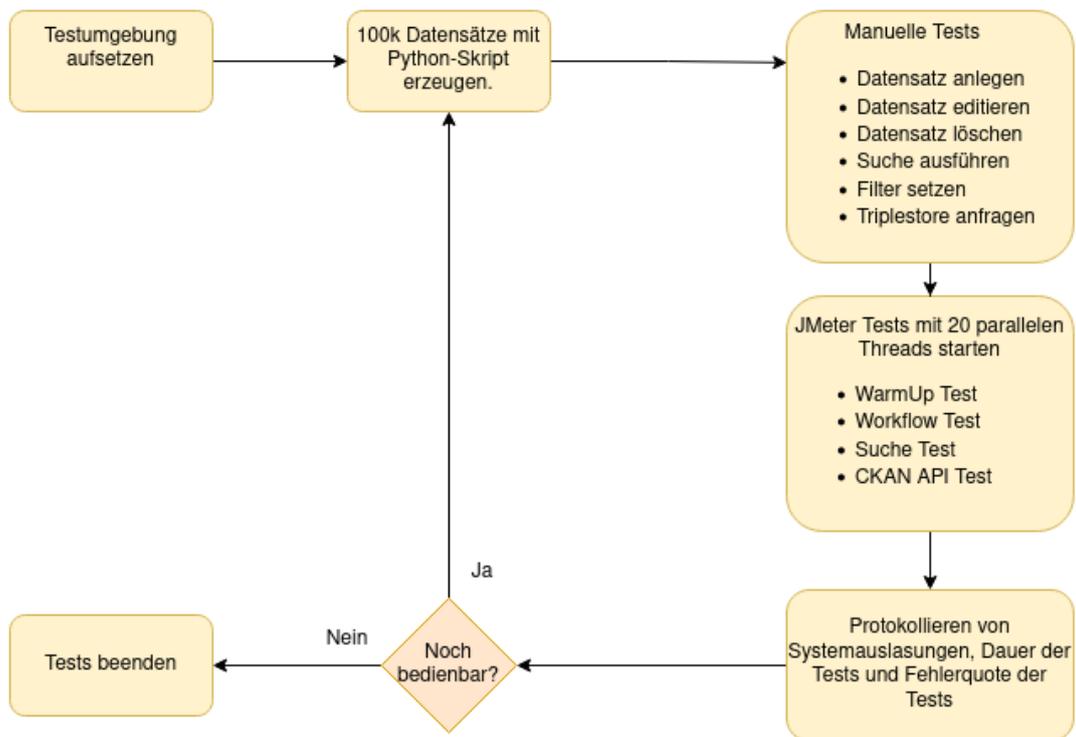


Abbildung 33: Ablaufplan der Schwellenwertanalyse

Die in der Grafik genannten JMeter Testgruppen enthalten die folgenden Abfragen:

- WarmUp Test: Es werden die meisten in den anderen Testgruppen später genutzten Seiten aufgerufen, um ein eventuelles Caching zu forcieren, so dass die Caches bereits aufgebaut sind.
- Suche Test: Es werden als Gast verschiedene Suchen (mit und ohne Filterung) im Portal ausgeführt.
- Workflow Test: Die Aufrufe werden mit einem im Portal eingeloggten Benutzer ausgeführt. Bei den ausgeführten Tests handelt es sich um die gleichen Tests wie in der Testgruppe „Suche Test“. Darüber hinaus werden noch einzelne Informationsseiten, Metadaten-Detailansichten und die Blogliste aufgerufen.
- CKAN API Test: An die CKAN-API wird eine Abfrage mit einer leeren Suche, zur Ausgabe aller Informationen zu einem Datensatz, für die erste Seite des RDF-Endpunkts und für die Liste aller Kategorien ausgeführt.

Ende der Messungen

Nachdem auf der Testumgebung 2.300.000 Metadatensätze eingespielt waren (siehe Abbildung 34), wurde das Experiment beendet, obwohl keine erheblichen Performance-Einbußen zu beobachten waren. Der angestrebte Wert einer mittleren sechsstelligen Datensatzanzahl wurde damit um ein Weites übertroffen wurde und der Beweis wurde erbracht, dass die Umgebung in Bezug auf diesen Indikator für die Verarbeitung der zukünftig erwarteten Datensatzmengen geeignet ist.

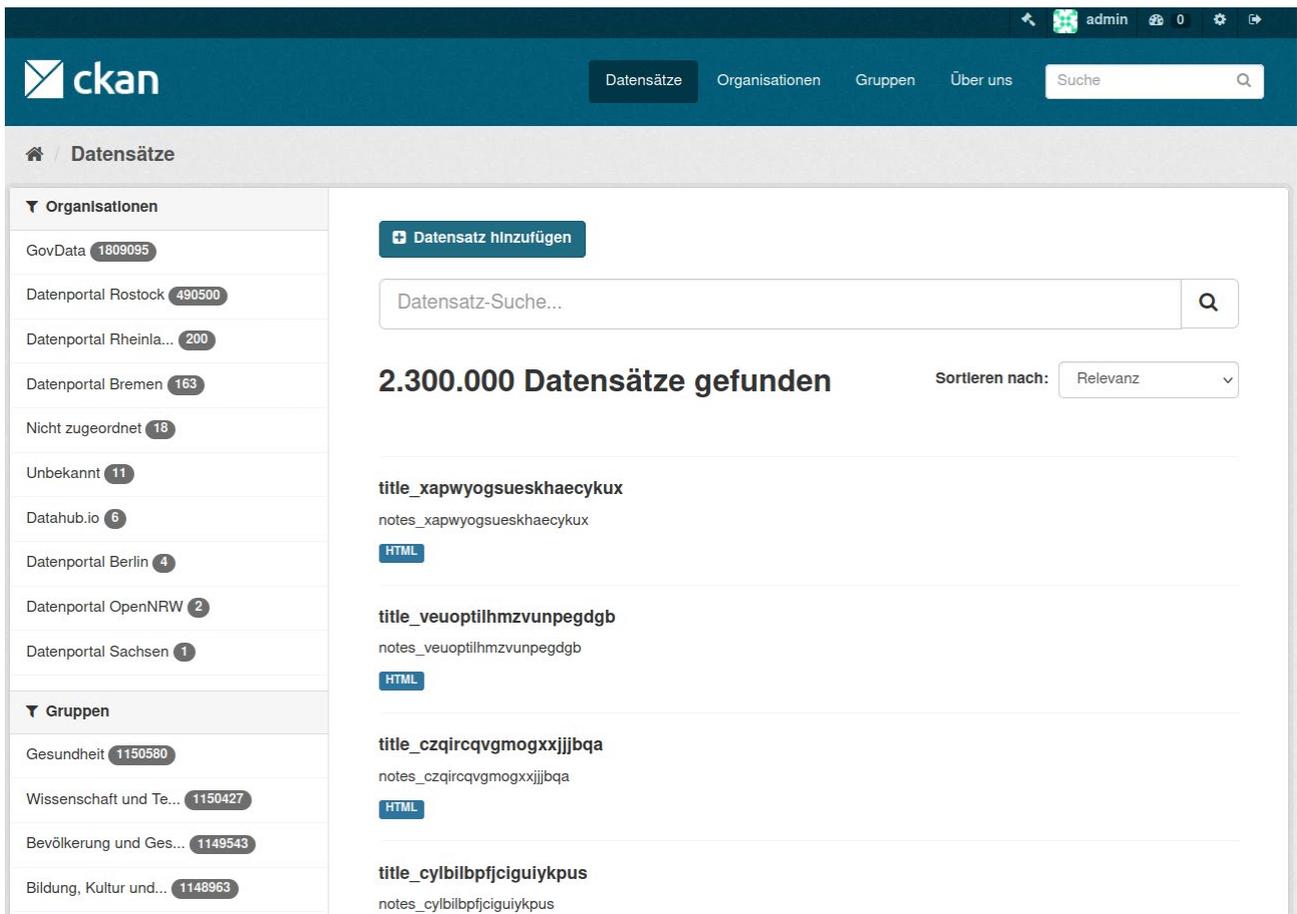


Abbildung 34: CKAN Backend mit 2,3 Millionen Datensätzen

Ergebnisse

Beim Erzeugen der Metadatenätze fällt auf, dass die Dauer zum Erstellen neuer Datensätze kontinuierlich

ansteigt (siehe Abbildung 35). Für die ersten 100.000 Datensätze wurden 10 Stunden zur Generierung benötigt. Bis 1.200.000 Datensätzen steigt dieser Wert nur gering bis auf ca. 11 Stunden an. Ab dann wird das Erstellen neuer Datensätze mit jedem Schritt deutlich langsamer. Für die letzten 100.000 Datensätze wurde bereits über 14 Stunden benötigt. Dieser Effekt ist auch beim manuellen Testen der Umgebung festzustellen. Das Anlegen neuer Metadatenätze wird mit zunehmender Datenanzahl langsamer.

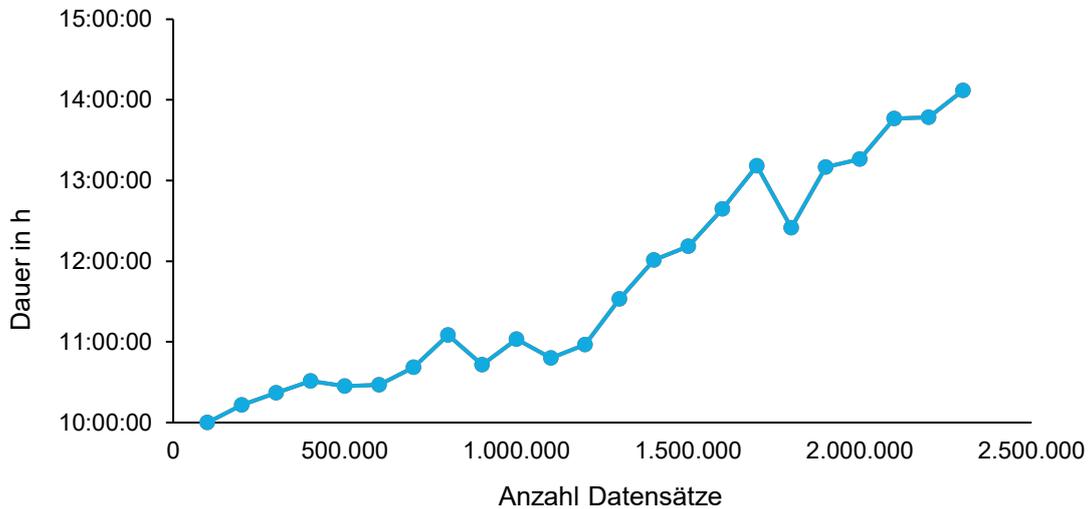


Abbildung 35: Erzeugungsdauer für neue Test-Datensätze (je 100.000)

Der Festplattenbedarf zum Abspeichern der Daten in CKAN steigt konstant mit ca. 2,3 GB pro 100.000 Datensätzen an. Bei 2.300.000 Datensätzen werden bereits über 50 GB benötigt (siehe Abbildung 36).

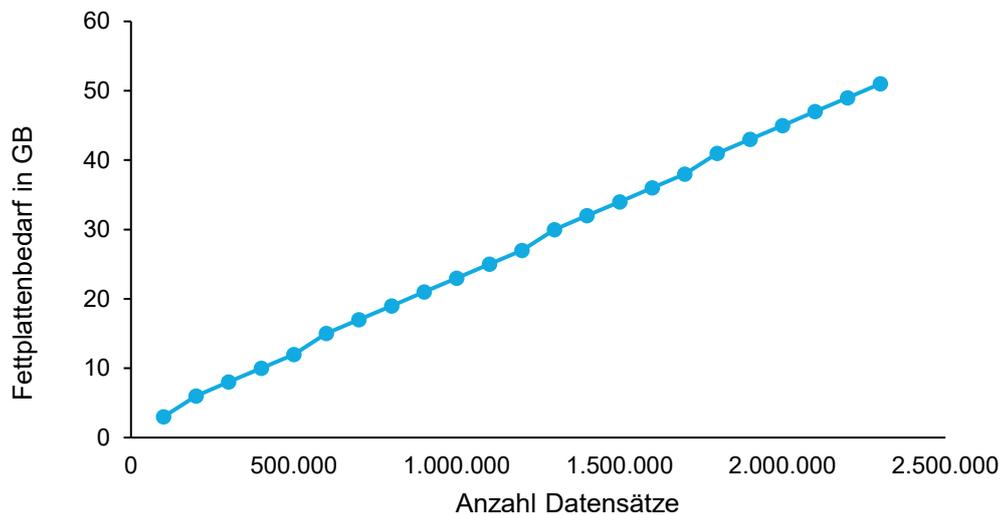


Abbildung 36: Festplattenbedarf von CKAN

Die Auslastung der 3 CPUs beim Ausführen der JMeter Tests ist sehr hoch. Bereits ab 200.000 Datensätzen wurde ein Wert von durchschnittlich über 90% CPU-Auslastung gemessen (siehe Abbildung 37). Dies könnte

darauf hindeuten, dass eine Erhöhung der Anzahl an verfügbaren CPUs oder der Einsatz leistungsfähigerer CPUs die Ergebnisse der Performance- und Lasttests noch etwas verbessert werden könnten.

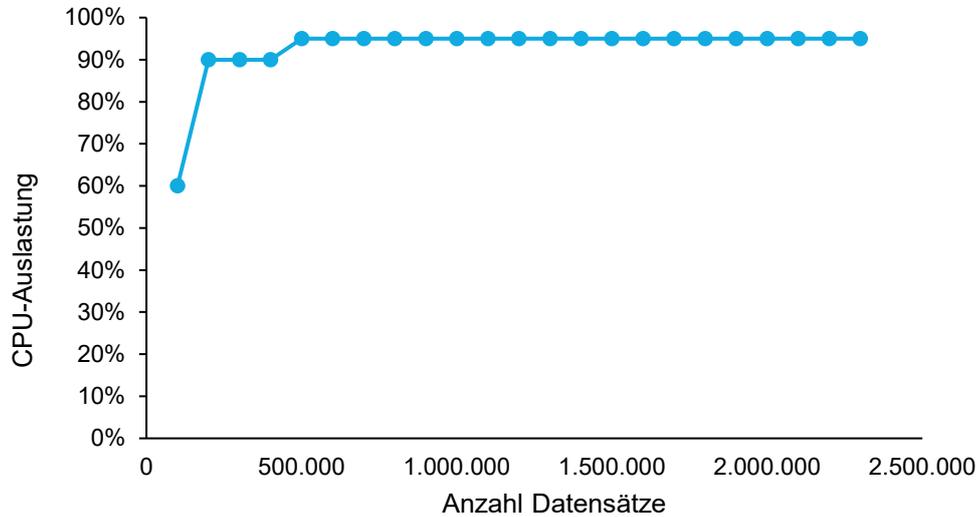


Abbildung 37: Durchschnittliche CPU Auslastung beim Ausführen der JMeter Tests

Die Dauer zum Ausführen aller JMeter Tests steigt mit zunehmender Datenmenge an (siehe Abbildung 38). Dabei kommt es zu Schwankungen, die möglicherweise darauf zurückzuführen sind, dass auf der Umgebung selbst noch andere Prozesse unregelmäßig die Ressourcen der Umgebung nutzen. Ein anderer Einfluss auf die Schwankungen könnte sein, dass die JMeter Tests unter Umständen nicht immer mit derselben Synchronität ausgeführt werden. Als Trend ist deutlich zu erkennen, dass die Dauer zunimmt.

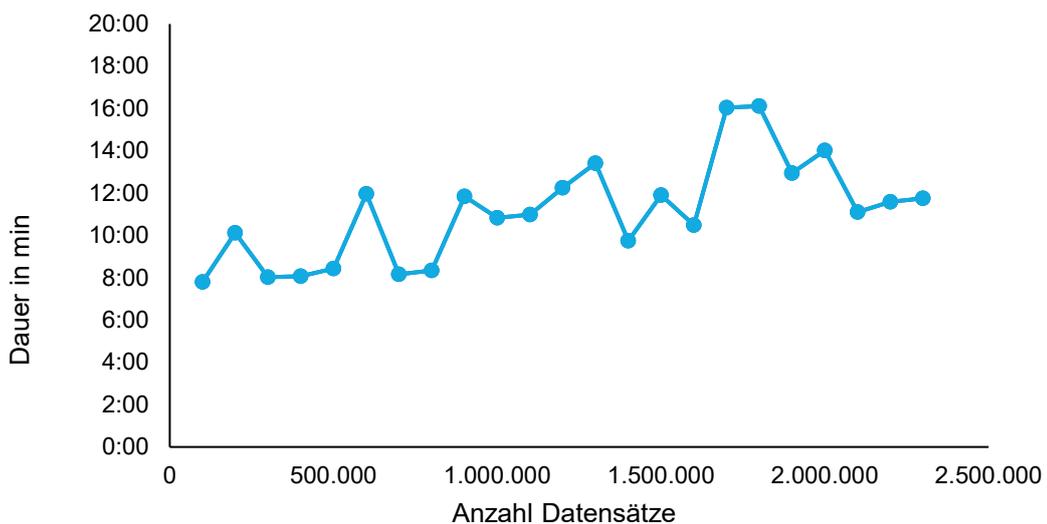


Abbildung 38: Dauer der Ausführung von JMeter Tests

Bei einem Durchlauf der JMeter Tests werden insgesamt 300 Anfragen gestellt. Mit zunehmender Datenmenge im System steigt auch die Anzahl fehlerhafter Anfragen (siehe Abbildung 39). Eine fehlerhafte Anfrage bedeutet,

dass auf eine Anfrage keine Antwort zurückgeliefert wird oder diese nicht die korrekte Antwort enthält. Die Quote steigt jedoch nicht konstant, sondern ist kleineren Schwankungen ausgesetzt. Es ist eine hohe Korrelation zur Ausführungszeit zu erkennen. Dementsprechend sind die Schwankungen mit denselben Faktoren wie bei der Ausführungszeit zu begründen. Als Trend ist auch hier ein leichter Anstieg der Fehlerquote zu erkennen, was darauf hinweist, dass viele Daten auch für geringfügig erhöhte fehlerhafte Anfragen sorgen kann. Der Zuwachs der Quote ist jedoch nicht sehr groß und bewegt sich selbst bei 2.300.000 Datensätzen mit etwa 0,07% noch in einem akzeptablen Bereich.

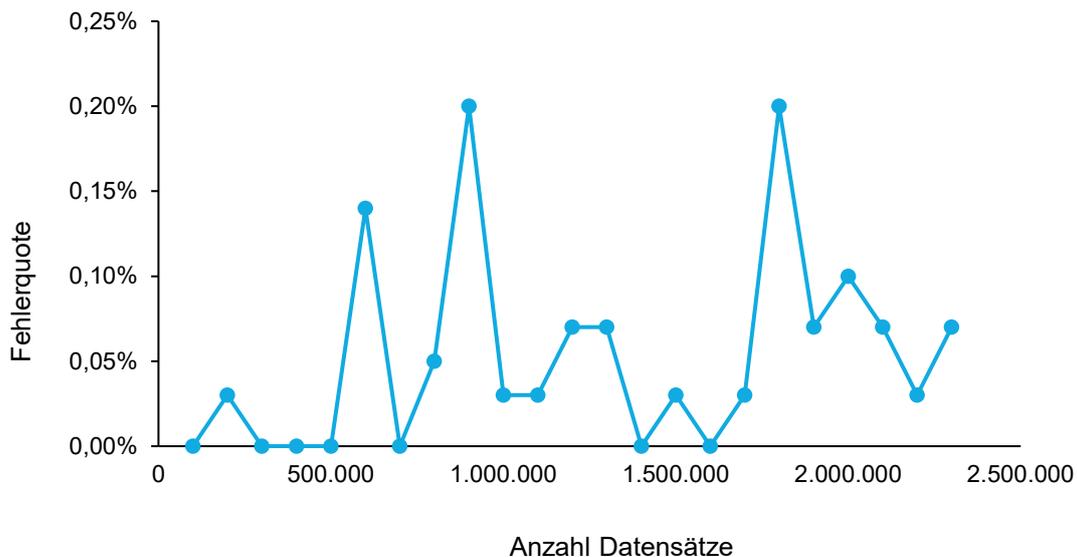


Abbildung 39: Fehlerquote der JMeter Tests

Analyse

Zu Beginn des Experiments wurde aufgrund von Erfahrungswerten einiger Betreiber anderer Open Data Portale davon ausgegangen, dass die aktuelle Systemarchitektur mit CKAN als Anwendung für die Datenhaltung nur mit bis zu etwa einer halben Millionen Datensätze stabil und performant einsetzbar ist. Dies entspricht einem Wert den das europäische Datenportal 2017 bei ähnlichen Experimenten ermittelt hat. Die im Projekt DataGo durchgeführten Tests zeigen nun, dass die CKAN-Komponente in der aktuellen Version 2.9 mit einer sehr viel höheren Anzahl von Datensätzen ohne Performanceverlust betrieben werden kann. Dies liegt an mehreren Faktoren. Als erstes spielen modernere Versionen von Softwarekomponenten eine große Rolle. Perspektivisch ist zu erwarten, dass die bereits angekündigte CKAN Version 3⁵² weitere wesentliche Optimierungen bereithalten wird. Ein weiterer Punkt ist, dass die Testinstanz der CKAN Umgebung nicht eins zu eins mit einer Produktivumgebung verglichen werden kann. Eine produktive Umgebung sammelt im Lauf der Jahre Datenartefakte in der Datenbank und zusätzliche Daten wie Historien an. Dies könnte die Umgebung bei einem Betrieb über mehrere Jahre hinweg zusätzlich verlangsamen. Da dies aber nicht in realistischer Zeit simuliert werden kann und die Erfahrungen der letzten Jahre aus dem Betrieb des GovData-Portals keine größeren

⁵² Datopian Tech, „CKAN v3“, Datopian Tech, 4. Feb. 2022, zuletzt abgerufen am 25.11.2022, <https://tech.datopian.com/ckan-v3/>

Effekte aufzeigten, sind die Auswirkungen als eher gering einzustufen.

4.1.3.5 Testen des Triplestores

Vorgehen

Der Triplestore wurde gesondert behandelt, da hier im Vorfeld die Annahme getroffen wurde, dass dieser die großen Datenmengen besser verträgt als CKAN. Nachdem die CKAN-Tests abgeschlossen waren, wurde einmal eine Re-Indizierung aller sich in CKAN befindlichen Datensätze ausgeführt, um die Daten in den Triplestore zu schreiben. Dieser Vorgang dauerte bei 2.300.000 Datensätzen etwa 56 Stunden. Der Fokus beim Testen des Triplestores lag darin, zu schauen, an welchen Stellen die bestehenden Anwendungen angepasst werden müssen, welche mit dem Triplestore kommunizieren, wenn die Anzahl der Datensätze stark ansteigt.

Analyse

Bei 2.300.000 Datensätzen im Triplestore ist aufgefallen, dass einige der SPARQL-Beispielabfragen im Portal nicht mehr funktionieren, da diese eine zu lange Laufzeit haben und ein Timeout verursachen. Ein ähnliches Problem ist beim Skript zum Erzeugen der Metriken festzustellen, bei dem einige Abfragen aufgrund eines Timeouts abbrechen. Dieses Problem ist einfach zu lösen, indem die Dauer bis zum Timeout entsprechend erhöht wird. Anfragen, die nicht über alle Datensätze iterieren, haben aber bei vielen Datensätzen im Triplestore, z.B. 1.000.000, noch immer dieselbe Reaktionszeit wie auf Umgebungen mit eher wenigen Datensätzen, z.B. 50.000. Dies trifft auf Anfragen zu, welche beispielsweise auf einen konkreten Datensatz abzielen oder ein Limit (üblicherweise im Bereich von 10 bis 1.000) enthalten, das dafür sorgt, dass nur eine bestimmte Menge an Triples zurückgeliefert und damit auch ausgewertet wird.

4.1.3.6 Prognose der Harvestingdauer

Vorgehen

Da zukünftig viele neue Endpunkte an das Portal angebunden werden sollen, müssen auch die möglichen Auswirkungen auf das Harvesting untersucht werden. Die Harvester werden aktuell alle zwei Tage ausgeführt. Falls viele neue Quellen angebunden werden, könnte es dazu führen, dass das Harvesting so lange dauert, dass zwei Tage nicht mehr ausreichen. Die Folge wäre, dass das Harvesting seltener stattfinden kann oder der Harvesting-Prozess optimiert werden muss.

Für eine bessere Einschätzung der Harvestingdauer muss betrachtet werden, aus welchen Teilen der Harvestingvorgang besteht (siehe Abbildung 40). Die folgenden Schritte werden dabei ausgeführt:

- Einlesen der Datensätze vom Endpunkt
- Parsen des serialisierten Graphen in ein Graph-Objekt
- Serialisierung des Graph-Objektes eines Datensatzes
- Validierung eines einzelnen serialisierten Datensatzes
- Validierungsergebnis im Triplestore speichern
- Übersetzung der Datensätze im Graph-Objekt in ein Dataset-Dictionary (CKAN)
- Dublettenprüfung (Löschen von „älteren“ Duplikaten)
- Ermittlung der nicht mehr mitgelieferten Datensätze und Entfernung aus dem Datenbestand
- Mapping der alten DCAT-AP.de-Lizenzen auf die aktuellen Lizenzen

- Falls für den Harvester aktiviert, Prüfung der Datensätze auf vorhandene Ressourcen (Distributionen) und Skip der Datensätze ohne Ressourcen
- Aktualisierung des Datensatzes in der CKAN-Datenbank

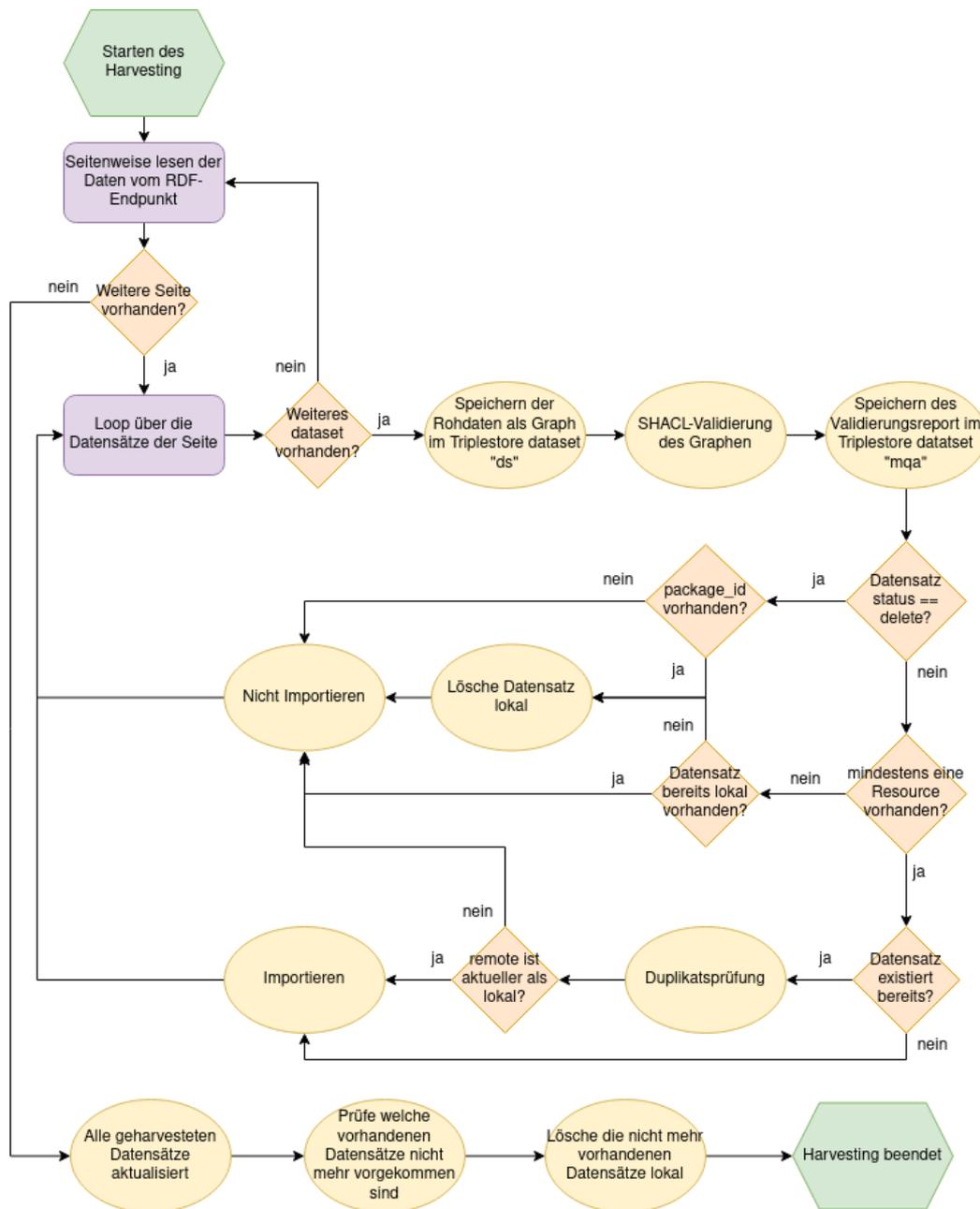


Abbildung 40: Harvestingablauf bei GovData

Um die Harvestingdauer einzelner Harvester einfacher bestimmen zu können, wird in diesem Kapitel stets mit einer konstanten durchschnittlichen Dauer pro Harvester gerechnet. Aktuell beträgt diese für die 16 bisher vorhandenen Harvester ca. 37,5 Minuten bei einer durchschnittlichen Anzahl von knapp 4.000 Datensätzen. Der exakte Wert für die Harvestingdauer neu anzuschließender Endpunkte ist schwer zu bestimmen, da die zu erwartenden Datenmengen unbekannt sind. Für neue Harvester wird deshalb näherungsweise davon ausgegangen, dass die durchschnittliche Harvestingdauer gleich bleibt. Dadurch kann die Gesamtdauer nach

der Anbindung neuer Harvesting-Quellen linear hochgerechnet werden.

Wie in Abbildung 41 zu sehen ist, benötigen die 16 aktuellen Harvester etwa 10 Stunden für einen kompletten Harvestingdurchlauf. Wenn pro neuem Harvester die ermittelten 37,5 Minuten hochgerechnet werden, ergibt sich ein linearer Anstieg für die Gesamtharvestingdauer.

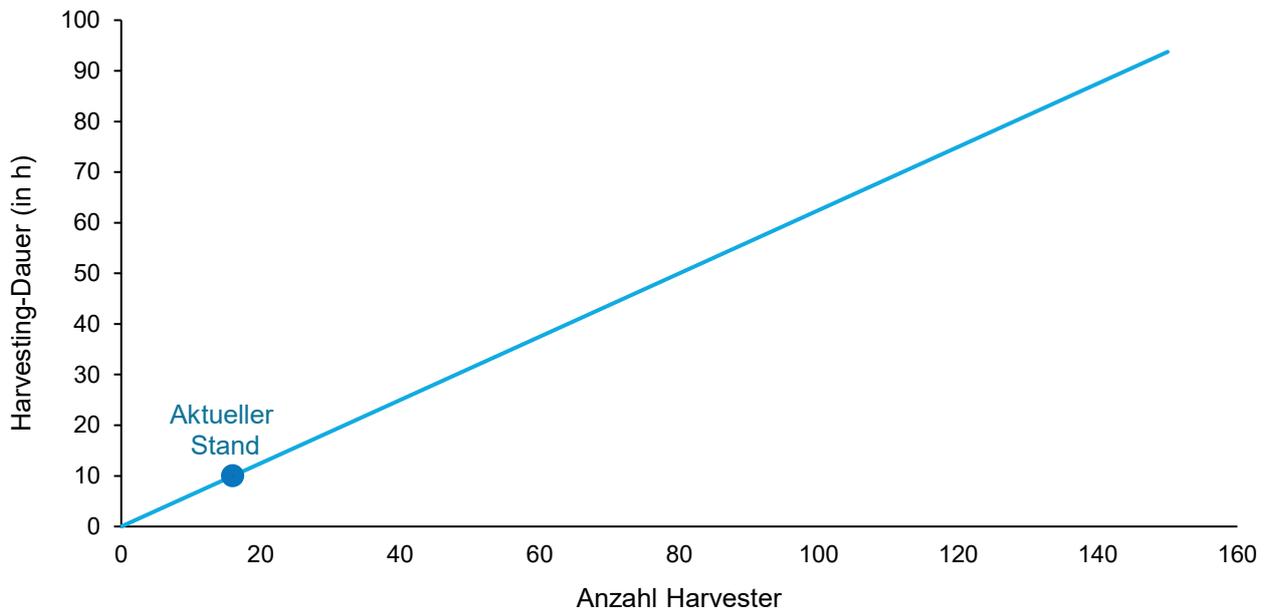


Abbildung 41: Hochrechnung der Harvestingdauer (in Stunden)

Analyse

Als realistisches Ziel für die Anzahl an Harvestern in den nächsten zwei bis drei Jahren wird aktuell von etwa 100 Harvestern ausgegangen. Das entspräche einer Gesamtharvestingdauer von 62,5 Stunden bzw. etwa 2,6 Tagen. Somit könnten eine Aktualisierung des Datenbestandes durch das Harvesting nur noch alle 3 Tage erfolgen. Um den Aktualisierungsrhythmus beibehalten oder sogar verkürzen zu können, müssten in der bestehenden Infrastruktur demnach in Bezug auf das Harvesting Optimierungen zur Verbesserung der Gesamtlaufzeit vorgenommen werden.

4.1.3.7 Fazit zur Schwellenwertanalyse

Durch die Schwellenwertanalyse konnte festgestellt werden, dass die aktuelle Umgebung durchaus in der Lage ist, mit großen Datenmengen umzugehen. Es konnte eine erheblich größere Datenmenge in die Umgebung eingespielt werden, als angenommen wurde, und die Umgebung verhielt sich bis etwa 1.000.000 Datensätze sehr stabil, bevor ein langsamer Abfall der Performance zu erkennen war.

Für den Suchindex und das Portal gibt es hinsichtlich der Performance keine Bedenken, da die Suchverfahren (Solr oder Elasticsearch) für eine große Anzahl an Dokumenten optimiert sind und die meisten darzustellenden Inhalte im Portal in einem Cache vorgehalten werden. Die erforderlichen Änderungen für den Triplestore bestehen lediglich aus Anpassungen von Timeouts und einigen neuen Beispielqueries für den SPARQL-Endpunkt. Die Komponente, welche unter wachsender Datenlast voraussichtlich Performance-Einbußen verursacht, ist der CKAN mit der relationalen Datenbank für die Datenspeicherung.

Um das Harvesting weiterhin regelmäßig ausführen zu können, müssen hier mittelfristig Optimierungen stattfinden. Da das CKAN-Harvesting bisher nicht besonders effizient implementiert ist, findet sich hier viel Potential für Verbesserungen. Lösungsansätze zur Optimierung des CKAN-Harvestings sind in Kapitel 4.1.5.2 zu finden.

4.1.4 Untersuchung neuer Anwendungskandidaten

4.1.4.1 Vorgehen

Im Rahmen des Projektes wurden zunächst 25 Anwendungen ermittelt, die für die Zusammensetzung einer neuen GovData Systemarchitektur im Gesamten oder den Austausch einzelner Module geeignet sein könnten. Diese Anwendungen wurden anschließend einzeln untersucht um festzustellen, wie gut sie auf die Anforderungen von GovData zugeschnitten sind, und ob sie sich potentiell für eine neue Systemarchitektur eignen. Um diese Eignung zu bestimmen, wurden Kriterien festgelegt, welche bei der Untersuchung bewertet werden sollten. So wurde der Kreis der zu betrachtenden Anwendungen sukzessive eingeschränkt. Anhand der gewonnenen Erkenntnisse wurden schließlich zwei mögliche Varianten für eine zukünftige Systemarchitektur beschrieben und ein Konzept für die Systemmigration entwickelt.

4.1.4.2 Auswahl der Kandidaten

Zu Beginn der Untersuchung wurden verschiedene Anwendungen recherchiert, die Gesamtsysteme oder einzelne Funktionalitäten von Open Data Portalen abbilden können. So wurden 25 Anwendungen ermittelt, die anschließend analysiert wurden.

4.1.4.3 Kriterien

Um die Untersuchung verschiedener Anwendungen besser miteinander vergleichen zu können wurden Kriterien festgelegt, die bei der Untersuchung zur Bewertung herangezogen werden sollten. In Tabelle 8 sind diese Kriterien aufgelistet, die jeweils einem der Bereiche allgemeine Informationen, funktionale Anforderungen, Sicherheitsanforderungen und Entwicklungsanforderungen zuzuordnen sind. Zu jedem Kriterium ist zum besseren Verständnis und einer möglichst einheitlichen Interpretation eine kurze Beschreibung angegeben. In der dritten Spalte ist eine Priorität für das Kriterium festgelegt. Je höher der Wert desto wichtiger ist das Kriterium für die Bewertung. Weitere Informationen zu der Bewertung finden sich in Kapitel 4.1.4.4.

Kriterium	Beschreibung	Priorität
Hersteller	Von welchem Hersteller wird das Produkt vertrieben?	0

Versionsnummer	Welche Versionsnummer des Produktes liegt dieser Untersuchung zugrunde?	0
Open Source	Ist das Produkt unter einer Open Source Lizenz erhältlich oder wird es ausschließlich kommerziell vertrieben?	3
Lizenzen	Unter welchen Lizenzen kann das Produkt eingesetzt werden?	1
Aktuelle Weiterentwicklung	Wird die Software aktuell noch weiterentwickelt? Gibt es einen Release Fahrplan?	10
Erfahrungswerte	Hat jemand uns Bekanntes Erfahrungen mit der Software?	4
Gibt es eine Community	Wie groß und aktiv ist die Community? Gibt es Bugtracker, Forum, Release Logs, etc.	7
Voraussetzungen	Welche speziellen Hard-/Softwareanforderungen liegen vor?	4
Wiederverwendbarkeit aktueller Anwendungen	Welche Teile der aktuellen Lösung können in Kombination mit der Software wiederverwendet/integriert werden?	5
Verwendete Datenbanken	Welche Datenbank(en) verwendet die Software?	3
Erweiterbarkeit	Kann die Software durch Erweiterungen oder neue Services einfach um neue Funktionen erweitert werden? Gibt es bereits Erweiterungen vom Entwickler/der Community dafür?	4
Skalierbarkeit	Wie skaliert die Software mit wachsenden Datenmengen und wachsender Nutzerzahl? Welche Probleme können dabei auftreten?	6
Bekannte Systemgrenzen /Schwellenwerte	Gibt es bekannte Grenzen der Software die uns beeinflussen könnten? Hier können auch die ermittelten Schwellenwerte einfließen.	5
Programmiersprache/n	Welche Programmiersprachen können im Zusammenspiel mit der Software verwendet werden?	3
Dokumentation	Wie gut/ausführlich ist die vorhandene Dokumentation der Software? Findet sich ein ungeübter Nutzer schnell ein?	9
Konfigurationsmöglichkeiten	Gibt es Anpassungsmöglichkeiten der Software oder handelt es sich um eine Standardlösung? Welche Möglichkeiten für ein besseres Handling gibt es?	5
Umgebung aufsetzen	Wie viel Aufwand ist es eine Umgebung mit der Software aufzusetzen, so dass zumindest eine minimale Anwendung läuft?	3
Nutzerverwaltung	Ist eine Nutzerverwaltung in die Software integriert? Kann das bestehende GovData Nutzerset damit abgebildet werden?	4
Sub-Accounts	Können Sub-Accounts von Haupt-Accounts selbst angelegt und verwaltet werden?	2
Mehrsprachigkeit	Ist Mehrsprachigkeit der Inhalte in die Software integriert?	2
SEO	Welche SEO Optionen sind möglich?	2
Blogfunktion	Ist ein Blog in die Software integriert?	2
Forum	Ist ein Forum in die Software integriert?	2
Feed	Ist es möglich einen Feed für Änderungen zu abonnieren?	2
Kommentarfunktion	Ist eine Kommentarfunktion für Datensätze in die Software integriert?	1
Bewertungssystem	Ist eine Bewertung für Datensätze in die Software integriert? Worauf basiert die Bewertung?	1

Redaktionelle Inhalte	Ist ein redaktionelles Management des Contents möglich? Wenn die Anwendung mit einem CMS kombiniert ist, welches wurde verwendet?	3
Bietet Dataset-API an	Ist eine API zum Abfragen der Datensätze vorhanden? Wenn ja, was für eine?	3
Harvesting	Gibt es eine Möglichkeit Daten zu harvesten? Wenn ja, über welche Schnittstellen und in welchen Formaten? Sind Harvester flexibel hinzufübar?	5
Manuelle Datenerfassung/-verwaltung	Gibt es die Möglichkeit, Metadaten manuell zu erfassen und zu verwalten?	4
RDF Endpunkt	Gibt es Optionen um Daten als RDF darzustellen und zu veröffentlichen?	4
Verschlüsselung vertraulicher Daten	Können persönliche Daten wie z.B. Passwörter verschlüsselt gespeichert werden? Welche Verschlüsselungsalgorithmen können angewendet werden?	6
Schutz vor korrupten Eingaben	Gibt es Schutzmechanismen? Z.B. gegen maschinelle oder korrupte Eingaben.	6
Backups	Können Backups gemacht/eingespielt werden? Wie viel Aufwand ist das?	6
Bemerkungen	Was sollte sonst noch erwähnt werden?	0
Score	Berechneter Wert zum Vergleichen der Anwendungen. Berechnet sich aus der Summe aller Gewichtungen, welche mit ihrer Priorität multipliziert werden. Je höher der Wert desto besser.	0
Backend-Score	Berechneter Wert zum Vergleichen der Anwendungen. Dieser Wert wird, wie der Score berechnet aber es fließen nur die Felder in die Gewichtung ein welche auch einen Bezug zum Daten-Backend haben. Da einige untersuchten Anwendungen ein CMS mitliefern und andere nicht ist dieser Backend-Score in dem Fall besser zum Vergleichen geeignet als der Score.	0

Tabelle 8: Bewertungskriterien zur Untersuchung verschiedener Anwendungen

4.1.4.4 Berechnung des Scores

Um verschiedene Anwendungen nach der Analyse besser vergleichen zu können, wird ein Score für jede Anwendung berechnet, welcher angibt, wie gut die Software auf die gegebenen Anforderungen passt. Für die Berechnung des Scores wird zunächst eine Priorität für jedes Kriterium festgelegt. Diese erhält einen Wert von 0 bis 10 und gibt an, wie wichtig das Kriterium für die Einstufung ist. Ein hoher Wert entspricht dabei einer hohen Relevanz (siehe Tabelle 8). Der Wert der Priorität ist für alle Anwendungen identisch.

Als zweite Variable wird die Gewichtung eines Kriteriums pro Anwendung festgelegt. Während der Untersuchung einer Anwendung wird pro Kriterium ein Wert von 0 bis 5 festgelegt, welcher angibt, wie gut die Anwendung das entsprechende Kriterium erfüllt. Je höher der Wert, desto besser deckt die Anwendung das Kriterium ab (Bewertung jeder Anwendung siehe Anhang 7.9.3).

Mit diesen zwei Variablen lässt sich für jede Anwendung ein Score berechnen, durch die Summe der Multiplikationen aller Prioritäten mit ihren Gewichtungen.

$$\text{Score} = \sum_k^{\text{Kriterien}} p_k * g_k$$

Wobei p_k die Priorität des Kriteriums ist und g_k die Gewichtung. Dieser Score ist keineswegs alleinentscheidend für die Auswahl der finalen Software. Da die unterschiedlichen Anwendungen verschiedene Funktionen besitzen, muss der Score nicht mit der Güte der Software übereinstimmen. Er soll lediglich als Indikator dienen, wie gut die Anwendung die Gesamtheit der Kriterien erfüllt.

Der Backend-Score berechnet sich auf demselben Rechenweg, allerdings werden dabei nur die für das Backend relevanten Kriterien in die Berechnung einbezogen. Nutzerverwaltung, Sub-Accounts, Mehrsprachigkeit, SEO, Blog, Forum, Feed, Kommentarfunktion, Bewertungsfunktion und Redaktionelle Inhalte werden hier nicht in die Berechnung mit einbezogen.

4.1.4.5 Voruntersuchung

HINWEIS: Bei dieser Version des Berichtes handelt es sich um eine gekürzte Version. Der vollständige Bericht enthält Inhalte, die aus wettbewerbsrechtlichen Gründen nicht für eine uneingeschränkte Veröffentlichung geeignet sind. Der folgende Abschnitt konnte aus diesem Grund leider nicht veröffentlicht werden.

4.1.5 Neue Systemarchitektur

Aus den bisherigen Erfahrungen und den neu gewonnenen Erkenntnissen aus der Untersuchung potentieller Anwendungen sowie den Antworten aus den Befragungen der Pilotinstitutionen wurden verschiedene Anpassungen und Verbesserungen der Systemarchitektur entwickelt. Wie die Änderungen an der Systemarchitektur aussehen, wird in diesem Kapitel näher beschrieben. Der Weg zu einer neuen Systemarchitektur wurde in mehrere Schritte und Varianten unterteilt, die unabhängig voneinander umgesetzt werden können.

HINWEIS: Bei dieser Version des Berichtes handelt es sich um eine gekürzte Version. Der vollständige Bericht enthält Inhalte, die aus wettbewerbsrechtlichen Gründen nicht für eine uneingeschränkte Veröffentlichung geeignet sind. Der folgende Abschnitt konnte aus diesem Grund leider nicht veröffentlicht werden.

4.1.5.1 bisheriges CMS

4.2 Prozessoptimierung für zukünftige Anbindungen

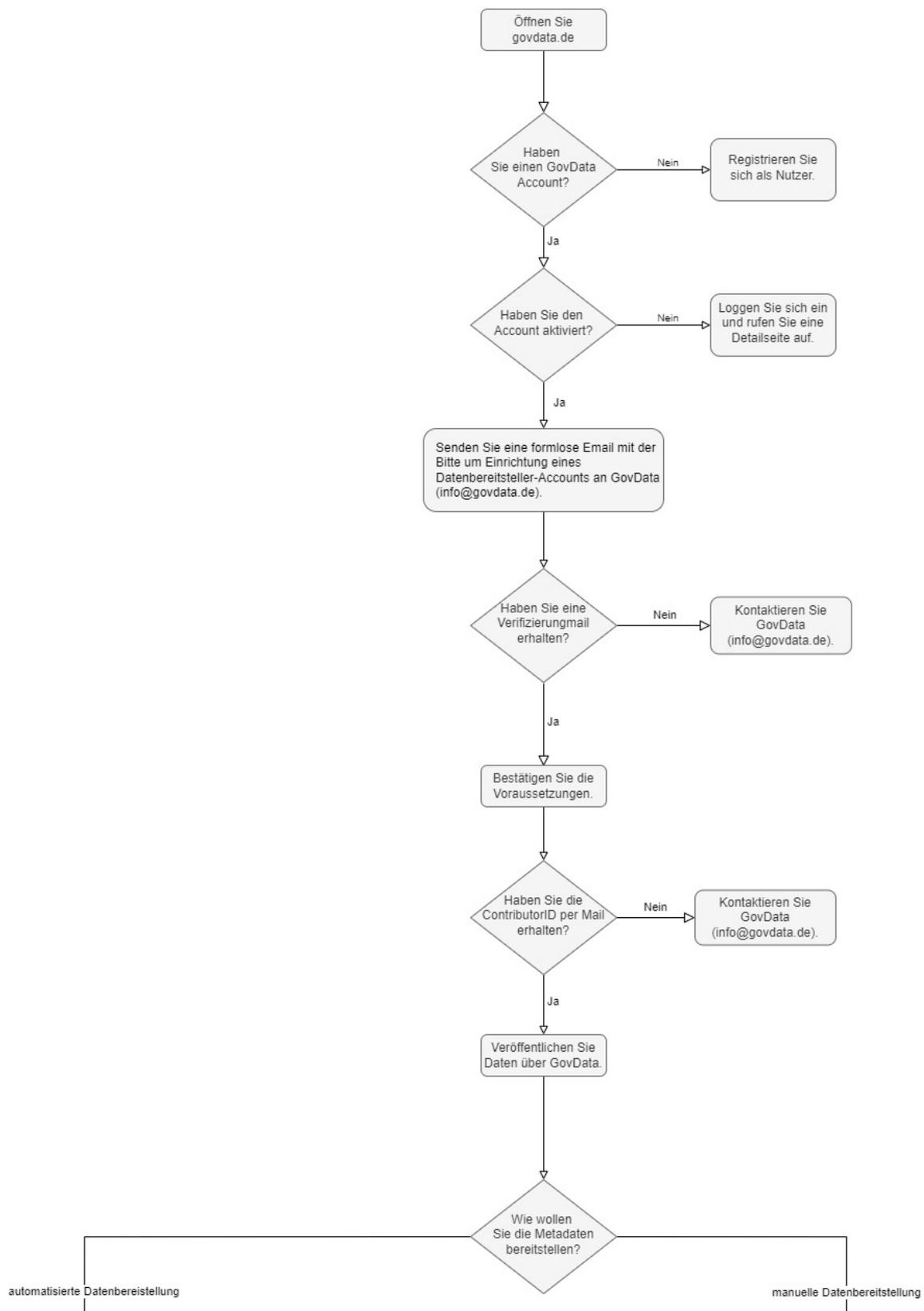
In Abschnitt 3 wurden die Ergebnisse der Personalbedarfsermittlung dargestellt, die im Rahmen des Teilprojektes 2 auf Grundlage von Pilotierungen, Untersuchungen, Zeiterfassungen etc. ermittelt wurde. Neben der Erfassung der Aufwände, die während der Anbindung von datenbereitstellenden Stellen entstehen, war ein zentrales Ziel des Teilprojektes die Analyse und Verbesserung der bestehenden Prozesse, In diesem Abschnitt

werden auf Grundlage dieser Untersuchungen Vorschläge entwickelt, die dazu beitragen sollen, den Anbindungsprozess zu verbessern.

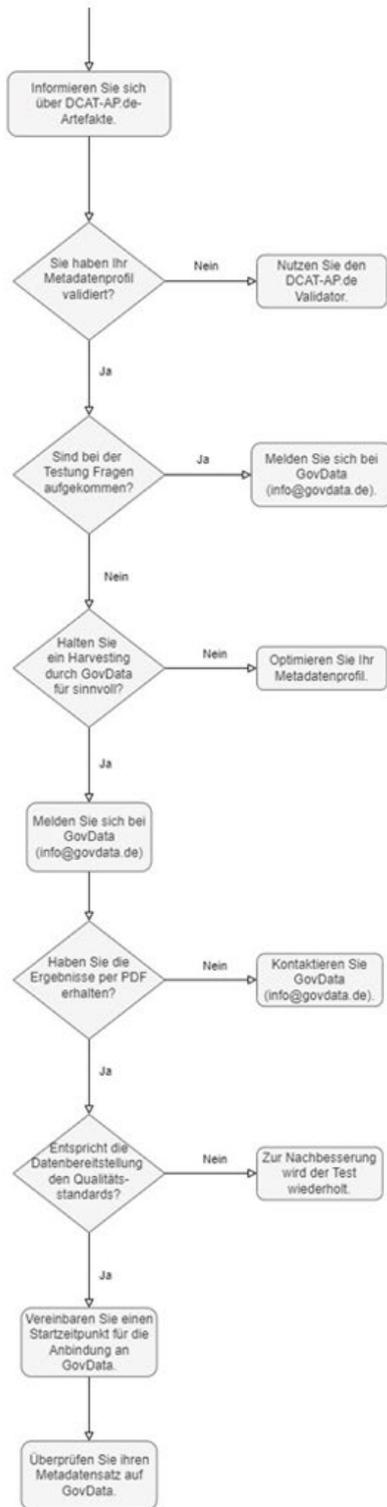
4.2.1 Analyse der Ist-Prozesse

Der Prozess der Anbindung an GovData kann zur Veranschaulichung in zwei Teile gegliedert werden. Der erste Teil betrifft die Anmeldung im Portal und die Freischaltung eines Datenbereitsteller-Accounts, dieser ist unabhängig von der Art der technischen Anbindung. In der Folge unterscheiden sich die notwendigen Schritte deutlich, abhängig von der Art der technischen Anbindung. Der Prozess ist in einem Flussdiagramm (siehe Abbildung 47) dargestellt.

Nachfolgend werden die Charakteristika und Problemfelder der jeweiligen Prozessschritte erläutert. Dabei wird im Hinblick auf die automatische Datenbereitstellung abermals entsprechend der gebildeten Gruppen zwischen automatischer Anbindung und manueller Anbindung, inkl. Mapping unterschieden. Hinsichtlich der vorbereitenden Schritte (siehe Abbildung 47) und in Bezug auf die manuellen Anbindungen wird zudem auf das Dokument „Step-by-Step-Anleitung_GovData 1.1“ zum besseren Verständnis der Prozessschritte verwiesen.



2a. Weitere Schritte für die automatisierte Datenbereitstellung



2b. Weitere Schritte für die manuelle Datenbereitstellung

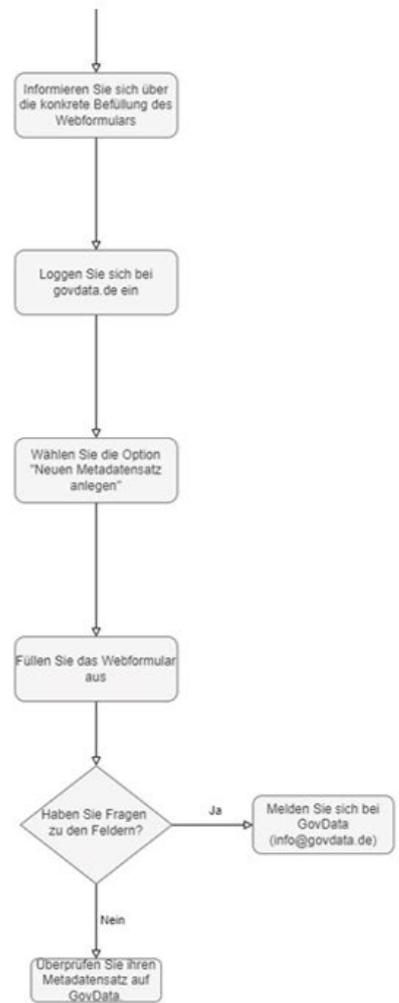


Abbildung 42: Erstellung eines Datenbereitsteller-Accounts und einer ContributorID

4.2.1.1 Erstellen eines Datenbereitsteller-Accounts und einer ContributorID

Die Schritte zur Erstellung des Datenbereitsteller-Accounts und der ContributorID sind für alle Arten der Datenbereitstellung an das GovData Portal erforderlich.

Der Datenbereitsteller-Account ist eine besondere Rolle für Nutzer des Portals, der zum Bereitstellen und Bearbeiten von Metadaten über eine Schnittstelle oder ein manuelles Erfassungsformular berechtigt. Darüber hinaus können Nutzer mit der Rolle im GovData Portal ihre Metadaten bearbeiten und die Qualität der Metadaten überwachen.

Die ContributorID ist eine URI, die eindeutige Kennzeichnung der Datenbereitstellenden (dcatde:contributorID) ermöglicht. Zur Unterstützung der Kommunikation im Portalverbund, und um maschinenverarbeitbare Herkunftsangaben zu ermöglichen, pflegt die GKSt GovData eine Liste der Datenbereitstellenden des Portals GovData.de⁵³. Sie enthält die ContributorIDs des GovData.de Portals, die die Datenbereitstellenden zwingend in den Metadaten angeben müssen.

4.2.1.1.1 Prozessdarstellung

Der gegenwärtige Prozess zur Erstellung des Datenbereitsteller-Accounts und der ContributorID wird in Abbildung 48 dargestellt.

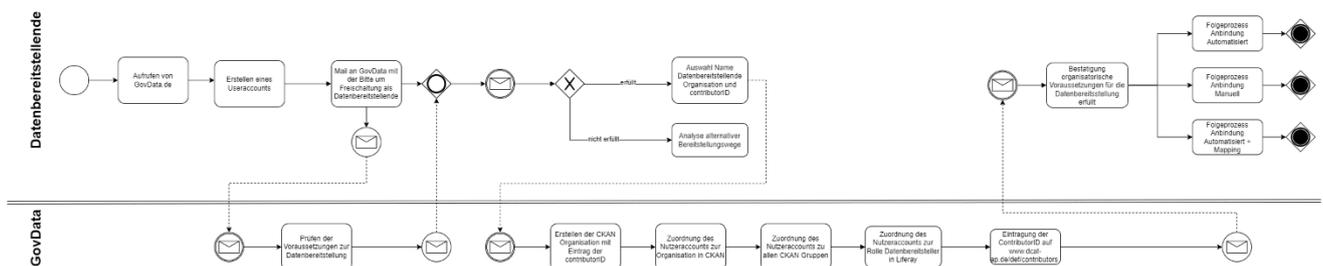


Abbildung 43: Aktueller Prozess für die Erstellung des Datenbereitsteller-Accounts

Bei der Evaluation der Pilotierungen wurde deutlich, dass der Prozess zur Erstellung des Datenbereitsteller-Accounts und der ContributorID vor allem auch auf Seiten GovData und den beteiligten Dienstleister manuellen Aufwand erzeugt, der bei einer deutlich höheren Anzahl neuer Datenbereitstellender bevorzugt durch Automatisierungen und Prozessoptimierung gemindert werden sollte. Bei der Kommunikation sowie bei der Prozessoptimierung sollten die Punkte Erstellung Datenbereitsteller-Account und Erstellung contributorID nicht als zwei getrennte Prozesse angesehen werden. Die Erwartungen der Nutzer sind, dass sie sich als neuer Datenbereitsteller unter „Anmelden / Konto erstellen“ auf GovData.de registrieren und dann, ggf. nach einer kurzfristigen Freigabe zeitnah mit der Datenbereitstellung beginnen können. An diesen Bedarfen der User sollte sich ein optimierter Prozess orientieren.

⁵³ DCAT-AP.de, „Liste der Datenbereitsteller“, zuletzt abgerufen am 21.11.2022, <https://www.dcat-ap.de/def/contributors/>

4.2.1.1.2 Informationsbedarf

Die Evaluation der Pilotierungen im Rahmen von Data Go hat gezeigt, dass die Informationsangebote auf der Webseite des Portals und auf DCAT-AP.de nicht ausreichen, um Datenbereitstellenden ohne zusätzliche Beratung zu erläutern, ob und wie Sie für die Datenbereitstellung auf Govdata.de berechtigt werden können und welche Schritte sie hierfür übernehmen können. Häufig wurde z.B. der Account nicht durch Login und Seitenaufruf aktiviert. Dieser Schritt erschien für viele User ungewöhnlich, was wiederum Kommunikationsaufwand und Verzögerungen zur Folge hatte.

Im Rahmen von Data Go wurden die Informationsangebote auf GovData.de um eine Step-by-Step-Anleitung⁵⁴ ergänzt. Diese Anleitung hilft potenziellen Datenbereitstellenden. Allerdings gibt es hier auch weitere Optimierungsbedarfe, da Nutzer die Informationen auf der Webseite teilweise nicht finden. Häufig verstehen Nutzer auch nicht, warum sie einen Datenbereitsteller-Account bzw. eine contributorID benötigen.

4.2.1.2 Gruppe „Automatisierte Anbindung“

4.2.1.2.1 Prozessdarstellung

Der gegenwärtige Prozess der Anbindung für automatische Datenbereitstellung ist in Abbildung 49 dargestellt.

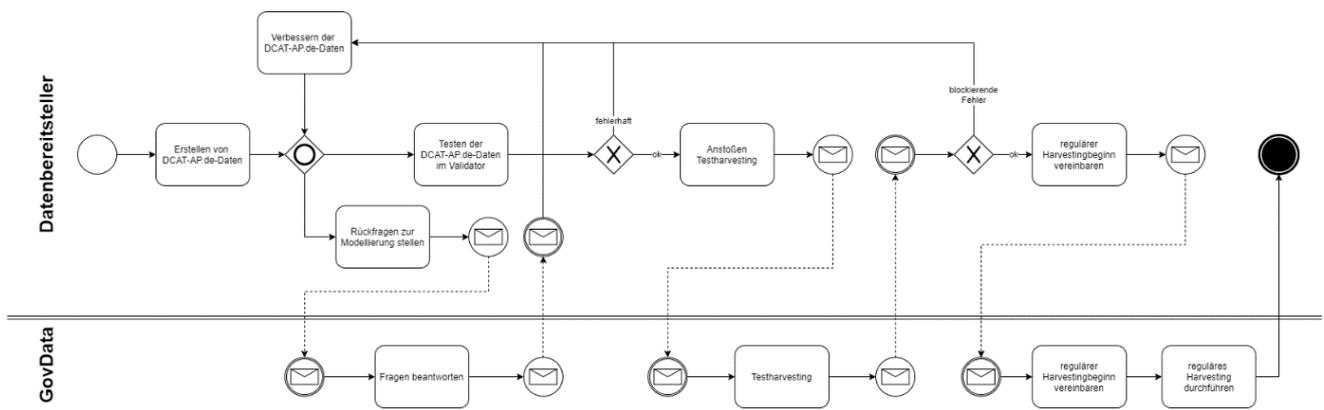


Abbildung 44: Prozessschritte zur automatischen Datenanlieferung

Probleme und Aufwandstreiber finden sich dabei vorrangig im ersten Drittel des Prozesses, aber auch das Testharvesting kann zum „Nadelöhr“ werden.

Die automatisch angebotenen Kommunen besitzen häufig eigene Portale und arbeiten in der Regel mit IT-Dienstleistern zusammen oder haben eigene technische Expertise. Aus diesem Grund wurde in dieser Gruppe weniger Wert auf persönliche Absprachen gesetzt. Fragen konnten zumeist auch per E-Mail oder durch Nutzung vorhandener Dokumentation beantwortet werden. Dennoch wurde von den Pilotinstitutionen ein erstes Beratungsgespräch mit Inhalten zu DCAT-AP.de, dem Harvestingprozess, bei dem auch anstehende Aufgaben und ein Zeitplan diskutiert wurden, sehr positiv bewertet.

⁵⁴ GovData, „Anleitung zur Datenbereitstellung auf GovData“, zuletzt abgerufen am 21.11.2022, <https://www.govdata.de/web/guest/datenbereitstellungaufgovdata>

4.2.1.2.2 Fehlender „Single Point of Truth“

Als großes Problem erwies sich für datenbereitstellende Institutionen die Informationsbeschaffung. Zwar wurde im Laufe des Projektes eine „Anleitung zur Datenbereitstellung auf GovData“ im Sinne der genannten Step-by-Step-Anleitung in der Informationssektion der Website bereitgestellt, diese wurde aber von teilnehmenden Institutionen als zu wenig sichtbar bezeichnet und nicht zur Kenntnis genommen.

Darüber hinaus erfüllt diese Sektion nicht die Anforderung eines „Single Point of Truth“, da hier technische Informationen bspw. zur CKAN-Plattform oder Schnittstellendokumentation nicht oder nur rudimentär bereitgestellt werden. Mehrere Institutionen berichteten dementsprechend von der Notwendigkeit, Informationen selbst zusammensuchen zu müssen. Vor diesem Hintergrund kam insbesondere den IT-Dienstleistern der Verwaltungen eine zentrale Rolle zu.

4.2.1.2.3 „Nadelöhr“ Testharvesting

Im bisherigen Prozess erwies sich das Testharvesting als „Nadelöhr“. Das Testharvesting musste durch GovData beim Dienstleister SEITENBAU beauftragt werden. SEITENBAU kommentierte und übermittelte die Ergebnisse an GovData, die sie prüfte und dann an die datenbereitstellenden Institutionen weiterleitete. Dies führt zu zwei miteinander verbundenen Problemen:

1. Der zeitliche und organisatorische Aufwand ist eine Belastung der personellen Ressourcen von GovData und kann punktuell zu Verzögerungen im Ablauf führen, worauf GovData selbst auf der Website hinweist.
2. Die datenbereitstellende Institution ist abhängig von der Bearbeitung durch GovData. Sie kann nicht autonom handeln, da sie keine direkte Rückmeldung zur Qualität ihrer Daten erhält. Somit kann sie nicht direkt Anpassungen vornehmen, um erneut testen zu können.

4.2.1.2.4 Fehlende Autonomie

Die Frage der Autonomie der Datenbereitstellenden berührt auch die Frage des Know-how auf Seiten der Behörden. Hier ist es nicht allein die bereits genannte Abhängigkeit von IT-Dienstleistern hinsichtlich Plattform und Schnittstellen von Relevanz, sondern auch die Fähigkeit, die Ergebnisse des Testharvestings korrekt interpretieren zu können. Dies betrifft ebenfalls die Ergebnisse des Validators, der ansonsten als Werkzeug für die eigenständige Prüfung der Datenqualität den Datenbereitstellenden größere Autonomie verschaffen kann.

4.2.1.3 Gruppe „Automatische Anbindung, inkl. Mapping“

4.2.1.3.1 Prozessdarstellung

Der Prozess der Anbindung von Institutionen mit einem anderen Metadatenstandard wurde erst im Rahmen des Projektes eingeführt. Vorher bestand ein solcher nicht. Grundsätzlich orientiert sich der Prozess an der automatischen Bereitstellung (siehe 4.2.1.2.1). Der zentrale Unterschied zur automatischen Datenbereitstellung liegt in der Notwendigkeit eines Mappings der Datenfelder von den spezifischen von Forschungseinrichtungen genutzten Metadatenformaten auf den für die Anlieferung an GovData notwendigen DCAT-AP.de-Standard. Sofern ein solches Mapping vorliegt, können die Metadaten wie bei der automatischen Bereitstellung durch

GovData geharvestet werden.

4.2.1.3.2 Vielfalt der Metadatenstandards

DCAT-AP.de ist im Forschungsbereich weitgehend unbekannt. Hier haben sich andere Standards, vielfach aus dem bibliothekarischen Bereich kommend, durchgesetzt. Hierzu zählen insbesondere die XML-Standards DataCite, Dublin Core und MARC.

Auf Basis der Erkenntnisse zu den vorliegenden Fällen besteht eine hohe Wahrscheinlichkeit, dass Forschungsinstitutionen allgemein ihre Daten nicht in DCAT-AP.de ausliefern können. Tatsächlich war der Standard, so haben die Feedback-Befragungen gezeigt, Forschungsinstitutionen häufig unbekannt.

Vor dem Hintergrund, dass die automatische Anbindung die für GovData präferierte Lösung ist und eine manuelle Eingabe von Metadaten für Forschungsinstitutionen keine dauerhafte Lösung darstellt, bildet ein Mapping auf DCAT-AP.de die einzig sinnvolle Herangehensweise.

4.2.1.4 Gruppe „Manuelle Anbindung“

4.2.1.4.1 Prozessdarstellung

Der gegenwärtige Prozess der manuellen Datenbereitstellung ist in Abbildung 50 dargestellt.

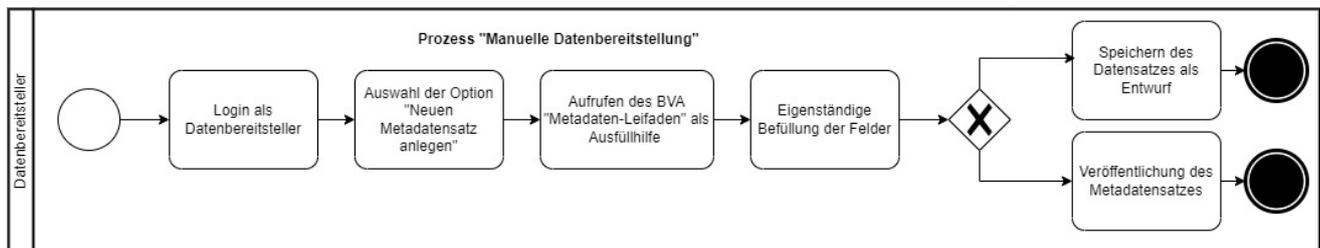


Abbildung 45: Prozessschritte zur manuellen Datenbereitstellung

4.2.1.4.2 Das GovData-Formular

Die Gruppe „Manuelle Anbindung“ hat die in den Einzelterminen und dem Workshop erarbeiteten Erkenntnisse zu den für Nutzer:innen problematischen Aspekten des Einpflegens von Metadaten im GovData-Formular in verschiedenen „Handlungsfeldern“ gesammelt. Diese unterscheiden nach funktionellen Ebenen, die im Rahmen des Kapitels „4.2.2 Optimierungs- und Automatisierungspotenziale“ ebenfalls zur Strukturierung verwendet werden. Der grundsätzliche Prozess des manuellen Datenbereitstellens wird in Abbildung 50 sowie im Dokument „Step-by-Step-Anleitung 1.1“ beschrieben. Das Formular selbst ist in Abbildung 51 dargestellt.

Metadatensatz hinzufügen



Mit * gekennzeichnete Felder sind Pflichtfelder.

Datenbereitsteller *

Sonstige

Ausgeblendet

Datensatz ist ausgeblendet und nur für Mitglieder der Organisation sichtbar

Titel *

Dieses Feld ist erforderlich

Beschreibung *

Dieses Feld ist erforderlich

GovData-ContributorID *

<http://dcat-ap.de/def/contributors/govData>

Kontakte

Autor

Name

Email

Webseite

Adresse

Adressat

Zusatz

Straße

Stadt

PLZ

Land

Ansprechpartner

Name

Email

Webseite

Adresse

Adressat

Zusatz

Straße

Stadt

PLZ

Land

Veröffentlichende Stelle

Name	<input type="text"/>	
Email	<input type="text"/>	
Webseite	<input type="text"/>	
Adresse	Adressat	<input type="text"/>
	Zusatz	<input type="text"/>
	Straße	<input type="text"/>
	Stadt	<input type="text"/>
	PLZ	<input type="text"/>
	Land	<input type="text"/>

Urheber

Name	<input type="text"/>	
Email	<input type="text"/>	
Webseite	<input type="text"/>	
Adresse	Adressat	<input type="text"/>
	Zusatz	<input type="text"/>
	Straße	<input type="text"/>
	Stadt	<input type="text"/>
	PLZ	<input type="text"/>
	Land	<input type="text"/>

Kategorien

- | | |
|--|--|
| <input type="checkbox"/> Bevölkerung und Gesellschaft | <input type="checkbox"/> Bildung, Kultur und Sport |
| <input type="checkbox"/> Energie | <input type="checkbox"/> Gesundheit |
| <input type="checkbox"/> Internationale Themen | <input type="checkbox"/> Justiz, Rechtssystem und öffentliche Sicherheit |
| <input type="checkbox"/> Landwirtschaft, Fischerei, Forstwirtschaft und Nahrungsmittel | <input type="checkbox"/> Regierung und öffentlicher Sektor |
| <input type="checkbox"/> Regionen und Städte | <input type="checkbox"/> Umwelt |
| <input type="checkbox"/> Verkehr | <input type="checkbox"/> Wirtschaft und Finanzen |
| <input type="checkbox"/> Wissenschaft und Technologie | |

Schlagwörter

Ein oder mehrere Werte erlaubt, mehrere Werte mit Komma getrennt angeben: Wert1, Wert2, Wert3

Webseite

Qualitätssicherungsprozess URI

Kalender-Daten

Veröffentlicht: Datum wählen



Aktualisiert: Datum wählen



Ressourcen

Ressource 1

Webseite *

Name

Beschreibung

Format

Sprache

Ein oder mehrere Werte erlaubt, mehrere Werte mit Komma getrennt angeben: Wert1, Wert2, Wert3

Lizenz *

Namensnennungstext für "By"-Clauses

Aktualisierungsdatum

Verfügbarkeit

Ressource hinzufügen

Zeitbezug

Von: Bis:

Geopolitische Abdeckung URI

Ein oder mehrere Werte erlaubt, mehrere Werte mit Komma getrennt angeben: Wert1, Wert2, Wert3

Rechtsgrundlage für Zugangseröffnung

Ein oder mehrere Werte erlaubt, mehrere Werte mit Komma getrennt angeben: Wert1, Wert2, Wert3

Ebene der geopolitischen Abdeckung URI

Abdeckung Text

Ein oder mehrere Werte erlaubt, mehrere Werte mit Komma getrennt angeben: Wert1, Wert2, Wert3

Raumbezug

Mit [Geojson.io](#) kann das aktuelle Polygon bearbeitet oder ein neues erzeugt werden. Bitte folgendes Schema einhalten:

```

{
  "type": "Polygon",
  "coordinates": [
    [
      [11.598079, 51.29537],
      [11.598079, 53.89386],
      [6.654667, 53.89386],
      [6.654667, 51.29537],
      [11.598079, 51.29537]
    ]
  ]
}

```

Speichern

Abbrechen

Abbildung 46: Formular von GovData (mehrtellig)

4.2.1.4.3 Probleme und Aufwandstreiber

Grob zusammengefasst lassen sich folgende Aspekte identifizieren:

- Das Formular ist statisch aufgebaut.
- Das Formular muss vollständig von Hand befüllt werden.
- Das Formular sieht keine „Führung“ der Nutzer:innen vor.
- Verschiedene Datenfelder sind unverständlich oder missverständlich.
- Im Formular selbst finden sich keine Hilfestellungen.

Eine genauere Beschreibung der spezifischen Probleme und Aufwandstreiber findet sich im Dokument „Lastenheft GovData Formular“ (siehe Anhang 7.8).

4.2.2 Optimierungs- und Automatisierungspotentiale

4.2.2.1 Erstellen eines Datenbereitsteller-Accounts und einer ContributorID

4.2.2.1.1 Informationsangebote

Die Pilotierung zeigte auf, dass viele Institutionen Interesse haben, ihre Metadaten für GovData bereitzustellen, hierfür aber noch zusätzliche Informationsangebote benötigt werden. Ein Informationsbereich auf dem Portal sollte es potenziellen Datenbereitstellenden ermöglichen, mit geringem Zeitaufwand die folgenden Fragen zu klären:

- Ist die Institution berechtigt ihre Metadaten bei GovData.de bereitzustellen?
- Welcher Weg der Bereitstellung der Metadaten ist für die Institution am besten geeignet?
- Was gibt es bei der Bereitstellung der Metadaten zu beachten?

Hierfür kann der im Rahmen von Data Go entstandene Step-by-Step-Anleitung⁵⁵ (Datenbereitstellung GovData – GovData) ergänzt werden. Hier sollten auch die Hinweise unter Kapitel 4.2.2.2.3 Berücksichtigung finden.

4.2.2.1.2 Automatisierungen

Die Erstellung des Datenbereitsteller-Accounts und der ContributorID sollten für die Skalierung so automatisiert werden, dass neue Datenbereitstellende kurz nach der Erstellung eines Nutzeraccounts mit der Metadatenbereitstellung bzw. dem Testen der Metadaten beginnen können. GovData sollte durch das Heben von Automatisierungspotenzialen entlastet werden, um die Aufnahme neuer Datenbereitstellender skalierbarer zu gestalten.

4.2.2.2 Gruppe „Automatische Anbindung“

4.2.2.2.1 Motivation

Aus den Ergebnissen der Feedback-Gespräche zeigt sich, dass bei den Kommunen insgesamt ein Interesse an Open Data besteht und viele bereits über eigene Strategien für die Zugänglichkeit ihrer Daten verfügen. Die Anbindung an GovData über ein Landesportal wäre meist der präferierte Weg gewesen. In Ermangelung dieser Möglichkeit wurde der direkte Anschluss gewählt, da man sich davon eine bessere Sichtbarkeit und Zugänglichkeit der Daten verspricht.

Die bessere Sichtbarkeit im engeren Sinne ließe sich noch dadurch fördern, dass man neue und bestehende Datenbereitsteller auf GovData hervorhebt. Zum Beispiel durch:

- Hervorheben auf der Startseite, wenn ein:e Datenbereitstellende:r hinzukommt
- Blogposts/Interviews mit neuen Datenbereitstellern
- „Datensatz des Monats“ – Als Kategorie oder Blogpost, in Zusammenarbeit mit (neuen)

⁵⁵ GovData, „Anleitung zur Datenbereitstellung auf GovData“

Datenbereitstellenden

Im weiteren Sinne kann man den Wunsch nach Sichtbarkeit auch so auslegen, dass die Datenbereitstellenden untereinander besser miteinander vernetzt werden und so ein Ideen- und Erfahrungsaustausch stattfinden kann. Um die Vernetzung zu stärken, könnte die Kontaktpflege mit den Institutionen intensiviert werden. Dafür wären neue Veranstaltungsangebote notwendig, an denen auch interessierte Institutionen teilnehmen könnten, die noch keine Datenbereitsteller sind.

4.2.2.2.2 Dokumentation

Grundsätzlich wurde die Dokumentation des Standards gelobt, allerdings war es nicht immer auf Anhieb gelungen, die Dokumentation zu finden und die wesentlichen Punkte herauszufiltern. Die Optimierungspotentiale müssen bezogen auf folgende Aspekte getrennt betrachtet werden:

- Allgemeine Prozessinformationen
- Dokumentation der Anbindungsvarianten (insbesondere CKAN)
- Dokumentation des Validators
- Dokumentation des Standards

Da die eigentliche Arbeit zur Anbindung in weiten Teilen durch die Institutionen geleistet werden muss, finden sich die meisten Prozessverbesserungen im Bereich der Dokumentation. Wenn die Institutionen genauer wissen, was von ihnen erwartet wird und Hilfestellung bei der Umsetzung erhalten, erfolgt die Anbindung schneller.

4.2.2.2.3 Allgemeine Prozessinformationen

Die Informationsseite zur Datenbereitstellung auf GovData wurde wenig beachtet. Wünschenswert wäre es, wenn sie präsenter gemacht wird. Derzeit findet man sie unter „Informationen > Datenbereitstellung GovData“. Sie auf die oberste Ebene zu heben, erlaubt es auch, die Inhalte wiederum auf Unterseiten zu verteilen. Da der Umfang der „Anleitung zur Datenbereitstellung auf GovData“ zunehmen wird, wenn man sie zu einem „Single Point of Truth“ ausbaut, wird eine stärkere Gliederung die Übersichtlichkeit erhöhen. Eine mögliche Gliederung wäre:

- Einleitung und Übersicht
- Verwaltungstechnische Grundlagen
- Erstellen eines Datenbereitsteller-Accounts
- Manuelle Datenbereitstellung
- Automatische Datenbereitstellung

Die Grundlage für den Ausbau der Seite kann die Präsentation „DCAT-AP.de in aller Kürze“ sein, die im Rahmen von Data Go erstellt wurde.

Die Informationen sollten medial aufgelockert werden. Neben den bestehenden Visualisierungen wäre ein Video denkbar, das den Prozess und den zu erwartenden Aufwand kurz beschreibt.

4.2.2.2.4 Dokumentation der Anbindungsvarianten

Die meisten Institutionen, die sich automatisch anbinden, verwenden eine Portalsoftware, die früher bereits an

GovData angeschlossen wurde. Im Rahmen von DataGo wurden CKAN-, DKAN-, DUVA- und InGrid-Portale erfolgreich an GovData angeschlossen, wobei CKAN am häufigsten verwendet wird. Die Dokumentationen dieser Anbindungsvarianten sollte erstellt bzw. konsolidiert werden.

Für alle Anbindungsvarianten sollten Step-by-Step-Anleitungen in Zusammenarbeit mit den bereits angeschlossenen Institutionen erstellt werden. Zukünftigen Datenbereitstellenden, die eine CKAN-, DKAN-, DUVA- und InGrid-Portal nutzen, würde so die Anbindung wesentlich erleichtert.

Eine weitere Hilfestellung wäre es, Erfahrungsberichte von bereits angeschlossenen Datenbereitstellenden zu veröffentlichen. Nutzer:innen gleicher Portale sollten (nach Absprache) Kontakt mit erfolgreich angebotenen Datenbereitstellenden aufnehmen können, damit diese als Multiplikator und Musterimplementierung fungieren können.

4.2.2.2.5 Erweiterung um eine Push-Anbindung

In einem Fall wurde der Wunsch nach einer Push-API als Alternative zum aktuell ausschließlich möglichen Pull-Harvesting geäußert, bei der die Datenbereitstellenden gezielt die Metadaten im DCAT-AP.de-Format liefern. Damit hätten die Datenbereitstellenden eine größere Autonomie, die Häufigkeit der Aktualisierung selbst zu steuern. Entsprechende Informationen müssten auf der GovData-Website angeboten werden. Eine Erweiterung des Portals um eine Push-API sollte grundsätzlich geprüft werden, ist aber offenbar auch bei anderen Datenbereitstellenden, wie z.B. Forschungseinrichtungen kein regulärer Use Case. Dennoch kann eine solche Lösung sinnvoll sein, wenn dadurch die Anbindung für verschiedene Institutionen verbessert werden kann.

4.2.2.2.6 CKAN-Erweiterung

Da CKAN die am häufigsten anzuschließende Software ist und GovData selbst eine Erweiterung ckanext-dcatde entwickelt hat, sind an dieser Stelle besonders einflussreiche Verbesserungen möglich.

Während technische Details auf GitHub verbleiben können, sollte auf GovData im Bereich der „Anleitung zur Datenbereitstellung auf GovData“ eine High-Level-Step-by-Step-Anleitung bereitstehen, die erklärt, wie man ein CKAN-Portal für das Harvesten durch GovData vorbereitet.

Es sollte geprüft werden, ob die Erweiterung ckanext-dcatde selbst in Zusammenspiel mit der Erweiterung ckanext-dcat sinnvoll in Richtung einer „Full-Service“-Lösung weiterentwickelt werden kann, die es Datenbereitstellenden ermöglicht ohne individuelle Entwicklungsarbeiten gültiges DCAT-AP.de zu erzeugen. Unter Bezugnahme der Erfahrungen der Pilotinstitutionen aus Data Go sollte weiter dokumentiert werden, welche Änderungen an den Metadaten und welche Anpassungen am System notwendig sind.

4.2.2.2.7 Dokumentation des Validators

Die stetige Verbesserung des Validators, bis hin zum „Release“, hat sich im Rahmen der Anbindung von Datenbereitstellenden als sehr hilfreich herausgestellt. Wenn ein:e Datenbereitstellende seine Metadaten erfolgreich validiert hat, war das anschließende Testharvesting meist ohne besondere Auffälligkeiten.

Um den Validator weiter zu verbessern, sollten die Fehlermeldungen konstant optimiert werden.

Die Nutzerbasis des Validators sollte erweitert werden. Dafür sollte seine Dokumentation verbessert werden. Dies kann zum Beispiel auch durch eine Videodemonstration erfolgen.

Langfristig kann darüber nachgedacht werden, die Integration des Validators in das Redaktionssystem, wie von der Stadt Frankfurt bereits vorgenommen, allen CKAN-Nutzern zur Verfügung zu stellen.

4.2.2.2.8 Dokumentation des Standards

Die Webversion des Standards wurde sehr gut aufgenommen. Allerdings wurde kritisiert, dass sie schwer auffindbar und sehr umfangreich ist.

Die Spezifikation⁵⁶ und das Konventionenhandbuch⁵⁷ sollten direkt von der Startseite⁵⁸ erreichbar sein.

Die Webseite von DCAT-AP.de sollte seine Funktion in Abgrenzung zu GovData besser erklären und auf die relevanten Unterseiten für Datenbereitsteller verlinken.

Die Forderung, den Umfang der Spezifikation zu verringern, kann mit Blick auf die Systematik von DCAT-AP.de nicht erfüllt werden. Allerdings sollte sich das Konventionenhandbuch stärker auf wesentliche Anforderungen für die Datenanlieferung an GovData konzentrieren.

Die Empfehlungen, die die Präsentation „DCAT-AP.de in aller Kürze“ zu den zu verwendenden Eigenschaften macht, sollten sich in einem Kapitel im Konventionenhandbuch widerspiegeln.

Die 41 Konventionen, die das Konventionenhandbuch formuliert, sollten auf ihre Notwendigkeit und Verständlichkeit hin überprüft werden.

4.2.2.2.9 Kontinuierliche Verbesserung der Metadaten

Für die kontinuierliche Verbesserung der Metadaten nach der erfolgreichen Anbindung an GovData wurde hauptsächlich die Erweiterung des Dashboards gewünscht. Eine E-Mail zum Stand der Metadatenqualität wäre nach den befragten Datenbereitstellenden maximal monatlich oder bei signifikanten Änderungen notwendig.

Die Informationen des Dashboards sollten dabei Fehlerkorrekturen auf Datensatzebene ermöglichen. Wenn möglich sollten Fehlermeldungen dabei auf die Vorgaben der Spezifikation oder des Konventionenhandbuchs verweisen. Idealerweise können sie maschinenlesbar abgerufen werden.

4.2.2.2.10 Prozessoptimierung des Testharvestings

Mit der Weiterentwicklung des Validators kommt dem Testharvesting weniger die Prüfung der Richtigkeit der DCAT-AP.de-RDF-Daten zu. Stattdessen wird vorrangig getestet, ob der Harvester die Katalogdatei erfolgreich abholen kann. Der bisher variabel gehandhabte Zeitpunkt für das Testharvesting („wenn der Validator kaum noch Fehler zeigt“) hat sich für die Pilotinstitutionen als ausreichend genau herausgestellt.

Nach dem erfolgten Testharvesting ließe sich der Prozess verschlanken, wenn SEITENBAU das Protokoll direkt an den/die Datenbereitstellenden sendet. Um die Anbindung zur „Automatischen Anbindung“ noch skalierbarer zu gestalten, sollte der Test mit dem Validator für die Anbindung ausreichen und keinen verpflichtenden Test durch den Dienstleister GovDats erfordern.

⁵⁶ DCAT-AP.de, „DCAT-AP.de Spezifikation 2.0“

⁵⁷ DCAT-AP.de, „DCAT-AP.de Konventionenhandbuch 2.0“

⁵⁸ DCAT-AP.de, „Herzlich willkommen auf DCAT-AP.de“, zuletzt abgerufen am 21.11.2022, <https://www.dcat-ap.de/>

4.2.2.3 Gruppe „Automatische Anbindung, inkl. Mapping“

4.2.2.3.1 Motivation

In allen betrachteten Forschungsinstitutionen bestand eine hohe intrinsische Motivation, offene Daten bereitzustellen, zum einen aus der Erwartung der Verschärfung rechtlicher Vorgaben, zum anderen um den eigenen Daten höhere Sichtbarkeit zu verleihen und damit Öffentlichkeitswirksamkeit zu erreichen. Dass das Thema auch in fachlichen Vereinigungen und behördlichen Austauschkreisen zusehends diskutiert wird, kam hinzu.

Ein grundlegendes Verständnis für Metadaten und Harvesting bestand bei den Institutionen aufgrund der Tatsache, dass sie ihre Metadaten bereits selbst in andere Portale ausspielen und vielfach auch mit bibliothekarischen Themen beschäftigt sind.

Um die Bedeutung von GovData im Forschungsdatenumfeld zu erhöhen, sollten zum einen Forschungsinstitutionen auf der Website klarer adressiert sowie die Spezifika von Forschungsdaten stärker reflektiert und zum anderen der Mehrwert einer Veröffentlichung auf GovData z.B. durch Darstellung von Use Cases herausgestellt werden.

Hinsichtlich DCAT-AP.de bestand bei Forschungsinstitutionen großer Informationsbedarf aufgrund der Tatsache, dass innerhalb der Institutionen andere Metadatenformate in Verwendung sind und eine Beschäftigung mit DCAT-AP.de aufgrund bisher fehlenden Kontakts zu GovData noch nicht stattgefunden hatte. Hierzu könnten die in der Gruppe „Manuelle Anbindung“ erarbeiteten Handreichungen (siehe 3.2.3.6.3) dienen.

4.2.2.3.2 Mapping als Microservice

Das Mapping selbst sollte als Microservice aus GovData heraus aufgesetzt werden (siehe Abbildung 52). Ein Microservice in diesem Kontext ist ein separates Modul, in dem ein standardisierter Prozess – hier: das Mapping von DataCiteXML auf DCAT-AP.de – ausgeführt wird. Damit wäre eine nachnutzbare und anpassbare Lösung für zukünftige Anwendungszwecke geschaffen.

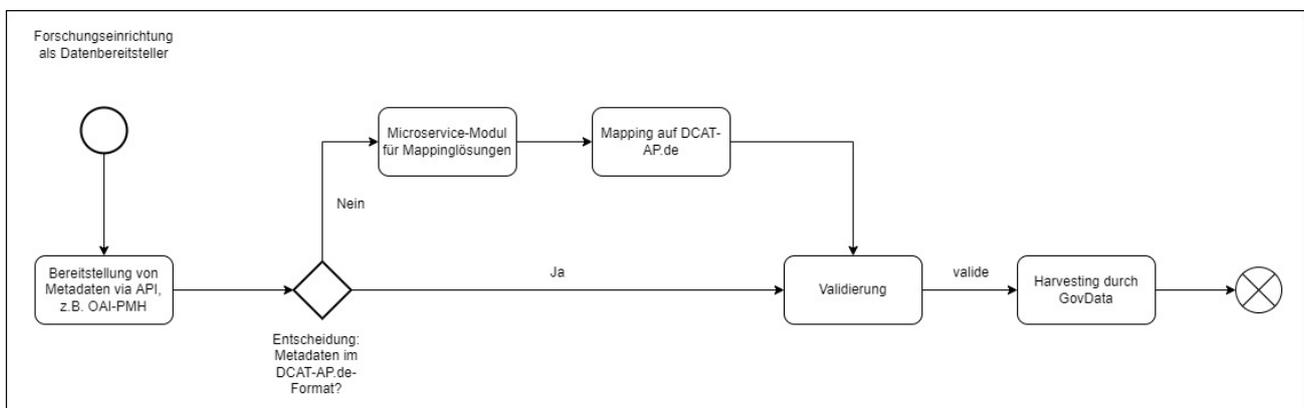


Abbildung 47: Einbindung des Mapping-Microservice in den Prozess

4.2.2.3.3 Mapping anderer Metadatenformate

Innerhalb der Pilotierung fiel auf Basis der Voraussetzungen der teilnehmenden Institutionen die Entscheidung, DataCiteXML als den Standard anzusehen, für den ein Mapping aufgrund der Verbreitung und der bereitgestellten Datenfelder am sinnvollsten erschien. Die Arbeit an einem Mapping für DataCiteXML wurde von den einzelnen Einrichtungen sehr begrüßt und stellt einen wichtigen Schritt hin zur Interoperabilität des Portals dar. Die Argumente gegen das ebenfalls weit verbreitete DublinCore werden auch in Zukunft ihre Gültigkeit behalten, möglicherweise wird sich auf Basis zukünftig anzubindender Forschungsinstitutionen jedoch die Sinnhaftigkeit der Erstellung weiterer Mappings zeigen.

4.2.2.3.4 Beschränkung auf DCAT-AP.de als Metadatenformat

Um die Zugänglichkeit des Portals für Forschungsinstitutionen zu verbessern, kam vereinzelt der Vorschlag, GovData möge das Spektrum der akzeptierten Metadatenstandards über DCAT-AP.de hinaus erweitern.

Aufgrund der Tatsache, dass DCAT-AP.de für die Anlieferung an GovData auf Beschluss des IT-Planungsrates verbindlich vorgeschrieben ist, erübrigt sich eine Diskussion der praktischen Vor- und Nachteile einer solchen Erweiterung. An dieser Stelle ist vielmehr auf die weitere Entwicklung des Mappings zu verweisen (siehe 4.2.2.3.2 und 4.2.2.3.3).

Eine praktische Lösung könnte auch im Falle solcher Forschungsinstitutionen bestehen, die ihre Daten bereits in übergeordnete Portale, wie im Falle OpenAgrar, ausspielen. Damit wäre für GovData eine gesammelte Anbindung für untergeordnete Institutionen wie bei der Anbindung über Länderportale möglich und die Normierung auf ein für ein Mapping sinnvolles Metadatenformat möglicherweise eher gegeben.

4.2.2.3.5 Schnittstellentest und Feedback zur Datenqualität

Unter dem Begriff „Schnittstellentest“ wurde im Einzelfall auch der Wunsch geäußert, dass eine automatische Validierung hinsichtlich der Frage stattfindet, ob im Rahmen eines Mappings valides DCAT-AP.de erzeugt wird und insbesondere die verpflichtenden Angaben vorhanden sind. Dieser Wunsch schließt an die stärkere Einbindung des Validators bei der automatisierten Einbindung an (siehe 4.2.2.2.7), dem die Transformation eines anderen Metadatenformats in DCAT-AP.de lediglich vorgelagert ist. Eine besondere Anforderung aus Sicht von Forschungsinstitutionen über die bereits gegebene Funktionalität des Validators hinaus ist hier nicht festzustellen.

Darüber hinaus wurde die persönliche Betreuung beim Anbindungsprozess allseits als sehr wichtig für die Identifikation mit dem Portal und den Erfolg der Anbindung betrachtet.

4.2.2.3.6 API-Katalog der Datensätze

Aus Sicht der Forschungsinstitutionen ergab sich zudem die Frage der Nachnutzbarkeit der Daten, bspw. für Forschungsprojekte. Hier kam im Einzelfall der Vorschlag auf, auf GovData einen durchsuchbaren Katalog der APIs bereitzustellen, über die die Datensätze bei den datenbereitstellenden Stellen abgerufen werden können. So wäre für Datennutzende ein schnellerer Überblick gewährleistet, um zu beurteilen, welche Datensätze für ihre eigene Nachnutzung automatisiert einbindbar wären, um eine breitere Verwendung multipler Datensätze aus unterschiedlichen Quellen zu ermöglichen.

4.2.2.3.7 *Strategische Positionierung*

Es wird klar, dass innerhalb des Forschungsbereichs andere Metadatenstandards verbreitet sind als bei den Verwaltungsdaten. Diese Lücke sollte dringend geschlossen werden, um GovData auch für andere Forschungsinstitutionen zu einer zentralen Anlaufstelle zu machen. Hierfür sollte der bereits aufgebaute Erstkontakt mit dem Nationale Forschungsdaten Infrastruktur e.V. (NFDI) als zentralem Verein für die Nutzung von Forschungsdaten in Deutschland dringend erweitert werden, da eine Anbindung der einzelnen NFDI-Konsortien einen großen Gewinn für GovData und Open Data allgemein darstellen würde. Dennoch sollte berücksichtigt werden, dass der Wartungsaufwand bei der Anbindung weiterer Standards konstant zunimmt.

4.2.2.4 **Gruppe „Manuelle Anbindung“**

4.2.2.4.1 *Optimierungsperspektiven für das GovData-Formular*

Die Erkenntnisse in dieser Gruppe beziehen sich maßgeblich auf mögliche Änderungen am GovData-Formular zum manuellen Einpflegen von Metadaten. Diese sind somit zu großen Teilen im Bereich der Optimierungspotentiale zu verorten, auch wenn Einzelaspekte durchaus in Richtung einer höheren Automatisierung zielen.

Folgende Bereiche für Verbesserungen lassen sich ausmachen:

- Das Formular sollte dynamischer aufgebaut sein und höhere Komplexität nur optional anbieten.
- Das Formular sollte semantisch verknüpfte Felder automatisch befüllen.
- Das Formular sollte sowohl im Gesamtaufbau wie auch bei einzelnen Datenfeldern die Nutzer:innen stärker führen.
- Einzelne Datenfelder sollten überarbeitet werden.
- Im Formular selbst sollten sich Hilfestellungen für die Befüllung finden.

Eine genauere Beschreibung der spezifischen Anforderungen findet sich im Dokument „Lastenheft“ (siehe Anhang 7.11).

4.2.2.4.2 *Art und Effekt der Änderungen*

Die aufgeführten Ergebnisse lassen eine grobe Unterscheidung in zwei Dimensionen zu: Zum einen im Hinblick auf die Art der vorgeschlagenen Änderungen, zum anderen im Hinblick auf den Effekt der vorgeschlagenen Änderungen.

Im Hinblick auf ersteres finden sich in den Handlungsfeldern sowohl Vorschläge, die letztlich Anpassungen allein auf textlicher Ebene bedeuten oder innerhalb des bestehenden Formulars kleinere Erweiterungen oder Streichungen beinhalten. Demgegenüber stehen Vorschläge, die mit einem grundsätzlichen Struktureingriff in das Formular gleichzusetzen sind und auch Fragen der Plattform-Technik berühren.

Im Hinblick auf zweitgenannte Dimension finden sich in den Vorschlägen sowohl solche, die den Nutzungskomfort erhöhen und damit mehr und bessere Metadaten erwarten lassen, als auch solche, die eine Lösung für die Einzelfälle bereitstellen, in denen Datenbereitstellende den Prozess des Dateneinpfegens bisher nicht zu Ende führen konnten oder gar nicht erst begonnen haben.

In beiden Dimensionen ist nicht immer eindeutig zu trennen zwischen den jeweiligen Werteausprägungen. Diese

sind entweder unter Einbeziehung der Voraussetzungen der technischen Plattform zu bewerten (Struktureingriff) oder einzelfallabhängig (Effekt).

4.2.2.4.3 Empfehlungen

Im Hinblick auf die Art der vorgeschlagenen Änderungen sollte GovData und der technische Dienstleister eine gemeinsame Analyse zur Umsetzbarkeit und möglicherweise auch zum zeitlichen Umsetzungshorizont vornehmen. Hierfür sei auf den Anhang des Dokuments „Lastenheft“ (siehe Anhang 7.11) verwiesen.

Im Hinblick auf den Effekt der Änderungen sollten Anforderungen zu Aspekten, die Institutionen aktuell das Einpflegen von Daten unmöglich machen, prioritär behandelt werden. Aufgrund der nicht zu klärenden Verallgemeinerbarkeit der Wichtigkeit der darüberhinausgehenden Anforderungen lässt sich keine weitere Priorisierung ableiten. Hier sollte nach dem Grundsatz verfahren werden: Was zeitnah umgesetzt werden kann, sollte umgesetzt werden.

Darüber hinaus bietet es sich an, eine Nachnutzung des Formulars für z.B. Landesportale anzudenken, um auch an diesen im Harvesting von GovData vorgelagerten Stellen DCAT-AP 2.0-Konformität in der Breite sicherzustellen.

4.2.2.5 Pflege und Weiterentwicklung

Im Rahmen des Projekts hat GovData zusammen mit seinem Kooperationspartner mehrere technische Hilfsmittel und Dokumentationen für den Regelbetrieb angefertigt. Im Falle der Weiterentwicklung von DCAT-AP.de wird es notwendig sein, die angefertigten Liefergegenstände auf die damit einhergehenden Vorgaben hin zu aktualisieren. Dabei können die Intervalle stark voneinander abweichen. Technische Komponenten wie das Datacite-Mapping oder Harvesting-Schnittstellen stehen in direkter Abhängigkeit zum DCAT-AP.de-Metadatenstandard. Sie sollten spätestens nach Veröffentlichung einer Folgeversion von DCAT-AP.de 2.0 angepasst werden.

4.2.2.6 Optimiertes Prozessmodell

In den vorangegangenen Kapiteln wurden bereits unterschiedliche Verbesserungsvorschläge beschrieben, davon einige mit dem Potential für Automatisierung und Prozessoptimierung. Im Rahmen des Projektes konnten erste kleinere Schritte bereits realisiert werden. So trugen die Verbesserungen am Validator⁵⁹ dazu bei, dass weniger Testharvestings auf Seiten GovDatas durchgeführt werden mussten als zunächst erwartet. Weitergehende Prozessoptimierungen konnten mit Blick auf den kurzen Projektzeitraum nicht im Rahmen von Data Go realisiert werden. Grundlegende Überarbeitungen sollten erst nach der Entscheidung über einem möglichen Wechsel der Portalsoftware, entsprechend den Ergebnissen aus Teilprojekt 3 bewertet, konkretisiert, priorisiert und umgesetzt werden.

Im Folgenden werden optimierte Prozesse für die Anbindungen skizziert, deren Grundlage neben den Erfahrungen der Geschäfts- und Koordinierungsstelle GovData und seiner Dienstleister vor allem das Feedback

⁵⁹ DCAT-AP.de, „DCAT-AP.de Validator“, zuletzt abgerufen am 21.11.2022, <https://www.itb.ec.europa.eu/shacl/dcat-ap.de/upload>

der Pilotinstitutionen (siehe Kapitel 4.2.1 und 4.2.2) bildet. Im Rahmen von Data Go wurde von neuen Datenbereitstellenden häufig der Bedarf an individuellen Beratungsgesprächen aufgezeigt. Entsprechend wurde darauf geachtet, dass neben den Automatisierungen auch Kanäle für Rückfragen und Austauschformate bestehen.

Zunächst werden die Schritte zur Erstellung des Datenbereitsteller-Accounts und der ContributorID beschrieben, welche alle neuen Datenbereitstellenden unabhängig von der Art der Anbindung durchlaufen müssen. Die weiteren Prozessschritte bis zur Anbindung werden dann in Abhängigkeit der gewählten Variante beschrieben.

4.2.2.6.1 Erstellung Datenbereitsteller-Account und ContributorID

Bei der Pilotierung wurde deutlich, dass die aktuellen Prozesse zur Erstellung des Datenbereitsteller-Accounts und der ContributorID aufwendig sind und nicht den Erwartungen der Nutzer entsprechen. Diese Probleme könnten mit einem Registrierungsverfahren für neue Datenbereitstellende gelöst werden, in dessen Verlauf alle nötigen Informationen innerhalb eines Schrittes auf GovData.de eingetragen werden. Der ungefähre Ablauf dabei könnte der folgende sein: Nach der Erstellung eines Accounts können sich Nutzer (im eingeloggten Zustand) als Datenbereitstellende:r registrieren. Sie geben hierzu den gewünschten Namen der datenbereitstellenden Organisation ein. Aus diesem Namen generiert das Portal automatisch eine ContributorID gemäß den Vorgaben von DCAT-AP.de. Diese ContributorID kann von dem/der Nutzer:in akzeptiert oder bearbeitet werden.

In einem nächsten Schritt wählen neue Datenbereitstellende die gewünschte Art der Anbindung an GovData aus. Sie können dazu bisher aus den folgenden Optionen wählen:

- Automatische Metadatenbereitstellung DCAT-AP.de über URL
- Automatische Metadatenbereitstellung DataCite über URL
- Manuelle Metadatenerfassung über Formular auf GovData.de
- Bitte um Rücksprache via Telefon / Mail / Videokonferenz

Bei der Abfrage sind alle relevanten Informationen über die Art der Anbindung und deren Voraussetzungen nutzerfreundlich aufbereitet einzusehen (Siehe Anforderungen aus 4.2.1). Im Rahmen von Data Go wurde von vielen Pilotinstitutionen ein Austauschtermin vor der Anbindung als wichtig empfunden. Zwar ist davon auszugehen, dass verbesserte Informationsbereitstellung einige der Fragen abdecken kann, dennoch sollte die Option an der Stelle durch eine Auswahl eröffnet werden. Diese Option kann auch genutzt werden, wenn neue Datenbereitstellende ihre Metadaten auf andere Weise anbieten möchten, z.B. über ein anderes Metadatenschema. Zum Abschluss des Dialogs bestätigt der neue Datenbereitstellende die Beantragung des Zugangs zur Datenbereitstellung.

Nach Absendung des Antrags erhält GovData eine E-Mail mit den Informationen über den Antrag und der Option zur Freigabe oder Rückfrage an die antragsstellende Organisation. Dies sollte im Regelfall innerhalb eines Werktags erfolgen.

Mit der Freigabe GovDatas werden die Prozesse zur Erstellung des Datenbereitsteller-Accounts und der ContributorID automatisiert angestoßen. Mit der aktuellen Systemsoftware CKAN in Kombination mit dem

Frontend Liferay umfasst dies insbesondere die folgenden Schritte:

- CKAN: Erstellen einer CKAN-Organisation entsprechend der Angaben des neuen Datenbereitstellenden
- CKAN: Zuordnen des/der Datenbereitsteller-Accounts zur CKAN-Organisation
- CKAN: Hinzufügen des Datenbereitsteller-Accounts zu allen CKAN-Gruppen (entsprechen den DCAT-Themen)
- Liferay: Hinzufügen des Datenbereitsteller-Accounts zu der Rolle Datenbereitsteller
- DCAT-AP.de - Liste der Datenbereitstellenden⁶⁰: Hinzufügen der neuen contributorID entsprechend der Angaben des neuen Datenbereitstellenden

Um eine bessere Skalierbarkeit zu erreichen, sollten die angegebenen Schritte automatisiert werden. Eine genauere Untersuchung der Möglichkeiten sollte dann entsprechend der Entscheidung über die zukünftige Infrastruktur erfolgen.

Nach Freigabe durch GovData erhalten die neuen Datenbereitstellenden alle erforderlichen Informationen zu den nächsten Schritten zur Anbindung per Mail. Damit ist der Prozess zur Erstellung der ContributorID und des Datenbereitsteller-Accounts abgeschlossen (siehe Abbildung 53). Die nächsten Schritte erfolgen danach abhängig von der gewählten Art der Anbindung.

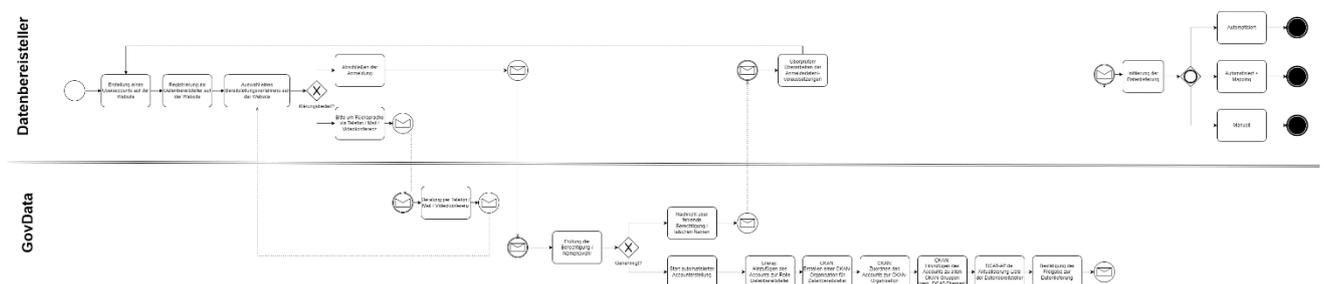


Abbildung 48: Datenbereitsteller-Account Soll

4.2.2.6.2 Manuelle Anbindung

Die neuen Datenbereitstellenden, die die manuelle Anbindung gewählt haben, erhalten mit der Freigabe-Mail alle nötigen Informationen, um die Daten manuell bei GovData einzutragen (siehe hierzu auch die Hinweise aus 4.2.1.4). Nach der Freigabe können die Datenbereitstellenden ihre Metadaten über das Erfassungsformular eingeben und bearbeiten. Damit ist die Anbindung abgeschlossen. Für die Optimierung des Prozesses der Metadatenerfassung über das Formular selbst, wurde aus dem Feedback der Pilotinstitutionen in Data Go ein Lastenheft angefertigt (siehe Anhang 7.11).

4.2.2.6.3 Automatische Anbindung DCAT-AP.de

Die neuen Datenbereitstellenden erhalten mit der Freigabe-Mail alle Informationen für die Voraussetzungen zur automatisierten Datenbereitstellung und zu den nächsten Schritten.

⁶⁰ DCAT-AP.de, "Liste der Datenbereitsteller"

Die neuen Datenbereitstellenden testen ihre Metadaten beim DCAT-AP.de-Validator mit einem Profil, das die Voraussetzungen zur Datenbereitstellung an GovData abschließend prüft. Die Fehlermeldungen des Validators werden in maschinenlesbarer und natürlichsprachlicher Variante auf die übermittelten DCAT-AP.de-Metadaten zurückgeführt. Die neuen Datenbereitstellenden erkennen mögliche Nachbesserungsbedarfe direkt anhand der von ihnen übermittelten Metadaten. Wenn die Prüfung erfolgreich ist, kann der getestete Endpunkt für das produktive Harvesting in das Profil des Datenbereitsteller-Accounts auf GovData.de übernommen werden. Über den Eintrag wird GovData per Mail informiert. GovData kann die Metadaten optional prüfen, den Endpunkt für das produktive Harvesting freigeben oder alternativ die Datenbereitstellenden kontaktieren, um mögliche Nachbesserungsbedarfe zu klären. Der Prozess ist schematisch in Abbildung 54 dargestellt.

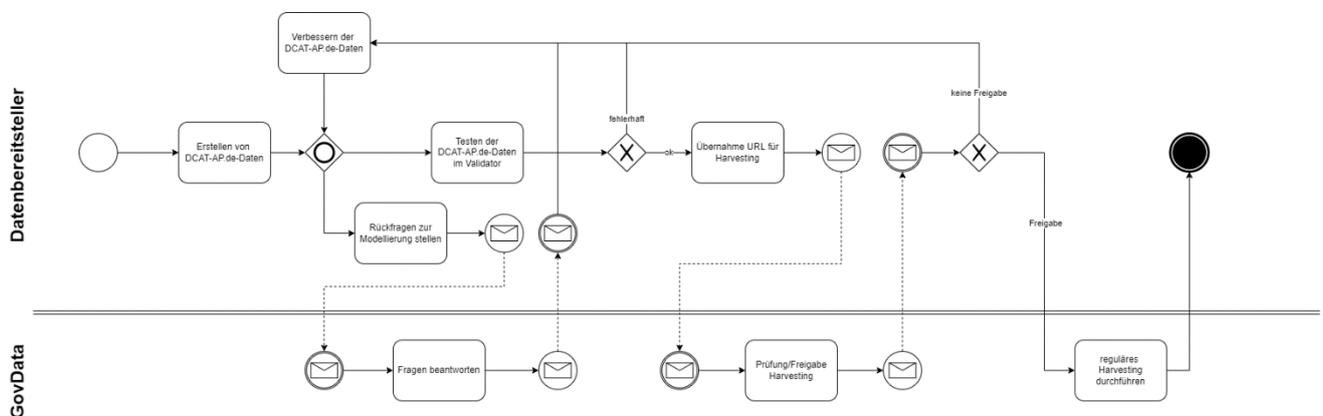


Abbildung 49: Automatische Anbindung Soll

4.2.2.6.4 Automatische Anbindung DataCite-XML

Die neuen Datenbereitstellenden erhalten mit der Freigabe-Mail alle Informationen für die Voraussetzungen und zu den nächsten Schritten zur automatisierten Datenbereitstellung mit DataCite-XML. Im Gegensatz zu der automatischen Anbindung über DCAT-AP.de stellen die neuen Datenbereitstellenden die Metadaten nach DataCite-XML beschrieben bereit, was in einem weiteren Schritt auf DCAT-AP.de gemappt werden muss.

Die neuen Datenbereitstellenden erhalten ein Online-Zugang zu einem Microservice-Mapping mit angebundenem DCAT-AP.de-Validator. Ihre DataCite-XML-Metadaten werden durch den Microservice auf DCAT-AP.de gemappt. Die daraus resultierenden DCAT-AP.de-Metadaten werden im Anschluss, analog zum Verfahren der automatischen Anbindung, mit dem DCAT-AP.de Validator getestet. Die Fehlermeldungen des Validators werden in maschinenlesbarer und natürlichsprachlicher Variante auf die Original DataCite-XML-Metadaten zurückgeführt. Hierzu müsste eine Erweiterung des DCAT-AP.de Validators entwickelt werden, die die Validierungsergebnisse zu DCAT-AP.de auf die gemappten DataCite Metadaten überträgt. Die neuen Datenbereitstellenden erkennen mögliche Nachbesserungsbedarfe direkt anhand den von ihnen übermittelten Metadaten. Wenn die Prüfung der gemappten Metadaten durch den Validator erfolgreich ist, kann der getestete DataCite-Endpoint für das Mapping und das anschließende Harvesting in das Profil des Datenbereitstelleraccounts auf GovData.de übernommen werden. Nach der Eintragung des Endpunktes wird GovData per Mail informiert. GovData kann die Metadaten optional prüfen, und ggf. die Datenbereitstellenden

kontaktieren, um mögliche Nachbesserungsbedarfe zu klären. Wenn die Metadaten den Vorgaben entsprechen, kann GovData die zusätzliche Quelle für das produktive Harvesting freigeben. Der Prozess ist schematisch in Abbildung 55 dargestellt.

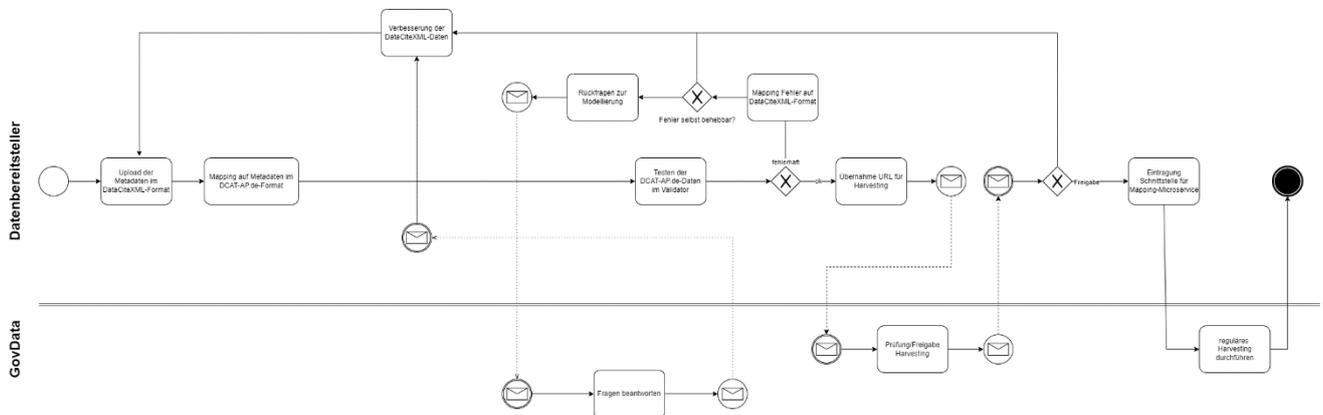


Abbildung 50: Automatische Anbindung, inkl. Mapping Soll

4.2.3 Fazit

Die Pilotierungen in TP2 dienten in erster Linie dazu, den Aufwand für die Anbindung an GovData zu ermitteln. Die Anbindung der Pilotinstitutionen und die Ermittlung der Aufwände ist mit der Differenzierung nach Art der Anbindung „Automatisch“, „Manuell“ und „Automatisch, inkl. Mapping“ gelungen.

Darüber hinaus sollte die begleitende Analyse Optimierungs- und Automatisierungspotenziale aufzeigen und nach Möglichkeit auch bereits umsetzen. In den Feedbackgesprächen äußerten die Vertreter:innen der Institutionen viele wertvolle Ideen, die in diesem Bericht dargestellt wurden. Im Rahmen von Data Go konnten mit neuen Informationsangeboten und Dokumentationen wie z.B. der Step-by-Step-Anleitung bereits erste Punkte umgesetzt werden und noch viele weitere Potenziale aufgezeigt werden.

Bei der Erstellung von DCAT-AP.de konformen Metadaten zeigte sich, dass der Validierung eine zentrale Rolle zukommt. Der DCAT-AP.de Validator konnte während des Data Go-Projektes von der Beta Phase in den ersten Produktivbetrieb überführt werden und stellte sich als sehr hilfreich für die teilnehmenden Institutionen heraus. Bei der Validierung gibt es, wie in anderen Bereichen der Anbindung, noch Automatisierungs- und Optimierungspotenziale. Diese können ergänzt mit qualifizierten Beratungsangeboten zur Skalierbarkeit des GovData-Portals hinsichtlich neuer Datenbereitstellender beitragen.

Bei den Pilotierungen wurde auch deutlich, dass GovData aktuell wenig attraktiv für die Datenbereitstellung für Forschungsinstitutionen ist. Für deren Anbindung scheint ein Mapping von Metadatenstandards aus dem Bereich Forschung unabdingbar. Im Rahmen von Data Go wurde daraufhin ein Mapping von Data-Cite-XML zu DCAT-AP.de weiterentwickelt. Wenn sich GovData dem Bereich Wissenschaft und Forschung aktiv öffnen will, sollten GovData neben der Implementierung des Mappings proaktiv auf diese Institutionen zugehen, deren Bedarfe erkennen und zusätzliche Anreize schaffen, wie bspw. Forschungsdaten im Portal ausweisen oder sogar auch die Marke „GovData – Verwaltungsdaten transparent, offen und frei nutzbar“ überarbeiten.

Abseits der Punkte, die das Portal betreffen wurde im Austausch mit den Datenbereitstellenden deutlich, dass

es auf allen Ebenen noch Bedarfe nach zielgruppenspezifischen Informationsangeboten gibt. Dies reicht von allgemeinen Fragen nach den Gründen von Open Data über juristische Fragen nach einer rechtssicheren Lizenzempfehlung hin zu Forderungen nach verbindlichen standardisierter Datenbereitstellungen. GovData sollte seine Rolle als bundesweite Informationsstelle für Open Data fortführen und nach Möglichkeit ausbauen.

5 Erstellung Kostenmodell

Ziel des Teilprojekts 4 war die Erstellung eines verursachergerechten Kostenmodells basierend auf den für die nächsten Jahre prognostizierten Daten der Teilprojekte 1 bis 3, um entstehende Mehrkosten durch Personalbedarf und Systemarchitektur abzudecken. In der finalen Ausarbeitung des Kostenmodells wurden die Systemarchitekturkosten (Ergebnis des Teilprojekts 3) für die Berechnung der Mehrkosten nicht berücksichtigt, da diese - laut Aussage der FITKO - bereits in die Finanzplanung für 2023 Eingang gefunden haben. Daher basieren die finalen Ergebnisse des Teilprojekts 4 auf dem ermittelten Personalbedarf, der aus den Ergebnissen der Teilprojekte 1 und 2 resultiert.

5.1 Zusätzlicher Mehraufwand

Die in Teilprojekt 2 erarbeiteten benötigten Personalaufwände (siehe 3.4) umfassen die Kapazitäten, die insgesamt benötigt werden. Da die GKSt GovData bisher im Rahmen ihrer bestehenden 2,32 VZÄ Anbindungen durchgeführt hat, mussten diese zunächst von den benötigten Aufwänden abgezogen werden. Hierfür wurde basierend auf den insgesamt 14 automatischen und manuellen Anbindungen der Jahre 2016 bis 2021 die durchschnittlich aufgebrauchten jährlichen internen Personalkapazitäten errechnet. Diese belaufen sich auf 0,02 VZÄ für automatische Anbindungen und 0,01 VZÄ für manuelle. Diese Kapazitäten bleiben ab 2023 bestehen, und wurden daher von den Ergebnissen der Personalbedarfsermittlung in jedem Szenario abgezogen (siehe Abbildung 56).

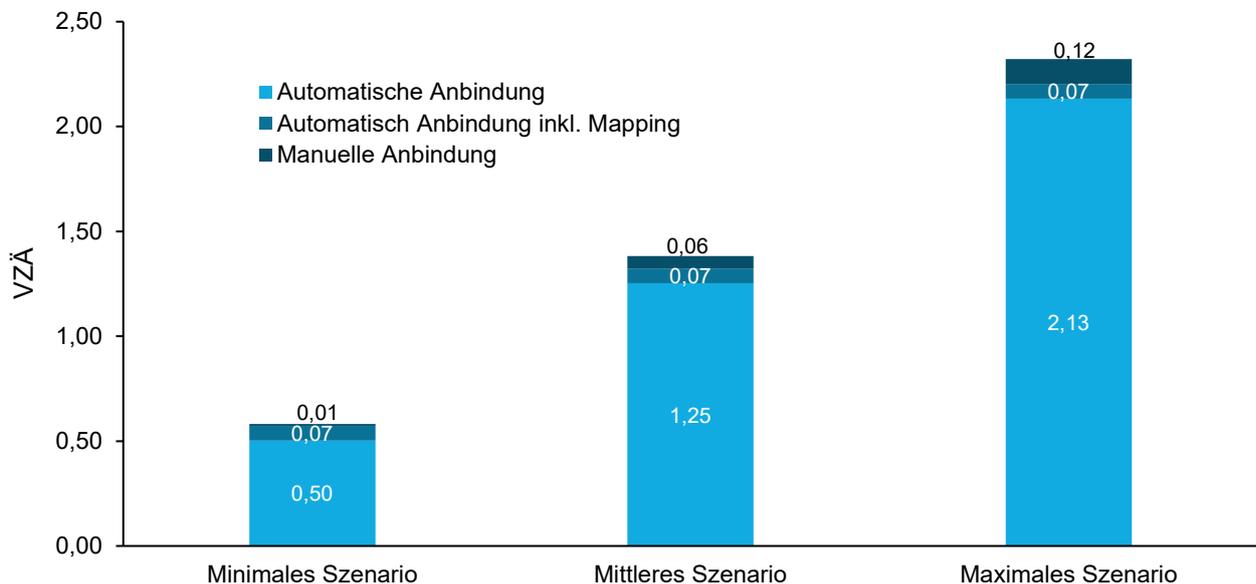


Abbildung 51: Prognostizierter zusätzlicher jährlicher Personalbedarf (2023-2025) nach Art der Anbindung und Szenario (in VZÄ)

Dieses Diagramm berücksichtigt die durch die GKSt GovData geleisteten und bereits mit Budget versehenen jährlichen VZÄ für die Anbindung von Institutionen. Der prognostizierte jährliche Personalbedarf variiert je nach Szenario, da hier auch jeweils unterschiedlich viele anzubindende Institutionen erwartet werden. Für die Begründung des Personalbedarfs wird das mittlere Szenario verwendet, während das minimale und maximale die Spanne darstellen.

5.2 Geschätzte Mehrkosten

Die geschätzten Mehrkosten für das zusätzlich benötigte Personal setzen sich zusammen aus dem Personalbedarf und den jeweiligen Kosten für ein Vollzeitäquivalent (VZÄ), differenziert nach internem und externem Personaleinsatz. Die Differenzierung, welche Aufgaben intern übernommen werden können bzw. bei welchen externe Expertise herangezogen werden muss, erfolgte basierend auf den Erfahrungswerten der GKSt GovData und unterscheidet sich je nach Aufgabe und somit Art der Anbindung. Je nach Szenario können zwischen 57,6% und 58,4% der Aufgaben durch interne Mitarbeitende bedient werden. Diese Anteile ergeben sich aus den unterschiedlichen Aufgabenzusammenstellungen je Anbindungsart in Kombination mit den unterschiedlichen Verhältnissen jeder Anbindungsart zur Zahl der Gesamtanbindungen. Wie in Kapitel 2.5 erläutert, dienen das minimale, mittlere und maximale Szenario der Darstellung einer validen Spanne einer erwartbaren Anzahl anzubindender Stellen und des daraus entstehenden Personalbedarfs inkl. der Personalkosten. Als ausschlaggebend für alle Empfehlungen ist jedoch das mittlere Szenario anzusehen.

Die jährlichen Gesamtkosten für den Personalbedarf i. H. v. 1,37 VZÄ im mittleren Szenario belaufen sich auf 251.647,33 €, mit einer Spanne von 108.948,15 € für 0,59 VZÄ im minimalen bis 422.744,67 € für 2,32 VZÄ im maximalen Szenario (siehe Tabelle 9). Die Kosten für eine interne Stelle, berechnet nach E14 TV-H, belaufen sich pro VZÄ und Jahr auf 106.961,00 €. Für eine:n externe:n Dienstleister:in auf Senior Consultant Level mit einem durchschnittlichen Tagessatz von 1.350 €⁶¹ ist mit 285.120,00 € pro VZÄ und Jahr zu rechnen.

⁶¹ Bundesverband Deutscher Unternehmensberatungen, "Studie: Honorare im Consulting steigen leicht an", zuletzt abgerufen am 21.10.2022, <https://www.bdu.de/newsletter/ausgabe-12020/studie-honorare-im-consulting-steigen-leicht-an/>

		Szenario								
		Minimal			Mittel			Maximal		
		Anzahl	VZÄ	Kosten in €	Anzahl	VZÄ	Kosten in €	Anzahl	VZÄ	Kosten in €
Gesamt	Gesamt	23	0,59	108.948,15	51	1,37	251.647,33	86	2,32	422.744,67
	Intern		0,33	34.806,71		0,79	84.236,92		1,34	143.746,64
	Extern		0,26	74.141,44		0,59	167.410,41		0,98	278.998,03
Anbindungsart										
<i>Automatisch</i>	Gesamt	18	0,51	93.346,53	44	1,25	229.266,73	75	2,13	391.325,43
	Intern		0,29	30.652,53		0,71	76.014,73		1,22	130.100,43
	Extern		0,22	62.694,00		0,54	153.252,00		0,92	261.225,00
<i>Automatisch inkl. Mapping</i>	Gesamt	2	0,07	13.343,92	2	0,07	13.343,92	2	0,07	13.343,92
	Intern		0,03	3.703,79		0,03	3.703,79		0,03	3.703,79
	Extern		0,03	9.640,13		0,03	9.640,13		0,03	9.640,13
<i>Manuell</i>	Gesamt	2	0,01	2.257,70	5	0,06	9.036,68	9	0,12	18.075,32
	Intern		0,004	450,38		0,04	4.518,40		0,09	9.942,41
	Extern		0,006	1.807,31		0,02	4.518,28		0,03	8.132,91

Tabelle 9: Zusätzliche Personalkosten pro Jahr nach Anbindungsart und je Szenario im Überblick

Die Tabelle listet für die Anbindungsarten (insgesamt und einzeln), wie viele Institutionen sich je geschätztem Szenario anbinden werden, wie viele VZÄ dafür benötigt werden und welche Kosten - differenziert nach intern und extern - diese VZÄ jeweils verursachen (in € pro Jahr). Durch gerundete Werte kann es zu abweichenden Summen kommen. Das mittlere Szenario, von welchem zentral ausgegangen wird, ist zur Hervorhebung grau hinterlegt. Werte < 0,01 werden zur besseren Nachvollziehbarkeit mit drei Nachkommastellen dargestellt.

5.3 Kostenmodell

Zur Deckung der durch zusätzlich benötigtes Personal entstehenden Mehrkosten erforderte der Projektauftrag die Entwicklung und Darstellung eines verursachergerechten Kostenmodells. Dies resultierte in einem Kostenmodell, in dem die Kosten pro Anbindung einer einzelnen Institution berechnet und dem Bund bzw. den jeweiligen Bundesländern in Rechnung gestellt werden (Kostenmodell 1, siehe Kapitel 5.3.1). Zur Entwicklung des entsprechenden Kostenmodells wurde zum Auftakt des Teilprojekts 4 mit der GKSt GovData ein Workshop durchgeführt. In diesem wurde sich auch kritisch mit den Potenzialen und Risiken eines verursachergerechten Kostenmodells auseinandergesetzt. In diesem Zusammenhang wurde zusätzlich zum verursachergerechten Kostenmodell ein alternativer Ansatz entwickelt, nach dem die Kosten entsprechend des GovData Verteilungsschlüssels durch den Bund und Bundesländer, die der Verwaltungsvereinbarung beigetreten sind, getragen werden würden (Kostenmodell 2, siehe Kapitel 5.3.2).

Als Ergebnis der Abwägung von Vor- und Nachteilen beider Modelle wird die Umsetzung des Kostenmodells 2 empfohlen. Im Folgenden werden beide Modelle aufgeführt und ihre Vor- und Nachteile beschrieben.

5.3.1 Kostenmodell 1: Individuelle Abrechnung pro Anbindung

Das aus dem Projektauftrag hervorgehende Kostenmodell „Individuelle Abrechnung pro Anbindung“ sieht vor, dass Bund und Bundesländern jede Anbindung von Institutionen Ihres Zuständigkeitsbereichs einzeln in Rechnung gestellt wird. Demnach wäre dem Prinzip der verursachergerechten Kostenverteilung Rechnung getragen. Dabei zu beachten ist jedoch, dass durch die individuelle Abrechnung neben den Personalkosten zusätzliche Verwaltungsaufwände (u. a. für Controlling, Rechnungslegung, Vertragserstellung, Regelung zum Umgang mit Abbrüchen, Mahnwesen etc.) entstehen, die hier in Form eines Verwaltungsgemeinzuschlags berücksichtigt wurden. Der Verwaltungsgemeinzuschlag wurde auf Basis der Erfahrungswerte der GKSt GovData und in Abstimmung mit der FITKO auf eine Pauschale von 5% der Gesamtpersonalkosten festgelegt, die für die jeweiligen Anbindungen zusätzlich anfallen.

Die Preise für eine einzelne Anbindung ergeben sich damit wie folgt:

$$\begin{aligned} & \textbf{Personalkosten} \\ & \textit{(Personalkosten je Anbindungsart pro Jahr / prognostizierte Anzahl anzubindender Institutionen pro} \\ & \textit{Jahr)} \\ & + \\ & \textbf{Verwaltungsgemeinzuschlag} \\ & \textit{(Gesamte Personalkosten pro Jahr * 5\%)} \end{aligned}$$

Damit würde eine automatische Anbindung im mittleren Szenario 5.457,32 € (minimales Szenario: 5.422,76 €, maximales Szenario: 5.463,45 €) kosten, eine automatische Anbindung inklusive Mapping 6.918,67 € (minimales Szenario: 6.908,80 €, maximales Szenario: 6.917,74 €), und eine manuelle Anbindung 2.054,05 € (minimales Szenario: 1.365,69 €, maximales Szenario: 2.254,15 €)⁶². Tabelle 10 gibt eine Übersicht über die Preise inklusive ihrer Zusammensetzung.

Dieses Kostenmodell bietet den zentralen Vorteil, dass die Abrechnung gemäß in Anspruch genommener Leistung erfolgen und es verursachergerecht ist. Im Vergleich zum alternativen Kostenmodell 2 ist es allerdings aufwändiger und kostenintensiver, weil es mit zusätzlichen Verwaltungsaufwänden verbunden ist. Weiterhin entsteht durch die Nennung der Kosten Preiselastizität, die im Rahmen des Projekts nicht berücksichtigt werden kann. Die Reaktionen von Institutionen, Bund und Ländern auf eine expliziten Preisnennung können also nicht eingeschätzt werden. Es besteht das Risiko, dass Bundesländer und der Bund ihre Institutionen dazu

⁶² Die unterschiedlichen Kosten für die gleichen Anbindungsarten zwischen den Szenarien kommen vor allem durch zwei Faktoren zustande: (1) Unterschiede in den Gesamtkosten abhängig vom Szenario und (2) Beeinflussung der gerundeten Ergebnisse durch die Nachkommastellen der Zwischenergebnisse.

anzuhalten, sich nicht anzubinden, um Ausgaben zu senken. Dies widerspricht einerseits dem Open Data-Prinzip und sorgt andererseits für eine geringere Planungssicherheit. Infolgedessen kann es zu einer unvollständigen Auslastung des eingestellten Personals kommen.

Szenario		Anbindungsart		
		Automatisch	Automatisch inkl. Mapping	Manuell
Minimal	Gesamt	5.422,76	6.908,80	1.365,69
	Personalkosten	5.185,92	6.671,96	1.128,85
	Verwaltungsgemeinzuschlag	236,84	236,84	236,84
Mittel	Gesamt	5.457,32	6.918,67	2.054,05
	Personalkosten	5.120,61	6.671,96	1.807,34
	Verwaltungsgemeinzuschlag	246,71	246,71	246,71
Maximal	Gesamt	5.463,45	6.917,74	2.254,15
	Personalkosten	5.217,67	6.671,96	2.008,37
	Verwaltungsgemeinzuschlag	245,78	245,78	245,78

Tabelle 10: Kosten pro Anbindung nach Anbindungsart und Szenario in €

Die Tabelle enthält die Kosten in €, die den Bundesländern und dem Bund pro Anbindung in Rechnung gestellt werden würden, differenziert nach Art der Anbindung und Szenario. Die Kosten setzen sich zusammen aus den Personalkosten (Gesamtkosten pro Anbindung pro Jahr geteilt durch die prognostizierte Anzahl anzubindender Institutionen) und einem Verwaltungsgemeinkostenzuschlag (5% der Gesamtkosten). Durch gerundete Werte kann es zu abweichenden Summen kommen. Das mittlere Szenario, von welchem zentral ausgegangen wird, ist zur Hervorhebung grau hinterlegt.

5.3.2 Kostenmodell 2: Umlage auf Bundesländer

Als Ergebnis des Workshops zu Beginn von Teilprojekt 4 wurde das Kostenmodell „Umlage auf Bundesländer“ definiert. Dieses sieht vor, die entstehenden Personalkosten auf die der Verwaltungsvereinbarung beigetretenen Bundesländer und den Bund zu verteilen. Die Höhe der Umlage je Entität richtet sich nach dem GovData Verteilungsschlüssel, der auf dem Königsteiner Schlüssel basiert. Da Sachsen-Anhalt und Saarland der Verwaltungsvereinbarung bisher nicht beigetreten sind, wurden die für sie errechneten Beträge anteilig auf die restlichen Beteiligten umgelegt. Die Beiträge variieren im mittleren Szenario zwischen 2.038,46 € für Bremen bis 46.248,75 € für den Bund, mit einem durchschnittlichen Beitrag von 16.776,49 €. Die Beträge für alle Teilnehmenden in allen drei Szenarien sind in Tabelle 11 dargestellt.

Dieses Kostenmodell bietet entscheidende Vorteile. Zunächst werden durch die einfachere administrative Umsetzung zusätzliche Verwaltungskosten vermieden. Infolgedessen fallen auch geringere Gesamtkosten als im Kostenmodell 1 an. Zudem wird auf die Differenzierung nach Anbindungstyp verzichtet. Dadurch werden

potenzielle Abgrenzungsschwierigkeiten, wie z. B., ob ein umfangreiches Mapping vorliegt, das die Mehrkosten in voller Höhe rechtfertigt, vermieden. Auch wirkt dieser Ansatz motivierender für die Bundesländer und den Bund, ihre Institutionen zur Anbindung anzuhalten, da die Kosten bereits gezahlt sind. Als Nachteil ist anzuführen, dass durch die pauschale Abrechnung die Möglichkeit besteht, dass die Kosten für einzelne Bundesländer in Relation zu den anzubindenden Institutionen anteilig höher ausfallen können.

Bund / Bundesland	Anteil nach GovData Verteilungsschlüssel 2015	Kosten Je Szenario (in €)		
		Minimal	Mittel	Maximal
Bund	17,79%	20.022,92 €	46.248,75 €	77.693,70 €
Baden-Württemberg	10,72%	12.066,30 €	27.870,62 €	46.820,11 €
Bayern	12,79%	14.398,12 €	33.256,64 €	55.868,14 €
Berlin	4,27%	4.802,19 €	11.092,05 €	18.633,64 €
Brandenburg	2,49%	2.803,50 €	6.475,49 €	10.878,24 €
Bremen	0,78%	882,53 €	2.038,46 €	3.424,42 €
Hamburg	2,14%	2.408,92 €	5.564,10 €	9.347,18 €
Hessen	6,11%	6.881,44 €	15.894,68 €	26.701,62 €
Mecklenburg-Vorpommern	1,63%	1.832,48 €	4.232,65 €	7.110,47 €
Niedersachsen	7,72%	8.693,37 €	20.079,86 €	33.732,35 €
Nordrhein-Westfalen	17,33%	19.501,26 €	45.043,82 €	75.669,53 €
Rheinland-Pfalz	3,96%	4.458,47 €	10.298,14 €	17.299,94 €
Saarland	0,99%	-	-	-
Sachsen	4,10%	4.609,85 €	10.647,79 €	17.887,32 €
Sachsen-Anhalt	2,22%	-	-	-
Schleswig-Holstein	2,80%	3.151,32 €	7.278,89 €	12.227,88 €
Thüringen	2,16%	2.435,46 €	5.625,39 €	9.450,15 €

Tabelle 11: Für Bund und Länder entstehende Kosten durch Umlegung der Mehrkosten

In der Tabelle sind Bund und alle Bundesländer enthalten sowie ihr Anteil an der Finanzierung von GovData entsprechend dem Verteilungsschlüssel von 2015. Die Anteile von Saarland und Sachsen-Anhalt, die der Verwaltungsvereinbarung (noch) nicht beigetreten sind, wurden anteilig auf die restlichen Teilnehmenden verteilt, um den vollständigen Ausgleich der Mehrkosten zu gewährleisten. Durch gerundete Werte kann es zu abweichenden Summen kommen. Das mittlere Szenario, von welchem zentral ausgegangen wird, ist zur Hervorhebung grau hinterlegt. Drei weitere Varianten der Kostenverteilung finden sich in Anhang 7.7.

5.4 Fazit

Durch die zu erwartende höhere Anzahl anzubindender Institutionen in den nächsten Jahren entstehen neue Bedarfe für Personal, das eingestellt bzw. beauftragt werden muss. Die entstehenden Mehrkosten in Höhe von 251.647,33 € für 1,37 zusätzliche VZÄ im mittleren Szenario, welches als zentral für Empfehlungen und Berechnungen angenommen wurde, teilen sich in interne und externe Personalkosten. Im Rahmen einer Anbindung durchgeführte Aufgaben können zu ca. 60% (entspricht 0,82 VZÄ) durch internes Personal abgedeckt werden, das z.T. schon durch die vorhandenen Kapazitäten der GKSt GovData abgedeckt ist (0,02 VZÄ). Für ca. 40% der Aufgaben müssen externe Dienstleistungen in Anspruch genommen werden (entspricht 0,59 VZÄ)⁶³. Somit belaufen sich die errechneten zusätzlichen Kosten in den Jahren 2023 bis 2025 für interne Leistungen auf 84.236,92 € pro Jahr und für die externen Dienstleistungen auf 167.410,41 € pro Jahr.

Die Mehrkosten können durch eines von zwei alternativen Kostenmodellen abgedeckt werden:

1. Ein individuelles Kostenmodell, in dem jede Anbindung in Rechnung gestellt wird, abgedeckt werden. Zusätzlich entstehen durch die individuellen Abrechnungen Verwaltungsaufwände, die pauschal mit 5% der Gesamtkosten festgelegt wurden. Somit würde im mittleren Szenario eine automatische Anbindung 5.457,32 €, eine automatische mit Mapping 6.918,67 € und eine manuelle 2.054,05 € kosten. Die Kosten für die Anbindung einer Institution würden entsprechend dem Bund bzw. dem entsprechenden Bundesland in Rechnung gestellt werden.

Dieses Kostenmodell würde eine verursachergerechte Variante darstellen, aber gleichzeitig kostenintensiver und aufwändiger sein. Es besteht das Risiko negativer Effekte von Preiselastizität durch die Benennung eines Preises pro Anbindung, was weiterhin zu geringerer Planungssicherheit und einer möglichen unvollständigen Auslastung des eingestellten Personals führen kann.

2. Ein gemeinschaftliches Kostenmodell, bei dem sich alle der Verwaltungsvereinbarung beigetretenen Partner an den Kosten beteiligen. Diesem läge der Verteilungsschlüssel GovDatas zugrunde, und Anteile noch nicht beigetretener Bundesländer (Saarland und Sachsen-Anhalt) würden auf die restlichen umgelegt werden. Die Kosten für die einzelnen Beitragszahlenden würden hierbei im mittleren Szenario im Mittel 16.776,49 € pro Jahr betragen, mit einer Spanne von 2.038,46 € (Bremen) bis 46.248,75 € (Bund).

Dieses Kostenmodell ist einfacher in seiner administrativen Umsetzung, wodurch geringere Verwaltungskosten zu erwarten sind. Außerdem ist innerhalb dieses Modells keine exakte Differenzierung zwischen Anbindungsarten notwendig, womit mögliche Abgrenzungsschwierigkeiten vermieden werden. Letztlich kann es auch eine größere motivierende Wirkung für die Beteiligten haben, ihre Institutionen zu einer Anbindung anzuhalten. Als negativen Aspekt ist zu vermerken, dass die Kosten für einzelne Beteiligte in Relation zu ihren anzubindenden Institutionen anteilig höher ausfallen könnte.

Aufgrund der überwiegenden Vorteile des zweiten Kostenmodells wird die Umsetzung einer Umlage auf die Bundesländer und den Bund empfohlen.

⁶³ Summen können aufgrund von Rundungen von Dezimalstellen von dem Gesamtergebnis abweichen.

6 Zusammenfassung und Ausblick

Im Rahmen des vorgelegten Berichts wurde der Beschluss 2021/48 des IT-Planungsrates umgesetzt. Basierend auf der in Teilprojekt 1 erhobenen Datengrundlage wurde eine Personalbedarfsermittlung sowie eine Planung der zukünftigen Systemarchitektur durchgeführt, um schließlich zwei Optionen für mögliche Kostenmodelle zu erarbeiten.

Fraglos ist der Blick in die Zukunft immer mit Unwägbarkeiten verbunden. Dies gilt insbesondere in einem Umfeld, das unter dem Einfluss von Regulatorik steht, die stetig weiterentwickelt wird. Zusammen mit der Datenanalyse ermöglichte die Online-Umfrage jedoch eine Annäherung an die zu erwartenden Mengengerüste zukünftiger Metadatenätze sowie neuer Datenbereitstellender. So lässt sich insgesamt festhalten, dass in den nächsten drei Jahren mit einem anhaltenden Wachstum der über GovData bereitgestellten Metadatenätze gerechnet werden kann (siehe Kapitel 2). Nach dem als maßgeblich angesehenen Szenario wird sich die Gesamtzahl der über GovData verfügbaren Metadaten in den nächsten drei Jahren gegenüber dem Stand von Juni 2022 nahezu verdoppeln. Zusätzlich werden gemäß den Ergebnissen der Umfrage die direkten Anbindungen an GovData stark zunehmen, sodass bis 2025 mit insgesamt knapp 200 direkt an GovData angebotenen Datenbereitstellenden zu rechnen ist. Dies spiegelt auch das allgemeine Empfinden sowohl der Projektbeteiligten als auch der interviewten Expertinnen und Experten wider, demnach das Thema Open Data zunehmend an Aufmerksamkeit und politischem sowie gesellschaftlichem Interesse gewinnt.

Durch die erwartete Zunahme der direkt an GovData anzubindenden Datenbereitstellenden ist ebenfalls mit Mehraufwänden auf Seiten der GKSt bzw. FITKO zu rechnen, die sich insbesondere in einem veränderten Personalbedarf niederschlagen. Um diesen zielführend und methodisch valide zu bemessen, wurde auf Grundlage des Organisationshandbuchs des Bundes (siehe Kapitel 3) eine Personalbedarfsermittlung durchgeführt. Berücksichtigt wurden dabei auch die unterschiedlichen Aufwände je nach Anbindungstyp, die im Rahmen der durchgeführten Anbindung der dreizehn Pilotinstitutionen erhoben wurden. Demnach ist im maßgeblichen Szenario für die in den Jahren 2023 – 2025 erwarteten Anbindungen mit insgesamt 1,40 VZÄ zu rechnen. 0,82 VZÄ davon sind intern zu erwarten, die Aufwände der restlichen 0,59 VZÄ werden von externen Dienstleistern erbracht.

Mit Blick auf die technische Infrastruktur von GovData wurde festgestellt, dass sowohl der Suchindex als auch das Portal dem antizipierten Wachstum durchaus gewachsen ist. Jedoch hat sich gezeigt, dass es zu Performance-Einbußen beim CKAN-Harvesting kommen kann, die mittelfristig optimiert werden müssen. Somit wird es notwendig sein, Systemanpassungen vorzunehmen, wie z.B. die Ablösung von Liferay sowie der damit einhergehenden Ersetzung entsprechender Funktionalitäten. Im Fokus der Überlegungen steht dabei auch der Austausch des CKAN durch Piveau (siehe Kapitel 4).

Zur Beantwortung der Ausgangsfrage – wie die zusätzlich erwarteten Aufwände auch auf jene Länder umgelegt werden können, die offene Daten nicht über eine zentrale Schnittstelle bereitstellen, wurden basierend auf den Ergebnissen der Teilprojekte 1-3 zwei Kostenmodelle entwickelt. Ausgehend von den ermittelten

Mehraufwänden wurden die erwarteten Mehrkosten i.H.v. insgesamt 251.647,33 € p.a. in Variante 1 des Kostenmodells ein verursachergerechter Ansatz entwickelt, der sowohl die Anbindungsart als auch die zusätzlichen Kosten für Verwaltungsmehraufwände berücksichtigt (siehe Kapitel 5.3.1). Variante 2 sieht eine Finanzierung der Kosten über den bisherigen Verteilungsmechanismus vor, da eine Anbindung der zu Beginn des Projekts noch nicht an GovData angebotenen Bundesländer derzeit anvisiert wird (siehe Kapitel 5.3.2). Somit würden bürokratische Zusatzprozesse (z.B. zur Buchhaltung und Rechnungslegung) entfallen und weitere Mehrkosten vermieden werden. Entsprechend empfehlen die Autoren dieses Berichts auch die Umlegung der Mehrkosten im Sinne der Variante 2 des Kostenmodells.

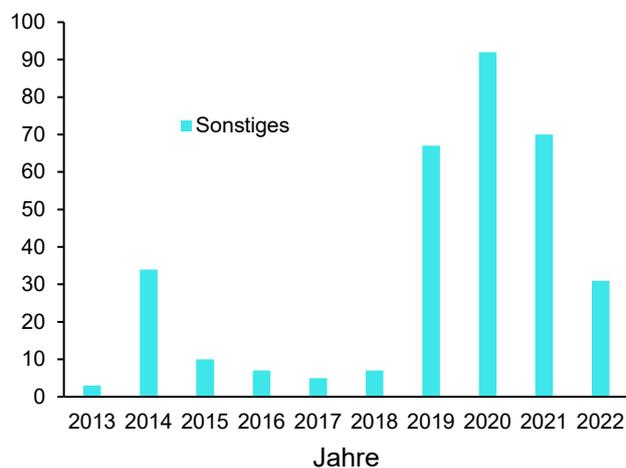
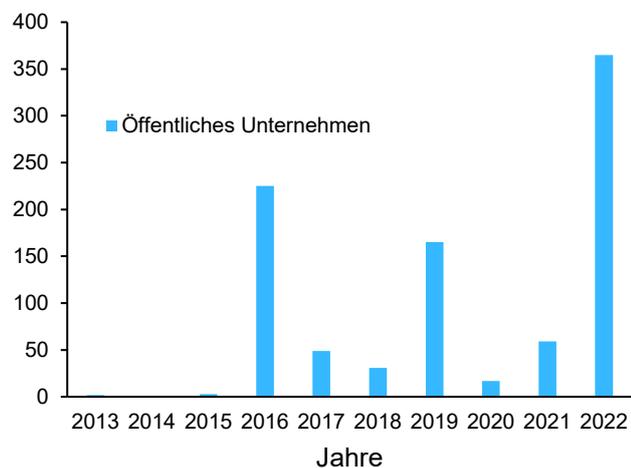
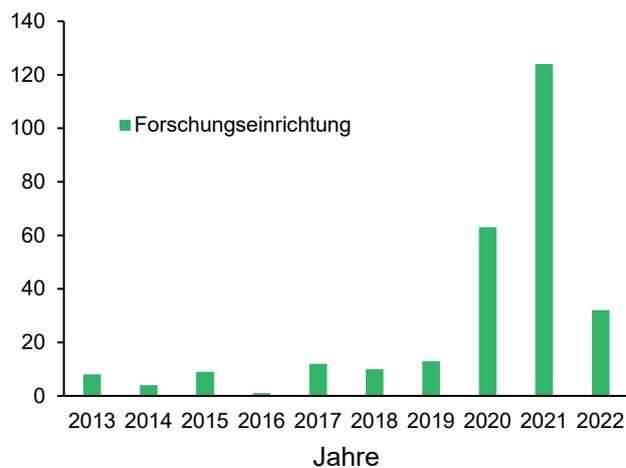
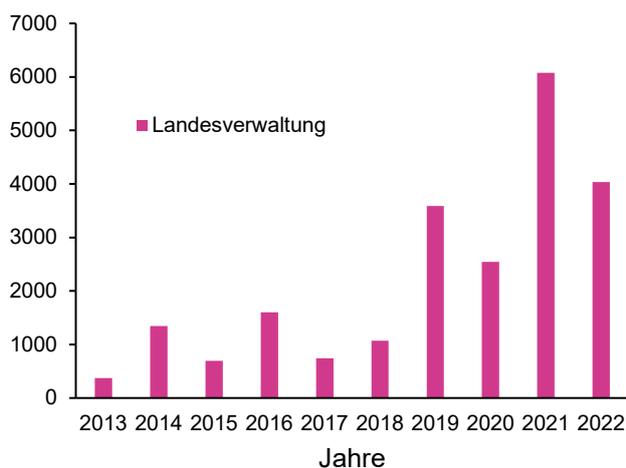
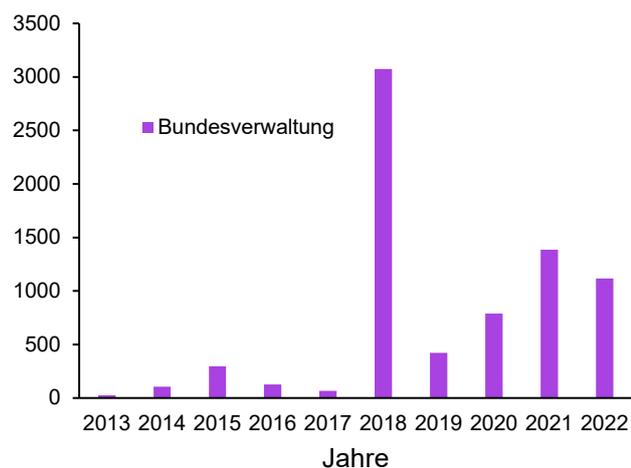
GovData wird ab dem 1. Januar 2022 in den Betrieb der FITKO überführt. Mit diesem Bericht wurden auch Vorschläge für mögliche Verbesserungen an der technischen Infrastruktur und den bestehenden Prozessen erarbeitet. Die Entscheidung über die Umsetzung dieser Vorschläge obliegt somit der FITKO in Abstimmung mit dem IT-Planungsrat. Bei der Durchführung des Berichtes und insbesondere bei der Entwicklung von Verbesserungsvorschlägen, als auch der Kostenmodelle, wurde die FITKO bereits eng eingebunden, um einen möglichst reibungsfreien Übergang zu gewährleisten.

Open Data gewinnt zunehmend an Bedeutung. Dies spiegeln auch die Ergebnisse dieses Berichts wider. Der Gesetzgeber hat dies ebenfalls mit der Einführung des DNG gewürdigt. Weitere Anpassungen der Rechtslage sind zudem nicht ausgeschlossen, wie die derzeitigen Gespräche über die Einführung einer – bislang noch nicht vorhandenen und über die Verpflichtungen des §12a EGovG hinausgehende – Veröffentlichungspflicht von Daten zeigen. Auch wenn eine solche Verpflichtung in den Prognosen dieses Berichts keine Berücksichtigung finden konnten, so sieht die GKSt GovData dennoch den Grundstein gelegt, dass GovData auch den Anforderungen der Zukunft gewachsen ist.

7 Anhang

7.1 Entwicklung der Veröffentlichung von Metadatensätzen auf GovData

Entwicklung der Veröffentlichung von Metadatensätzen auf GovData je Institutionsart



7.2 Fragenkatalog Online-Umfrage

Nr.	Gruppe	Fragentext	Typ	Antwortoptionen
1.1	Demografie	*Zu welcher Art von Institution zählt die Organisation, der Sie angehören?	Single Choice	Bundesverwaltung, Landesverwaltung, Kommunalverwaltung, Forschungseinrichtung, öffentliches Unternehmen, Sonstiges
1.2	Demografie	*Welchem Teil der Bundesverwaltung ist Ihre Behörde zuzuordnen?	Dropdown	Oberste Bundesbehörde, obere Bundesbehörde, Bundesmittelbehörde, Bundesunterbehörde, Anstalt des öffentlichen Rechts, Körperschaft des öffentlichen Rechts, Stiftung des öffentlichen Rechts
1.3	Demografie	*Welchem Bundesland ist Ihre Organisation schwerpunktmäßig zuzuordnen?	Dropdown	Baden-Württemberg, Bayern, Brandenburg, Bremen, Berlin, Hamburg, Hessen, Mecklenburg- Vorpommern, Niedersachsen, Nordrhein-Westfalen, Rheinland- Pfalz, Saarland, Sachsen, Sachsen- Anhalt, Schleswig-Holstein, Thüringen
2.1	Anbindung an ein Open Data Portal	*Ist Ihre Organisation bereits direkt an das GovData Portal angebunden? (Wenn Sie nicht direkt, sondern z.B. über ein Landes- oder Fachportal an GovData angebunden sind, klicken Sie hier bitte auf "Nein")	Single Choice	Ja, Nein, Weiß ich nicht
2.2	Anbindung an ein Open Data Portal	*Ist Ihre Organisation an ein Open Data Landesportal oder ein Fachportal angebunden? Beispiele sind etwa Open.NRW (Landesportal) oder mCLOUD (Fachportal des BMVI).	Single Choice	Ja, Nein, Weiß ich nicht
2.3	Anbindung an ein Open Data Portal	*Derzeit sind bereits acht Landesportale an GovData angeschlossen bzw. liefern Daten zu. Zusätzlich gibt es drei Fachportale, die Daten an GovData liefern. Bitte wählen Sie aus der Liste die Landes- oder Fachportale aus, an die Ihre Organisation angebunden ist:	Multiple Choice	Landesportal Berlin – „Berlin Open Data“, Landesportal Brandenburg – „Datenadler“, Landesportal Bremen – „Transparenzportal Bremen“, Landesportal Hamburg – „Transparenzportal Hamburg“, Landesportal Nordrhein-Westfalen – „Open.NRW“, Landesportal Rheinland-Pfalz – „Open-Government-Data-Portal Rheinland-Pfalz“, Landesportal Sachsen – „Open Data Portal“,

				Landesportal Schleswig-Holstein – „Open Data Schleswig Holstein“, Fachportal „Regionaldatenbank“, Fachportal Geodateninfrastruktur Deutschland – „GDI-DE“, Fachportal „mCLOUD“, Fachportal Destatis (Bundesamt für Statistik), Sonstiges, Weiß ich nicht
2.4	Anbindung an ein Open Data Portal	*Ist Ihre Organisation an ein Open Data Fachportal angebunden? Ein Beispiel für ein Fachportal ist etwa die mCLOUD (Fachportal des BMVI) oder der GDI-DE (Geodateninfrastruktur Deutschland)	Single Choice	Ja, Nein, Weiß ich nicht
2.5	Anbindung an ein Open Data Portal	*Derzeit gibt es drei Fachportale, die Daten an GovData liefern. Bitte wählen Sie aus der Liste die Fachportale aus, an die Ihre Organisation angebunden ist:	Multiple Choice	Fachportal „Regionaldatenbank“, Fachportal Geodateninfrastruktur Deutschland – „GDI-DE“, Fachportal „mCLOUD“, Fachportal Destatis (Bundesamt für Statistik), Sonstiges, Weiß ich nicht
3.1	Art der Anbindung	*Auf welche Art ist Ihre Organisation an GovData bzw. das Landes- oder Fachportal angebunden?	Single Choice	Automatisierte Anbindung (z.B. via Harvesting), manuelle Dateneingabe (z.B. via Eingabemaske auf GovData bzw. dem Landes- oder Fachportal), Weiß ich nicht
4.1	Angebundene Datensätze	*Wie viele Metadatenätze stellt Ihre Organisation derzeit bereits auf GovData oder über ein Landes- oder Fachportal bereit? Bitte schätzen Sie die aktuelle Gesamtzahl der bereitgestellten Datensätze.	Dropdown	0, 1-10, 11-50, 51-100, 101-250, 251-500, >500, Weiß ich nicht
4.2	Angebundene Datensätze	*Wie viele Metadatenätze plant Ihre Organisation in den nächsten 3 Jahren schätzungsweise auf GovData oder über ein Landes- oder Fachportal jährlich neu bereitzustellen? Bitte wählen Sie für jedes Feld einen Wert. Für den Fall, dass sie z.B. planen, erst ab 2025 Daten veröffentlichen, wählen Sie für die Jahre davor bitte den Wert "0" aus.	Array	Je Jahr 2023 bis 2025: 0, 1-10, 11-50, 51-100, 101-250, 251-500, >500, Weiß ich nicht
5.1	Motivation und Hemmnisse	Erlebt Ihre Organisation regelmäßig Hindernisse beim Bereitstellen von Metadaten auf GovData, die den Prozess erschweren oder vereinzelt unmöglich machen?	Single Choice	Ja, Nein, Weiß ich nicht, Keine Antwort
5.2	Motivation und Hemmnisse	Welche Hindernisse erleben Sie bei der Datenbereitstellung?	Multiple Choice	Fehlende Personalkapazitäten, Bedenken bzgl. Datenschutz/Sicherheit, Mangelnde Datenqualität (Datenformate, Unvollständigkeit), Unklare/fehlende Standardisierungsprozesse,

				Mangelnde Anwenderfreundlichkeit der Eingabemaske für Daten im Portal, Unzureichende Unterstützung durch GovData bei Problemen, Weiß ich nicht, Sonstiges
6.1	Anderweitige Veröffentlichung	*Veröffentlicht Ihre Organisation auf eine andere Art und Weise Daten im Sinne von Open Data?(z.B. Geodaten, Statistiken zur Bevölkerungsentwicklung, wirtschaftliche Kennzahlen, Literaturkataloge, ...)	Single Choice	Ja, Nein, Weiß ich nicht
6.2	Anderweitige Veröffentlichung	*Plant Ihre Organisation, in den nächsten drei Jahren Daten im Sinne von Open Data zu veröffentlichen?	Single Choice	Ja, Nein, Weiß ich nicht
6.3	Anderweitige Veröffentlichung	*Wie viele Daten im Sinne von Open Data hat Ihre Organisation schätzungsweise derzeit bereits veröffentlicht? Bitte schätzen Sie die Gesamtzahl der von Ihrer Organisation veröffentlichten Daten.	Dropdown	0, 1-10, 11-50, 51-100, 101-250, 251-500, >500, Weiß ich nicht
6.4	Anderweitige Veröffentlichung	*Wie viele Daten im Sinne von Open Data plant Ihre Organisation schätzungsweise in den nächsten drei Jahren jährlich neu zu veröffentlichen? Bitte wählen Sie für jedes Feld einen Wert. Für den Fall, dass sie z.B. planen, erst ab 2025 Daten veröffentlichen, wählen Sie für die Jahre davor bitte den Wert "0" aus.	Array	Je Jahr 2023 bis 2025: 0, 1-10, 11-50, 51-100, 101-250, 251-500, >500, Weiß ich nicht
6.5	Anderweitige Veröffentlichung	*Was hindert Ihre Organisation daran, auch in Zukunft Daten im Sinne von Open Data zu veröffentlichen?	Multiple Choice	Open Data ist keine Priorität bei uns im Haus, Fehlende Personalkapazitäten, Fehlende Expertise, Bedenken beim Datenschutz/Sicherheit, Zu hohe Kosten, Geringer Nutzen, Technische Schwierigkeiten, Rechtliche Hürden, Weiß ich nicht, Sonstiges
7.1	Schnittstellen	*Existieren hinsichtlich der Daten, die Sie schon veröffentlichen bzw. bei denen eine Veröffentlichung geplant ist, bereits Schnittstellen, die eine automatisierte Anbindung ermöglichen oder erleichtern würden?(z.B. CKAN API, CSW)	Single Choice	Ja, Nein, Weiß ich nicht
7.2	Schnittstellen	*Plant Ihre Organisation in Zukunft die Entwicklung und Nutzung von Schnittstellen zur automatisierten Veröffentlichung von Daten?	Single Choice	Ja, Nein, Weiß ich nicht
7.3	Schnittstellen	*Welche Schnittstellen sind dies?	Multiple Choice	CKAN API, CSW, Sonstiges, Weiß ich nicht
8.1	Metadatenstandards	*Verwendet Ihre Organisation einen einheitlichen Standard zur Beschreibung offener Daten (Metadatenstandard)?	Single Choice	Ja, Nein, Weiß ich nicht
8.2	Metadatenstandards	*Welche dieser Metadatenstandards verwendet Ihre Organisation zur Beschreibung der Metadaten?	Multiple Choice	DCAT-AP.de – DCAT-AP Version für Deutschland,

				DataCite Metadata Schema, DCAT-AP – Data Catalog Vocabulary Application Profile, Dublin Core, EAD – Encoded Archival Description, ISO 19115 (INSPIRE), MARC – Machine-Readable Cataloging Standard, METS – Metadata Encoding & Transmission Standard, MODS – Metadata Object Description Schema, OAI-ORE – Open Archives Initiative Object Reuse and Exchange, Sonstiges, Weiß ich nicht
8.3	Metadatenstandards	*Verfügt Ihre Organisation über ein eigenes Open Data Portal zur Veröffentlichung der Daten?	Single Choice	Ja, Nein, Weiß ich nicht
8.4	Metadatenstandards	*Auf welcher Technologie baut das Open Data Portal Ihrer Organisation auf?	Multiple Choice	Apache FUSEKI, ArcGIS, CKAN, Dataverse, DKAN, DUVA, Entryscape, Fiware, Frictionless, GeoNetwork, GeoNode, GeoServer, Government Site Builder (GSB), InGrid, JKAN, Liferay, Open Data Soft, Piveau, PKAN, Socrata, SPIP, Swirrl, Virtuoso, Zenodo, Sonstiges, Weiß ich nicht
8.5	Metadatenstandards	*Sie haben angegeben, dass Ihre Organisation plant, in den nächsten 3 Jahren Daten zu veröffentlichen. Nutzt ihre Organisation bereits einen einheitlichen Standard zur Beschreibung der zu veröffentlichenden Daten (Metadatenstandard)?	Single Choice	Ja, Nein, Weiß ich nicht
8.6	Metadatenstandards	*Welche dieser Metadatenstandards verwendet Ihre Organisation zur Beschreibung der Metadaten?	Multiple Choice	Apache FUSEKI, ArcGIS, CKAN, Dataverse, DKAN, DUVA, Entryscape, Fiware, Frictionless, GeoNetwork, GeoNode, GeoServer, Government Site Builder (GSB), InGrid, JKAN, Liferay, Open Data Soft, Piveau, PKAN, Socrata, SPIP, Swirrl, Virtuoso, Zenodo, Sonstiges, Weiß ich nicht
9.1	Zukünftige Anbindung	*Besteht seitens Ihrer Organisation die Absicht, in den nächsten drei Jahren ein eigenes Open Data Portal zur Veröffentlichung von Daten aufzubauen?	Single Choice	Ja, Nein, Weiß ich nicht
9.2	Zukünftige Anbindung	*Auf Basis welcher Technologie möchte Ihre Organisation in Zukunft ein Open Data Portal aufbauen?	Multiple Choice	Apache FUSEKI, ArcGIS, CKAN, Dataverse, DKAN, DUVA, Entryscape, Fiware, Frictionless, GeoNetwork, GeoNode, GeoServer, Government Site Builder (GSB),

				InGrid, JKAN, Liferay, Open Data Soft, Piveau, PKAN, Socrata, SPIP, Swirrl, Virtuoso, Zenodo, Sonstiges, Weiß ich nicht
9.3	Zukünftige Anbindung	*Besteht seitens Ihrer Organisation die Absicht, sich in Zukunft direkt an GovData anzubinden?	Single Choice	Ja, Nein, Weiß ich nicht
9.4	Zukünftige Anbindung	*Was ist die Hauptmotivation Ihrer Organisation, sich in Zukunft an GovData anzubinden?	Multiple Choice	Die Anbindung ist geplant, um einer bundesrechtlichen Verpflichtung nachzukommen (z.B. § 12 EGovG, §7 IV DNG, IWG), Die Anbindung ist geplant, um einer landesrechtlichen Verpflichtung nachzukommen (z.B. EGovG, Open Data Verordnungen oder Transparenzgesetze der Länder), Die Anbindung ist geplant, weil eine uns weisungsbefugte Behörde dies angeordnet hat, Die Anbindung ist freiwillig geplant, Weiß ich nicht, Sonstiges
9.5	Zukünftige Anbindung	*Besteht seitens Ihrer Organisation die Absicht, sich in Zukunft direkt an ein Open Data Landesportal oder ein Fachportal anzubinden? Beispiele sind etwa Open.NRW (Landesportal) oder mCLOUD (Fachportal)	Single Choice	Ja, Nein, Weiß ich nicht
9.6	Zukünftige Anbindung	*An welches Landes- oder Fachportal möchte Ihre Organisation sich in Zukunft anbinden?	Multiple Choice	Landesportal Berlin – „Berlin Open Data“, Landesportal Brandenburg – „Datenadler“, Landesportal Bremen – „Transparenzportal Bremen“, Landesportal Hamburg – „Transparenzportal Hamburg“, Landesportal Nordrhein-Westfalen – „Open.NRW“, Landesportal Rheinland-Pfalz – „Open-Government-Data-Portal Rheinland-Pfalz“, Landesportal Sachsen – „Open Data Portal“, Landesportal Schleswig-Holstein – „Open Data Schleswig Holstein“, 9.7 Fachportal „Regionaldatenbank“, Fachportal Geodateninfrastruktur Deutschland – „GDI-DE“, Fachportal „mCLOUD“, Fachportal Destatis (Bundesamt für Statistik), Sonstiges,

				Weiß ich nicht
9.7	Zukünftige Anbindung	*Besteht seitens Ihrer Organisation die Absicht, sich in Zukunft direkt an ein Open Data Fachportal anzubinden? Ein Beispiel ist etwa die mCLOUD (Fachportal des BMVI)	Single Choice	Ja, Nein, Weiß ich nicht
9.8	Zukünftige Anbindung	*An welches Fachportal möchte Ihre Organisation sich in Zukunft anbinden?	Multiple Choice	Fachportal „Regionaldatenbank“, Fachportal Geodateninfrastruktur Deutschland – „GDI-DE“, Fachportal „mCLOUD“, Fachportal Destatis (Bundesamt für Statistik), Sonstiges, Weiß ich nicht
9.9	Zukünftige Anbindung	*Was ist die Hauptmotivation Ihrer Organisation, sich in Zukunft an ein Open Data Landes- oder Fachportal anzubinden?	Multiple Choice	Die Anbindung erfolgt, um einer bundesrechtlichen Verpflichtung nachzukommen (z.B. § 12 EGovG, §7 IV DNG, IWG), Die Anbindung erfolgt, um einer landesrechtlichen Verpflichtung nachzukommen (z.B. EGovG, Open Data Verordnungen oder Transparenzgesetze der Länder), Die Anbindung erfolgt, weil eine unweisungsbefugte Behörde dies angeordnet hat, Die Anbindung erfolgt freiwillig, Sonstiges, Weiß ich nicht
9.10	Zukünftige Anbindung	*Was hindert Ihre Organisation daran, sich an GovData oder ein Open Data Landes- bzw. Fachportal anzubinden?	Multiple Choice	Open Data ist keine Priorität bei uns im Haus, Fehlende Personalkapazitäten, Fehlende Expertise, Bedenken beim Datenschutz/Sicherheit, Zu hohe Kosten, Geringer Nutzen, Technische Schwierigkeiten, Rechtliche Hürden, Weiß ich nicht, Sonstiges

Tabelle 12: Fragenkatalog Online-Umfrage

*Der Fragenkatalog, der für die Online-Umfrage genutzt wurde, enthält 37 Fragen. Keine Teilnehmende:r hat alle Fragen beantwortet, sondern verfolgte einen Pfad basierend auf den jeweils gegebenen Antworten. Pflichtfragen sind mit * markiert.*

7.3 Glossar Online-Umfrage

Begriff	Definition / Erklärung
Anbindung (an GovData/Landesportal)	Eine technische Anbindung ermöglicht es voneinander unabhängigen Systemen, Daten miteinander auszutauschen. Bei der Anbindung an GovData geht es darum, dass Metadatenätze von einer Quelle übertragen werden. Die Anbindung an GovData erfolgt in der Regel über eine "automatisierte Anbindung". Alternativ können Metadatenätze auch manuell eingetragen werden.
automatisierte Anbindung	Eine automatisierte Anbindung ermöglicht das maschinelle Austauschen von Daten zwischen zwei Systemen via Schnittstellen und ohne manuelle Aufwände.
CKAN	CKAN ist ein Open Source Datenkatalog, der häufig das Grundgerüst für Open Data bildet. Auch GovData nutzt CKAN als zentralen Metadatenkatalog.
Daten	Daten bilden die einzelnen Werte innerhalb eines Datensatzes. Unter Daten wird in der IT jedwede Art von Information verstanden, die durch einen Computer interpretierbar ist und weiterverarbeitet werden kann. Daten sind allgemein eine digitale Repräsentation, bzw. Beschreibung, von Information.
Datenportal	Eine Software-Lösung (in der Regel eine Webseite) welche einen Katalog durchsuchbarer Metadaten und herunterladbarer Datensätze über eine Benutzeroberfläche und eine Schnittstelle anbietet.
Datensatz	Eine Sammlung inhaltlich zusammenhängender Daten in einer Datei. Der Datensatz enthält die verwertbaren Informationen für die User. Beschrieben wird ein Datensatz durch Metadaten.
DCAT-AP.de	DCAT-AP.de ist das gemeinsame deutsche Metadatenmodell zum Austausch offener Verwaltungsdaten von Bund und Ländern.
Datennutzungsgesetz (DNG)	Das Datennutzungsgesetz (Gesetz für die Nutzung von Daten des öffentlichen Sektors) löst das Informationsweiterverwendungsgesetz (IWG) ab. Es weitet die Bereitstellung offener Verwaltungsdaten der Bundesverwaltung aus, vereinfacht und verbessert die Nutzungsmöglichkeiten bereitgestellter öffentlich finanzierter Daten.
E-Government-Gesetz (EGovG)	Das EgovG (Gesetz zur Förderung der elektronischen Verwaltung) dient dem Ziel, die elektronische Kommunikation mit der Verwaltung zu erleichtern und Bund, Ländern und Kommunen zu ermöglichen, einfachere, nutzerfreundlichere und effizientere elektronische Verwaltungsdienste anzubieten.
Fachportal	Über Fachportale können Behörden offene Daten bereitstellen, die spezielle Themengebiete umfassen, z.B. Mobilitäts-, Statistik- oder Geodaten. Ein Beispiel von einem Fachportal ist die GDI-DE (Geodateninfrastruktur Deutschland).
GovData.de	Das Datenportal für Deutschland. Bietet einen einheitlichen, zentralen Zugang zu offenen Verwaltungsdaten aus Bund, Ländern und Kommunen. Ziel ist es, diese Daten an einer Stelle auffindbar und so einfacher nutzbar zu machen.
Harvesting	Ein „Abernten“ von Daten. Eine Computer Software Technik, mit der Informationen von Webseiten bzw. aus Metadatenkatalogen extrahiert/ kopiert werden können.
Informationsfreiheitsgesetz (IFG)	Jede natürliche Person hat gegenüber den Behörden des Bundes einen Anspruch auf einen freien Zugang zu amtlichen Informationen. Das IFG (Gesetz zur Regelung des Zugangs zu Informationen des Bundes) legt die Voraussetzungen fest, unter denen amtliche Informationen zugänglich gemacht werden sollen. Das Gesetz ist 2006 in Kraft getreten.
Landesportal	In Deutschland werden Daten, die z.B. im Zuge des behördlichen Verwaltungshandelns der Landesverwaltungen entstehen, teilweise über Open Data-Portale der jeweiligen Länder zur Verfügung gestellt. Diese Landesportale sind zugleich Open Data- und Informations-Plattformen. Ein Beispiel ist Open.NRW, das Landesportal der Landesregierung Nordrhein-Westfalen.
Manuelle Dateneingabe	Bei der manuellen Dateneingabe müssen z.B. Informationen zum Datensatz manuell über eine Maske (Formular) eingetragen werden.
Metadaten	Die beschreibenden Metadaten liefern die nötigen Informationen, um den Inhalt des Datensatzes auffindbar zu machen. Neben obligatorischen Feldern, wie Angaben zum

	Datenbereitstellenden und Titel des Datensatzes, können auch weitere Metainformationen beigefügt werden, die eine noch präzisere Darstellung der Daten ermöglicht und so bei der Wieder- und Weiterverwendung der Daten eine optimale Grundlage darstellen.
Metadatensatz	Ein Metadatensatz ist eine zusammenhängende Darstellung der Metadaten eines einzelnen Datensatzes nach einem bestimmten Muster/Standard.
Metadatenstandard	Metadatenstandards bieten Konventionen zur Beschreibung von Daten und beschreiben/definieren somit die Struktur von Metadaten. Durch einen festgelegten Metadatenstandard können Metadaten einfacher und umfassender ausgetauscht werden.
Offene Daten	Offene Daten sind frei zugängliche Rohdaten, die zur weiteren Verarbeitung und Aufbereitung öffentlich bereitgestellt werden. Diese Daten müssen maschinenlesbar sein.
Open Data	Open Data ist ein Prinzip, bei dem, durch die Bereitstellung offener Daten aus Wirtschaft, Wissenschaft und Verwaltung ein Mehrwert generiert werden kann. Daten können Grundlage für Innovation und Wirtschaftswachstum sein, indem neue Geschäftsmodelle auf Basis frei zugänglicher Rohdaten oder Applikationen für Computer und mobile Endgeräte entwickelt werden.
Open Data Portal	Ein Open Data Portal ist eine Software-Lösung (meistens eine Webseite), welche offene Daten über einen durchsuchbaren Katalog bereitstellt. Beispiele von Open Data Portalen in Deutschland sind z.B. „Berlin Open Data“ und „Open Data Portal Sachsen“.
Rohdaten	Bezeichnen die bei einer Erhebung unmittelbar gewonnenen Daten, die weder das Ergebnis einer Bearbeitung anderer Daten sind, noch eine Bearbeitung nach der Erhebung erfahren haben. Eine Ausnahme bilden Daten, die aus sachlichen oder rechtlichen Gründen bearbeitet werden müssen, wenn ohne diese Bearbeitung eine Veröffentlichung nicht möglich wäre.
Schnittstelle	Eine Schnittstelle ist die Verbindungsstelle zwischen zwei Systemen an der der Austausch von Daten erfolgt. Eine API ist eine Schnittstelle.
Technologie (für Open Data Portal)	In Bezug auf folgende Fragen: „Auf welcher Technologie baut das Open Data Portal Ihrer Organisation auf?“, „Auf Basis welcher Technologie möchte Ihre Organisation in Zukunft ein Open Data Portal aufbauen?“ Open Data Portale werden meist mit einer (Meta-)Datenkatalog-Software gebaut. Beispiele hierfür sind CKAN und InGRID. Solch eine Software kann Daten und Metadaten in unterschiedlichen Formaten bereitstellen und bildet oft das Grundgerüst für Open Data. GovData nutzt beispielsweise CKAN als Katalogsystem.

Tabelle 13: Glossar Online-Umfrage

Um das Verständnis der an der Umfrage Teilnehmenden zu erhöhen, wurde ein Glossar erstellt, das unterschiedliche Begriffe mit Bezug zu Open Data und Informationstechnologie beinhaltet.

7.5 SPARQL Abfrage für alle verfügbaren Metadaten auf GovData

Die folgende Abfrage wurde geschrieben, um die in für die Datenanalyse (Kapitel 2.2) benötigten Daten zu erhalten.

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX dcatde: <http://dcat-ap.de/def/dcatde/>

SELECT * WHERE {
  # To evaluate a single dataset uncomment the following line and replace the
  # placeholder URI with the URI of the dataset
  #BIND(<https://geoportal.sachsen.de/md/a4adc6dc-3f04-4e9f-853c-9e7ab9e6f65d> AS
  ?dataset)
  {
    # Get dataset values
    SELECT ?dataset ?p ?o WHERE {
      ?dataset ?a dcat:Dataset .
      ?dataset ?p ?o .
      FILTER(?p IN (dcatde:contributorID, dct:title, dcat:theme, dct:modified,
dct:issued))
    }
  }
  UNION
  {
    # Get foaf:name from dct:publisher in dataset
    SELECT ?dataset ?p ?o WHERE {
      ?dataset ?a dcat:Dataset .
      ?dataset dct:publisher ?b .
      ?b ?p ?o .
      FILTER(?p IN (foaf:name))
    }
  }
  UNION
  {
    # Get distribution values
    SELECT ?dataset ?p ?o WHERE {
      ?dataset ?a dcat:Dataset .
      ?dataset dcat:distribution ?b .
      ?b ?p ?o .
    }
  }
}
```

```
        FILTER(?p IN (dct:license, dct:format))
    }
}
UNION
{
    # Calculate distribution count per dataset
    SELECT ?dataset ("DistributionCount" as ?p) (count(?b) as ?o) {
        ?dataset ?a dcat:Dataset .
        ?dataset dcat:distribution ?b
    } GROUP BY ?dataset
}
}
ORDER BY asc(?dataset)
```

7.6 Zuordnung ContributorID

ContributorID	Zugeordneter Bereitsteller
http://dcat-ap.de/def/contributors/schleswigHolstein	Schleswig Holstein
http://dcat-ap.de/def/contributors/freistaatSachsen	Freistaat Sachsen
http://dcat-ap.de/def/contributors/transparenzportalHamburg	Transparenzportal Hamburg
Statistisches Amt für Hamburg und Schleswig-Holstein	Statistisches Amt fuer Hamburg und Schleswig-Holstein
http://dcat-ap.de/def/contributors/landBrandenburghttps://www.dcat-ap.de/def/contributors/20211108#landBrandenburg	Land Brandenburg
http://dcat-ap.de/def/contributors/destatis	Statistisches Bundesamt
http://dcat-ap.de/def/contributors/bundesministeriumDesInnernUndHeimat	Bundesministerium des Innern und Heimat
http://dcat-ap.de/def/contributors/bundesamtFuerJustiz	Bundessamt fuer Justiz
http://dcat-ap.de/def/contributors/bundesministeriumDesInnernUndHeimathttp://www.bka.de/	Bundesministerium des Innern und Heimat
http://dcat-ap.de/def/contributors/bundesministeriumFuerFamilieSeniorenFrauenUndJugend	Bundesministerium fuer Familie Senioren Frauen und Jugend
http://dcat-ap.de/def/contributors/derBundesbeauftragteFuerDenDatenschutzUndDieInformationsfreiheit	Der Bundesbeauftragte fuer den Datenschutz und die Informationsfreiheit
http://dcat-ap.de/def/contributors/bundesamtFuerJustizhttps://www.bundesjustizamt.de	Bundesjustizamt
http://dcat-ap.de/def/contributors/deutschesPatentUndMarkenamt	Deutsches Patent und Markenamt
http://dcat-ap.de/def/contributors/bundesministeriumFuerErnaehrungUndLandwirtschaft	Bundesministerium fuer Ernaehrung und Landwirtschaft
http://dcat-ap.de/def/contributors/bundesanstaltFuerArbeitsschutzUndArbeitsmedizin	Bundesanstalt fuer Arbeitsschutz und Arbeitsmedizin
http://dcat-ap.de/def/contributors/bundesministeriumDerFinanzen	Bundesministerium der Finanzen
http://dcat-ap.de/def/contributors/bundesministeriumFuerArbeitUndSoziales	Bundesministerium fuer Arbeit und Soziales
http://dcat-ap.de/def/contributors/bundesministeriumFuerWirtschaftUndKlimaschutz	Bundesministerium fuer Wirtschaft und Klimaschutz
http://dcat-ap.de/def/contributors/bundesamtFuerMaterialforschungUndPruefung	Bundesamt fuer Materialforschung und Pruefung
http://dcat-ap.de/def/contributors/bundesministeriumDerVerteidigung	Bundesministerium der Verteidigung
http://dcat-ap.de/def/contributors/bundessortenamt	Bundessortenamt
http://dcat-ap.de/def/contributors/generalzolldirektion	Generalzolldirektion
http://dcat-	Bundesamt fuer

ap.de/def/contributors/bundesinstitutFuerBauStadtUndRaumforschungImBundesamtFuerBauwesenUndRaumordnung	Bauwesen und Raumordnung
http://dcat-ap.de/def/contributors/dieBeauftragteDerBundesregierungFuerKulturUndMedien	Die Beauftragte der Bundesregierung fuer Kultur und Medien
http://dcat-ap.de/def/contributors/bundesverwaltungsamt	Bundesverwaltungsamt
http://dcat-ap.de/def/contributors/robertKochInstitut	Robert Koch Institut
http://dcat-ap.de/def/contributors/bundeszentralamtFuerSteuern	Bundeszentralamt fuer Steuern
http://dcat-ap.de/def/contributors/bundesarchiv	Bundesarchiv
http://dcat-ap.de/def/contributors/auswaertigesAmt	Auswaertiges Amt
http://dcat-ap.de/def/contributors/bundesministeriumDesInnernUndHeimat	Bundesministerium des Innern und Heimat
http://www.geodatenzentrum.de/geodaten/gdz?l=metadaten_bkg	
http://dcat-ap.de/def/contributors/stadtNuernberg	Stadt Nuernberg
http://dcat-ap.de/def/contributors/bundesversicherungsamt	Bundesversicherungsamt
http://dcat-ap.de/def/contributors/bundesministeriumDerFinanzenbmf_static	Bundesministerium der Finanzen
http://dcat-ap.de/def/contributors/govData	GovData
http://dcat-ap.de/def/contributors/bundesamtFuerVerbraucherschutzUndLebensmittelsicherheit	Bundesamt fuer Verbraucherschutz und Lebensmittelsicherheit
http://dcat-ap.de/def/contributors/itzBund	ITZBund
http://dcat-ap.de/def/contributors/bundesministeriumFuerWirtschaftlicheZusammenarbeitUndEntwicklung	Bundesministerium fuer wirtschaftliche Zusammenarbeit und Entwicklung
http://dcat-ap.de/def/contributors/maxRubnerInstitut	Max Rubner Institut
http://dcat-ap.de/def/contributors/openNRW	Open.NRW
http://dcat-ap.de/def/contributors/friedrichLoefflerInstitut	Friedrich Loeffler Institut
http://dcat-ap.de/def/contributors/bundesausgleichsamt	Bundesausgleichsamt
http://dcat-ap.de/def/contributors/bundesanstaltFuerMaterialforschungUndPruefung	Bundesanstalt fuer Materialforschung und Pruefung
http://dcat-ap.de/def/contributors/bundesministeriumFuerBildungUndForschung	Bundesministerium fuer Bildung und Forschung
http://dcat-ap.de/def/contributors/bundesamtFuerWirtschaftUndAusfuhrkontrolle	Bundesamt fuer Wirtschaft und Ausfuhrkontrolle
http://dcat-ap.de/def/contributors/openNRWhttps://www.offenesdatenportal.de	Open.NRW
http://dcat-ap.de/def/contributors/openNRW{http://dcat-ap.de/def/contributors/openNRW}	Open.NRW
http://dcat-ap.de/def/contributors/openNRWhttps://www.landesdatenbank.nrw.de/	Open.NRW
http://dcat-ap.de/def/contributors/openNRWhttps://opendata.ruhr/	Open.NRW
http://dcat-ap.de/def/contributors/openNRWhttp://offenedaten.aachen.de	Open.NRW
http://dcat-ap.de/def/contributors/openNRW Familie Flüchtlinge und OpenNRW MKFF11-Ministerium für Kinder	Open.NRW
http://dcat-ap.de/def/contributors/openNRWhttps://opendata.meerbusch.de	Open.NRW
http://dcat-ap.de/def/contributors/openNRWhttps://opendata.gelsenkirchen.de	Open.NRW
http://dcat-ap.de/def/contributors/openNRW Gesundheit und Soziales OpenNRW MAGS1-Ministerium für Arbeit	Open.NRW
http://dcat-ap.de/def/contributors/openNRWmarcomaas-Stadtverwaltung Hennef	Open.NRW

http://dcat-ap.de/def/contributors/openNRWqualisnrw-QUA-LiS NRW	Open.NRW
http://dcat-ap.de/def/contributors/openNRWkettler-Stadt Arnsberg	Open.NRW
http://dcat-ap.de/def/contributors/openNRWOpenNRWLZG-LZG	Open.NRW
http://dcat-ap.de/def/contributors/openNRWbendesjg-Landesbetrieb StraÙenbau Nordrhein-Westfalen	Open.NRW
http://dcat-ap.de/def/contributors/openNRWOpenNRWMIK1-Ministerium des Innern des Landes NRW	Open.NRW
http://dcat-ap.de/def/contributors/openNRWOpenNRWMSB1-Ministerium für Schule und Bildung NRW (MSB)	Open.NRW
http://dcat-ap.de/def/contributors/openNRWvolkergronau900-Landesbetrieb StraÙenbau Nordrhein-Westfalen	Open.NRW
http://dcat-ap.de/def/contributors/openNRWBLEodolter-Ministerium für Wirtschaft Nordrhein-Westfalen	Open.NRW
http://dcat-ap.de/def/contributors/openNRWmichaelknoben-Stadt Mönchengladbach	Open.NRW
http://dcat-ap.de/def/contributors/openNRWWilkening-Stadt Minden	Open.NRW
http://dcat-ap.de/def/contributors/openNRW Digitalisi InnovationOpenNRWWMIDE1-Ministerium für Wirtschaft	Open.NRW
http://dcat-ap.de/def/contributors/openNRWbrar0405-Bezirksregierung Arnsberg	Open.NRW
http://dcat-ap.de/def/contributors/openNRWOpenNRWTest1-Ministerium für Tests	Open.NRW
http://dcat-ap.de/def/contributors/openNRW Bau Gleichstellun Kommunalesbaiekaro-Ministerium Heimat	Open.NRW
http://dcat-ap.de/def/contributors/openNRWhttps://www.ig.nrw.de	Open.NRW
http://dcat-ap.de/def/contributors/openNRWKarbowiak-Landesbetrieb StraÙenbau Nordrhein-Westfalen	Open.NRW
http://dcat-ap.de/def/contributors/openNRWBerbigB-MAGS NRW	Open.NRW
http://dcat-ap.de/def/contributors/openNRWOpenNRWSEOEST1-Stadt Soest	Open.NRW
http://dcat-ap.de/def/contributors/openNRWhttps://www.opendata-oeplv.de/	Open.NRW
http://dcat-ap.de/def/contributors/openNRWOpenNRWVM1-Ministerium für Verkehr des Landes Nordrhein-Westf	Open.NRW
http://dcat-ap.de/def/contributors/openNRWkarrierenrw-Landesamt für Finanzen NRW	Open.NRW
http://dcat-ap.de/def/contributors/openNRWOpenNRWSTK1-Staatskanzlei Nordrhein-Westfalen	Open.NRW
http://dcat-ap.de/def/contributors/openNRWNRWTourismus-Tourismus NRW e.V.	Open.NRW
http://dcat-ap.de/def/contributors/openNRWPDliskaschneider-PD - Berater der öffentlichen Hand	Open.NRW
http://dcat-ap.de/def/contributors/openNRWseVRaS-Verkehrsverbund Rhein-Sieg GmbH	Open.NRW
http://dcat-ap.de/def/contributors/openNRWOpenNRWLIA1-Landesinstitut für Arbeitsgestaltung	Open.NRW
http://dcat-ap.de/def/contributors/openNRWOpenNRWPOLIZEI1-Polizei NRW	Open.NRW
http://dcat-ap.de/def/contributors/openNRWHSK552-Hochsauerlandkreis	Open.NRW
http://dcat-ap.de/def/contributors/openNRWOpenNRWWMKW1-Ministerium für Kultur und Wissenschaft des Landes	Open.NRW
http://dcat-ap.de/def/contributors/openNRWPierreSebrantke-Ministerium des Innern des Landes NRW	Open.NRW
http://dcat-ap.de/def/contributors/openNRWstutzpublicplan-publicplan GmbH (MAGS - Referat für Landesrecht Pf	Open.NRW
http://dcat-ap.de/def/contributors/openNRWBeykirch-Ministerium der Finanzen Nordrhein-Westfalen	Open.NRW
http://dcat-ap.de/def/contributors/openNRWhttps://opendata.dormagen.de	Open.NRW
http://dcat-ap.de/def/contributors/openNRWhttps://opendata.rhein-kreis-neuss.de	Open.NRW
http://dcat-ap.de/def/contributors/openNRWOpenNRWTest2-Kommune für Tests	Open.NRW
http://dcat-ap.de/def/contributors/openNRWben.riep-Stadt Ennigerloh	Open.NRW
http://dcat-ap.de/def/contributors/openNRWdfoerster-cosinex GmbH	Open.NRW
http://dcat-ap.de/def/contributors/openNRWhttps://www.offenedaten-wuppertal.de/	Open.NRW
http://dcat-ap.de/def/contributors/openNRWhttps://opendata.dortmund.de/Informationsportal	Open.NRW
http://dcat-ap.de/def/contributors/deutscheNationalbibliothek	Deutsche Nationalbibliothek
http://dcat-ap.de/def/contributors/openGovernmentDataPortalRheinlandPfalz	Open Government Data Portal Rheinland Pfalz
http://dcat-ap.de/def/contributors/berlinOpenData	Berlin Open Data
http://dcat-ap.de/def/contributors/gdiDE	GDI-DE
http://dcat-ap.de/def/contributors/mCloud	mCloud
http://dcat-ap.de/def/contributors/landBrandenburg	Land Brandenburg
http://dcat-ap.de/def/contributors/openNRWhttps://offenedaten-koeln.de/	Open.NRW
http://dcat-ap.de/def/contributors/openNRWhttps://offenedaten.kdvz-frechen.de/	Open.NRW
http://dcat-ap.de/def/contributors/openNRWhttps://open-data.bielefeld.de/	Open.NRW
http://dcat-ap.de/def/contributors/openNRWhttps://opendata-duisburg.de/	Open.NRW
http://dcat-ap.de/def/contributors/openNRWhttps://opendata.bonn.de/	Open.NRW

http://dcat-ap.de/def/contributors/openNRWhttps://opendata.duesseldorf.de/	Open.NRW
http://dcat-ap.de/def/contributors/openNRWhttps://opendata.essen.de/	Open.NRW
http://dcat-ap.de/def/contributors/schleswigHolsteinhttp://dcat-ap.de/def/contributors/schleswigHolstein	Schleswig Holstein
http://dcat-ap.de/def/contributors/openNRWhttps://opendata.stadt-muenster.de/	Open.NRW
http://dcat-ap.de/def/contributors/genesisDestatishttp://dcat-ap.de/def/contributors/genesisDestatis	Statistisches Bundesamt
http://dcat-ap.de/def/contributors/openDataHRO	Open Data Hansestadt Rostock
http://dcat-ap.de/def/contributors/regionalstatistikhttp://dcat-ap.de/def/contributors/genesisDestatis	Regionalstatistik
http://dcat-ap.de/def/contributors/regionalstatistikhttp://dcat-ap.de/def/contributors/regionalstatistik	Regionalstatistik
http://dcat-ap.de/def/contributors/transparenzportalBremen	Transparenzportal Bremen

Tabelle 14: Zuordnung ContributorID

7.7 Kostenmodell 2: Umlage auf Bundesländer – Variationen

Abhängig davon, welche Bundesländer der Verwaltungsvereinbarung beigetreten sind, ändern sich die Kosten, die jedes einzelne Bundesland für die durch das zusätzliche Personal entstehende Mehrkosten tragen müsste. Da ein Beitritt Sachsen-Anhalts und des Saarlands grundsätzlich möglich ist, wurden drei zusätzliche Versionen berechnet: (1) mit Sachsen-Anhalt, aber ohne Saarland, (2) mit Saarland, aber ohne Sachsen-Anhalt und (3) mit allen Bundesländern.

Bund / Bundesland	Anteil nach GovData Verteilungsschlüssel 2015	Kosten Je Szenario (in €)		
		Minimal	Mittel	Maximal
Bund	17,79%	19.574,71 €	45.213,46 €	75.954,50 €
Baden-Württemberg	10,72%	11.796,19 €	27.246,73 €	45.772,03 €
Bayern	12,79%	14.075,82 €	32.512,18 €	54.617,51 €
Berlin	4,27%	4.694,69 €	10.843,75 €	18.216,52 €
Brandenburg	2,49%	2.740,74 €	6.330,53 €	10.634,72 €
Bremen	0,78%	862,77 €	1.992,83 €	3.347,77 €
Hamburg	2,14%	2.354,99 €	5.439,54 €	9.137,94 €
Hessen	6,11%	6.727,40 €	15.538,87 €	26.103,89 €
Mecklenburg-Vorpommern	1,63%	1.791,46 €	4.137,90 €	6.951,30 €
Niedersachsen	7,72%	8.498,77 €	19.630,37 €	32.977,24 €
Nordrhein-Westfalen	17,33%	19.064,72 €	44.035,50 €	73.975,65 €
Rheinland-Pfalz	3,96%	4.358,67 €	10.067,61 €	16.912,67 €
Saarland	0,99%	-	-	-
Sachsen	4,10%	4.506,66 €	10.409,43 €	17.486,90 €
Sachsen-Anhalt	2,22%	2.438,84 €	5.633,21 €	9.463,27 €
Schleswig-Holstein	2,80%	3.080,78 €	7.115,95 €	11.954,15 €
Thüringen	2,16%	2.380,94 €	5.499,47 €	9.238,60 €

Tabelle 15: Für Bund und Länder entstehende Kosten durch Umlegung der Mehrkosten – Variante mit Sachsen-Anhalt, aber ohne Saarland.

In der Tabelle sind Bund und alle Bundesländer enthalten sowie ihr Anteil an der Finanzierung von GovData entsprechend dem Verteilungsschlüssel von 2015. In dieser Variante wurde der Anteil des Saarlands anteilig auf die restlichen Teilnehmenden verteilt, um den vollständigen Ausgleich der Mehrkosten zu gewährleisten. Durch gerundete Werte kann es zu abweichenden Summen kommen. Das mittlere Szenario, von welchem zentral ausgegangen wird, ist zur Hervorhebung grau hinterlegt.

Bund / Bundesland	Anteil nach GovData Verteilungsschlüssel 2015	Kosten Je Szenario (in €)		
		Minimal	Mittel	Maximal
Bund	17,79%	19.821,21 €	45.782,83 €	76.910,99 €
Baden-Württemberg	10,72%	11.944,74 €	27.589,84 €	46.348,43 €
Bayern	12,79%	14.253,07 €	32.921,60 €	55.305,31 €
Berlin	4,27%	4.753,81 €	10.980,31 €	18.445,92 €
Brandenburg	2,49%	2.775,25 €	6.410,25 €	10.768,65 €
Bremen	0,78%	873,64 €	2.017,92 €	3.389,92 €
Hamburg	2,14%	2.384,65 €	5.508,04 €	9.253,01 €
Hessen	6,11%	6.812,11 €	15.734,55 €	26.432,62 €
Mecklenburg-Vorpommern	1,63%	1.814,02 €	4.190,01 €	7.038,84 €
Niedersachsen	7,72%	8.605,79 €	19.877,57 €	33.392,52 €
Nordrhein-Westfalen	17,33%	19.304,80 €	44.590,04 €	74.907,22 €
Rheinland-Pfalz	3,96%	4.413,56 €	10.194,39 €	17.125,65 €
Saarland	0,99%	1.097,57 €	2.535,16 €	4.258,84 €
Sachsen	4,10%	4.563,41 €	10.540,52 €	17.707,12 €
Sachsen-Anhalt	2,22%	-	-	-
Schleswig-Holstein	2,80%	3.119,57 €	7.205,56 €	12.104,69 €
Thüringen	2,16%	2.410,92 €	5.568,72 €	9.354,94 €

Tabelle 16: Für Bund und Länder entstehende Kosten durch Umlegung der Mehrkosten – Variante mit Saarland, aber ohne Sachsen-Anhalt.

In der Tabelle sind Bund und alle Bundesländer enthalten sowie ihr Anteil an der Finanzierung von GovData entsprechend dem Verteilungsschlüssel von 2015. In dieser Variante wurde der Anteil Sachsen-Anhalts anteilig auf die restlichen Teilnehmenden verteilt, um den vollständigen Ausgleich der Mehrkosten zu gewährleisten. Durch gerundete Werte kann es zu abweichenden Summen kommen. Das mittlere Szenario, von welchem zentral ausgegangen wird, ist zur Hervorhebung grau hinterlegt.

Bund / Bundesland	Anteil nach GovData Verteilungsschlüssel 2015	Kosten Je Szenario (in €)		
		Minimal	Mittel	Maximal
Bund	17,79%	19.381,88 €	44.768,06 €	75.206,28 €
Baden-Württemberg	10,72%	11.679,99 €	26.978,32 €	45.321,13 €
Bayern	12,79%	13.937,16 €	32.191,90 €	54.079,48 €
Berlin	4,27%	4.648,44 €	10.736,93 €	18.037,07 €
Brandenburg	2,49%	2.713,74 €	6.268,17 €	10.529,96 €
Bremen	0,78%	854,27 €	1.973,19 €	3.314,79 €
Hamburg	2,14%	2.331,80 €	5.385,96 €	9.047,92 €
Hessen	6,11%	6.661,12 €	15.385,80 €	25.846,74 €
Mecklenburg-Vorpommern	1,63%	1.773,82 €	4.097,14 €	6.882,82 €
Niedersachsen	7,72%	8.415,05 €	19.436,99 €	32.652,38 €
Nordrhein-Westfalen	17,33%	18.876,92 €	43.601,71 €	73.246,91 €
Rheinland-Pfalz	3,96%	4.315,73 €	9.968,44 €	16.746,07 €
Saarland	0,99%	1.073,25 €	2.478,97 €	4.164,45 €
Sachsen	4,10%	4.462,26 €	10.306,89 €	17.314,64 €
Sachsen-Anhalt	2,22%	2.414,81 €	5.577,71 €	9.370,05 €
Schleswig-Holstein	2,80%	3.050,43 €	7.045,85 €	11.836,39 €
Thüringen	2,16%	2.357,48 €	5.445,29 €	9.147,59 €

Tabelle 17: Für Bund und Länder entstehende Kosten durch Umlegung der Mehrkosten – Variante mit allen Bundesländern

In der Tabelle sind Bund und alle Bundesländer enthalten sowie ihr Anteil an der Finanzierung von GovData entsprechend dem Verteilungsschlüssel von 2015. Durch gerundete Werte kann es zu abweichenden Summen kommen. Das mittlere Szenario, von welchem zentral ausgegangen wird, ist zur Hervorhebung grau hinterlegt.

7.8 Gesamtspezifikation GovData

- Systemspezifikationen: Gesamtsystemspezifikation (Pflichtenheft) –

GovData.de

Version: 1.3

Projektbezeichnung	GovData
Projektleiter	Ralph Nörenberg (SEITENBAU)
Verantwortlich	Anforderungsanalytiker (AN)
Erstellt am	25.06.2015
Zuletzt geändert	05.04.2023 13:13
Bearbeitungszustand	<input type="checkbox"/> in Bearbeitung
	<input type="checkbox"/> vorgelegt
	<input checked="" type="checkbox"/> fertig gestellt
Dokumentablage	
V-Modell-XT Version	1.4

Weitere Produktinformationen

Erzeugt	[nicht beteiligt]	Prüfer
	[nicht beteiligt]	QS-Verantwortlicher
	[nicht beteiligt]	Systemarchitekt
	[nicht beteiligt]	Systemintegrator
Erzeugung	Initial	

Änderungsverzeichnis

Datum	Version	Geänderte Kapitel	Beschreibung der Änderungen	Autor	Zustand
25.06.2015	0.1	Alle	Initiale Produkterstellung	Ralph Nörenberg	i.B.
21.07.2015	0.2	Alle	Erstellung User Stories und Inhalte der restlichen Kapitel erstellt	Alexander Rothbauer	i.B.
31.07.2015	1.0	Alle	Überarbeitung gemäß Kommentare, fehlende Kapitel ergänzt	Alexander Rothbauer	i.B.
25.11.2015	1.1	3.2.7, 3.2.1.11	Ortssuche der Kartensuche als eigene User Story erfasst Konkretisierung Hilfetexte	Ralph Nörenberg	vorgelegt
04.12.2015	1.2	3.2.1.7, 3.2.1.16, 3.2.21.3, 3.2.28.1, 4.1.1, 5.2, 7	Ergänzungen um relevante Themen für den Datenschutz	Ralph Nörenberg	vorgelegt
17.11.2022	1.3	Alle	Gesamtspezifikation an den aktuellen Stand anpassen	Ralph Nörenberg Antje Göldner	fertig gestellt

Prüfverzeichnis

Die folgende Tabelle zeigt einen Überblick über alle Prüfungen – sowohl Eigenprüfungen wie auch Prüfungen durch eigenständige Qualitätssicherung – des vorliegenden Dokumentes.

Datum	Geprüfte Version	Anmerkungen	Prüfer	Neuer Produktzustand
-------	------------------	-------------	--------	----------------------

7.8.1 Ausgangssituation und Zielsetzung

Über GovData bieten öffentliche Stellen aus Bund, Ländern und Kommunen Daten der Verwaltung an. So soll insbesondere Verwaltungsmitarbeitern, Bürgern, Unternehmen und Wissenschaftlern die Möglichkeit gegeben werden, über einen zentralen Einstiegspunkt auf Daten und Informationen der öffentlichen Verwaltung in Deutschland zugreifen zu können. Ziel ist es, dass diese Daten aus der Verwaltung besser genutzt und weiterverwendet werden, so dass durch neue Ideen sowie Kombination und Analyse neue Erkenntnisse aus den vorhandenen Daten gewonnen und neue Anwendungsfelder erschlossen werden können.

Das GovData-Portal umfasst einen in ein Content Management System integrierten Datenkatalog von Metadatensätzen zu offen bereitgestellten Daten aus verschiedenen Verwaltungsebenen.

2015 ist GovData im Rahmen einer Verwaltungsvereinbarung zwischen Bund und Ländern in den Regelbetrieb überführt worden. Die Betreuung des Portals ist dabei vom Bundesministerium des Innern zu einer Geschäfts- und Koordinierungsstelle im Land Hamburg übertragen worden.

Die Verantwortung für den Betrieb und die Weiterentwicklung des GovData Portals wird in 2023 an die FITKO übergehen.

7.8.2 Funktionale Anforderungen

7.8.2.1 Content Management System

7.8.2.1.1 Grundlage CMS

GovData ist auf Grundlage der Open Source Portalsoftware „Liferay“ in der Community-Variante realisiert, welches u.a. Content Management System (CMS) Funktionalitäten bereitstellt und über Portlets erweiterbar ist. Für die Anforderungen, die nicht durch ein bereits vorhandenes Liferay-Portlet abgedeckt werden können, entwickelt der AN eigene Portlets, die dann in das Portal eingebunden werden.

7.8.2.1.2 Grundsätzliche Funktionen des Webportals

Durch die in der eingesetzten Portalsoftware Liferay enthaltenen CMS-Funktionalitäten ist eine strukturell optimierte Darstellung für externe Suchmaschinen möglich.

7.8.2.1.3 SEO

In Liferay sind die generierten URLs standardmäßig „deep-link“-fähig. Soweit möglich wird versucht, die URLs auch menschenlesbar zu gestalten. Bereits im Prototyp enthaltene, menschenlesbare URLs werden erhalten, um die Funktionsfähigkeit bestehender externer Links zu erhalten.

Die eingesetzte Portalsoftware Liferay unterstützt die parallele Nutzung von mehreren Tabs/Fenstern in einem Web Browser.

Mit der eingesetzten Portalsoftware Liferay können Inhalte von mehreren Stellen referenziert werden. Duplicate Content wird vermieden.

7.8.2.1.4 Mehrsprachigkeit

Das Liferay-CMS unterstützt die Mehrsprachigkeit für den Redakteur sowie das Erstellen und Verwalten von mehrsprachigen Seiten. Dabei müssen keine echten Sprachen verwendet werden; auch eine Seite oder einzelne Inhalte in leichter Sprache können von den Redakteuren als Sprachversion verfasst werden. Über eine Sprachauswahl, die auf jeder Seite des Portals verfügbar ist, kann ein Nutzer des Portals jederzeit auf eine andere Sprachversion umstellen. Aktuell gibt es keine weitere Sprache neben der Sprache „Deutsch“, weshalb die Sprachauswahl für die Nutzer des Portals ausgeblendet ist.

7.8.2.1.5 Blog

Als Basis für die Blog-Funktionalitäten kommt das im Standardumfang des Liferay-Portals enthaltene Blog-Portlet zum Einsatz. Dieses ermöglicht es der Redaktion Blog-Beiträge zu verfassen, zu bearbeiten und auf den Portalseiten zu veröffentlichen. Die Zuordnung von Schlagworten wird ebenfalls bereits von diesem Blog-Portlet bereitgestellt wie auch Kommentar-, Pingback- und Trackback-Funktionalitäten.

7.8.2.1.6 Übergreifende Inhalte

Übergreifende Informationsseiten werden als gewöhnliche Portalseiten im CMS angelegt und gepflegt. Dies betrifft die folgenden Seiten:

- Impressum
- Kontakt (mit Webformular)
- Nutzungsbedingungen
- Datenschutzerklärung.

Diese Informationsseiten werden auf den Portalseiten im Header- und Footer-Bereich so verlinkt, dass sie von allen Portalinhalten (Informationsseiten, Blog, Suche, Ergebnisseite usw.) aus direkt aufgerufen werden können. Die Kontaktseite beinhaltet ein Webformular, mit dem Portalnutzer direkt aus dem Portal Anfragen an die Geschäfts- und Koordinierungsstelle per E-Mail senden kann.

Die Verhinderung von maschinellen Eingaben durch Bots erfolgt über die Umsetzung des Honeypot-Verfahrens, siehe Kapitel 7.8.2.1.15, Abschnitt „Schutz vor maschinellen Eingaben bei der Registrierung“.

7.8.2.1.7 Content Syndication und Social-Media-Angebote

Die Links zu den Social-Media-Kanälen von GovData werden als Buttons im Footer angezeigt und ermöglichen es den Nutzern, zu diesen Social-Media-Kanälen zu gelangen.

Die Platzierung im Footer verbessert die Benutzerfreundlichkeit, da bei sehr kleinen Bildschirmgrößen (Viewports), z.B. bei Smartphones, eine Anzeige der Link-Buttons am rechten Bildschirmrand störend wirken könnte. Folgende Abbildung zeigt den Footer mit der Verlinkung zu den Social-Media-Kanälen:



Abbildung 52: Footer mit Verlinkung zu Social-Media Kanälen

Zum Weiterverbreiten von Inhalten der GovData-Website befindet sich auf jeder Seite, ein Sharing-Bereich, von dem aus sämtliche Inhalte der GovData-Website geteilt werden können. Dieser Bereich befindet sich oberhalb

des Footers. Die datenschutzrechtlichen Maßgaben werden durch die Verwendung der vom Heise-Verlag entwickelten und bereitgestellten Shariff-Bibliothek (<https://github.com/heiseonline/shariff>) berücksichtigt. Der Nutzer hat die Möglichkeit, Inhalte per E-Mail, Facebook, Twitter und Diaspora zu teilen.

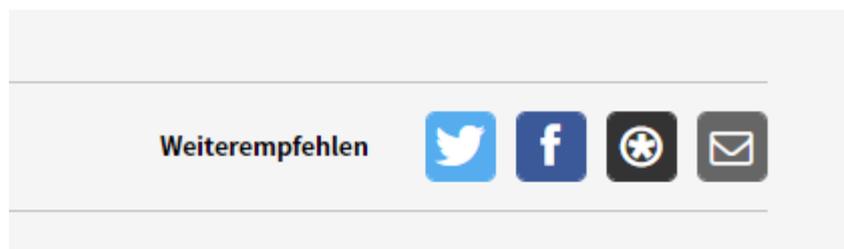


Abbildung 53: Interaktionselemente zum Weiterverbreiten

7.8.2.1.8 Pflege redaktioneller Inhalte und strukturelle Anpassungsmöglichkeiten

Die Erstellung und Pflege von redaktionellen Inhalten (Portalseiten) erfolgt mittels der Funktionalitäten des Liferay-CMS. Neben der rein inhaltlichen Pflege bietet das CMS auch Funktionen für Redakteure an, mit denen Anpassungen an der Seitenstruktur bzw. Seitenhierarchie vorgenommen werden können.

7.8.2.1.9 Verwaltung von Dateien

Die Bereitstellung eines zentralen Dateipools erfolgt mit dem im Liferay-CMS beinhalteten zentralen Datei-Repository, in das Medien wie Dokumente, Audio- und Videodateien sowie Bilder hochgeladen und zentral verwaltet werden können. Diese im Repository gespeicherten Dateien können von den Redakteuren in redaktionelle Inhalte der Portalseiten eingebunden werden.

7.8.2.1.10 WYSIWYG-Editor

Die eingesetzte Portallösung umfasst einen webbasierten WYSIWYG-Editor (Echtbilddarstellung), der den Nutzer beim Erstellen neuer bzw. beim Bearbeiten bestehender Inhalte unterstützt. Der Editor stellt u.a. Funktionen zur Textformatierung, für Listen, Tabellen und Hyperlinks sowie dem Einfügen von Medien zur Verfügung.

7.8.2.1.11 Benutzerverwaltung

Das GovData-Portal kann sowohl von nicht-registrierten als auch von registrierten Nutzern mit unterschiedlichem Funktionsumfang genutzt werden. Registrierten Nutzern können weitere Detail-Rollen zugewiesen werden, die den erweiterten Funktionsumfang der registrierten Nutzer steuern.

Bei der Verwaltung der Nutzer sind zwei Nutzergruppen zu unterscheiden, Nutzer der CKAN REST API und Portalnutzer (z.B. Bürger).

Die Nutzer der CKAN REST API sind schreibende Datenbereitsteller, die Metadaten im Push-Verfahren in den Metadatenkatalog von GovData übertragen und eigene Datensätze pflegen. Der Zugriff der Nutzer über die CKAN REST API wird über API Schlüssel abgesichert. Die Verwaltung der API Keys für den schreibenden Zugriff auf die CKAN REST APIs erfolgt über die Oberfläche in CKAN.

Die Benutzerverwaltung und die Zuordnung von Rollen erfolgt zentral im Liferay Portal. Dazu werden die in Liferay vorhandenen Funktionen verwendet und entsprechend erweitert, wo notwendig.

Die eingesetzte Liferay-Technologie verfügt über ein Modell und Framework zur Benutzerverwaltung und zur Implementierung von Rollen und Rechten. Dieses generische Modell erlaubt es, alle in der Leistungsbeschreibung dargestellten Anforderungen abzubilden. Durch das konsequente Nutzen des Liferay Berechtigungsmodells steht dem AG eine zukunftsfähige Berechtigungslösung und Benutzerverwaltung zur Verfügung, die jederzeit einfach erweitert und ergänzt werden kann.

Die Benutzer- und Rollenverwaltung der GovData-Nutzer erfolgt im Liferay Portal Backend. Dazu werden die Mechanismen, die Liferay bietet, eingesetzt und das allgemeine Rollen- und Rechte-Modell von Liferay (siehe nachfolgende Abbildung) verwendet.

Abbildung 54: Allgemeines Liferay-Rollen- und Rechte-Modell

Zur Abbildung der Rollen „Nicht-registrierter Nutzer“, „Registrierter Portalnutzer“, „Datenbereitsteller“, „Redakteur“, „Chefredakteur“ und „Geschäfts- und Koordinierungsstelle (GuK)“ werden in Liferay entsprechende Rollen angelegt, die frei konfigurierbar sind. Rollen (und Berechtigungen) werden in anwendungsspezifischen Portlets zur Berechtigungsprüfung in GovData genutzt. Die Rollen werden in einem Portlet auf einen im Quelltext verwendeten Rollenqualifier abgebildet. Dadurch ist der Liferay-Rollenname unabhängig von der Rollenbezeichnung im Quelltext der Anwendung.

Jeder Rolle können Berechtigungen zugeordnet werden. Gleichzeitig kann die Berechtigungsprüfung aber auch direkt auf Rollenbasis erfolgen. Weitere Rollen können jederzeit eingeführt werden.

Nutzern können nicht nur Rollen zugewiesen werden, sondern Nutzer können auch zu Gruppen zusammengefasst werden. Gruppen unterstützen dabei, die Benutzerverwaltung deutlich zu vereinfachen.

Gastnutzer mit der Rolle „Nicht-registrierter Nutzer“ können auf dem Liferay Portal die folgenden Funktionen nutzen: Suchen von Metadaten, Abfragen von weiteren Inhalten sowie Kommentierung von Inhalten und Metadaten. Ein Gastnutzer hat über die CKAN REST API die Möglichkeit nach Metadaten zu suchen und weitere Informationen abzufragen. Über das Kontaktformular im Liferay Portal kann ein Nutzer der Rolle „Nicht-registrierter Nutzer“ eine Anwendung mit entsprechender Metadaten-Beschreibung vorschlagen.

Diese Rollenkonfiguration kann durch Nutzer mit der Rolle „Geschäfts- und Koordinierungsstelle“ über die Liferay-Dialoge erweitert oder eingeschränkt werden.

Die Nutzer werden in Liferay verwaltet. Jeder Bürger kann sich über den Registrierungsprozess ein Nutzerkonto anlegen. Im Rahmen des Projekts wird der Registrierungsprozess von Liferay über entsprechende „Hooks“ erweitert. Ist ein Nutzer über den Registrierungsprozess angelegt, hat dieser die Rolle „registrierter Portalnutzer“. Ein Nutzer mit der Rolle „Geschäfts- und Koordinierungsstelle“ kann über die Benutzerverwaltung anderen Nutzer weitere Rollen zuordnen oder Nutzer sperren bzw. wieder aktivieren.

Neben den in der Leistungsbeschreibung benannten Rollen wird eine Rolle „Admin“ eingeführt. Dieser Rolle sind alle Rechte zugewiesen, sodass sie für die Pflege und den Support des Portals genutzt werden kann.

Jeder Nutzer kann sein Nutzerkonto mittels Double-Opt-Mechanismus selbst löschen.

Die E-Mail-Adresse wird als identifizierendes Attribut genutzt, da der Nutzer über einen Double-Opt-in-Prozess selbst nachweisen kann, dass ihm die Adresse gehört. Liferay nutzt den Algorithmus „PBKDF2WithHmacSHA1/160/128000“ für die Passwortverschlüsselung, mit dem 160 Bit lange Schlüssel in 128.000 Runden generiert werden.

Nicht-registrierter Nutzer (Rolle)

Ein nicht-registrierter Nutzer kann über die Weboberfläche oder die CKAN-API- Schnittstellen nach Metadaten suchen und Metadaten sowie alle weiteren Inhalte abrufen. Zudem kann ein nicht-registrierter Nutzer in der Standard-Einstellung Inhalte von GovData kommentieren. Ein nicht-registrierter Nutzer kann ferner mittels dem Kontaktformular eine Anwendung zur Veröffentlichung vorschlagen.

Registrierter Portalnutzer

Ein registrierter Portalnutzer verfügt über alle Rechte eines nicht-registrierten Nutzers. Darüber hinaus kann er nach Anmeldung ohne Angabe der eigenen E-Mail-Adresse Inhalte von GovData kommentieren. Ein registrierter Portalnutzer verfügt über ein persönliches Profil, dessen Inhalte er nach Anmeldung eigenständig pflegen kann. Auch kann ein Portalnutzer einen Account jederzeit über die Weboberfläche selbständig löschen.

Datenbereitsteller

Bei einem Datenbereitsteller handelt es sich um einen Portalnutzer. Er verfügt daher über die gleichen Rechte wie ein registrierter Portalnutzer. Darüber hinaus kann ein Datenbereitsteller Metadatensätze einstellen bzw. eigene eingestellte Metadatensätze editieren oder löschen. Der Datenbereitsteller kann mittels einer Ein- und Ausblenden-Funktion über die Sichtbarkeit eigener eingestellter Metadatensätze für Nutzer der Plattform entscheiden.

Redakteur

Bei einem Redakteur handelt es sich um einen registrierten Portalnutzer und verfügt über die gleichen Rechte. Er kann zusätzlich redaktionelle Inhalte im „Bibliotheks“-Bereich und Blog schreiben und editieren sowie die Webseitenstruktur des „Bibliotheks“-Bereich strukturell anpassen. Ein Redakteur kann Dateien in den Dateienpool hochladen und vorhandene Dateien verwalten. Auch kann ein Redakteur über die Kommentierungs- und Bewertungseinstellungen für die erstellten Inhalte entscheiden. Ein Redakteur kann Kommentare

moderieren und löschen. Darüber hinaus kann der Redakteur über Anpassungen des Design Templates das Design der grafischen Benutzeroberfläche verändern.

Chefredakteur

Ein Chefredakteur ist ein Redakteur und verfügt über die gleichen Rechte. Darüber hinaus kann ein Chefredakteur redaktionelle Inhalte freigeben, veröffentlichen und löschen. Ferner kann er die global geltenden Kommentierungs- und Bewertungseinstellungen ändern. Der Chefredakteur kann außerdem die übergreifenden Informationen (Impressum, Nutzungsbedingungen, Datenschutzerklärung) erstellen, editieren und löschen.

Geschäfts- und Koordinierungsstelle

Die Geschäfts- und Koordinierungsstelle kann die Rollen Datenbereitsteller, Autor, Redakteur und Chefredakteur Portalnutzern zuweisen. Außerdem hat die Geschäfts- und Koordinierungsstelle Zugriff auf die Kennzahlen aus dem Report und kann diese tagesaktuell einsehen, verarbeiten und weiterleiten. Darüber hinaus kann die Geschäfts- und Koordinierungsstelle Anwendungen erstellen, pflegen und löschen.

Fachgruppe (Rolle)

Die Rolle Fachgruppe umfasst alle Berechtigungen eines Portalnutzers. Zusätzlich werden Nutzern dieser Rolle redaktionelle Berechtigungen in einem speziellen Bereich des GovData-Portals eingeräumt. Hier können Fachgruppen-Nutzer redaktionelle Seiten erstellen und Dokumente und Dateien bereitstellen, die nur von anderen Nutzern dieser Rolle sowie Nutzern der Rolle „Geschäfts- und Koordinierungsstelle“ eingesehen werden können. Alle übrigen Portalnutzer erhalten keine Zugriffsberechtigung auf diesen Bereich.

Forumsteilnehmer (Rolle)

Bei einem Forumsteilnehmer handelt es sich um einen Portalnutzer. Er verfügt daher über die gleichen Rechte wie ein registrierter Portalnutzer. Die Rolle Forumsteilnehmer besitzt keine besonderen Berechtigungen. Mit der Rolle können Benutzer für die Anzeige der Forumsseite berechtigt werden. Alle übrigen Portalnutzer erhalten keine Zugriffsberechtigung auf diesen Bereich.

Systemadministrator

Der Systemadministrator hat die Berechtigung zur betrieblichen Verwaltung des Gesamtsystems.

7.8.2.1.12 Redaktionsprozess

Die für das Portal eingesetzte Applikation (Liferay Portal) besteht aus verschiedenen Produkten.

Zum einen aus der gleich benannten Kernapplikation, zum anderen aus Liferay CMS, dem darauf aufbauenden Enterprise Content Management System.

Liferay bietet mit seinen Portlets und der zentralen Benutzerverwaltung das CMS als integralen Bestandteil der GovData Anwendung an.

Liferay trennt Content (Artikel, Dokumente, Bilder usw.) von der Navigation bzw. den Seiten. Der gesamte Webcontent kann losgelöst von den Seiten angelegt werden. Erst über den Einsatz von Portlets auf den jeweiligen Seiten wird dynamisch der Content angezeigt. Zum Beispiel wird der Inhalt des „Impressum“

redaktionell als Artikel angelegt und auf der Seite Impressum wird ein Portlet „Webcontent Anzeige“ platziert und explizit der Artikel „Impressum“ ausgewählt.

Durch diese Architektur ist ein nahtloser Übergang je nach Rolle und Berechtigung vom Nutzer der Seite zum Redakteur einzelner Inhalte und Artikel möglich. Die Redakteure können ihre Artikel per Drag-And-Drop einfach zusammenstellen. Fließtexte werden mit einem WYSIWYG-Editor erstellt. Dabei können auch Inhalte per Copy-and-paste aus anderen Programmen, wie Microsoft Word übernommen, werden. Der Editor unterstützt auch echtes „inline editing“. Das bedeutet es muss kein extra Editorfenster geöffnet werden, sondern der Text kann direkt innerhalb der Artikel geschrieben werden.

Die Artikel können im Backend zentral in einer übersichtlichen Ordnerstruktur verwaltet werden. Gelöschter Inhalt landet zunächst im Papierkorb und wird erst nach einem definierten Zeitraum tatsächlich gelöscht. Somit sind aus Versehen gelöschte Inhalte wiederherstellbar.

Jeglicher Inhalt in Liferay unterliegt einer eigenen Versionsverwaltung. Somit können auch ältere und gelöschte Versionen mit wenigen Klicks wiederhergestellt werden.

Der Redaktionsprozess wird von Liferay unterstützt. Falls Inhalte vor einer Freigabe erst durch einen Chefredakteur geprüft werden müssen, ist dies durch den im Standardumfang enthaltenen Vier-Augen-Veröffentlichungsprozess möglich. Weitere Workflows (Mehraugenprinzip) können einfach hinzugefügt werden. Der zuständige Chefredakteur bekommt seine zu genehmigenden Inhalte im Portal-Backend innerhalb des Control-Panel gelistet.

Jeder Inhalt kann zeitgesteuert veröffentlicht werden. Es können sowohl Erscheinungs- als auch Ablaufdatum vom Redakteur individuell gepflegt werden. Zusätzlich gibt es eine Wiedervorlagefunktion, die den Redakteur zum eingestellten Termin daran erinnert den Inhalt zu überarbeiten.

Den Inhalten können Kategorien oder Tags hinzugefügt werden, die wiederum an anderer Stelle als Filter zur Veröffentlichung herangezogen werden.

Soweit gewünscht, kann auf jeder Seite getrennt, jeglicher Content zum Kommentieren oder Bewerten freigegeben werden.

Neben den Standardinhalten bietet Liferay die Möglichkeit weitere Inhalte aus anderen Quellen über spezielle Portlets auszugeben, im Backend zu editieren und die Ausgabe zu konfigurieren.

Portlets sind Funktionsbausteine; jede Seite ist aus verschiedenen Portlets aufgebaut. Diese kleinen Funktionsbausteine sind die Grundbausteine des gesamten Portals. Die häufigsten Portlets dienen zur Anzeige von Webcontent und zur Darstellung von Inhaltslisten. Erweiterte Funktionalitäten bieten Mediengalerie-, Lesezeichen-, Wiki-, Foren-, Blog- oder Kalender-Portlets. Portlets sind beliebig auf der Benutzeroberfläche kombinierbar und werden von einem Portalserver angezeigt und verwaltet. Jedes Portlet ist eine Applikation für sich und kann individuell konfiguriert werden. Für GovData wurde beispielsweise ein Portlet für die Anzeige der Metadatensätze speziell entwickelt. Die Nutzer mit der Rolle „Datenbereinsteller“ können über dieses Portlet ihre Metadatensätze bearbeiten und die jeweiligen Einstellungen zur Sichtbarkeit, Anzeige der Kommentarfunktion und auch alle anderen Feldinhalte der CKAN-Daten für den jeweiligen Metadatensatz ändern. Über das Portlet werden die Daten direkt in die CKAN-Basis zurückgeschrieben.

Eine besondere Liferay-Funktionalität sind die verbundenen Assets. Ein Asset beschreibt im Kontext der Content Management Systeme einen Datensatz oder eine Datei, zum Beispiel Bilder oder Dateianhänge. Über verbundene Assets können zu einem Artikel ein oder mehrere Dokumente oder Bilder aus der zentralen Ablage

hinzugefügt werden.

Liferay unterstützt den Redakteur auch bei der Aufbereitung des Inhalts für andere Auflösungen, dem sogenannten responsive Design. Die Redakteure können je Seite, die im Vorfeld definierten Auflösungen ansehen. Das responsive Verhalten der Seite wird im Theme definiert.

7.8.2.1.13 Feeds

Liferay verfügt über die Möglichkeit, redaktionelle Feeds über das Redaktionssystem zu erstellen und zu verwalten. Über einen geführten Dialog kann der zuständige und berechtigte Redakteur einen neuen Feed erstellen bzw. bestehende Feeds bearbeiten oder löschen. Beim Erstellen wird der im Feed zu liefernde Inhalt / Inhaltstyp ausgewählt. Dabei werden auch alle anderen Feed-Parameter bestimmt. Das sind unter anderem die Anzahl und Sortierung der Beiträge. Ebenso wird in diesem Schritt das Format des neuen Feeds ausgewählt. Die zur Auswahl stehenden Formate sind ATOM 1.0, RSS 1.0 und RSS 2.0. So definierte Feeds werden dem Nutzer auf der Portalseite mit einem Feed-Icon angezeigt, über das der Feed abonniert werden kann.

Bei der Suchfunktion werden individuelle Feeds bereitgestellt. Diese Feeds stehen bei jeder Anzeige eines Suchergebnisses bzw. einer Trefferliste nach Facettierung (Auswahl von Filtermöglichkeiten) zum Abonnement zur Verfügung. Bei jeder Suchanfrage kann die Suchzusammenstellung mit allen Facetten als Feed durch den Nutzer abonniert werden. Die Feeds stehen in zwei Formaten zur Verfügung, ATOM 1.0 und RSS 2.0. Feeds können somit durch jeden Besucher der Seite im Rahmen der GovData-Portalsuche individuell definiert werden.

7.8.2.1.14 Ein- und Ausblenden von Metadatenätzen

Das Ein- und Ausblenden von Metadatenätzen erfolgt im System wie folgt:

- **Einblenden von Metadatenätzen**

Die Anzeige der Metadatenätze wird durch ein Flag bei den CKAN Daten geregelt. Über ein Liferay Anzeige-Portlet für die Metadatenätze können abhängig von der Rolle des Redakteurs bzw. des Datenbereitstellers diese Metadatenätze direkt über das CMS editiert werden. Hierbei wird auch das Flag zur Anzeige der Metadatenätze gesetzt. Durch die Speicherung der Information in der CKAN Datenstruktur ist auch das Auslesen der Daten direkt über die CKAN Schnittstelle möglich und die Datensätze sind im täglichen RDF Dump abrufbar.

- **Ausblenden von Metadatenätzen**

Auf dem gleichen Weg kann auch die Sichtbarkeit von Metadatenätzen verhindert werden. Der angemeldete Redakteur oder Datenbereitsteller kann die ihm zugeordneten Metadatenätze über das, für die Metadatenätze entwickelte Portlet, im CMS Backend erreichen und direkt bearbeiten. Die Information wird direkt zum CKAN Datensatz geschrieben und unterbindet somit auch die Auslieferung über die CKAN Schnittstellen. Ausgeblendete Metadatenätze werden in der Suche nicht berücksichtigt, ausgenommen es handelt sich bei dem Suchenden Nutzer um den Eigentümer des Datensatzes. Bei Aufruf einer Detailseite eines ausgeblendeten Datensatzes wird der Datensatz ebenfalls nicht angezeigt und es erscheint eine Meldung „Der Datensatz ist zur Zeit nicht verfügbar“. Dies kann z.B. dann der Fall sein, wenn ein Nutzer ein Bookmark auf einen Datensatz gespeichert hat, der später ausgeblendet wird. Wird der Datensatz wieder einblendend, ist er auch unter derselben URL wieder erreichbar, sofern sich nicht die ID des Datensatzes ändert.

7.8.2.1.15 *Gastzugang, Registrierung und Login*

Die auf dem GovData-Portal bereitgestellten Inhalte sind sowohl für angemeldete wie auch nicht angemeldete Benutzer zugänglich. Bestimmte Funktionen, z.B. die Kommentarfunktion, können jedoch über die Berechtigungseinstellungen im CMS auch auf angemeldete Nutzer beschränkt werden, falls dies gewünscht ist. In diesem Fall wird dem nicht angemeldeten Nutzer ein Hinweis angezeigt, dass die Funktion erst nach Anmeldung genutzt werden kann.

Die Registrierung eines Nutzers erfolgt im Self-Care-Verfahren über die Registrierungsfunktion des Portals durch den Nutzer selbst. Nach Eingabe relevanter Pflichteingaben (Benutzername, Vor- und Nachname, E-Mail-Adresse und Passwort inkl. Wiederholung) in ein Webformular wird der Double-Opt-In-Prozess gestartet. Dem Nutzer wird dabei eine E-Mail mit einem Bestätigungslink zugesendet, über den der Nutzer seine Identität bzgl. der E-Mail-Adresse bestätigt. Solange der Opt-In-Prozess vom Nutzer nicht abgeschlossen wurde, ist kein Login möglich. Nach Klicken des Bestätigungslinks wird der Login für den Nutzer freigeschaltet und der Nutzer kann sich daraufhin am Portal anmelden.

Ebenso hat der Nutzer die Möglichkeit, seinen Account im Portal wieder zu löschen. Dies erfolgt über eine Funktion, die in der Kontoverwaltung des Nutzers verlinkt ist. Auch die Löschung des Accounts durch den Nutzer selbst erfolgt immer über einen Double-Opt-Out-Prozess. Wie bei der Registrierung erhält der Nutzer einen Bestätigungslink per E-Mail. Erst nach Klicken des Links wird der Account des Nutzers tatsächlich gelöscht, wobei alle persönlichen Daten des Nutzers aus dem System gelöscht werden.

Schutz vor maschinellen Eingaben bei der Registrierung

Das CMS hält eine Captcha-Lösung bereit. Da eine herkömmliche Captcha-Lösung jedoch wegen mangelnder Maschinenlesbarkeit nicht barrierefrei sein kann, werden Formulare mit einer Kombination aus mehreren Methoden vor Bot-Zugriffen und maschinellen Eingaben geschützt. Dabei kommen das sogenannte „Honeypot“-Verfahren sowie zusätzliche Vorkehrungen, wie IP-Filter, Content-Filter und Timestamp-Management zum Einsatz.

Bei der Honeypot-Lösung wird ein weiteres Inputfeld definiert, welches für den menschlichen Nutzer unsichtbar ist und für Vorlesesoftware entsprechend gekennzeichnet wird. Das Formular wird backendseitig nur dann verarbeitet, wenn dieses Feld nicht befüllt wurde. Dabei wird kein Fehler ausgegeben, so dass ein Bot nicht erkennen kann, dass das Formular nicht verarbeitet wurde. Kombiniert mit einem Timestamp-Management, bei dem eine gewisse Dauer für das Befüllen von Formularen in Bezug zum Seitenaufruf gesetzt wird, lassen sich so Bots wirkungsvoll daran hindern, Formulare zu missbrauchen.

Nutzeraccount

Jeder Nutzer erhält nach der Registrierung im GovData-Portal ein persönliches Profil, in dem er optional persönliche Daten eingeben und ändern sowie ein Profilbild einstellen kann.

Jedem Nutzer steht eine Funktion zur Verfügung, um seinen Account im Double-Opt-Out-Verfahren löschen zu können (siehe auch Kapitel 7.8.2.1.15) Nach Löschung werden die Profilinformationen gelöscht und die vom Nutzer erstellten Beiträge anonymisiert.

Datenfilterung

Sämtliche Nutzereingaben werden vom System serverseitig validiert bzw. gefiltert. Damit wird verhindert, dass Nutzereingaben die Sicherheit und den Betrieb des Portals beeinträchtigen können. Es werden dazu Maßnahmen gegen übliche Angriffstechniken wie SQL- oder Code-Injection, Cross-Site-Scripting, Cross-Site-Request-Forgery usw. implementiert.

Um dem Nutzer bereits früh eine Rückmeldung zu eventuellen Fehleingaben geben zu können, werden auch clientseitig Validierungsmaßnahmen ergriffen. Der Nutzer wird dabei zeitnah darauf hingewiesen, wenn seine Eingabe einen Fehler enthält.

Sowohl bei der server- wie auch clientseitigen Validierung werden dem Nutzer verständliche Informationen zur Fehleingabe rückgemeldet, um die Korrektur der Daten möglichst einfach zu gestalten.

7.8.2.1.16 Verschlüsselung vertraulicher Daten

Die vollständige Kommunikation zwischen Client und Portal-Server erfolgt SSL verschlüsselt per HTTPS. Dadurch ist sichergestellt, dass vertrauliche Daten immer in verschlüsselter Form übertragen werden.

7.8.2.1.17 Kommentarfunktionalität

Die Anforderungen an die Kommentarfunktion werden in Form einer Erweiterung des in Liferay bestehenden Kommentar-Portlets umgesetzt. Generell kann bei jedem Metadatensatz, Blog und jeder Informationsseite die Kommentarfunktion redaktionell aktiviert werden.

Die Funktion kann sowohl von angemeldeten als auch von unangemeldeten Nutzern verwendet werden. Die Kommentare unterscheiden sich in der Darstellung, da bei den angemeldeten Nutzern zusätzlich der Accountname und, falls vorhanden, ein Profilbild ausgegeben wird. Über die zentrale Verwaltung der Kommentare in Liferay können diese durch die Redaktion moderiert werden. Genauso können über diese Kommentarverwaltung die Eigenschaften der Kommentare konfiguriert werden, wie beispielsweise, ob eine generelle Freigabe durch den Moderator erfolgen muss oder nur angemeldete Nutzer Kommentare abgeben dürfen.

Schutz vor maschineller Eingabe von Kommentaren

Um die Kommentarfunktion vor maschineller Eingabe durch Bots zu verhindern, wird die Kommentarfunktion durch das HoneyPot-Verfahren abgesichert, siehe auch Kapitel 7.8.2.1.15, Abschnitt „Schutz vor maschinellen Eingaben bei der Registrierung“.

Benachrichtigung über neue Kommentare

Die GovData-Redaktion wird über alle eingehenden Kommentare informiert. Falls ein neuer Kommentar einem Metadatensatz zugeordnet werden kann, wird die in den Metadaten angegebene Kontaktstelle darüber ebenfalls per E-Mail in Kenntnis gesetzt.

7.8.2.1.18 Bewertung

Derzeit ist die Bewertung von Metadaten nicht umgesetzt, da die aktuelle Version von CKAN die bisher dazu genutzten API-Methoden entfernt hat.

7.8.2.1.19 *Interne und externe Links*

Webseiten-Verweise werden auf den Portalseiten einheitlich und eindeutig für den Nutzer als solche erkennbar gemacht. Eine Unterscheidung zwischen Links innerhalb des Portals und Links auf externe Webseiten unterscheiden sich über ein vorangestelltes, unterschiedliches Icon. Ebenfalls wird bei externen Links im Tooltip des Verweises ein Vermerk „Verweis auf externe Website:“ vorangestellt.

7.8.2.1.20 *Nutzung auf mobilen Endgeräten*

Die Umsetzung des responsive Webdesigns erfolgt mittels eines „liquid adaptive“ Layouts, bei dem die Darstellung von der zur Verfügung stehenden Bildschirmgröße (Viewport) abhängig ist. Dabei wird der verfügbare Platz optimal ausgenutzt.

Die grafische und funktionale Konsequenz von „liquid adaptive“ ist, dass das Layout stets die gesamte Bildschirmbreite einnimmt und die Breite der einzelnen Elemente von der zur Verfügung stehenden Auflösung abhängig ist. Der Inhalt ist jedoch maximal 1.600 Pixel breit, um auch auf großen Bildschirmen eine gute Lesbarkeit zu gewährleisten.

Die Benutzeroberfläche mit allen Elementen passt sich also immer an die verfügbare Breite des jeweiligen Anzeigebereichs „liquid“, nahtlos ohne Zwischenstufen, an. Zusätzlich werden sogenannte Breakpoints bestimmt, die definieren, ab wann neben der fließenden Breitenanpassung der Seiten auch strukturelle „adaptive“ Anpassungen (z.B. Schriftgröße, Rasterbreite, Anzahl Rasterspalten) vorgenommen werden.

Die Anordnung und Darstellung von Elementen ist ebenfalls von der Auflösung abhängig. So wird beispielsweise bei Tablets im Landscape-Mode (quer) aus einem 3-spaltigen Aufbau im Portrait-Mode (hochkant) ein 2-spaltiger Aufbau, bei noch kleineren Bildschirmgrößen, wie bei Smartphones, ein 1-spaltiger.

Bereits bei der Gestaltung der Seiten wird die Wirkung des Layouts auf unterschiedlichen Bildschirmgrößen berücksichtigt und die Sichtbarkeit und Anordnung der Elemente orientiert sich an einer optimalen Nutzerfreundlichkeit. Dennoch wird für alle Endgeräte dieselbe Codebasis verwendet, sodass eine einfache Wartbarkeit gewährleistet ist.

Alle Interaktionselemente reagieren auch auf Touch, sodass eine Bedienung der GovData-Website auch über mobile Endgeräte ohne Einschränkung möglich ist. Die Bedienung bei kleineren Auflösungen ist angepasst, aber uneingeschränkt möglich. So ist ab einer bestimmten Auflösung (in etwa Tablet Hochformat) das Menü nicht mehr permanent sichtbar, sondern kann über ein Icon geöffnet werden. In der Suche erkennt der Nutzer, dass Filter aktiv sind. Details zur Filterauswahl kann er mit Klick auf eine Schaltfläche anzeigen lassen (z.B. bei Ansicht auf einem Smartphone).

7.8.2.1.21 *Browserkompatibilität*

GovData wird so implementiert, dass die Portalseiten auf allen Browsern mit einem Marktanteil von mehr als fünf Prozent uneingeschränkt und unverfälscht dargestellt werden.

7.8.2.2 **Metadatenkatalog**

Die wichtigste, zentrale Komponente von GovData ist der zugrundeliegende Metadatenkatalog, der als Basis für die zentralen Funktionen des Portals dient.

Für die Offenlegung von Datenbeständen der öffentlichen Verwaltung („Open Government Data“) in Europa hat es sich bisher bewährt, die Metadaten-Struktur von CKAN (Comprehensive Knowledge Archive Network) des

OFKN (Open Knowledge Foundation Network) zu verwenden. Im Bereich Open Data stellt der Open-Source-Metadatenserver des CKAN den de-facto-Standard für Datenkatalog-Software dar.

CKAN definiert standardisierte Schnittstellen und Datenformate zum Austausch von Metadaten. Das einzige Pflichtfeld im CKAN stellt das Datenfeld „name“ dar. Dieses sollte sowohl für Nutzer als auch „URL-friendly“ lesbar sein. Alle weiteren Felder sind optional.

Da die Flexibilität des CKAN-Datenformats eine Reihe von Schwierigkeiten bei der Umsetzung und Integration verschiedener Metadatenkataloge mit sich bringt und viele Entwickler und Datenanbieter genaue Anweisungen verlangen, welche Informationen in welcher Form zu liefern sind, hat die Geschäftsstelle GovData das Metadatenmodell DCAT-AP.de entwickelt und etabliert.

Für Deutschland existiert mit DCAT-AP.de ein Standard für den Metadaten austausch zwischen deutschen Open Data Portalen. DCAT-AP.de ist eine standardkonforme deutsche Ableitung von DCAT-AP dem Standard für den Datenaustausch auf europäischer Ebene. DCAT-AP.de wurde mit dem Beschluss des IT-Planungsrates vom 28.06.2018 ab 2019 als gemeinsame verbindliche Grundlage für den Metadaten austausch zwischen deutschen Open Data-Portalen festgelegt.

Als Datenquelle für den CKAN-Server dienen sogenannte „Harvester“. Für das Metadatenmodell DCAT-AP.de wurde ein RDF-Harvester entwickelt, der zu den bereits unterstützten „DCAT-AP“-Eigenschaften auch die spezifischen Eigenschaften von DCAT-AP.de unterstützt. Somit können relativ einfach neue RDF-Endpunkte angebunden werden, die Metadaten entsprechend dem Metadatenmodell DCAT-AP.de bereitstellen.

Eine zusätzliche Datenquelle ist die direkte Eingabe von Metadaten durch den Datenbereitsteller über ein Webformular. So können auch Datenbereitsteller mit wenigen Datensätzen ihre Daten über die GovData-Schnittstellen zugreifbar machen, ohne selbst Infrastruktur für die Bereitstellung von Metadaten betreiben zu müssen.

Durch einen Kompatibilitätscheck wird beim automatisierten Import von Metadaten aus anderen Portalen (Harvesting) sowie bei der manuellen Erfassung von Datensätzen für jeden Metadatenatz ein Kompatibilitätsstatus ermittelt und gespeichert. Dabei werden die im TripleStore gespeicherten original angelieferten RDF-Graphen mittels einer RDF-Validierung geprüft und das Ergebnis selbst wieder in einem Ergebnisgraphen gespeichert.

Die Anzahl an Metadatenätzen, die die mit dem RDF-Validator geprüften Eigenschaften erfüllen oder eben nicht erfüllen werden im Portal auf einem Metadatenqualitäts-Dashboard dargestellt, welches aktuell nur für Datenbereitsteller, Redakteure sowie die Mitarbeiter der Geschäfts- und Koordinierungsstelle zugänglich ist (siehe Kapitel 7.8.2.2.4). Optional kann die Zugriffsberechtigung für diese Seite auf alle GovData-Nutzer erweitert werden.

Die dargestellten Datensätze können nach Datenbereitsteller gefiltert werden.

Konsumenten der Datensätze, sogenannte Anwendungen, können über eine webbasierte Benutzerschnittstelle der Redaktion zur Veröffentlichung vorgeschlagen werden. Wird eine Anwendung als Konsument eines Datensatzes von der Redaktion akzeptiert, werden die Informationen zu dieser Anwendung als Teil des zum Datensatz zugehörigen Metadatenatz im gespeichert und stehen somit öffentlich zur Verfügung.

Die in den Metadatenätzen enthaltenen Verweise auf Daten werden regelmäßig automatisch auf Erreichbarkeit geprüft. Die Prüfung beschränkt sich dabei auf eine technische Prüfung der Erreichbarkeit. Eine inhaltliche Prüfung erfolgt nicht.

Ist ein Verweis bei dem letzten Durchlauf des Verweisprüfers nicht abgerufen werden wird im Portal auf der Metadaten-Detailansicht ein entsprechender Hinweis auf die fehlende Verfügbarkeit dargestellt. Die Anzeige des Hinweises kann auch redaktionell unterbunden werden.

Die nicht verfügbaren Verweise werden im Portal in einer Übersichtsseite dargestellt. Um eine übersichtliche Darstellung zu erreichen und das Finden relevanter Einträge zu erleichtern, können die dargestellten Datensätze nach verschiedenen Kriterien sortiert und gefiltert werden. Dazu gehören unter anderem die Kriterien

- ID
- Name des Metadatenatzes
- Kontaktperson
- URL
- Fehler
- Anzahl der Versuche

Um eine einfache Integration des Datenbestandes von GovData in andere Open Data Portale zu ermöglichen wird eine RDF-Schnittstelle bereitgestellt, die alle Metadaten entsprechend dem Metadatenmodell DCAT-AP.de liefert.

7.8.2.2.1 Eigenschaften und Funktionen des Metadatenkatalogs

GovData verfügt über einen webbasierten Katalog von Metadaten zu Daten. GovData speichert diese Metadaten entsprechend der Spezifikation des Metadatenmodells DCAT-AP.de (<https://www.dcat-ap.de/def/>) und stellt diese zentral zur Verfügung.

Daneben beinhaltet das Portal Metadaten zu Anwendungen, die mit eben diesen Datensätzen arbeiten. Anwendungen können mittels des Kontaktformulars auf GovData zur Veröffentlichung der Redaktion vorgeschlagen werden.

7.8.2.2.2 Gesamter Metadatenkatalog als Export-Datei

Um den Metadatenkatalog des Portals zum Download zur Verfügung zu stellen, wurde ein Dienst erstellt, der einmal wöchentlich ausgeführt wird und den Datenbestand in Form eines RDF-Dumps exportiert. Die erzeugte Datei wird serverseitig bis zum nächsten erfolgreichen Export gespeichert. Auf der Benutzeroberfläche des Portals wird die URL bereitgestellt. Die URL bleibt dabei immer unverändert, d.h. der aktuelle Export ist immer unter derselben URL erreichbar. Somit kann zum einen der Download per Link aus der Weboberfläche erfolgen, zum anderen kann die URL kopiert und für automatisierte Downloads verwendet werden, um die Daten beispielsweise in andere Datenportale einzubinden.

7.8.2.2.3 Kompatibilitätscheck

Die im Metadatenkatalog gespeicherten Metadaten entsprechen aufgrund der Heterogenität der Datenquellen nicht zwangsläufig vollständig den im Metadatenmodell DCAT-AP.de definierten Datenformaten. Die gespeicherten und für Nutzer sichtbaren Metadaten werden deshalb bei jeder Aktualisierung hinsichtlich ihrer Kompatibilität mit dem definierten Metadatenmodell DCAT-AP.de geprüft und das Ergebnis dieser Überprüfung auf einem Metadatenqualitäts-Dashboard im Portal veröffentlicht.

Der Zugang zu dem Metadatenqualitäts-Dashboard ist dabei aktuell auf Datenbereitsteller, Redakteure und die Geschäfts- und Koordinierungsstelle beschränkt.

Es besteht allerdings die Möglichkeit, das Metadatenqualitäts-Dashboard auch für alle anderen Nutzern von GovData zugänglich zu machen.

Die Validierung der RDF-Graphen der Metadaten erfolgt mittels SHACL-Validierung und wurde so implementiert, dass diese als Teil des Harvesting-Prozesses, wie auch bei der Speicherung von Metadaten über das Bearbeitungsformular im Portal ausgeführt wird.

7.8.2.2.4 Metadatenqualitäts-Dashboard

Das Metadatenqualitäts-Dashboard zeigt anhand von Diagrammen für die geprüften Eigenschaften jeweils an wie viele der Metadaten diese erfüllen und welche nicht. Aktuell werden die folgenden Eigenschaften geprüft:

- Keywords
- ContributorID

Darüber hinaus zeigt das Metadatenqualitäts-Dashboard auch die TOP 5 der genutzten Lizenzen und Dateiformate der im Portal enthaltenen Metadaten an.

Um die Qualität der Metadaten auf einem Dashboard darzustellen, müssen die Datensätze zunächst mit dem ITB SHACL-Validator analysiert werden. Dieser Prozess geschieht während dem Harvesting. Der Ablauf der Analyse beim Harvesting ist auf der folgenden Abbildung zu sehen.

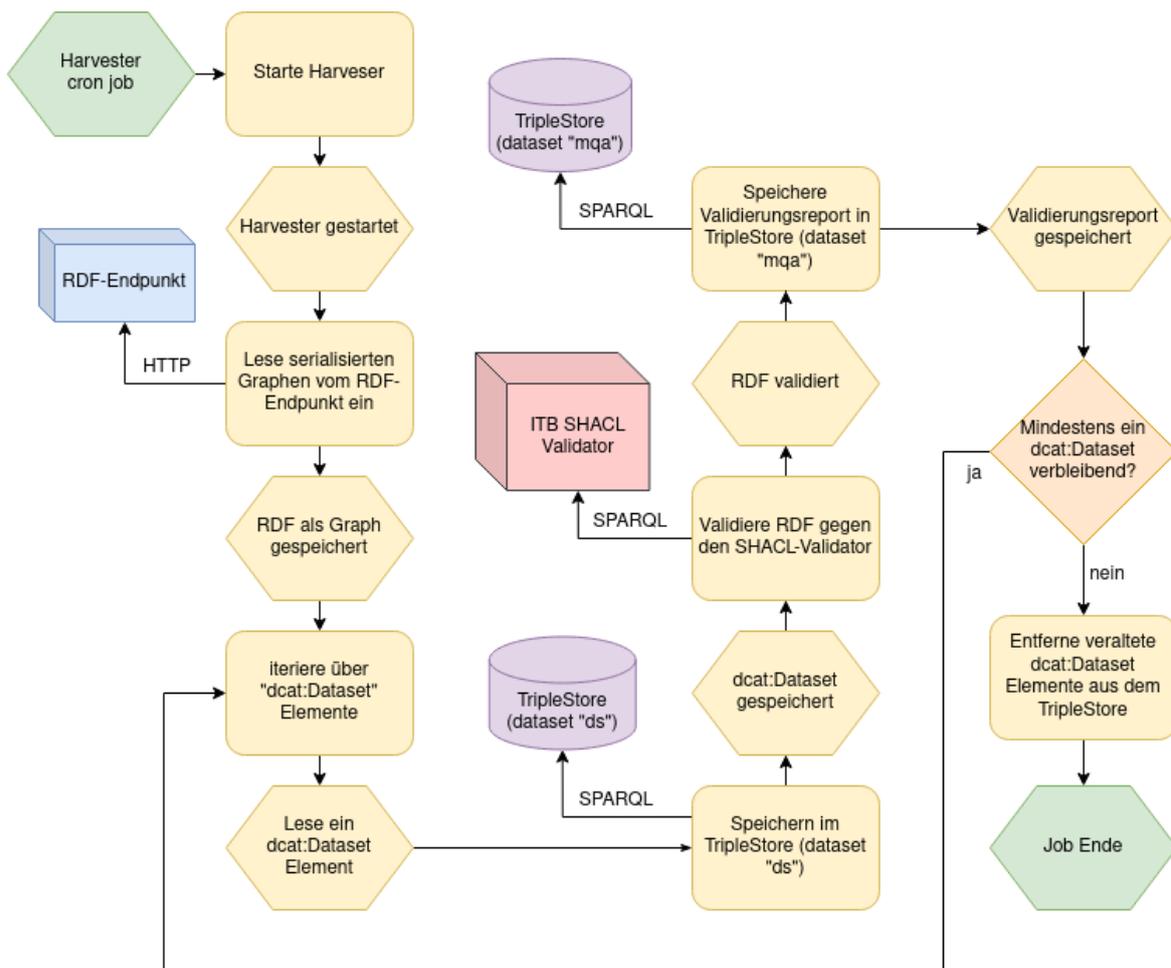


Abbildung 55: SHACL-Validierungsprozess beim Harvesting

Bei diesem Prozess wird zunächst das RDF des ausgewählten Endpunkts eingelesen und über alle darin vorhandenen Datensätze ("dcat:Dataset"-Elemente) iteriert. Die Datensätze werden im TripleStore gespeichert und anschließend an den ITB SHACL Validator geschickt. Dieser führt die SHACL-Validierung nach einem vorgegebenen Schema aus. Als Resultat wird vom SHACL-Validator ein Validierungs-Report zurückgeschickt. In diesem steht, ob die im Schema definierten Bedingungen von dem RDF-Datensatz erfüllt sind oder nicht. Dieser Report wird in einen weiteren TripleStore geschrieben.

Um nun die Anzeigen für das Dashboard zu erhalten, werden bei einem Aufruf des Dashboards die Validation-Reports im TripleStore analysiert und es werden die prozentualen Werte für die Anzeige im Dashboard zurückgeliefert. Diese Werte werden in einem Cache gespeichert um Rechenleistung zu sparen.

Zudem werden noch weitere Metriken über den aktuellen gesamten Datenbestand im Dashboard zur Metadatenqualität angezeigt. Diese werden aus dem TripleStore und dem Suchindex zusammengetragen und in einem eigenen Index abgespeichert. Das Erzeugen dieser Metriken geschieht mit einem eigenen Cron-Job, welcher in der folgenden Abbildung näher beschrieben ist.

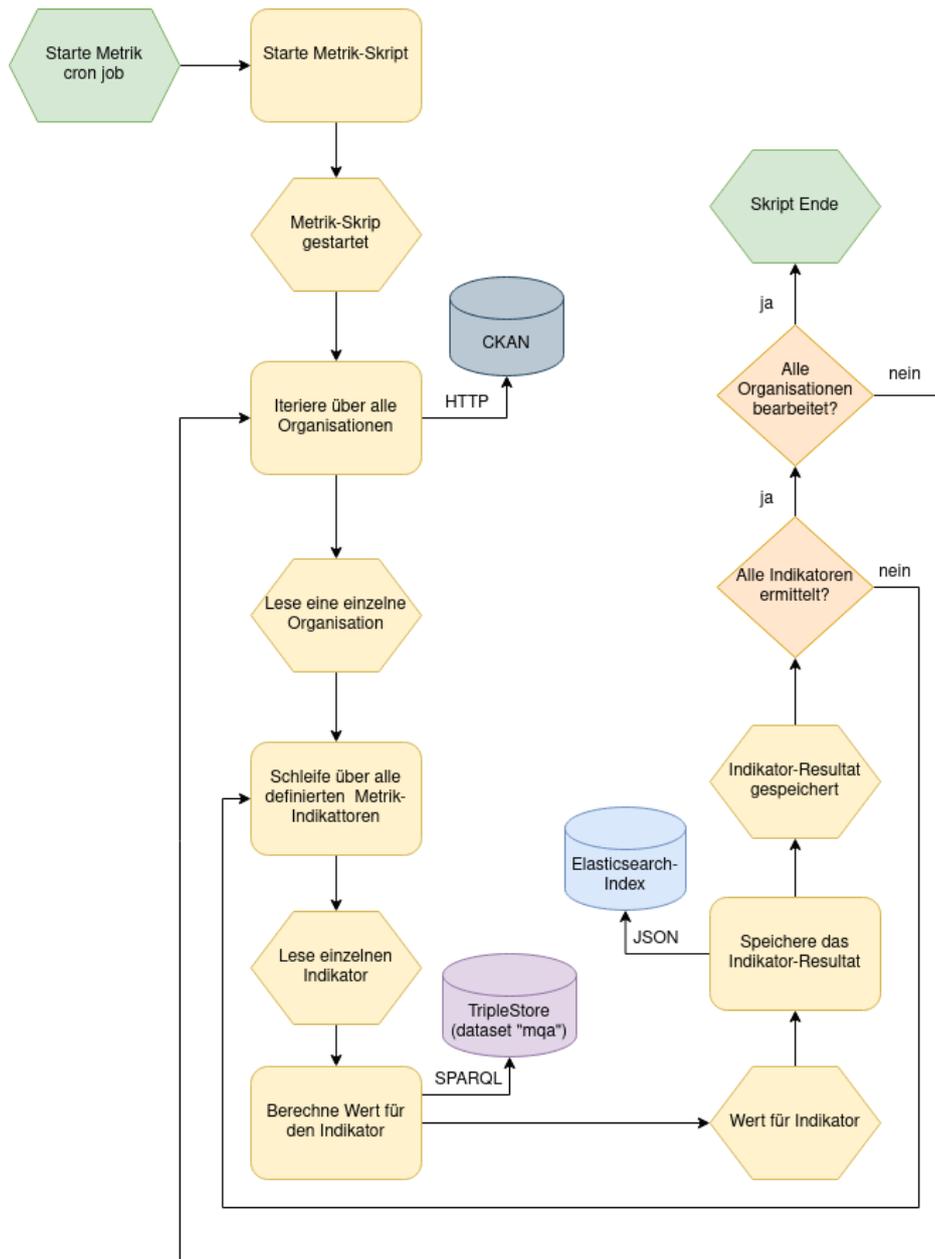


Abbildung 56: Prozess der Metrikchecks bei der SHACL-Validierung

7.8.2.2.5 Dead Link Check

Funktionalität des Dead Link Checks

Der Dead-Link-Check ist über einen Cron-Job realisiert, der in konfigurierbaren Intervallen (z.B. alle zwei Wochen) die in den Metadaten des gesamten Metadatenkatalogs von GovData enthaltenen Verweise auf Verfügbarkeit prüft. Wird in einem Metadatensatz ein nicht erreichbarer Verweis erkannt wird zum Datensatz gespeichert, dass ein bestimmter Verweis nicht funktioniert.

Die Anzahl der Prüfungen wird ebenfalls gespeichert. Ist ein Link nach mehreren Prüfungen weiterhin nicht erreichbar, wird der Verweis auf der Metadaten-Detailseite ausgeblendet und stattdessen ein Text „Ressource derzeit nicht verfügbar“ angezeigt. Das Ausblenden und die Anzeige des Hinweises kann redaktionell deaktiviert werden.

Dead-Link-Check-Webseite

Für die Anzeige wurde eine Portalseite eingerichtet, auf der berechnete Nutzer die Ergebnisse des Dead-Link-Checks einsehen können. Die Berechtigungen können über die Berechtigungseinstellungen im Portal jederzeit geändert werden. Initial wird der Zugang zu dieser Portalseite auf Datenbereinsteller, Redakteure und die Geschäfts- und Koordinierungsstelle beschränkt.

Auf dieser Portalseite werden die fehlgeschlagenen Prüfungen in Listenform aufbereitet. Die Liste stellt zur besseren Übersichtlichkeit eine Reihe von Sortier- und Filtermöglichkeiten bereit (in Klammern sind jeweils die Namen entsprechend des CKAN-Metadatenschemas aufgeführt):

- Sortieren nach:
 - o ID
 - o Name des Metadatenatzes (name)
 - o Kontaktperson (maintainer)
- Filtern nach:
 - o Datenbereinsteller / Harvester-Quelle (metadata_harvested_portal)

7.8.2.2.6 Suchen und Filtern

Um eine Suche über Portalinhalte jederzeit direkt auf den Portalseiten ausführen zu können, wird den Nutzern des Portals ein zentrales Suchfeld bereitgestellt. Darüber kann mittels einer einfachen Schlagwortsuche nach Inhalten gesucht werden. Dieses Suchfeld umfasst auch eine Autosuggest-Funktion (Wortvorschlagsfunktion), die den Nutzer bei der Eingabe von Suchbegriffen unterstützt.

Die Ergebnisse werden in Form einer sortier-, filter- und blätterbaren Trefferliste dargestellt.

Die Funktionen, die für die Suche und Trefferlisten zur Verfügung stehen, sind in den folgenden Kapiteln detaillierter aufgeführt.

Webbasierte Suchfunktion

Suchen Grundfunktionalität

Über die zentrale Suchkomponente des GovData-Portals werden mehrere Datenquellen durchsuchbar gemacht. Die folgenden Datenquellen sind in die Suche integriert:

- Portalinhalte aus Liferay (Blogs und weitere redaktionelle Inhalte)
- CKAN Metadatenkatalog
- Anwendungen

Um eine schnelle, exakte und nutzerfreundliche portalweite Suche anbieten zu können, wurde eine zentrale Suchkomponente auf Basis von Elasticsearch realisiert.

Zur Anbindung der Portaleinhalte aus GovData an die Suche werden die durchsuchbaren Inhalte in den Index aufgenommen. Die Integration dieser Inhalte in den Index erfolgt über entsprechende Liferay-Erweiterungen. Dazu werden die von Liferay vorgesehenen Erweiterungspunkte genutzt. Die Anbindung an den Suchindex erfolgt über eine Messaging-Lösung, welche eine Lastentkopplung ermöglicht. Insbesondere dient die Messaging-Lösung aber der Ausfallsicherheit. Die über die Messaging-Lösung bereitgestellten Daten für den Suchindex werden in einen portalspezifischen Index in Elasticsearch abgelegt. Welche Inhalte des Portals

durchsuchbar gemacht werden, wird in einer Konfiguration festgelegt.

Die Inhalte aus dem Metadatenkatalog werden ebenfalls in einen Elasticsearch-Index integriert. Zur Realisierung der Suche auf Basis von Elasticsearch wird ein Index erstellt, der aus dem bestehenden Metadatenkatalog aufgebaut wird. Werden Daten für den Metadatenkatalog bereitgestellt, werden diese in den Elasticsearch-Index automatisch aufgenommen. Dazu wurde eine entsprechende CKAN-Anpassung erstellt. Damit steht sowohl die Möglichkeit zur Verfügung, den Index inkrementell fortzuschreiben als auch den gesamten Metadatenkatalog jederzeit komplett zu reindizieren.

Das Suchverfahren Elasticsearch ermöglicht die Suche über mehrere Indexe hinweg. Eine Auftrennung der Indexe in Portal und Metadatenkatalog erleichtert eine Re-Indizierung. Für den Nutzer der Suche ist diese Auftrennung in mehrere Indexe nicht transparent, die Suche erfolgt immer über ein zentrales Eingabefeld.

Bei der Eingabe der Suchbegriffe wird der Nutzer durch Vervollständigungsfunktionen unterstützt. Dabei macht das System den Nutzern auf Basis des Portalindex als auch auf Basis der Daten aus dem Metadatenkatalog-Index Vorschläge, die während des Eintippens unterhalb des Suchfeldes eingeblendet werden. Der Nutzer kann Vorschläge auswählen und per Klick in das Suchfeld übernehmen.

Wird eine Suche mit Tippfehlern abgeschickt, bietet das Suchverfahren dem Nutzer über eine „Meinten Sie“ Funktion entsprechende Korrekturvorschläge an. Diese werden auf der Ergebnisseite unterhalb des Suchfeldes dargestellt. Per Klick kann der Nutzer eine Suche über einen Korrekturvorschlag ausführen.

Die Such- und insbesondere die Trefferliste ist so aufgebaut, dass alle für die Suche relevanten Parameter in die URL als Query-Parameter codiert werden. Dieser Deep-Link Ansatz in der Suche führt dazu, dass keine Such-Parameter in der Nutzersession auf dem Server gespeichert werden müssen. Damit kann die Suche in mehreren Browser-Fenstern und Tabs parallel ohne Seiteneffekte genutzt werden.

Sucharten

Die einfache Suche erfolgt prinzipiell immer als Volltextsuche über alle im Suchindex vorhandenen Informationen.

Die vom Nutzer sequenziell in ein Suchfeld eingegebenen Suchbegriffe werden im Suchsystem standardmäßig mit einer AND-Verknüpfung interpretiert. Der Nutzer hat jedoch auch die Möglichkeit, selbst über AND-, OR- bzw. NOT-Operatoren explizit eine logische Suche nach den eingegebenen Regeln durchzuführen.

Bei der Sucheingabe kann der Nutzer auch entsprechende Wildcards nutzen. Unterstützt werden „*“ für beliebige Zeichen und „?“ für ein beliebiges Zeichen.

In der eingesetzten Suchmaschinentechologie Elasticsearch stehen, insbesondere zur Volltextsuche, eine Reihe an Möglichkeiten und Erweiterungen zur Unterstützung von linguistischen Analysen (in Elasticsearch als Language Analyzer bezeichnet) ebenso wie die Möglichkeit, unscharfe Suchen („Fuzzy Search“) auszuführen, zur Verfügung. Für Elasticsearch wurde eine Konfiguration definiert, über welche entsprechende Analysen für die einzelnen Felder des Suchindex sprachbezogen, sofern sinnvoll, hinterlegt werden können.

Die Suchergebnisse lassen sich über eine Typisierung (Inhaltstyp) per Klick filtern. Über diesen einfachen Filter kann der Nutzer z.B. steuern, ob in der Trefferliste nur Treffer bestimmter Inhaltstypen angezeigt werden sollen. Neben der Filterfunktion nach unterschiedlichen Inhaltstypen wie Apps, Dokumente, Datensatz, Portalinhalte usw. stehen dem Nutzer zur weiteren Filterung entsprechende Facetten im linken Seitenbereich zur Verfügung. Die Facetten-Filterung passt sich automatisch an den ausgewählten Typ zur Filterung an. Damit werden dem

Nutzer kontextbezogenen Filteroptionen angeboten. Die Darstellung der Facetten erfolgt über ein generisches Facetten-Portlet; die Darstellung ergibt sich aus den Daten im Suchindex.

Erweiterte Suche

Für die gezielte Suche nach einzelnen Feldern im DCAT-AP.de-Metadaten-Schema wird eine Expertensuche (erweiterte Suche) im GovData-Portal angeboten. Über diese kann der Nutzer eine komplexe Suchanfrage formulieren. Die erweiterte Suche ermöglicht das Suchen in einzelnen Feldern des Metadaten-Schema.

Zur Suche wählt der Nutzer zunächst über eine Auswahlliste das Feld aus, in dem gesucht werden soll. Danach gibt er in das zugehörige Eingabefeld seine Suchbegriffe ein. Handelt es sich beim Feld nicht um ein Textfeld, sondern um einen spezielleren Datentyp, werden dem Nutzer zur Definition der Suchanfrage entsprechende Eingabehilfen angeboten. Beispielsweise kann der Nutzer im Feld „Start-Datum“ das Datum zur Suche über ein entsprechendes Widget auswählen.

Über die erweiterte Suche kann in einigen Metadatenfeldern des Katalogs gezielt gesucht werden. Es kann auch in mehreren Metadatenfeldern gleichzeitig gesucht werden, wobei die Felder mit AND verknüpft werden.

Geografische Suchfunktion

Die eingesetzte Suchtechnologie Elasticsearch bietet Möglichkeiten zur Suche von Objekten mit Geodatenbezug an. Dazu werden die Geodaten im Suchindex in speziellen Geo-Feldern (als geo_point oder in einer JSON-Struktur GeoJSON) abgelegt. Zur eigentlichen Suche von Objekten mit Geodatenbezug bietet das Suchverfahren unterschiedliche Filter an.

Über den „Geo Bounding Box“-Filter können in der Suchanfrage durch die Angabe eines einfachen Rechtecks (Punkt links oben und Punkt rechts unten) Objekte, die in einem solchen Bereich liegen, gesucht werden. Diese Form der Geo-Suche wird genutzt, wenn der Nutzer in der Oberfläche einen Bereich frei wählt.

Ebenfalls ist über die erweiterte Suchfunktion („Expertensuche“) eine Ortssuche per Texteingabe möglich, die über textuelle Ortsangaben in den Metadaten sucht.

Suchergebnisse

Darstellung der Suchergebnisse

Die einzelnen Suchergebnisse werden jeweils mit einem Titel, einem kurzen Text, der Angabe der veröffentlichenden Stelle und der Kategorie dargestellt. Für den Nutzer sind ferner auf einen Blick die enthaltenen Dateiformate durch Symbole sowie das Datum der letzten Änderung ersichtlich. Im linken Seitenbereich befinden sich die Filtermöglichkeiten (z.B. Kategorien, Schlüsselwörter, Dateiformate). Die Filter funktionieren nach dem Prinzip „Click-to-Action“, d.h. sobald ein Filter gesetzt ist, werden die Suchergebnisse dynamisch gefiltert (ohne zusätzliches Klicken auf einen Absenden Button). Gleiches gilt bei der Verfeinerung der Sucheingabe.

In der Ergebnisliste werden dem Nutzer die folgenden Funktionen angeboten:

- Filterfunktionen: nach Dokumententyp, kontextbezogen (nach unterschiedlichen Facetten, z.B. Kategorie, Dateiformat)
- Sortierfunktionen: nach Relevanz, Datum der letzten Änderung, alphabetisch – jeweils auf- und absteigend

- Suchverfeinerung: über Sucheingabe oder die erweiterte Suche
- Geografischer Bezug (Raumbezug): wird bei den Treffern in Form von Karten angezeigt.

Die Anzahl der Treffer wird oberhalb der Filtermöglichkeiten angezeigt.

Entscheidend für eine gute Ergebnisliste ist eine geeignete Sortierung. Die Ergebnisliste ist immer nach der Relevanz der Suchtreffer sortiert. Die Bestimmung der Relevanz lässt sich im Suchverfahren durch vielfältige Möglichkeiten, z.B. durch Boosting zur Anfrage- oder Indizierungszeit, beeinflussen.

Eine zentrale Leistung bei der Realisierung der Suche stellt die Justierung der Relevanzbestimmung im Suchverfahren dar. Als Kriterien für die Relevanz werden folgende Parameter einbezogen:

- Häufigkeit des Vorkommens
- Anzahl der Suchbegriffe mit Treffer (bei Angabe von mehreren Suchbegriffen)
- Gewichtung (Boosting) der Suchfelder und Genauigkeit der Übereinstimmung
- Geografische Angaben

Für den direkten Einstieg in angezeigte Treffer werden unterschiedliche Funktionen, je nach Dokumententyp, angeboten. Beispielsweise ist bei Treffern mit hinterlegten Dokumenten ein Link zum direkten Abruf des Dokuments enthalten, um die Navigationswege möglichst kurz zu halten.

Sortierung der Suchergebnisse

Neben der Standard-Sortierung nach Relevanz werden dem Nutzer auf der Ergebnisseite weitere Möglichkeiten angeboten, die Sortierung der Trefferliste zu beeinflussen:

- Datum der letzten Änderung
- Titel
- Relevanz

Die Sortierungsmöglichkeiten werden jeweils auf- und absteigend angeboten.

Analyse und Optimierung der Suche

Das Suchverfahren ist so aufgebaut, dass die Suchanfragen, die über das Portal gestellt werden, entsprechend protokolliert werden. Die Protokollierung der Suchen wird anonymisiert abgelegt, ohne Bezug zu einem Nutzer, um den Anforderungen hinsichtlich des Datenschutzes gerecht zu werden. Die Protokollierung der Suchanfragen erfolgt in einem gesonderten Elasticsearch Index. Aus den Suchanfragen kann die Suche um Funktionen, wie z.B. die Anzeige von Empfehlungen erweitert werden, die aus den gespeicherten Suchverläufen ermittelt werden.

Suchergebnis-URI und weitere Darstellungsarten

Die Such- und insbesondere die Trefferliste sind so aufgebaut, dass alle für die Suche relevanten Parameter in die URL als Query-Parameter codiert werden. Dieser Deep-Link Ansatz in der Suche führt dazu, dass keine Such-Parameter in der Nutzersession auf dem Server gespeichert werden müssen. Um die Anforderungen bezogen auf Deep-Links auf allen Seiten erfüllen zu können, ist diese Form der Umsetzung zwingend notwendig. Wenn alle suchrelevanten Informationen in der URL codiert sind, kann eine URL als Link auf eine Trefferliste an einen Nutzer versendet werden. Der Nutzer kann sich diese Trefferliste durch Aufrufen der URL anzeigen lassen, sofern sich keine Änderungen im Datenbestand ergeben haben. Diese URLs auf eine Trefferliste können vom

Nutzer aufgrund der Struktur ohne weiteres als Browser-Bookmarks genutzt werden.

Neben der Darstellung der Trefferliste als HTML-Seite können die Treffer, sortiert nach Änderungsdatum, als RSS- oder Atom-Feed über die Seite abonniert werden.

Filter (facettierte Suche)

Filterkriterien

In der Suchergebnisliste wird eine Reihe von Filtern bereitgestellt, über die der Nutzer die Suchergebnisse weiter einschränken kann:

- Art des Ergebnisses (Typ des Datensatzes (type), „Blog-Beitrag“, „sonstiger Inhalt“)
- Offenheit der Lizenz (Alle, Freie Nutzung, Eingeschränkte Nutzung)
- Kategorie (groups)
- Schlagworte (tags)
- Formate (format)
- Lizenz (license)
- Raumbezug (spatial -> coordinates)
- Zeitbezug (temporal)
- Datenbereitsteller
- Typ (bei Anwendungen)
- System (bei Anwendungen)

Mit Ausnahme des Raum- und Zeitbezugs werden die genannten Filter in Form von Facetten umgesetzt. Dabei liefert die Suchkomponente mit dem Ergebnis die Feldwerte und die Anzahl der möglichen Treffer der definierten Facetten zurück.

Die von der Suchmaschine ermittelten Facettenwerte werden in Listenform links neben den Suchergebnissen aufgeführt. Per Klick auf einen Filterwert kann der Nutzer so komfortabel seine Suchergebnisse nach Wunsch eingrenzen.

Der Filter „Zeitbezug“ erfolgt über zwei Eingabefelder „Datum von“ und „Datum bis“, in die mit Unterstützung eines Datepickers (Kalender-Widgets) Eingaben erfolgen können.

Die Nutzung des Filters „Raumbezug“ führt den Nutzer auf eine Kartenansicht, wo er die gewünschte geographische Einschränkung vornehmen kann.

Eigenschaft "Click-to-Action" für Filter

Sämtliche Facetten-Filter in der Ergebnisliste werden so umgesetzt, dass diese per „Click-to-Action“ bedienbar sind, sich die Ergebnisliste beim Anklicken eines Filters also automatisch aktualisiert. Dies gilt sowohl für die Auswahl eines Filters wie auch beim Entfernen eines ausgewählten Filters.

Darstellung von Filtermerkmalen

Da sich die Facettenfilter aus den Inhalten des Metadatenkatalogs ergeben, können hieraus teilweise große Wertelisten entstehen. Um zu vermeiden, dass Filter im oberen Bereich die Filter im unteren Bereich stark nach unten drängen und dem Nutzer die Übersicht über die Filterkategorien verlorenght, werden initial auf der Ergebnisliste die Wertelisten der Filter zunächst „eingeklappt“ dargestellt, d.h. es werden nur die Top-5

Merkmale angezeigt. Über einen Link „Weitere anzeigen“ bei jeder Filterliste, die mehr als 5 Werte umfasst, kann sich der Nutzer die ausgeblendeten Werte anzeigen lassen. Im ausgeklappten Zustand wird der Link „Weitere anzeigen“ gegen „Weitere ausblenden“ ersetzt, womit die Liste wieder eingeklappt werden kann.

Bestimmte Facetten sollen immer voll ausgeklappt dargestellt werden; dies wird so umgesetzt, dass in der Portalkonfiguration hinterlegt werden kann, welche Facetten nicht eingeklappt werden sollen.

Die Reihenfolge der Filterwerte erfolgt bei Facetten standardmäßig nach der Anzahl der möglichen Treffer bei Auswahl des Filters. Sofern es für eine Facette, beispielsweise „Offenheit der Lizenz“, Sinn macht von der Standard-Reihenfolge abzuweichen, kann z.B. eine alphabetische oder eine festgelegte Reihenfolge implementiert werden.

7.8.2.2.7 Schnittstellen und Importfunktionen

Im GovData-Prototyp waren bereits eine Reihe von Schnittstellen vorhanden, die einerseits den Nutzern der Daten ermöglichen, Metadaten über standardisierte Schnittstellen zu suchen und abzurufen, und die andererseits den Datenbereitstellern verschiedene Möglichkeiten bieten, Metadaten sowohl aktiv an den Metadatenkatalog von GovData zu übergeben, als auch Massendaten für eine aktive Übernahme durch GovData (Harvesting) bereitzustellen.

Für die Abfrage von Metadaten durch die Datennutzer sowie für die aktive Übergabe von Daten durch die Datenbereitsteller (Push) kommt dabei die Webservice-Schnittstelle des CKAN-Servers (CKAN-API) zum Einsatz.

Der automatisierte Abruf von Daten aus den bereitstellenden Portalen durch GovData (Pull) erfolgt über sogenannte Harvester, die als Erweiterungen von CKAN implementiert sind.

Für die Abfrage von Metadaten durch die Nutzer wird der vom CKAN-Server zur Verfügung gestellte RDF-Endpunkt bereitgestellt. Seit der CKAN-Version 2.8.x steht nur noch die neuere Version 3 (die sogenannte „Action-API“) der Schnittstelle zur Verfügung.

Es wurde ein allgemeiner DCAT-AP.de-Harvester implementiert, mit dem alle entsprechend dem DCAT-AP.de-Metadatenchema bereitgestellten RDF-Endpunkte (RDF-XML, JSON-LD, Turtle) eingelesen werden können.

Automatisierter Import

Push-Verfahren

Die Übertragung von Metadaten im Push-Verfahren findet vom Datenbereitsteller aus direkt über die CKAN-API statt.

Um die Nutzung der Schnittstelle durch unautorisierte Dritte zu verhindern, wird die Schnittstelle über einen API-Key abgesichert. Dieser API-Key muss bei jedem Request gegen die geschützten Schnittstellenfunktionen mit übertragen werden. Die Bereitstellung eines solchen individuellen API-Keys erfolgt auf organisatorischem Weg über die Geschäfts- und Koordinierungsstelle des GovData-Portals. Der API-Key wird von CKAN automatisch für jeden Datenbereitsteller erzeugt und kann von der GKst GovData in der Benutzerverwaltung des Portals zum Zweck der Weitergabe an den Datenbereitsteller eingesehen werden.

Pull-Verfahren: Harvesting

Für den Pull-Import der Daten werden sogenannte „Harvester“ eingesetzt. Diese rufen die Daten aktiv vom

Server des Datenbereitstellers ab.

Über einen Scheduling-Mechanismus wird das Harvesting zeitgesteuert ausgeführt, sodass z.B. eine tägliche Aktualisierung realisiert werden kann.

Die importierten Metadatensätze können pro Quelle einem einzelnen Bereitsteller-Konto zugeordnet werden.

Der Ablauf eines Harvesting-Prozesses wird auf dem folgenden Schaubild vereinfacht dargestellt.

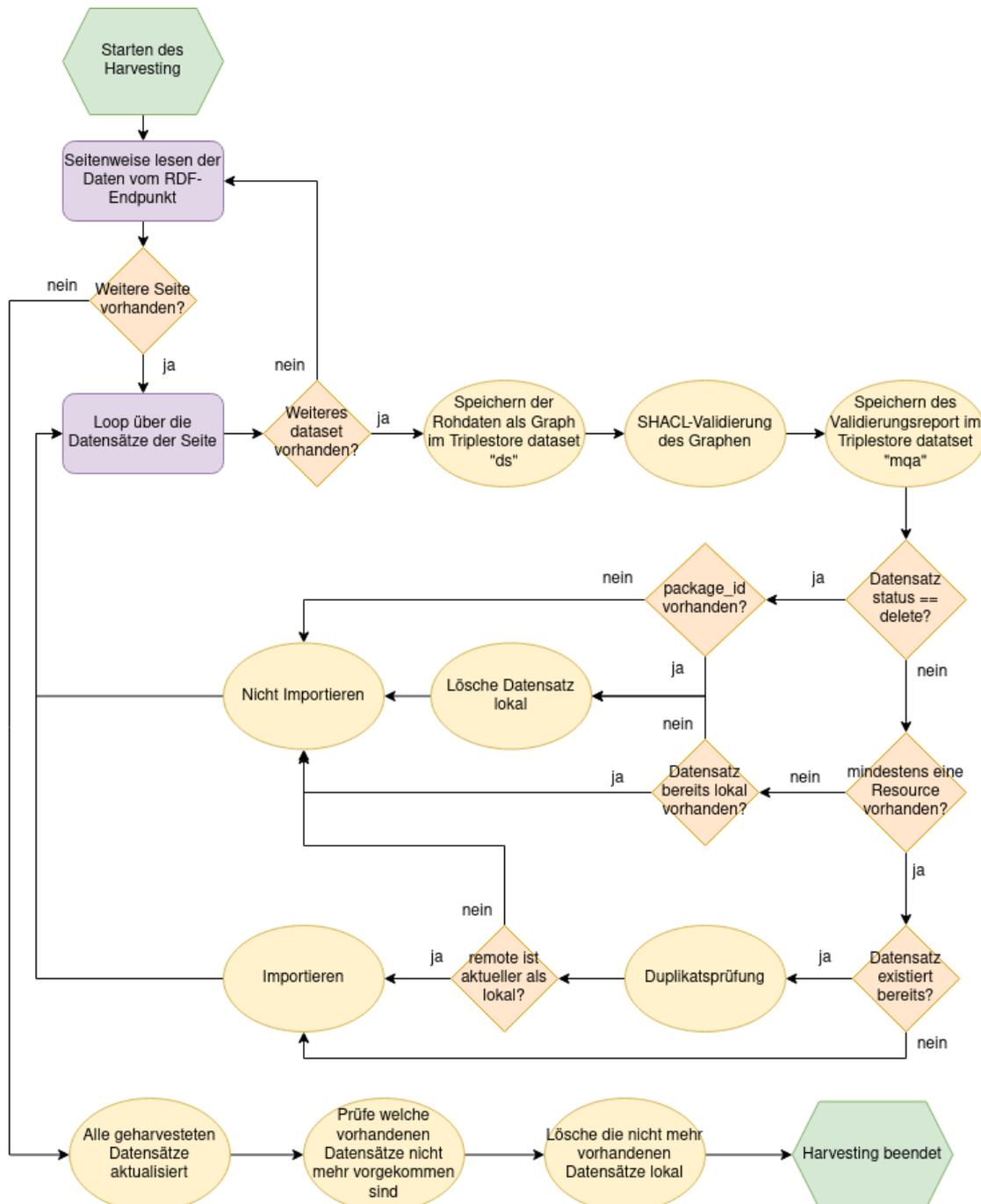


Abbildung 57: Harvesting Prozess

Zunächst wird der RDF-Endpoint Seite für Seite eingelesen. Auf jeder Seite befinden sich in der Regel mehrere Datasets. Diese werden nacheinander verarbeitet. Zunächst wird ein Dataset als RDF-Graph serialisiert und in

den TripleStore „ds“ geschrieben. Der Graph wird auch in den ITB-SHACL-Validator gegeben wo er analysiert wird. Dieser Prozess ist in Kapitel 7.8.2.2.4 genauer beschrieben.

Sobald dieser Vorgang beendet ist beginnt die „Import-Phase“ in der entschieden wird ob ein Dataset in die Datenbank geschrieben werden soll oder nicht. Zunächst wird geprüft, ob das Dataset über einen Statuswert „delete“ besitzt, welcher angibt, dass der Datensatz nicht importiert werden soll und falls er bereits vorhanden ist gelöscht werden soll. Als nächstes wird geprüft, ob der remote Datensatz über mindestens eine Ressource verfügt. Datensätze ohne gültige Ressource werden nicht importiert. Als nächster Schritt wird geprüft, ob der remote Datensatz bereits in der Datenbank vorliegt. Falls nicht handelt es sich um einen neuen Datensatz welcher importiert werden soll. Falls aber bereits eine Version des remote Datensatzes vorliegt muss zunächst geprüft werden, ob der lokale Datensatz überschrieben werden soll oder nicht. Dieser Vorgang ist im nächsten Abschnitt „Dublettenprüfung beim Import von Datensätzen“ genauer beschrieben.

Sobald alle geharvesteten Datensätze verarbeitet wurden wird geprüft welche bisher vorhandenen geharvesteten Datensätze nicht mehr vorhanden sind. Diese Datensätze werden gelöscht.

Dublettenprüfung beim Import von Datensätzen

Nachdem ein Datensatz geharvestet wird, muss geschaut werden, ob dieser Datensatz bereits in der Datenbank besteht, um zu entscheiden, ob der neue Datensatz importiert, aktualisiert oder verworfen werden soll. Dieser Schritt ist wichtig, um zu verhindern, dass Datensätze mehrfach vorkommen können.

Der Ablauf der Dublettenprüfung ist auf dem folgenden Schaubild abgebildet.

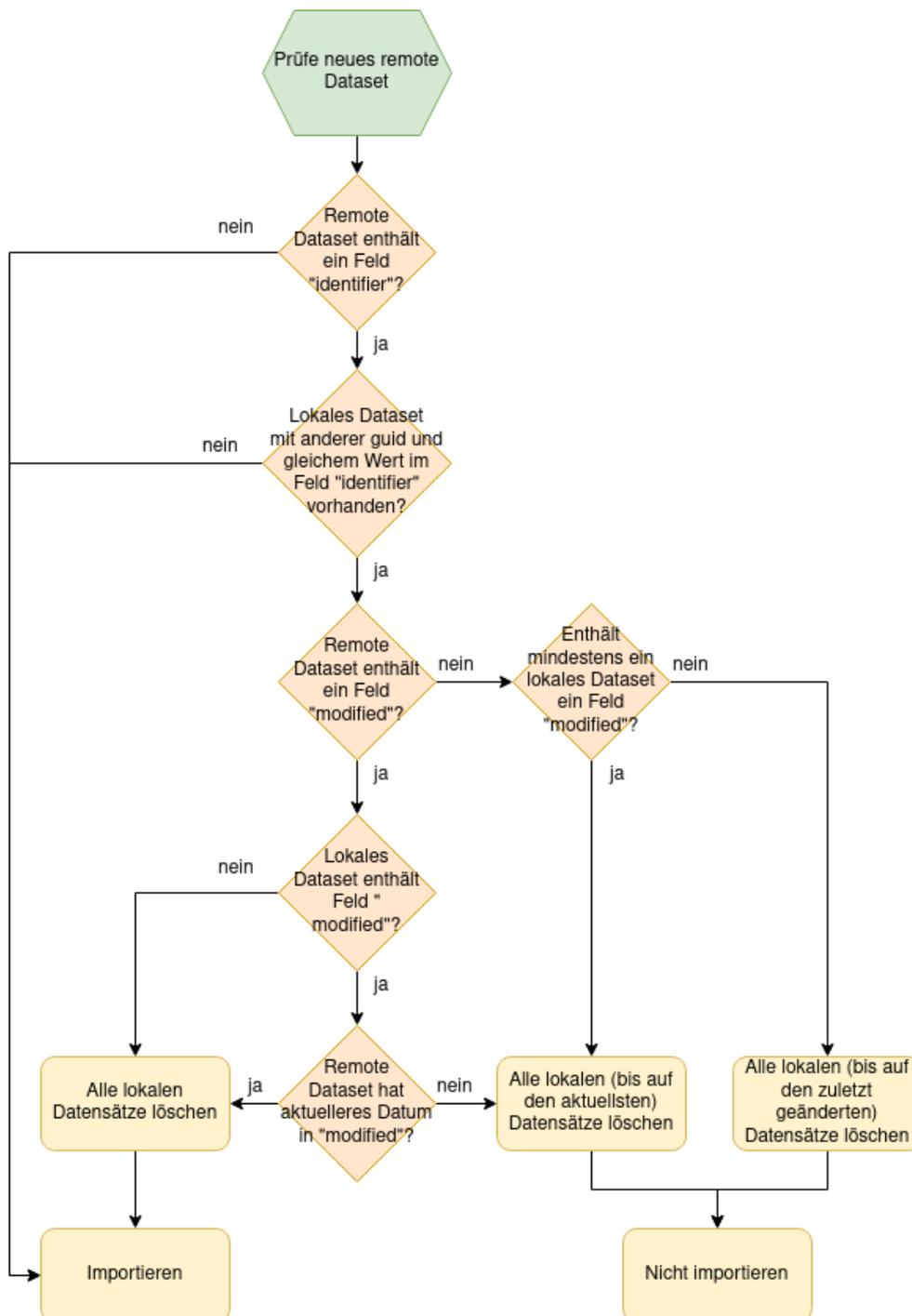


Abbildung 58: Dublettenprüfung bei Datensets

Manuelle Bereitstellung

Metadaten-Liste der Datenbereitsteller

Datenbereitsteller können sich eine Liste der von ihnen verwalteten Metadaten im GovData-Portal anzeigen lassen. Im Wesentlichen handelt es sich dabei um eine ähnliche Ansicht wie die Suchergebnisseite, jedoch ist automatisch der Filter „Datenbereitsteller“ so gesetzt, dass nur die vom Nutzer verwalteten Datensätze angezeigt werden.

Zur Navigation in der Liste und zur Einschränkung der angezeigten Datensätze stehen dieselben Funktionen

(Filter, Sortierung, Blätterfunktion) wie in der Suchergebnisseite zur Verfügung.

Manuelle Bereitstellung von Metadaten

Für Portal-Nutzer mit der Rolle „Datenbereitsteller“ stehen webbasierte Oberflächen zur manuellen Erstellung und Bearbeitung von Metadatenansätzen zur Verfügung.

Der Nutzer kann sich eine Liste mit von ihm verwalteten Metadatenansätzen anzeigen lassen. Durch Auswahl eines Eintrags in der Liste gelangt der Nutzer auf eine Detailseite, auf der die Details des Datensatzes dargestellt und bearbeitet werden können. Über einen Link auf der Übersichtsseite kann auch ein neuer Datensatz erfasst werden.

Das Webformular zur manuellen Eingabe eines Datensatzes umfasst mindestens die durch das DCAT-AP.de-Metadatenchema vorgegebenen Pflichtfelder. Darüber hinaus enthält es die folgenden Metadatenfelder:

- Webseite Datensatz (url)
- Schlagwörter (tags)
- Webseite veröffentlichende Stelle (publisher url)
- Adresse veröffentlichende Stelle (publisher address)
- Ansprechpartner (Feldgruppe gemäß DCAT-AP.de-Metadatenchema)
- Autor (Feldgruppe gemäß DCAT-AP.de-Metadatenchema)
- Urheber (Feldgruppe gemäß DCAT-AP.de-Metadatenchema)
- Zeitliches und Geografisches (Feldgruppe gemäß DCAT-AP.de-Metadatenchema)
- Räumliche Abdeckung (Feldgruppe gemäß DCAT-AP.de-Metadatenchema)
- Ressourcen (Feldgruppe gemäß DCAT-AP.de-Metadatenchema)
- Sprache der Ressourcen

Für jede Datenquelle können die Felder URL, Name, Beschreibung und Format (JSON, XML, XSL etc.) angegeben werden. Alle Daten werden bei der Eingabe auf Validität und Plausibilität geprüft. Die Validierung der Felder erfolgt anhand der Metadatenfelder und Wertemengen des DCAT-AP.de-Metadatenchemas.

7.8.2.3 Reporting

7.8.2.3.1 Webseitenreports bzw. Webdatenanalysen

Um die Popularität von GovData sowie Muster in der Nutzung des Datenportals erfassen zu können wird das Webseitenanalyse-Tool Matomo eingesetzt, um z.B. monatlich Webseitenreports bzw. Webdatenanalysen anfertigen.

Dabei beinhalten diese Berichte mindestens die folgenden Kennzahlen:

- Verfügbarkeit
- Performance
- Besucheranzahl
- Einstiegsseiten
- Ausstiegsseiten
- Durchschnittliche Verweildauer
- Absprungrate

- Geografische Herkunft
- Browserstatistik
- Auflösungsstatistik (Endgerät)
- Wiederkehrende Nutzer
- Verweisende Links
- Durchschnittliche Seitenaufrufe
- Pfadanalyse
- Keywords / Schlagworte (z.B. in Blog-Artikeln)
- Verwendete Suchparameter
- Aufgerufene Webseiten

7.8.2.3.2 Auswertung über Suchparameter

Die Erstellung der Webseitenreports bzw. Webdatenanalysen erfolgt mittels Matomo, eines für den Bereich Reporting vorgesehenen Open-Source-Produktes. Der Report kann als PDF-Dokument oder im HTML-Format generiert werden.

Matomo bietet über seine Reporting-HTTP-API die Möglichkeit die Daten in maschinenlesbarer Form abzurufen und zur Verfügung zu stellen. Die Abfragen an die Reporting-API für die in der Portal-Suche verwendeten Suchbegriffe können dabei an die Anforderungen angepasst werden (z.B. Aufruf der Blog-Artikelseiten, Metadatenatz-Detailseiten). Neben JSON stehen als weitere Formate XML, CSV, TSV, HTML, PHP und RSS zur Auswahl.

Die Abfragen an die Reporting-API können über die entsprechenden Methoden mit den gewünschten Parametern durchgeführt werden. Beispielsweise liefert die folgende Abfrage

```
http://beispielwebsite.org/?module=API&method=Actions.getSiteSearchKeywords&period=month&date=previous1&format=json
```

den Report der Suchbegriffe für die Webseite <http://beispielwebsite.org> für den letzten Monat im JSON-Format.

7.8.2.3.3 Gewährleistung HamDSG

Bei der Generierung der o.g. Kennzahlen kommen nur Werkzeuge und Technologien zum Einsatz, die eine Gewährleistung des HamDSG bzw. der DSGVO ermöglichen.

7.8.2.3.4 Anonymisierung von Nutzerdaten

Dateiformat und Prozess der Bereitstellung

Die Rohdaten der ermittelten Kennzahlen werden über Matomo ermittelt. Die datenschutzrechtlichen sensiblen Logdaten werden auf dem eigenen Server gespeichert, sodass die Einhaltung des HamDSG gewährleistet ist. Eine Anonymisierung der IP-Adressen ist durch Aktivierung eines Plugins „AnonymizeIP“ und Einstellungen zur Löschung der letzten beiden Oktette der IP-Adresse gewährleistet.

Alle Auswertungen, die das verwendete Produkt über die Oberflächen bereitstellt, können auch über die API abgerufen werden. Über diese API lassen sich alle angeforderten Kennzahlen abrufen. Mit den entsprechenden URL-Parametern können die Daten zyklisch und auch automatisiert abgeholt und an anderer Stelle weiterverwendet werden. Folgende Formate stehen zur Verfügung: XML, JSON, CSV, HTML, PHP, RSS und TSV. Aus Sicht des AN eignen sich für die gestellten Anforderungen die gängigen Formate XML und JSON im

Hinblick auf die Bereitstellung der Daten am besten.

Für die Kennzahlen Verfügbarkeit und Performance nutzt SEITENBAU ein weiteres Werkzeug, welches folgende Informationen bietet:

- Überwachung von Verfügbarkeit und Performance von verschiedenen Messpunkten aus
- Verschiedene Auswertungen zu Verfügbarkeit und Performance
- Versand von Reports per E-Mail in maschinenlesbarem CSV-Format
- API zum Abruf von Status- und Verfügbarkeitsinformationen.

Kennzahlen der Webdatenanalyse ad-hoc einsehen

Matomo ist als Software auf dem eigenen Server installiert. Das bedeutet, dass auch die datenschutzrechtlichen sensiblen Logdaten auf dem eigenen Server gespeichert werden.

Das Werkzeug verfügt über benutzerfreundliche Oberflächen und bietet dem Nutzer die Möglichkeit, die für ihn wichtigsten Analysen in einem Dashboard darzustellen und ad-hoc einzusehen. Dieses lässt sich individuell mithilfe vorhandener Widgets konfigurieren.

In verschiedenen Reitern kann sich der Nutzer folgende Daten anzeigen lassen:

- Besucher: z.B. Herkunftsseite, besuchte Seiten, Endgeräte, geografische Herkunft
- Aktionen: z.B. Einstiegs-, Ausstiegsseiten, Suchbegriffe (interne Suche auf der Webseite), aufgerufene Webseiten
- Verweise: z.B. Suchbegriffe (in Suchmaschinen), Browserstatistik, verweisende Links.
- Ziele: z.B. Auswertungen zum Ziel „Nutzerregistrierung“, sofern dieses Ziel festgelegt wurde.

Über die Eingabe eines Auswahlzeitraumes können die Daten auf bestimmte Zeiträume (Tag, Woche, Monat, Jahr oder Zeitspanne) eingeschränkt werden. Die Daten in der Ansicht „Besucher in Echtzeit“ sind aktuell; die übrigen Auswertungen werden einmal täglich erneuert.

Die Daten können über die umfangreiche API abgeholt und zur weiteren Verwendung gespeichert werden. Dafür stehen die Formate CSV, Excel, JSON, PHP und XML zur Verfügung. Darüber hinaus können Reports per E-Mail versendet werden. Der Versandzeitpunkt kann hierbei konfiguriert werden (z.B. wöchentlich, monatlich). Für den Report-Versand per E-Mail stehen das PDF- und das HTML-Format zur Verfügung.

7.8.3 Nicht-funktionale Anforderungen

7.8.3.1 Betriebsumgebung und Beschaffenheit des Rechenzentrums

Der Betrieb des angebotenen Systems findet in einem Rechenzentrum in Deutschland statt, das für eine leistungsfähige, redundante Anbindung an zentrale Internetknoten sowie für hoch verfügbare Anwendungen optimiert ist.

Das Rechenzentrum ist nach verschiedenen Standards zertifiziert, insbesondere liegen folgende Zertifizierungen vor:

- ISO 27001-Zertifizierung
- ITIL®-Zertifizierung
- Zertifizierung nach IT-Grundschutz-Katalog des BSI.

Sicherheit und Verfügbarkeit der Systeme sind durch eine Reihe von Maßnahmen auf allen Ebenen und durch

die Anwendung gängiger Standards und Methoden sichergestellt.

7.8.3.1.1 Bereitstellung einer betriebsbereiten Systemumgebung

GovData wird als Webanwendung in einer klassischen 3-Schicht-Architektur umgesetzt und unterhalten. Die Darstellung erfolgt dabei rein webbasiert, die Kommunikation mit dem Server erfolgt über HTTP/HTTPS.

Der Betrieb erfolgt in einer zweistufigen Systemarchitektur. Neben dem Produktivsystem wird eine zweite, unabhängige Instanz des Systems für Test-, Integrations- und Schulungszwecke bereitgestellt. Die Daten des Produktivsystems werden dabei in regelmäßigen Abständen oder bei Bedarf in pseudonymisierter Form in das Testsystem übernommen.

Darüber hinaus wird beim Auftragnehmer eine Systeminstanz für Entwicklung und Test von Systemanpassungen und Erweiterungen betrieben.

7.8.3.1.2 Test der Systemumgebung

Vor der Erklärung der Abnahme werden alle wesentlichen Funktionen von GovData und dessen einzelnen Komponenten auf verschiedene Qualitätskriterien hin getestet, insbesondere

- Fachliche Funktionalität
- Performance und Lastverhalten (bei Änderungen an der Server-Infrastruktur)
- Sicherheit
- Verfügbarkeit.

Der Auftragnehmer erstellt dazu eine Testfallspezifikation mittels einer durch den Auftragnehmer bereitgestellten Testmanagementlösung (beispielsweise dem Open-Source-Werkzeug TestLink), die mit dem Auftraggeber abgestimmt wird.

Die Tests werden weitestgehend automatisiert durchgeführt. Dazu werden unterschiedliche Testwerkzeuge eingesetzt, wie Selenium für die funktionalen Tests der Benutzeroberflächen oder JMeter für Last- und Performancetests. Die automatisierten Tests werden durch manuell durchgeführte explorative Tests ergänzt, um eine optimale Testabdeckung zu gewährleisten.

Zusätzlich zu den Tests vor der Installation einer neuen Version, werden Performance und Lastverhalten – insbesondere die Einhaltung der zugesicherten Antwortzeiten vor dem Einspielen neuer Software-Releases – durch automatisierte Tests überprüft. Diese werden allerdings nur nach Bedarf ausgeführt, oder wenn sich eine Änderung an der Server-Infrastruktur ergeben hat.

Die Ergebnisse aller Tests werden in Form eines Testprotokolls dokumentiert und dem Auftraggeber per Ticketsystem (JIRA) zur Verfügung gestellt.

7.8.3.2 Verfügbarkeit, Performanz und Skalierbarkeit

Die Verfügbarkeit der Systeme wird durch eine Reihe von Maßnahmen auf allen Ebenen und durch die Anwendung gängiger Standards und Methoden sichergestellt.

Die einzelnen Komponenten von GovData werden in virtuellen Servern innerhalb dieser Umgebung betrieben.

Als Betriebssystem kommt dabei die Long-Time-Support-Version von Ubuntu Linux zum Einsatz, bei der Sicherheitsupdates für einen Zeitraum von 5 Jahren nach Erscheinen bereitgestellt werden.

Durch die Nutzung von virtualisierten Systemen kann jede Komponente von GovData (CMS/Portal, Datenbank/Suchindex, CKAN, Microservices) in einem eigenen unabhängigen System betrieben werden. Die

Systeme sind dabei auf getrennte Netzwerkzonen für Anwendung und Datenbank verteilt.

Der Zugriff durch die Administratoren ist über eine Public-Key-Authentisierung mit individuellen, personalisierten Schlüsseln abgesichert.

Bei steigenden Anforderungen an die Leistungsfähigkeit von GovData kann das System durch Hinzufügen weiterer Instanzen ohne Anpassung der Softwarekomponenten problemlos skaliert werden. Die Skalierung kann dabei für die einzelnen Komponenten (CMS, Datenbank etc.) unabhängig erfolgen.

Durch die Trennung von Geschäftslogik und Datenbank wird eine unabhängige Skalierbarkeit der Schichten erreicht. Bei steigenden Anforderungen ist eine Skalierung auch auf das Zehnfache der aktuellen Besucherzahl problemlos möglich.

Um Cyber-Attacken proaktiv zu verhindern, ist der Anwendung eine Web-Application-Firewall vorgeschaltet, die auch für die verschlüsselte Kommunikation über https sorgt. Unter anderem wird hier eine Erkennung und Blockierung massenhaft eingehender Anfragen an die Benutzeroberflächen oder Webservice-API durchgeführt, um eine Überlastung der Anwendung durch fehlkonfigurierte Clients oder gezielte Angriffe (DoS-Attacken) zu verhindern.

Alle Systeme und Services sind in ein Monitoring-System eingebunden, das ständig die Verfügbarkeit der Services sowie weitere Parameter, wie Prozessor- oder Speicherauslastung, überwacht. Bei Ausfall eines Systems oder Dienstes oder beim Überschreiten definierter Grenzen der überwachten Parameter wird automatisch eine Benachrichtigung per E-Mail oder SMS ausgelöst. Dazu kommt das Open-Source-System Icinga 2 zum Einsatz. Dabei handelt es sich um einen Fork der verbreiteten Monitoringlösung Nagios, sodass eine breite Palette an Erweiterungen und Plugins sowie ein reichhaltiges Know-How der Entwickler- und Nutzercommunity verfügbar sind.

Die Konfiguration und Pflege der Systeme erfolgt automatisiert über Werkzeuge wie Ansible, die es erlauben, Updates für alle Instanzen einheitlich und kontrolliert durchzuführen und neue Systeminstanzen nach einem Ausfall oder bei einer Erweiterung schnell installieren zu können. Die Konfigurationsdaten werden dabei in einem Versionskontrollsystem gehalten, um ein konsistentes, nachvollziehbares Konfigurationsmanagement zu gewährleisten.

7.8.3.2.1 Beschaffenheit des Rechenzentrums

Das Rechenzentrum ist nach verschiedenen Standards zertifiziert, insbesondere liegen folgende Zertifizierungen vor:

- ISO 27001-Zertifizierung
- ITIL®-Zertifizierung
- Zertifizierung nach IT-Grundschutz-Katalog des BSI.

Alle Systeme werden automatisiert nächtlich auf Backup-Speicher gesichert. Die Sicherung erfolgt dabei in Form von Image-Backups, bei denen ein Abbild der gesamten virtuellen Maschine gesichert wird. Dadurch kann eine Wiederherstellung nach dem Ausfall eines Systems innerhalb kurzer Zeit durch Zurückspielen des gesamten Abbilds erfolgen.

7.8.3.2.2 IT-Sicherheit

- Das Rechenzentrum, in dem GovData betrieben wird, ist nach verschiedenen in der öffentlichen Verwaltung gängigen Standards zertifiziert, insbesondere liegen folgende Zertifizierungen vor:
- ISO 27001-Zertifizierung
- ITIL®-Zertifizierung
- Zertifizierung nach IT-Grundschutzkatalog des BSI.

Für den Betrieb des angebotenen Systems sind im Rahmen eines Informationssicherheits-Managements nach ISO 27001 die Verantwortlichkeiten für die Informationssicherheit festgelegt und dokumentiert. Ferner sind verbindliche Prozesse für den Umgang mit Sicherheitsvorfällen sowie eine kontinuierliche Überprüfung der Umsetzung und Wirksamkeit der Sicherheitsmaßnahmen definiert. Die Einhaltung der Prozesse sowie die kontinuierliche Überprüfung werden durch den Auftragnehmer gewährleistet.

Auf Basis der IT-Grundschutzkataloge des BSI werden die erforderlichen Sicherheitsmaßnahmen definiert und umgesetzt. Dabei werden auch die datenschutzrechtlichen Aspekte berücksichtigt, um die Einhaltung der Verpflichtungen des Auftraggebers nach dem Hamburger Datenschutzgesetz sicherzustellen.

Vom AN erfolgt bei Bedarf eine Zuarbeit bei der Erstellung der für die Abstimmung mit dem HmbBfDI notwendigen Dokumente.

Die Service- und Supportprozesse, insbesondere zum Event-, Incident- und Problem-Management orientieren sich an den Vorgaben der IT Infrastructure Library (ITIL) Service Operation.

7.8.3.3 Systemservice und Hotline

Bei SEITENBAU steht ein festes IT-Service- und Support-Team aus derzeit elf Mitarbeitern für interne IT- und Netzwerkstrukturen, Hosting sowie für den Support von Kundensystemen zur Verfügung. Die Service- und Supportprozesse, insbesondere zum Event-, Incident- und Problem-Management orientieren sich dabei an den Vorgaben der IT Infrastructure Library (ITIL) Service Operation.

Der Support kann über folgende Kanäle vom Auftraggeber kontaktiert werden:

- Anruf einer dedizierten Supporttelefonnummer,
- E-Mail an Support-E-Mail-Adresse (govdata-support@seitenbau.com) sowie
- Supportanfrage über ein vom Auftragnehmer bereitgestelltes, webbasiertes Ticketsystem (Atlassian JIRA).

Support-Anfragen von den Ansprechpartnern der geharvesteten Datenportale können vom Auftraggeber an den Auftragnehmer weitergeleitet werden können. Eingehende Supportanfragen werden vom Support-Mitarbeiter in das Ticketsystem eingetragen und gepflegt, um die Nachvollziehbarkeit zu gewährleisten.

Der Auftragnehmer gewährleistet die Erreichbarkeit der Supporthotline über die genannten Kanäle während der vereinbarten Servicezeiten (Montag bis Freitag von 9 bis 17 Uhr außer an Feiertagen in Hamburg/Berlin). Über die Supporttelefonnummer kann optional eine erweiterte Rufbereitschaft außerhalb der regulären Supportzeiten bereitgestellt werden. Während den angegebenen Servicezeiten stehen qualifizierte Mitarbeiter als feste Ansprechpartner zur Verfügung.

Alle Schritte einer Meldung von der Supportanfrage/Störungsmeldung bis zur Lösung/Störungsbeseitigung werden im bereitgestellten, webbasierten Ticketsystem als Ticket geführt und dokumentiert. Erfolgt eine Supportanfrage per Telefon oder E-Mail, wird das Ticket durch den Supportmitarbeiter erstellt, der die Meldung

entgegennimmt. Der Auftraggeber kann jederzeit auf das webbasierte Ticketsystem zugreifen und den Status der Tickets verfolgen.

Bei Eingang einer Supportanfrage erfolgt eine Rückmeldung beim meldenden Mitarbeiter des Auftraggebers. Der Kontakt erfolgt dabei über den Kanal, über den die ursprüngliche Anfrage erfolgte. Zu diesem Zeitpunkt wird auch die Kategorie bzw. mögliche Störungsklasse der Meldung/Störung festgelegt.

SEITENBAU klassifiziert die vom Auftraggeber gestellten Supportanfragen in zwei verschiedene Kategorien:

- Anfragen bzw. Änderungen
 - o Wird eine eingehende Meldung als Anfrage bzw. Änderung klassifiziert, wird dies im Ticketsystem JIRA dokumentiert. SEITENBAU bewertet die Anfrage oder Änderung und schätzt den zur Umsetzung notwendigen Aufwand. Gemeinsam mit dem Auftraggeber wird anschließend entschieden, welche Anfragen bzw. Änderungen im Rahmen der Weiterentwicklung umgesetzt werden.
- Störungen
 - o Wird eine eingehende Meldung als Störung identifiziert, wird diese im Rahmen des Supports in die nachfolgenden 3 Klassen eingeteilt und gemäß den geforderten Reaktions- und Wiederherstellungszeiten behoben.
 - o Klasse 1 - betriebsverhindernder Mangel:
 - Das betroffene Softwaresystem GovData oder wesentliche Komponenten davon sind vollständig ausgefallen, eine Nutzung des Systems ist nicht möglich.
 - o Klasse 2 - betriebsbehindernder Mangel:
 - Das betroffene Softwaresystem oder wesentliche Komponenten davon können nur noch eingeschränkt genutzt werden, sind aber im Wesentlichen verfügbar. Das Problem kann kurzzeitig mit organisatorischen oder sonstigen Hilfsmitteln umgangen werden.
 - o Klasse 3 - sonstiger Mangel:
 - Beeinträchtigungen bei Darstellung oder Funktion ohne wesentliche Auswirkungen auf die Nutzbarkeit des Systems.

Fällt eine Störung in die zuvor genannten Klassen, beginnt der Supportmitarbeiter innerhalb der vorgegebenen Reaktionszeiten mit der Fehleranalyse und erfragt gegebenenfalls weitere notwendige Informationen vom meldenden Ansprechpartner. Dabei wird er durch eine interne, kompakte Dokumentation mit allen notwendigen Informationen zu den betroffenen Systemen sowie zu den zuständigen Ansprechpartnern beim Auftragnehmer als auch beim Auftraggeber, unterstützt. Eine schnelle Fehleranalyse sowie Erkennung und Behebung typischer Fehlerursachen wird durch eigens erstellte Checklisten und Anleitungen der Mitarbeiter, welche die Systeme und die Anwendung kennen, sichergestellt. Diese Dokumentation wird kontinuierlich erweitert.

Nach einer ersten Fehleranalyse informiert der Supportmitarbeiter den Auftraggeber über deren Ergebnis und die voraussichtliche Dauer der Fehlerbehebung.

Für den Fall von Störungen, die nicht durch die Mitarbeiter des Supportteams behoben werden können, wird für jeden Zeitpunkt innerhalb des Supportfensters ein Verantwortlicher für den 3rd-Level-Support festgelegt, der über detailliertes Fachwissen des GovData-Systems verfügt. Kann das Problem nicht unmittelbar durch den Supportmitarbeiter behoben werden, leitet er die Meldung unverzüglich an diesen Verantwortlichen weiter, der

den Supportmitarbeiter bei der weiteren Störungsbearbeitung unterstützt, bei Bedarf mit dem entsprechenden Mitarbeiter des Auftraggebers direkten Kontakt aufnimmt und zusammen mit dem Ansprechpartner für den 2nd-Level-Support und dem Auftraggeber die weiteren Schritte zur Störungsbehebung definiert.

Das wesentliche Ziel des Incident Management ist die schnelle Wiederherstellung eines betriebsfähigen Zustandes. Diese wird durch die redundante Auslegung aller System- und Infrastrukturkomponenten sowie den Betrieb in einer virtualisierten Umgebung sichergestellt. Durch die Nutzung virtueller Maschinen und Image-basierter Datensicherung können die einzelnen Komponenten selbst bei vollständigem Ausfall innerhalb kurzer Zeit, d.h. in weniger als acht Stunden durch Zurückspielen der gesicherten Abbilder wiederhergestellt werden. Beim Ausfall von Hardwarekomponenten können die Supportmitarbeiter auf eine 24-Stunden-Hotline des Rechenzentrums zugreifen.

Der Supportprozess ist in folgender Abbildung dargestellt.

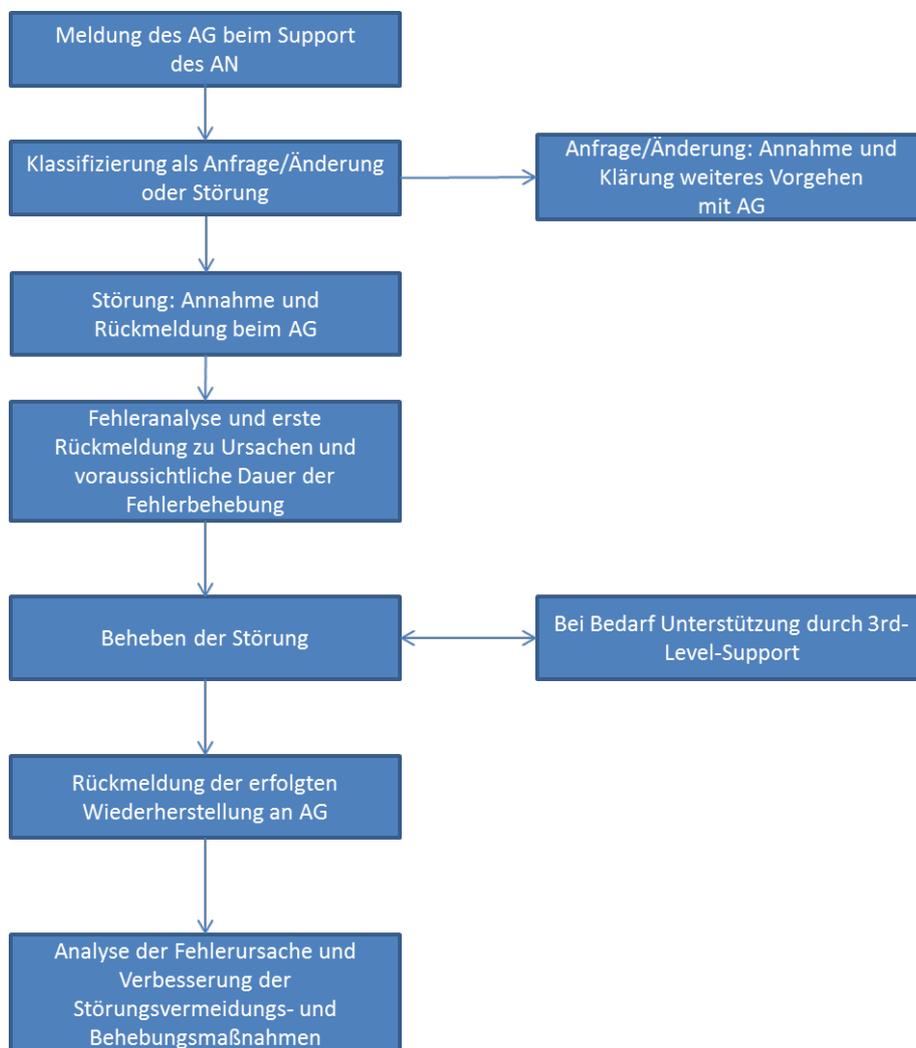


Abbildung 59: Supportprozess

Für den Betrieb von GovData stehen zwei feste Mitarbeiter des Support-Teams im Zuge des 2nd-Level-Supports als Ansprechpartner für die Mitarbeiter der Geschäfts- und Koordinierungsstelle zur Verfügung, die bestens mit der Systemumgebung von GovData vertraut sind. Über Dienstpläne und Vertreterregelungen wird generell sichergestellt, dass auch bei Abwesenheiten eine ständige Erreichbarkeit des Supports gewährleistet ist. Alle Mitarbeiter des Supportteams sind intensiv mit der Systemumgebung und den verwendeten Betriebs- und Middlewaresystemen vertraut und können Störungen innerhalb kurzer Zeit selbst beheben.

Im Rahmen eines Problem-Management-Prozesses werden die Fehlerursachen auf Basis der dokumentierten Störungsmeldungen analysiert und daraus Maßnahmen zur Fehlervermeidung und zur optimierten Fehleranalyse und -behebung abgeleitet, welche wiederum zur Verbesserung der künftigen Störungsbeseitigung dienen.

7.8.4 Lebenszyklusanalyse und Gesamtsystemarchitektur

7.8.4.1 Lebenszyklusanalyse

Das Projekt umfasst die Phasen Entwicklung, Datenmigration und Inbetriebnahme, Regelbetrieb, Wartung und Weiterentwicklung des neuen GovData-Portals.

7.8.4.2 Anwendungsarchitektur

Kernstück des Portals bildet die Community Edition der Portalsoftware Liferay in Version 7.4, die bereits weitreichende Funktionen wie CMS, Blog, Medienmanagement usw. bereitstellt. Über das Liferay-Portal wird im Wesentlichen die grafische Benutzeroberfläche realisiert, auf die über einen Web Browser zugegriffen werden kann. Die Funktionalität von Liferay wird über Portlets erweitert, die im Rahmen der Weiterentwicklung des GovData-Portals umgesetzt und integriert werden. Diese zu entwickelnden Portlets umfassen beispielsweise die Kartensuche, Ergebnislisten der Suche, Detailansichten der Metadatenätze, Dialoge zur Verwaltung der Metadaten usw.

Die zweite Kernkomponente des Portals bildet die Open Source Software CKAN in Version 2.9, die den gesamten Metadatenkatalog des Portals verwaltet und recherchierbar macht. Die Datensammlung erfolgt über einen Harvester, der die von anderen Open-Data-Katalogen bereitgestellten Daten (RDF-Endpunkt mit DCAT-AP.de) automatisiert abrufen und in den eigenen Metadatenkatalog aufnimmt und bestehende Daten aktualisiert.

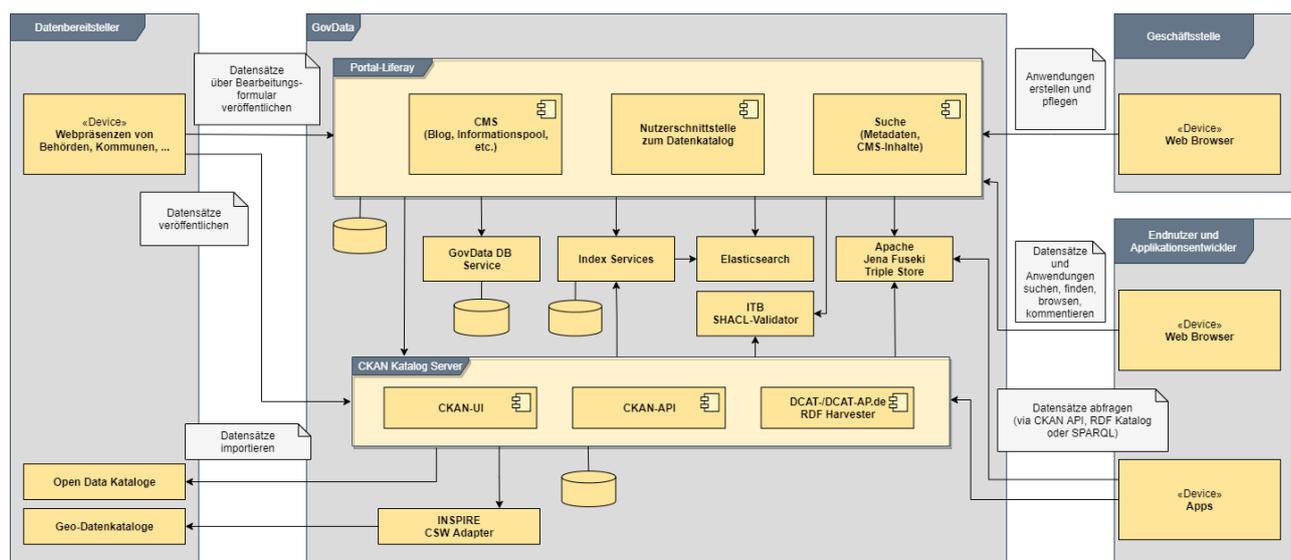


Abbildung 60: Übersicht Anwendungsarchitektur

CKAN exponiert eine API (RDF-Endpunkt und CKAN-JSON), über die die GovData-Benutzerschnittstelle oder auch externe Systeme (z.B. Apps) auf die Inhalte der Metadaten-Kataloge von GovData zugreifen und diese weiterverarbeiten können.

Neben CKAN werden alle öffentlichen Metadaten in einem TripleStore gespeichert. Im Gegensatz zu CKAN werden die originalen Metadaten als RDF verlustfrei gespeichert. Als TripleStore ist ein Apache Jena Fuseki in Version 4.5.0 installiert. Eine weitere Möglichkeit auf die Metadaten zuzugreifen wird über den Fuseki TripleStore geboten. Dieser besitzt ebenfalls eine Schnittstelle über die auf die Originaldaten zugegriffen werden kann und die über <https://www.govdata.de/sparql> exponiert wird.

Für die Suche über alle Inhalte des Portals ist ein Elasticsearch-Server in das Portal integriert. Über diesen werden sowohl die Informationsseiten des Portals wie auch der Metadatenkatalog indexiert. Um eine Indexierung des Metadatenkatalogs in Elasticsearch analog zur Indexierung durch CKAN zu ermöglichen, wurde CKAN um ein Plugin erweitert, welches sich direkt in den Indexierungsprozess einhängt und die Daten an den Indexer der Elasticsearch-Komponente weitergibt.

7.8.4.2.1 Anwendung Liferay

Passwort zurücksetzen

Passwörter werden niemals per E-Mail verschickt. Falls ein Benutzer sein Passwort nicht mehr weiß, kann über die „Passwort vergessen“-Funktion mit der Angabe der verwendeten E-Mail-Adresse eine E-Mail zum Zurücksetzen des Passworts angefordert werden.

Kennwortrichtlinien

Es gelten die folgenden Kennwortrichtlinien:

- a) Mindestlänge des Passworts: 6 Zeichen
- b) Minimale Anzahl Zahlen: 1
- c) Minimale Anzahl großer Buchstaben: 1
- d) Passwort-Historie: Keine

Speicherung von Passwörtern

Die Passwörter werden in der von Liferay genutzten PostgreSQL-Datenbank gespeichert. Dabei werden die Passwörter mit dem Verschlüsselungsalgorithmus SHA-1 (legacy) oder PBKDF2WithHmacSHA1/160/128000 verschlüsselt.

7.8.4.2.2 Anwendung CKAN

Speicherung von Passwörtern

Die Passwörter werden in der von CKAN genutzten PostgreSQL-Datenbank gespeichert. Dabei werden die Passwörter mit dem Verschlüsselungsalgorithmus PBKDF2 mit SHA-512 gespeichert. Lediglich für Benutzer, die in der CKAN-Benutzeroberfläche Aktionen durchführen dürfen, werden auch reale Passwörter vergeben.

Für die regulären Portalnutzer werden automatisiert Benutzerkonten angelegt, die ein generiertes Passwort vergeben, das aus einer Zufallszahl mit einer Länge von 16 Zeichen besteht.

7.8.4.3 Gesamtsystemarchitektur

GovData wird als Webanwendung in einer klassischen 3-Schicht-Architektur umgesetzt und unterhalten. Die Darstellung erfolgt dabei rein webbasiert, die Kommunikation mit dem Server erfolgt über HTTP/HTTPS.

Der Anwendung vorgeschaltet ist ein Apache Webserver, der für die Absicherung der Verbindung zu den Clients über https sorgt und eine Web Application Firewall (mod_security) bereitstellt, die verschiedene Cyber-Attacken erkennt und verhindert. Unter anderem werden Mechanismen zur Drosselung der Verarbeitung massenhaft eingehender Anfragen implementiert, um eine Überlastung der Anwendung durch fehlerkonfigurierte Clients oder gezielte Angriffe (DoS-Attacken) zu verhindern.

Ebenso erfolgen sämtliche Zugriffe der Applikation auf externe Systeme über einen abgesicherten Proxy für ausgehenden Datenverkehr.

In den darunterliegenden Schichten erfolgt eine Trennung von Geschäftslogik (Portalsoftware Liferay, CKAN Server und Harvester) und Datenhaltung (Datenbank, Solr, Elasticsearch), wodurch eine unabhängige Skalierbarkeit der Schichten erreicht wird. Bei steigenden Anforderungen ist so eine Skalierung auch auf das Zehnfache der aktuellen Besucherzahl problemlos möglich. Bei einem redundanten Betrieb mit mehreren Serverinstanzen kann zwischen den Schichten bei Bedarf eine Lastverteilung über jeweils vorgeschaltete Loadbalancer ermöglicht werden.

Der Betrieb erfolgt in einer zweistufigen Systemarchitektur: neben dem Produktivsystem wird eine zweite, unabhängige Instanz des Systems für Test-, Integrations- und Schulungszwecke bereitgestellt.

Beide Systeme werden in einem auf Sicherheit und Verfügbarkeit optimierten, ISO 27001 zertifizierten Rechenzentrum auf vom Rechenzentrumsbetreiber (Hetzner Online GmbH) bereitgestellter Hardware betrieben. Die Schichten werden in Form von virtualisierten Servern realisiert.

Das GovData-System wird dabei auf virtuellen Servern aufgesetzt; als Betriebssystem kommt Ubuntu Server 22.04 LTS zum Einsatz.

Darüber hinaus werden beim Auftragnehmer Systeminstanzen (Single-Server-Instanzen, nicht redundant ausgelegt) für Entwicklung und Test von Systemanpassungen und Erweiterungen betrieben. Diese Systeme sind fest in die interne Infrastruktur des Auftragnehmers integriert und können automatisiert erstellt, aktualisiert und auch wieder gelöscht werden. Die Entwicklungsserver-Instanzen sind direkt an die CI/CD-Infrastruktur des

Auftragnehmers angebunden und dienen zur Ausführung automatisierter und manueller Tests während der gesamten Entwicklung des Systems.

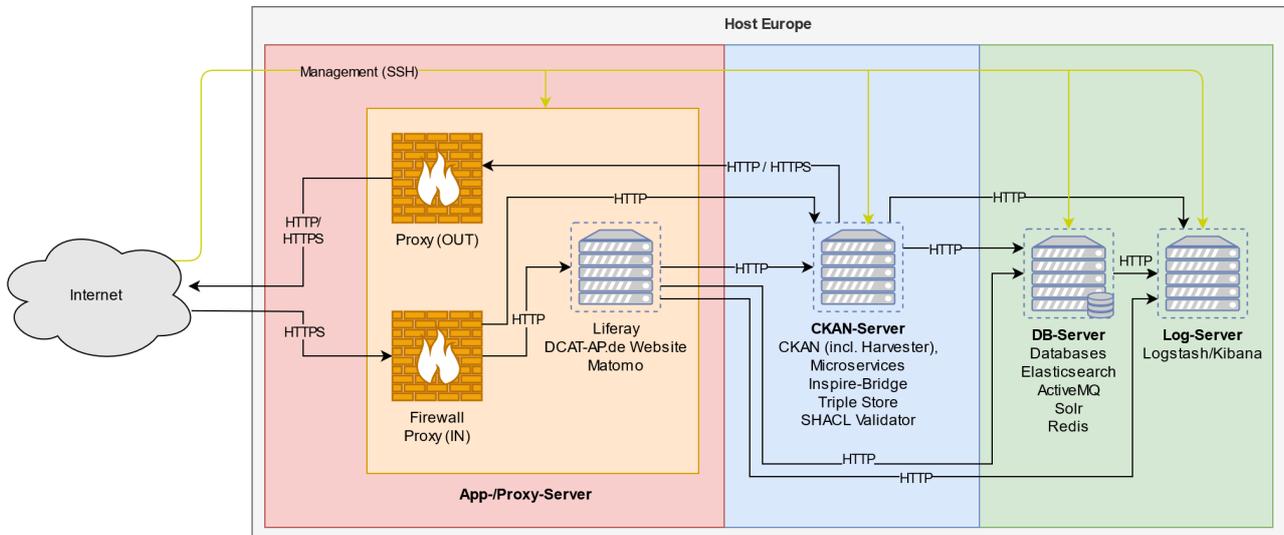


Abbildung 61: Gesamtsystemarchitektur

7.8.4.4 Einrichtung, Pflege und Betrieb des Portals

7.8.4.4.1 Einrichtung

Das Einrichten der Server auf der Produktions- und Testumgebung erfolgt über das Konfigurationsmanagement sowie weitere manuelle Schritte durch die Betriebsverantwortlichen des AN. Zum Umfang der Installationstätigkeiten zählen u.a.:

- Installation und grundlegende Konfiguration des Betriebssystems
- Installation und Konfiguration der Systemsoftware (Web- und Applikationsserver, Solr, Elasticsearch, Datenbank usw.)
- Einrichten der Replikation der PostgreSQL-Datenbank, Solr, Elasticsearch (nur bei einer redundanten Auslegung der Serverinstanzen)
- Einrichten der Firewalls
- Einrichten der Backup-Mechanismen
- Installation und Konfiguration der Portalsoftware

7.8.4.4.2 Pflege des Portals

Die Pflege des Portals umfasst unter anderem folgende Tätigkeiten:

- Regelmäßige Installation von OS-Updates
- In Ausnahmefällen (schwere Bugs) auch außerplanmäßige Installation von Sicherheitspatches

7.8.4.4.3 Betrieb

Der Betrieb umfasst u.a. die im Folgenden gelisteten Komponenten:

Monitoring

Alle physikalischen wie auch virtuellen Maschinen sind in das Monitoring des AG auf Basis der Software Icinga

eingebunden. Im Fehlerfall werden die Mitarbeiter automatisch im SEITENBAU Ticket-System, per Mail und/oder SMS benachrichtigt.

Backup

Erstellt werden Imagebackups der VM's zum Disaster Recovery (max 2 Tage zurück) sowie "normale" Daten-Backups mittels des „Bacula“-Forks „Bareos“.

Zentrale Logfileauswertung

Für Govdata ist ein zentrales Logfile-Management auf Basis des bewährten Elasticsearch/Logstash/Kibana Stacks eingerichtet.

7.8.4.5 Konfigurationsmanagement

Das Konfigurationsmanagement von GovData wird auf Basis folgender Open Source Tools aufgesetzt:

- Ansible (<http://www.ansible.com/home>)

7.8.4.5.1 Ansible

Ansible wird für das Konfigurationsmanagement der virtuellen Maschinen und für alle Prozesse, bei denen eine chronologisch korrekte Abfolge von Arbeitsschritten nötig ist, eingesetzt. Dies sind z.B. das Einspielen von Updates über alle Server, manuelles Starten/Stoppen/Restarten von Diensten usw. Weiterhin wird Ansible bei der Pflege des Portals für die Installation von OS-Updates eingesetzt.

7.8.4.6 Ausfallszenarien

7.8.4.6.1 VM defekt

Mögliche Ursache: Crash des Filesystems

Behebung:

- Variante 1: Zurückspielen des letzten Imagebackups. Falls nötig Synchronisieren von Datenbeständen und Einbinden in die Umgebung
- Variante 2: Automatisierte Neuinstallation der VM. Falls nötig Synchronisieren von Datenbeständen und Einbinden in die Umgebung

7.8.4.6.2 ESX-Server defekt I

Mögliche Ursache: Hardwareausfall an nicht redundanten Komponenten (z.B. Motherboard)

Behebung:

Die betroffenen VM's werden auf einen anderen ES-Server verschoben und wieder aktiviert. Wenn nötig werden die Daten synchronisiert und die VM wieder in den Regelbetrieb eingebunden.

7.8.4.6.3 ESX-Server defekt II

Mögliche Ursache: Fehler im RAID-Verbund - dadurch Datenverlust

Behebung:

- Variante 1: Zurückspielen des letzten Imagebackups. Falls nötig Synchronisieren von Datenbeständen und Einbinden in die Umgebung.

- Variante 2: Automatisierte Neuinstallation der VM. Falls nötig Synchronisieren von Datenbeständen und Einbinden in die Umgebung.

7.8.5 Schnittstellenübersicht

Das GovData-Portal nutzt eine Reihe von Schnittstellen von externen Diensten und bietet auch Schnittstellen für Dritte bereit. Die Schnittstellen umfassen die Benutzerschnittstelle des Portals, Schnittstellen für die Abfrage von Informationen durch Drittsysteme, Schnittstellen zur Datenbereitstellung und Visualisierung sowie Schnittstellen zu Social-Media-Angeboten. Abbildung 11 zeigt eine Übersicht aller Schnittstellen.

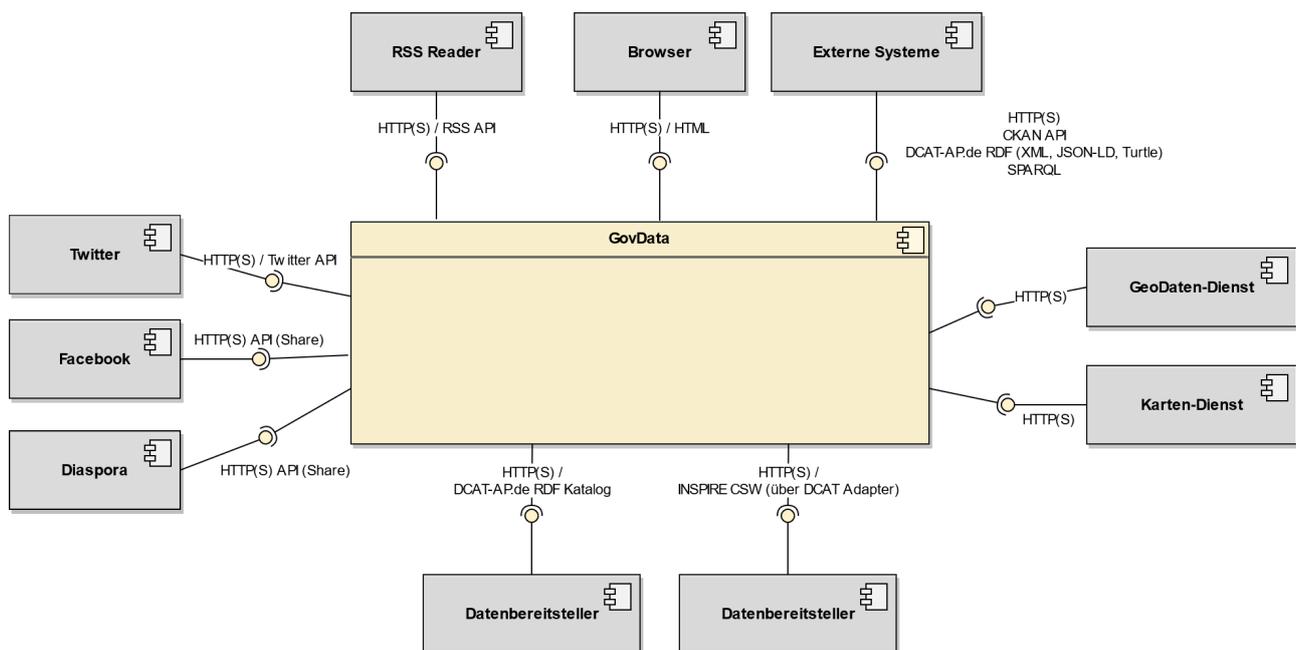


Abbildung 62: Übersicht der Schnittstellen des GovData-Portals

7.8.5.1 Angebotene Schnittstellen

Das Portal GovData bietet folgende Schnittstellen an:

- Die Browser-Schnittstelle in Form einer HTML-Oberfläche von GovData erlaubt es dem Nutzer, die Funktionen des GovData-Portals aufzurufen. Hier wird grundsätzlich zwischen zwei Benutzergruppen unterschieden: angemeldete (registrierte) und nicht-angemeldete Nutzer. Bis auf wenige Ausnahmen ist der Zugriff von nicht angemeldeten Nutzern im Wesentlichen lesender Natur. Angemeldete Nutzer werden wiederum in weitere Rollen unterteilt, die den Zugriff auf verschiedene weiterführende Funktionen steuern.
- Über eine RSS-Schnittstelle hat der Nutzer die Möglichkeit, verschiedene bereitgestellte Informationen in einen RSS-Reader einzubinden. Diese Schnittstelle steht allen Nutzergruppen gleichermaßen zur Verfügung.
- Über die CKAN API können einerseits Datenbereitsteller im Push-Verfahren Daten an den GovData Metadatenkatalog übertragen, andererseits können Nutzer bzw. Drittsysteme über die CKAN API Informationen aus dem Metadatenkatalog abrufen. Der schreibende Zugriff auf die CKAN-API ist dabei

über einen, für jeden Datenbereitsteller individuellen, API-Key bzw. API-Token abgesichert und steht somit nur autorisierten Nutzern zur Verfügung. Der lesende Zugriff auf die CKAN-API von GovData kann hingegen auch von nicht registrierten Nutzern erfolgen.

7.8.5.2 Benutzte Schnittstellen

GovData verwendet eine Reihe von Schnittstellen, um z.B. Metadaten von Datenbereitstellern im Pull-Verfahren zu sammeln und in den eigenen Metadatenkatalog aufzunehmen oder Schnittstellen zur Aufbereitung der Suche und Darstellung der Metadaten, z.B. über Geo-Daten.

7.8.5.2.1 Harvesting-Schnittstellen

Über die unterstützten Harvesting-Schnittstellen des GovData-Portals werden Metadaten von externen Datenbereitstellern gesammelt, bei Bedarf transformiert und in den eigenen Metadaten-Bestand des Portals übernommen. Hierfür wird ein allgemeiner DCAT-AP.de-Harvester verwendet, mit dem alle entsprechend dem DCAT-AP.de-Metadatenchema bereitgestellten RDF-Endpunkte (RDF-XML, JSON-LD, Turtle) abgerufen und in den Datenbestand von GovData übernommen werden können.

7.8.5.2.2 Geodaten-Dienste

Das GovData-Portal bindet eine Reihe von Geodaten-Diensten an, um Funktionen für geografische Suchoptionen bereitzustellen und um geographische Daten der Metadatenätze zu visualisieren.

- Aufbereitung von Geo-Daten:
 - o Zur Aufbereitung von Geo-Daten kommen u.a. Verfahren zur Ermittlung von Koordinaten zu Ortsnamen zum Einsatz. Hierfür kann beispielsweise der Geokodierungsdienst des Bundesamtes für Kartographie und Geodäsie (BKG) zum Einsatz kommen.
- Kartendienst:

Die Anbindung an einen Geo-Kartendienst (z.B. Openstreetmap) dient vor allem der visuellen Aufbereitung von geografischen Informationen im GovData-Portal. Bei der geografischen Suche wählt der Nutzer über die online verfügbare Kartenansicht des Portals einen gewünschten Kartenausschnitt aus, um die Suchergebnisse auf diesen geografischen Ausschnitt einzugrenzen.

7.8.5.2.3 Schnittstellen zu Social-Media-Angeboten

Alle Schnittstellen zu den Social-Media-Angeboten sind unter Berücksichtigung des aktuellen Datenschutzes umgesetzt, womit bei dem Aufruf einer Webseite keine Daten an die fremden Server übertragen werden. Eine Einbindung kann beispielsweise mit der „Shariff“-Bibliothek erfolgen, die vom Heise-Verlag veröffentlicht wird (<https://github.com/heiseonline/shariff>).

Folgende Dienste und Funktionen werden integriert:

- Facebook
 - o Die API des Social-Media-Angebots Facebook wird eingebunden, um dem Nutzer die „Teilen“- („Share“) Funktionen des Dienstes für Seiten des GovData-Portals bereitzustellen.
- Twitter
 - o Die REST-API des Social-Media-Angebots von Twitter wird zum Einen genutzt, um es den Nutzer im „2-Klick-Verfahren“ zu ermöglichen, von GovData aus Inhalte (Tweets) an den Dienst

zu übertragen, andererseits um von der Portalredaktion erstellte Twitter-Meldungen auf der Startseite des GovData-Portals darzustellen.

- Diaspora

Die API des Social-Media-Angebots von Disspora wird eingebunden, um dem Nutzer die „Teilen“-Funktion des Dienstes im „2-Klick-Verfahren“ für Seiten des GovData-Portals bereitzustellen.

7.8.6 Sicherheitsmaßnahmen

7.8.6.1 Organisation

- Einsatz einer Web-Application-Firewall

- o Einsatz einer Web-Application-Firewall (WAF) für virtuelles Patchen von Schwachstellen und Ausfiltern von Standard-Angriffen
- o Ab hohem Schutzbedarf detaillierte Anpassung der WAF an die betriebenen Anwendungen und Services

- Hinweise zu WAF siehe beispielsweise http://www.owasp.org/images/1/1b/Best_Practices_Guide_WAF.pdf

- Einbindung von ausschließlich vertrauenswürdigen Anwendungen in Portal

- o In das Portal dürfen ausschließlich vertrauenswürdige Anwendungen/Portlets eingebunden werden. Bei Einbindung von Anwendungen, die von anderen Institutionen betrieben werden, sollte für diese ein Sicherheitskonzept gefordert werden.
- o Die Einbindung von „Internet-Anwendungen“ ist nicht zulässig. Hiermit sind öffentliche Anwendungen bei Dritt-Anbietern gemeint, über die die Anwendungen des Innenministeriums und der Kundensysteme angegriffen werden könnten.

- Bezüglich der Authentifizierung müssen mindestens die folgenden Eigenschaften festgelegt und die Umsetzung geprüft werden:

- o Mindestlänge von Kennworten
- o Anforderungen an und Prüfung von Kennwortkomplexität
- o Anforderung an Kennwort-Historie
- o Sichere Speicherung von Kennworten
- o Sicherer Mechanismus zum Zurücksetzen von Kennworten

- Die folgenden Eigenschaften sollten geprüft und ggf. gefordert werden:

- o Fehlermeldungen bei der Fehleingabe von Passwörtern dürfen keinen Rückschluss auf verwendete Zugangsdaten erlauben. Es sollte der allgemeine Hinweis erfolgen, dass die Authentisierung fehlgeschlagen ist, aber nicht explizit ob der Benutzername oder das Kennwort falsch waren.

7.8.6.2 Kommunikation

- Aufteilung des Netzwerks in Zonen

- o Für die Applikationen und Datenbanken muss jeweils eine eigene Netzwerkzone vorgesehen werden.

- In den Zonen dürfen nur Systeme gemeinsam betrieben werden, die in einem engen funktionalen Zusammenhang stehen.
- Die Zonenübergänge müssen durch Technologien wie Sicherheitsgateways realisiert und die Kommunikationsverbindungen restriktiv gehandhabt werden.
- Anforderungen an Kommunikation GovData
 - Die Kommunikation vertraulicher Inhalte wie Anmeldedaten darf ausschließlich verschlüsselt erfolgen.
 - HTTPS muss unterstützt werden, aktuelle Algorithmen und Schlüssellängen sind zu nutzen. HTTPS kann auf Loadbalancern terminiert werden.
 - Die Kommunikation kann innerhalb der Netze von GovData unverschlüsselt erfolgen. Die Netze müssen durch geeignete Maßnahmen wie eine Segmentierung und Zugangskontrolle der physischen Verkabelung und Netzkomponenten geschützt werden.
- Kommunikation mit sozialen Netzwerken
 - Es darf ohne Zustimmung des Benutzers keine Kommunikation mit den Servern der externen Dienste erfolgen, mit der ein Benutzerprofil erstellt werden könnte.
 - Folgende Dienste werden angesprochen:
 - Weiterempfehlungsfunktion für die Social-Media-Angebote von Facebook, Twitter oder Diaspora
 - Twitter-API für die Anzeige der letzten Twittermeldung

7.8.6.3 Hardware und Software

- Sichere Liferay Grundkonfiguration
 - Für Liferay ist eine sichere Grundkonfiguration zu wählen, hierbei müssen mindestens die folgenden Punkte beachtet werden:
 - Erstellung eines eigenen Admin-Accounts sowie gesicherte Hinterlegung des Kennworts
 - Nutzung der Password-Hashing Funktion
 - Deaktivierung/Löschen von Standard-Usern
 - Sofern Verschlüsselung eingesetzt wird Nutzung von AES 128 oder aktuellere kryptographische Verfahren
 - Deaktivierung nicht erforderlicher Dienste und Schnittstellen
 - Deaktivierung nicht benutzter Remote APIs (WebDav, Sharepoint)
 - Bei der Grundkonfiguration können bei Bedarf verschiedene Klassen von Liferay-Instanzen berücksichtigt werden, beispielsweise falls Liferay extern und intern eingesetzt wird.
- Übernahme der Daten aus Produktion in die Testumgebung
 - Bei der Erzeugung von Test-Daten ist sicherzustellen, dass keine vertraulichen Daten aus der Produktivumgebung verwendet werden. Bei Tests mit Daten, die aus Produktivdaten abgeleitet sind, ist sicherzustellen, dass keine vertraulichen Informationen oder personenbezogenen Daten offengelegt werden.
- Nutzung einer sicheren Implementierung von Schutzfunktionen auf Code-Ebene

- Für Schutzfunktionen auf der Code-Ebene von Anwendungen sind auf Sicherheit geprüfte Implementierungen zu verwenden.
- Durch die Nutzung einer sicheren Implementierung von Schutzfunktionen kann die Angriffsfläche stark reduziert werden. Ein Beispiel für eine solche API ist OWASP Enterprise Security API (ESAPI, <http://www.owasp.org/index.php/ESAPI>)
- Hinweis: Diese Maßnahme dient dazu Sicherheitsfunktionen wie beispielsweise ein Output Encoding nicht selbst zu entwickeln, sondern Standard-Lösungen zu nutzen. Im konkreten Fall wird im Soll-Ist Vergleich geprüft ob Liferay derartige Funktionen integriert hat.
- Sichere Verwendung von Stored Procedures
 - Zur Vermeidung von SQL-Injection sind Stored Procedures mit Prepared Statements zu nutzen.
- Benutzer, Rollen und Berechtigungen in Anwendungen
 - Als Grundlage für die Autorisierungsprüfung müssen Benutzer, Rollen und Berechtigungen für jede Portalanwendung in einem Konzept festgelegt und auf dessen Basis umgesetzt werden. Dabei sind folgende Aspekte zu beachten:
 - Rollen- und Berechtigungskonzept
 - Minimale Nutzung von globalen Gruppen
- Architektur Betrieb auf mehreren Server-Instanzen
 - Die Portalanwendung muss so implementiert sein, dass sie zur Erhöhung der Verfügbarkeit gleichzeitig auf mehreren Server-Instanzen betrieben werden kann.

7.8.7 Lieferumfang

Der Lieferumfang umfasst die folgenden Liefergegenstände:

Artefakt	Beschreibung
Projekthandbuch	Das Projekthandbuch beinhaltet eine Kurzbeschreibung des Projekts, die Beschreibung des Tailoring-Ergebnisses, den grundlegenden Projektdurchführungsplan, die notwendige und vereinbarte Unterstützung des Auftraggebers sowie Organisation und Vorgaben für die Planung und Durchführung des Projekts und die anstehenden Entwicklungsaufgaben. Der Projektleiter muss dieses zentrale Produkt in Abstimmung mit den Schlüsselpersonen des Projekts erarbeiten.
Projektplan AN	Der Projektplan AN ist ein Auszug aus dem Projektplan des AN in einer für Statusberichte und Projekt- Jour-Fixe angemessenen Detailebene.
Gesamtsystemspezifikation (Pflichtenheft)	Die Gesamtsystemspezifikation (Pflichtenheft) ist das Pendant zu dem Auftraggeberprodukt Anforderungen (Lastenheft) auf Auftragnehmerseite. Sie wird vom Auftragnehmer in Zusammenarbeit mit dem Auftraggeber erstellt und stellt das zentrale Ausgangsdokument der Systemerstellung dar.
QS-Handbuch	Das QS-Handbuch beinhaltet eine Kurzbeschreibung der Qualitätsziele im Projekt, die Festlegung der zu prüfenden Produkte und Prozesse, die

	Organisation und Vorgaben für die Planung und Durchführung der Qualitätssicherung im Projekt sowie die Vorgaben für die Qualitätssicherung von externen Zulieferungen. Der QS-Verantwortliche muss dieses zentrale Produkt in Abstimmung mit den Schlüsselpersonen des Projekts erarbeiten.
Migrationskonzept	Das Migrationskonzept beschreibt, wie die Informationen aus dem Prototyp des Portals ohne Datenverlust in den Regelbetrieb übernommen werden. Ferner wird festgehalten, wie die bestehenden Schnittstellen zu Drittanbietern in das neue Portal integriert werden.
Software-Pakete	Hierbei handelt es sich um die zum Betrieb benötigten und vom AN entwickelten Software-Teile (Liferay-Portlets, CKAN-Erweiterungen, usw.). Ebenfalls enthalten sind in den Software-Paketen ggf. zusätzliche Skripte z.B. zur Ausführung von Datenmigrationen.
Release-Notes und Update-Dokumentationen	Zu jeder Softwarelieferung werden Release-Notes erstellt, die die an der Software erfolgten Änderungen dokumentieren. In der Update-Dokumentation sind alle Installations- und Konfigurationsschritte beschrieben, um die Software-Version erfolgreich in der Betriebsumgebung zu installieren.
Quellcode der entwickelten Software	Der Quellcode der vom AN entwickelten Programmteile wird über ein öffentliches Git-Repository veröffentlicht.
Prüfprotokoll Systemelement	Das Prüfprotokoll enthält die vom Prüfer verfassten Aufzeichnungen über den Verlauf der Prüfung, die Gegenüberstellung von Ist- und Soll-Ergebnissen, sowie die Analyse der identifizierten Ist-/Soll-Abweichungen und entsprechende Lösungsvorschläge.
Betriebshandbuch (BHB)	Dieses Dokument umfasst alle relevanten Informationen zur für den Betrieb des IT-Systems, also zur Einrichtung der Betriebsumgebung, Installation und Konfiguration der Portal-Software sowie für Betrieb des GovData-Portals.

Der Rohentwurf für den Prozess der Softwarelieferungen des IT-Systems ist wie folgt:

- Nach Fertigstellen der Softwareteile einer Iteration wird der Codestand im Versionskontrollsystem getagt, die Software-Pakete gebaut und auf einem dedizierten Testsystem des AN installiert.
- Auf diesem Testsystem prüft der QS-Verantwortliche des AN die Software gemäß der Spezifikation auf fachliche und technische Fehler und erstellt hieraus ein Prüfprotokoll.
- Nach Freigabe der Software durch den QS-Verantwortlichen erfolgt eine Übergabe der Software-Pakete an den Betrieb. Dazu werden relevante Informationen zur Paketinstallation, notwendige Anpassungen in der Betriebsumgebung oder an Konfigurationen des Portals und gegebenenfalls Anleitungen zu Datenmigrationen in einer gesonderten Update-Anleitung zusammengetragen und zusammen mit den Paketen an den Betrieb übergeben.

- Die Betriebsverantwortlichen installieren die Software in der Testumgebung und führen Smoke-Tests durch, um die grundsätzliche Funktionsfähigkeit des Systems sicherzustellen.
- Nach Freigabe durch den Betrieb erfolgen auf dem Test- und Schulungssystem weitere fachliche Tests durch den AN und AG.
- Nach Abnahme durch den AG kann die Installation auf dem Produktivsystem durch den Betriebsverantwortlichen erfolgen.
- Anschließend erfolgt eine weitere grobe Prüfung durch den AN, optional auch durch den AG, ob das Update erfolgreich war und das Portal uneingeschränkt zur Verfügung steht.

7.8.8 Abnahmekriterien

7.8.8.1 Prüfobjekt

Das Prüfobjekt dieser Spezifikation ist das endgültige System „GovData Portal“. Mit dem System zu liefernde Beschreibungen, Handbücher und sonstige Dokumente sind nicht Bestandteil dieser hier spezifizierten Abnahmekriterien, sondern werden vertraglich festgelegt.

Es ist jeder Systemlieferung eine eindeutig identifizierbare Version des Prüfobjektes und der Prüfspezifikation zuzuordnen. Ergeben sich im Laufe der Entwicklung aufgrund von Change Requests Änderungen an den Anforderungen, so ist die Definition der Abnahmekriterien im Rahmen des CR-Managements vom Auftraggeber entsprechend zu ändern, diese in der Dokumenthistorie zu vermerken und führt zu einer neuen Version der Gesamtspezifikation. Hierbei werden alle fachlichen Änderungen in der Gesamtspezifikation nicht nur in den Abnahmekriterien dokumentiert, sondern auch in den Beschreibungen der eigentlichen Anforderungen.

7.8.8.2 Klassifizierung der Abnahmekriterien

Die Anforderungen werden mittels der aufgeführten Abnahmekriterien überprüft. Falls Fehler auftreten, muss deren Schweregrad anhand von Fehlerklassen bemessen werden.

- Für jede umgesetzte Anforderung, die in der Lieferung enthalten ist, gibt es mindestens einen i.d.R. aber mehrere Systemtests. Diese Systemtests wurden ausgeführt.
- Die geplanten Ressourcen für die Tests wurden überschritten.
- Der Meilenstein für das Testende wurde erreicht.

Wird ein Kriterium erreicht, ist der Systemtest abgeschlossen. Ob der Systemtest erfolgreich abgeschlossen wird und eine Lieferung erfolgen kann, ergibt sich aus den folgenden Kriterien:

- Alle Systemtests wurden durchgeführt.
- Es gibt keine schweren bekannten Fehler (Kategorie 1 oder 2).

Sind beide Kriterien erfüllt, gilt der Systemtest als erfolgreich abgeschlossen und eine Lieferung kann ohne Vorbehalte erfolgen. Ist eines der beiden Kriterien nicht erfüllt, gilt der Systemtest als nicht erfolgreich abgeschlossen. Es kann dann nur eine Lieferung mit Vorbehalten erfolgen.

Der AG wird über den Abbruch von Testfällen beim Eintreten eines Testendkriteriums informiert.

In jedem Falle ist sicher zu stellen, dass zumindest die Grundfunktionalitäten der Anwendung durch erfolgreiche

Tests abgedeckt sind. Die Grundfunktionalitäten sind Funktionen, die den hauptsächlichen Betrieb behindern könnten, wie z.B. nach Metadaten suchen, Metadaten anzeigen, Harvesting und ähnliches. Nicht zu den Grundfunktionalitäten zählen zum Beispiel "Bewertungsdetails werden beim jeweiligen Metadatensatz angezeigt" oder "Suchende werden bei der Suche aktiv unterstützt".

Zudem hat der AG die Verpflichtung nach jeder Iteration den ausgelieferten Softwarestand zu testen.

7.8.8.2.1 Fehlerklassen

Die Fehlerklassen sind wie folgt definiert:

- **Fehlerklasse 1 = Kritische Fehler:**

Es liegt ein Zustand vor, der die Hauptfunktion des Produkts schwer beeinträchtigt und wegen der Auswirkungen für die Nutzer sofortige Gegenmaßnahmen erfordert, unabhängig von Tag und Uhrzeit, z.B.

- a. Betriebsunfähigkeit (völliger oder partieller Ausfall des Produkts)
- b. Reduzierte Übertragungs- und Datenverarbeitungskapazität, sodass erwartete Datenmengen nicht verarbeitet werden können.
- c. Ausfall von Notfallfunktionalitäten
- d. Sicherheitsrisiken oder Risiken von Sicherheitsverstößen

Fehler der Fehlerklasse 1 liegen auch vor, wenn Nebenfunktionen

- e. nicht ausführbar sind
- f. die zugrunde liegende Anforderung(en) nicht erfüllt wird
- g. zu Dateninkonsistenz führen

- **Fehlerklasse 2 = Schwere Fehler:**

Das Produkt bzw. eine Funktion ist nutzbar, aber es liegt ein Zustand vor, der die Funktionsfähigkeit, die Wartung oder die Administration des Produkts stark beeinträchtigt und zu Standarduhrzeiten behoben werden muss. Die Dringlichkeit ist niedriger als bei kritischen Zuständen, da die Bedienung durch den Nutzer weniger unmittelbar betroffen ist, z.B.

- a. reduzierte Leistungsfähigkeit (jedoch Verarbeitung der erwarteten Datenmenge)
- b. Ausfall der Wartbarkeit oder Administrierbarkeit des Produkts oder der Diagnosefähigkeit
- c. Einschränkung einer essentiellen Komponente oder Funktion
- d. Eingeschränkte Fähigkeit des Produkts, auf eine Fehlfunktion hinzuweisen

Funktionalitäten der Priorität B und C (s. Ziffer 3.2), die Fehler der Klasse 2 enthalten, sollten bis zu ihrer Behebung deaktiviert werden können. Liegen innerhalb einer Funktionalität mehrere Fehler der Klasse 2 vor, so dass die Funktion in ihrer Gesamtheit zu stark eingeschränkt wird, kann dies als Fehler der Klasse 1 gewertet werden.

- **Fehlerklasse 3 = Leichte Fehler:**

Andere Probleme, welche nicht als „kritisch“ oder „schwer“ eingestuft sind, z.B. Zustände, die die Funktion des Systems wenig oder gar nicht einschränken. Liegen innerhalb einer Funktionalität mehrere Fehler der Klasse 3 vor, so dass die Funktion in ihrer Gesamtheit zu stark eingeschränkt wird, kann dies als Fehler der Klasse 2 gewertet werden. Funktionalitäten der Priorität B und C, die Fehler der Klasse 3 enthalten, sollten bis zu ihrer Behebung deaktiviert werden können.

7.8.9 Abkürzungsverzeichnis

Abkürzung	Erklärung
AG	Auftraggeber
AN	Auftragnehmer
CKAN	Comprehensive Knowledge Archive Network, webbasiertes System zum Speichern und Verteilen von Daten
QS	Qualitätssicherung

7.9 Vergleichsmatrix

7.9.1 Matrix

HINWEIS: Bei dieser Version des Berichtes handelt es sich um eine gekürzte Version. Der vollständige Bericht enthält Inhalte, die aus wettbewerbsrechtlichen Gründen nicht für eine uneingeschränkte Veröffentlichung geeignet sind. Der folgende Abschnitt konnte aus diesem Grund leider nicht veröffentlicht werden.

7.9.2 Erläuterungen

Anmerkung: Priorität von 0 bis 10, wobei 10 die höchste Priorität.

	Kriterium	Beschreibung	Priorität	Backend-Priorität	KO-Kriterium
1	Hersteller	Von welchem Hersteller wird das Produkt vertrieben? Angaben möglichst mit Website und Kontakt	0	0	
2	Versionsnummer	Welche Versionsnummer des Produktes liegt dieser Untersuchung zugrunde?	0	0	
3	Open Source	Ist das Produkt unter einer Open Source Lizenz erhältlich? Ist das Produkt ausschließlich als Open Source Produkt erhältlich oder wird es auch kommerziell vertrieben? • offene Lizenz • kommerzielle Lizenz	3	3	
4	Lizenzen	Unter welchen Lizenzen kann das Produkt eingesetzt werden? (Auflistung der genauen Lizenzen)	1	1	
5	Aktuelle Weiterentwicklung	Wird die Software aktuell noch weiterentwickelt? Gibt es einen Releasefahrplan? • aktiv: aktuelle Entwicklung neuer Features und Releases • genutzt: Es gibt nur noch Sicherheitspatches • keine Updates: Der Letzte Release liegt lange zurück	10	10	Es finden keine Updates mehr statt.
6	Erfahrungswerte	Hat jemand uns bekanntes Erfahrungen mit der Software? • Jemand uns bekanntes setzt die Software bereits ein (Kontakte nennen) • Jemand uns bekanntes hat die Software ausprobiert (Kontakte nennen) • keine	4	4	
7	Gibt es eine Community	Sind die folgenden Punkte vorhanden? • Bugtracker: Es gibt einen aktiv genutzten Bugtracker • Forum o.Ä.: Es gibt eine Plattform für den Austausch • Releaselogs: Jeder Release ist auch protokolliert • Nutzerbasis: Gibt es 'gefühl' viele Nutzer? (groß/klein)	7	7	
8	Voraussetzungen	Welche speziellen Hard-/Software Requirements/Anforderungen liegen vor?	4	4	
9	Wiederverwendbarkeit aktueller Anwendungen	Welche Teile der aktuellen Lösung können in Kombination mit der Software wiederverwendet/integriert werden?	5	5	
10	Verwendete Datenbanken	Welche Datenbanken verwendet die Software?	3	3	
11	Erweiterbarkeit	Kann die Software durch Erweiterungen oder neue Services einfach um neue Funktionen erweitert werden? Gibt es bereits Erweiterungen vom Entwickler/der Community dafür?	4	4	
12	Skalierbarkeit	Wie skaliert die Software mit wachsender Datenmengen und wachsender Nutzerzahl? Welche Probleme können dabei auftreten?	6	6	
13	Bekannte Systemgrenzen /Schwellenwerte	Gibt es bekannte Grenzen der Software die uns beeinflussen könnten? Hier können auch die	5	5	

		ermittelten Schwellenwerte einfließen. • höher als Schwellenwert GovData • niedriger als Schwellenwert GovData			
14	Programmiersprache/n	Welche Programmiersprachen können im Zusammenspiel mit der Software verwendet werden?	3	3	
15	Dokumentation	Wie gut/ausführlich ist die vorhandene Dokumentation der Software? Findet sich ein ungeübter Nutzer schnell ein? • sehr gut: Ausführliche Doku, Tutorials, etc. • gut: Teilweise vorhandene Doku, Tutorials, etc. • schlecht: Nur vereinzelt Informationen über Google	9	9	Keine Dokumentation vorhanden.
16	Konfigurationsmöglichkeiten	Gibt es Anpassungsmöglichkeiten der Software oder Standardlösung? Welche Möglichkeiten für ein besseres Handling gibt es? • viele nützliche anpassbare Optionen • wenige nützliche anpassbare Optionen • nicht konfigurierbar	5	5	
17	Umgebung aufsetzen	Wie viel Aufwand ist es eine Umgebung mit der Software aufzusetzen, so dass zumindest eine minimale Anwendung läuft?	3	3	
18	Nutzerverwaltung	Ist eine Nutzerverwaltung in die Software integriert? Kann das bestehende GovData Nutzerset damit abgebildet werden? • Vorhanden • Durch Anbindung anderer Software möglich • Nicht Möglich	4	0	
19	Sub-Accounts	Können Sub-Accounts von Haupt-Accounts selbst angelegt und verwaltet werden? • Vorhanden • Durch Anbindung anderer Software möglich • Nicht Möglich	2	0	
20	Mehrsprachigkeit	Ist Mehrsprachigkeit der Inhalte in die Software integriert? • Vorhanden • Durch Anbindung anderer Software möglich • Nicht Möglich	2	0	
21	SEO	Welche SEO Optionen sind möglich? • Deep Link URLs • Keywordangaben • Titelangaben • Beschreibung angeben • Programmatische SEO	2	0	
22	Blogfunktion	Ist ein Blog in die Software integriert? • Vorhanden • Durch Anbindung anderer Software möglich • Nicht Möglich	2	0	
23	Forum	Ist ein Forum in die Software integriert? • Vorhanden • Durch Anbindung anderer Software möglich • Nicht Möglich	2	0	
24	Feed	Ist es möglich einen Feed für Änderungen zu abonnieren? • Vorhanden • Durch Anbindung anderer Software möglich • Nicht Möglich	2	0	
25	Kommentarfunktion	Ist eine Kommentarfunktion für Datensätze in die Software integriert? • Vorhanden • Durch Anbindung anderer Software möglich • Nicht Möglich	1	0	
26	Bewertungssystem	Ist eine Bewertung für Datensätze in die Software integriert? Worauf basiert die Bewertung? • Vorhanden • Durch Anbindung anderer Software möglich • Nicht Möglich	1	0	
27	Redaktionelle Inhalte	Ist redaktionelles Management des Contents möglich? Wenn die Anwendung mit einem CMS kombiniert ist, welches wurde verwendet? • Vorhanden • Durch Anbindung anderer Software möglich • Nicht Möglich	3	0	

28	Bietet Dataset-API an	Ist eine API zum Abfragen der Datensätze vorhanden? Wenn ja, was für eine? <ul style="list-style-type: none"> • Vorhanden • Durch Anbindung anderer Software möglich • Nicht Möglich 	3	3	
29	Harvesting	Gibt es eine Möglichkeit Daten zu harvesten? Wenn ja, über welche Schnittstellen und in welchen Formaten? Sind Harvester flexibel hinzufügar? <ul style="list-style-type: none"> • Harvesting mit allen bisherigen Formaten und Funktionen möglich • Durch Anbindung anderer Software möglich • Harvesting nur eingeschränkt möglich • Nicht Möglich 	5	5	
30	Manuelle Datenerfassung/-verwaltung	Gibt es die Möglichkeit, Metadaten manuell zu erfassen und zu verwalten? <ul style="list-style-type: none"> • Vorhanden • Durch Anbindung anderer Software möglich • Nicht Möglich • Möglichkeit für Entwürfe vorhanden, Metadaten als privat zu kennzeichnen? 	4	4	
31	RDF Endpunkt	Gibt es Optionen um Daten als RDF darzustellen und zu veröffentlichen? <ul style="list-style-type: none"> • Für Datensätze • Für den Gesamtkatalog. Ist dieser auch Filterbar? 	4	4	
32	Verschlüsselung vertraulicher Daten	Können persönliche Daten wie z.b. Passwörter verschlüsselt gespeichert werden? Welche Verschlüsselungsalgorithmen können angewendet werden? <ul style="list-style-type: none"> • Ausreichender Schutz ist vorhanden • Durch Erweiterung möglich • Nur mangelhafter Schutz ist möglich • Nicht möglich 	6	6	
33	Schutz vor korrupten Eingaben	Gibt es Schutzmechanismen? <ul style="list-style-type: none"> • Gegen maschinelle Eingaben (Honeypot) • Escaping von HTML 	6	6	
34	Backups	Können Backups gemacht/eingespielt werden? Wie viel Aufwand ist das ? <ul style="list-style-type: none"> • Unkompliziert • Machbar • Nicht möglich 	6	6	
35	Bemerkungen	Was sollte sonst noch erwähnt werden?	0	0	
36	Score	Berechneter Wert zum vergleichen der Anwendungen. Berechnet sich aus der Summe aller Gewichtungen welche mit ihrer Priorität multipliziert werden. Je höher der Wert desto besser.	0	0	
37	Backend-Score	Berechneter Wert zum vergleichen der Backends. Berechnet wie der Score mit dem Unterschied, dass Funktionalitäten die das CMS liefert nicht mitberechnet werden.	0	0	

7.9.3 Gewichtung je Anwendung

HINWEIS: Bei dieser Version des Berichtes handelt es sich um eine gekürzte Version. Der vollständige Bericht enthält Inhalte, die aus wettbewerbsrechtlichen Gründen einzelnen Marktteilnehmern schaden können und vor diesem Hintergrund nicht für eine uneingeschränkte Veröffentlichung geeignet sind. Der folgende Abschnitt konnte aus diesem Grund leider nicht veröffentlicht werden.

7.10 User Storys – Mindestanforderungen an Funktionsumfang der neuen Systemarchitektur

User Story B1.1: Als datenbereitstellende Person eigene Metadatensätze hinzufügen	
Akteure	Datenbereitstellende
Vorbedingungen	-
User Story	Als datenbereitstellende Person möchte ich neue Metadatensätze hinzufügen.
Akzeptanzkriterien	Als datenbereitstellende Person erwarte ich: <ul style="list-style-type: none"> - eine Funktion „Metadatensatz hinzufügen“ - ein Webformular, in das ich die Pflichtfelder und weitere Metadatenfelder ausfüllen kann - eine Validierung meiner eingegebenen Daten und verständliche Problem-Informationen bei negativem Validierungsergebnis - nach dem erfolgreichen Speichern ist der neue Metadatensatz im GovData-Metadatenkatalog vorhanden

User Story B1.2: Als datenbereitstellende Person eigene Metadatensätze bearbeiten	
Akteure	Datenbereitstellende
Vorbedingungen	Datensätze der datenbereitstellenden Person sind im Metadatenkatalog vorhanden.
User Story	Als datenbereitstellende Person möchte ich meine eigenen Metadatensätze bearbeiten.
Akzeptanzkriterien	Als datenbereitstellende Person erwarte ich: <ul style="list-style-type: none"> - eine Funktion „Bearbeiten“, über die ich in das Webformular des jeweiligen Datensatzes gelange - ein Webformular, in das ich die Pflichtfelder und weitere Metadatenfelder ausfüllen kann - eine Validierung meiner eingegebenen Daten und verständliche Problem-Informationen bei negativem Validierungsergebnis - nach dem erfolgreichen Speichern ist der geänderte Metadatensatz im GovData-Metadatenkatalog vorhanden

User Story B1.3: Als Datenbereitstellende Person eigene Metadatenätze ausblenden / als Entwurf abspeichern

Akteure	Datenbereitstellende
Vorbedingungen	Datensätze der datenbereitstellenden Person sind im Metadatenkatalog vorhanden.
User Story	Als datenbereitstellende Person möchte ich meine eigenen Metadatenätze ausblenden, so dass der Metadatenatz im GovData-Portal und über die API über die Suche nicht auffindbar ist und in der Detailansicht nicht angezeigt wird.
Akzeptanzkriterien	Als datenbereitstellende Person erwarte ich: <ul style="list-style-type: none"> - eine Funktion „Bearbeiten“, über die ich in das Webformular des jeweiligen Datensatzes gelange - ein Webformular, in dem eine Option zum Ausblenden des Metadatenatzes vorhanden ist - eine Validierung meiner eingegebenen Daten und verständliche Problem-Informationen bei negativem Validierungsergebnis - nach dem erfolgreichen Speichern ist der geänderte Metadatenatz im GovData-Metadatenkatalog nicht mehr auffindbar

User Story B1.4: Als Datenbereitstellende Person eigene Metadaten anzeigen

Akteure	Datenbereitstellende
Vorbedingungen	Datensätze der datenbereitstellenden Person sind im Metadatenkatalog vorhanden.
User Story	Als datenbereitstellende Person möchte ich meine eigenen Metadaten in einer sortier- und filterbaren Liste anzeigen lassen.
Akzeptanzkriterien	Als datenbereitstellende Person erwarte ich: <ul style="list-style-type: none"> - eine Funktion, über die alle meine eigenen Metadatenätze in Listenform angezeigt werden - in der Liste werden relevante Informationen des Datensatzes angezeigt - Sortier- und Filterfunktion für das Suchergebnis - Funktionen an jedem Datensatz, über die der Datensatz angezeigt oder bearbeitet werden kann.

User Story B2.1: Als Geschäfts- und Koordinierungsstelle Showcases hinzufügen

Akteure	Geschäfts- und Koordinierungsstelle
Vorbedingungen	-
User Story	Als GKSt möchte ich neue Showcases hinzufügen.
Akzeptanzkriterien	Als GKSt erwarte ich: <ul style="list-style-type: none"> - eine Funktion „Showcase hinzufügen“ - ein Webformular, in das ich die Pflichtfelder und weitere Datenfelder ausfüllen kann - eine Validierung meiner eingegebenen Daten und verständliche Problem-Informationen bei negativem Validierungsergebnis - nach dem erfolgreichen Speichern wird der neue Showcase im GovData-Portal angezeigt

User Story B2.2: Als Geschäfts- und Koordinierungsstelle Showcases bearbeiten

Akteure	Geschäfts- und Koordinierungsstelle
Vorbedingungen	Showcases sind in der Datenbank vorhanden.
User Story	Als GSKt möchte ich Showcases bearbeiten.
Akzeptanzkriterien	Als GKSt erwarte ich: <ul style="list-style-type: none"> - eine Funktion „Bearbeiten“, über die ich in das Webformular des jeweiligen Showcase gelange - ein Webformular, in das ich die Pflichtfelder und weitere Datenfelder ausfüllen kann - eine Validierung meiner eingegebenen Daten und verständliche Problem-Informationen bei negativem Validierungsergebnis - nach dem erfolgreichen Speichern wird der neue Showcase im GovData-Portal angezeigt

User Story B2.3: Als Geschäfts- und Koordinierungsstelle Showcases ausblenden / als Entwurf abspeichern

Akteure	Geschäfts- und Koordinierungsstelle
Vorbedingungen	Showcases sind in der Datenbank vorhanden.
User Story	Als GKSt möchte ich Showcases ausblenden, so dass der Showcase im GovData-Portal über die Suche nicht auffindbar ist und in der Detailansicht nicht angezeigt wird.
Akzeptanzkriterien	Als GKST erwarte ich: <ul style="list-style-type: none"> - eine Funktion „Bearbeiten“, über die ich in das Webformular des jeweiligen Showcase gelange - ein Webformular, in dem eine Option zum Ausblenden des Showcase vorhanden ist - eine Validierung meiner eingegebenen Daten und verständliche Problem-Informationen bei negativem Validierungsergebnis - nach dem erfolgreichen Speichern ist der geänderte Showcase im GovData-Portal nicht mehr auffindbar

User Story B2.4: Als Geschäfts- und Koordinierungsstelle Showcases anzeigen

Akteure	Geschäfts- und Koordinierungsstelle
Vorbedingungen	Showcases sind in der Datenbank vorhanden.
User Story	Als GKSt möchte ich Showcases in einer sortier- und filterbaren Liste anzeigen lassen.
Akzeptanzkriterien	Als GKSt erwarte ich: <ul style="list-style-type: none"> - eine Funktion, über die alle Showcases in Listenform angezeigt werden - in der Liste werden relevante Informationen des Showcase angezeigt - Sortier- und Filterfunktion für das Suchergebnis - Funktionen an jedem Showcase, über die der Showcase angezeigt oder bearbeitet werden kann.

7.11 Lastenheft GovData – Formular

Projektbezeichnung	DataGo TP2
Projektleiter	Christian Wittig
Erstellt am	20.07.2022
Letzte Änderung am	31.10.2022
Status	Abgeschlossen
Aktuelle Version	1.5

Änderungsverlauf

Nr.	Datum	Version	Geänderte Kapitel	Art der Änderung	Autor	Status
1	20.07.2022	1.1	Alle	Erstellung	Adam Kirschstein	Abgeschlossen
2	28.07.2022	1.2	Alle	Inhaltliche und sprachliche QS	Lukas Cladders	Abgeschlossen
3	31.10.2022	1.3	Alle	Finale QS nach Absprache mit Christian Wittig	Lukas Cladders	Abgeschlossen
5	02.11.2022	1.4	Alle	Eintragung der Umsetzbarkeit durch SEITENBAU	SEITENBAU	Abgeschlossen
4	03.11.2022	1.5	Alle	Finale Abnahme durch Christian Wittig	Christian Wittig	Abgeschlossen

7.11.1 Einleitung

Dieses Dokument beschreibt die funktionalen Anforderungen, die im Frontend für das GovData Metadaten-Formular umgesetzt werden sollen. Dabei werden nur Aspekte berücksichtigt, die für die Entwicklung des Frontends relevant sind. Funktionalitäten, die das Backend erbringt (z.B. die Kriterien, nach denen die Suchvorschläge für die Autovervollständigung im Backend generiert werden), sind nicht näher beschrieben.

Das Formular soll alle wichtigen Metadatenfelder unterstützen, die GovData für das Harvesting der Datensätze benötigt und den Datenbereitstellenden eine gute User Experience beim Einpflegen der Daten gewährleisten.

Erweitert wurde das Dokument um Stellungnahmen des technischen Dienstleisters SEITENBAU zur Umsetzbarkeit der Anforderungen.

7.11.2 Allgemeines

7.11.2.1 Ziel und Zweck des Dokuments

Ein Lastenheft dient u. a. als Orientierungshilfe für den IT-Dienstleister SEITENBAU als Auftragnehmer. Die Anforderungs- oder Kundenspezifikation bezeichnet ein Dokument, in welchem ein Auftraggeber seine Anforderungen gegenüber dem Auftragnehmer festhält. Es werden die genauen Anforderungen beschrieben,

die der Auftraggeber an die Erreichung des Projektzieles stellt.

Darauf basierend erstellt der Auftragnehmer das Pflichtenheft. Dieses behandelt detailliert alle Schritte, die zur Umsetzung des Projekts notwendig sind.

In diesem Fall wurden die Stellungnahmen des Dienstleisters SEITENBAU zur Umsetzbarkeit der Anforderungen direkt mit aufgenommen, um ein gesamtheitliches Bild zur Optimierbarkeit des aktuell bestehenden Formulars zur manuellen Dateneingabe auf GovData bereit zu stellen.

7.11.2.2 Ausgangssituation

Der Handlungsbedarf für die Optimierung des Formulars entstand daraus, dass bei der Analyse des Ist-Zustands für die manuelle Anbindung von Pilotinstitutionen an GovData, das Formular am häufigsten als Kritikpunkt im bestehenden Prozess genannt worden ist. Das aktuelle Formular wurde zunächst als temporärer Workaround implementiert, da eine automatisierte Anbindung an GovData bevorzugt wird. Nichtsdestotrotz wird das Formular von vielen Datenbereitstellenden seit Jahren regelmäßig verwendet und bietet vielen Organisationen einen niedrighschwelligem Einstieg in die Bereitstellung von Open Data.

7.11.2.3 Projektbezug

Die Anforderungserhebung für das GovData Formular ist ein Baustein im Teilprojekt 2 von Data Go. Data Go beschäftigt sich in diesem Teilprojekt mit der musterhaften Anbindung von Pilotinstitutionen an GovData, um im Anschluss ein Kostenmodell für den Regelbetrieb abzuleiten.

7.11.2.4 Verteiler und Freigabe

Rolle(n)	Name
Projektleiter	Christian Wittig / GovData
Consultant	Dr. Lukas Cladders /]init[
Consultant	Adam Kirschstein /]init[

7.11.3 Konzept

7.11.3.1 Ziel(e) des Anbieters

- Datenbereitstellenden ein userfreundliches und effektives Werkzeug an die Hand geben, um möglichst viele Metadaten in hoher Qualität für GovData bereitzustellen.

7.11.3.2 Ziel(e) und Nutzen des Anwenders

- Die datenbereitstellende Person will möglichst autonom einen reibungslosen und motivierenden Prozess beim Eintragen ihrer Metadaten erleben und sich vor allen Risikofaktoren absichern.

7.11.3.3 Zielgruppe(n)

- Neueinsteiger:
 - o Diese Zielgruppe beschreibt alle User, welche das Formular zum ersten Mal bedienen bzw. noch nicht mit dem generellen Umgang mit Metadaten in OpenData Portalen vertraut sind.
- Power-User:

- Diese Zielgruppe beschreibt alle Verantwortlichen in Institutionen, die regelmäßig das Formular verwenden und Wert auf Zeitersparnis/Effektivität legen.

7.11.3.4 Prozess

In Einzelinterviews und zwei UX-Workshops wurden alle Anforderungen der am DataGo-Projekt teilnehmenden Institutionen erhoben. Dieses Dokument konsolidiert die wichtigsten Erkenntnisse und stellt Handlungsfelder für die technische Weiterentwicklung dar.

7.11.4 Funktionale Anforderungen

Funktionale Anforderungen sind gewünschte Funktionalitäten oder Verhalten eines Systems bzw. Produkts. Die Daten wurden aus der Anforderungstabelle entnommen. Die Auflistung enthält lediglich die wesentlichsten Anforderungen je Handlungsfeld.

Je Anforderung wird eine „Priorität“ festgelegt, die sich an der höchsten Priorität eines innerhalb der Anforderung definierten Teilaspekts richtet. Innerhalb der „Beschreibung des Soll-Zustands“ wird gegebenenfalls eine Gewichtung von Teilaspekten vorgenommen. Die „Priorität“ wird in drei Stufen erfasst:

- „Priorität 1“: Die Anforderung ist im Hinblick auf den Komfort bei der Nutzung des Formulars eine begrüßenswerte Ergänzung, hat aber keine Auswirkung auf die generelle Nutzbarkeit.
- „Priorität 2“: Die Anforderung ist von besonderer Bedeutung für die generelle Nutzbarkeit des Formulars, verhindert jedoch nicht die Nutzung.
- „Priorität 3“: Die Anforderung hat zentrale Bedeutung für einige oder alle Nutzer:innen des Formulars und das Potential, eine Nutzung des Formulars zu verhindern.

Weiterhin wird für jede Anforderung die „Umsetzbarkeit“ definiert. Diese richtet sich nach dem zu erwartenden Aufwand für die Umsetzung und wurde durch den Dienstleister SEITENBAU vorgenommen:

- „Umsetzbarkeit 1“: Aufwand von bis zu 1 PT
- „Umsetzbarkeit 2“: Aufwand von 1-5 PT
- „Umsetzbarkeit 3“: Aufwand von über 5 PT

Diese Einstufungen sind als erste Orientierungshilfe für eine mögliche Priorisierung der Anforderungen in der Umsetzung zu verstehen. Vor der Umsetzung sollten die Anforderungen in Abhängigkeit der Systemarchitektur konkretisiert werden, was wiederum zu Abweichungen der genannten Schätzung führen kann.

7.11.4.1 Handlungsfeld: Architektur und Technik

7.11.4.1.1 Anforderung 4.1.1: Interaktive Führung

Nr. / ID	1002	Priorität	2	Umsetzbarkeit	3
-----------------	------	------------------	---	----------------------	---

Beschreibung des Soll-Zustands

Ziel des Formulars sollte es sein, die Nutzer:innen interaktiv durch den Prozess der Dateneingabe zu führen. Hierzu sollten stets nur die relevanten Teilabschnitte eingeblendet sein. Zu Beginn sollten nur die Felder angezeigt werden, die als Minimaleingabe benötigt werden (in Abstimmung mit den von GovData für den

automatischen Prozess definierten Pflichtfeldern). Bei Eingabe werden optional weitere ergänzende Datenfelder angezeigt. Designelemente sollten dazu beitragen, den User zu motivieren und Hindernisse bei der Befüllung abzubauen. Im Workshop wurde das Format eines „Interviews“ des Tools „WISO Steuer“ als Musterbeispiel für eine gute interaktive Führung durch ein statisches Formular genannt.

Vergleich mit bestehender Lösung

Aktuell fühlen sich User beim Öffnen des Formulars „erschlagen“ durch die Vielzahl an angezeigten Datenfeldern, bei denen zunächst unklar ist, welche davon befüllt werden müssen. Aufgrund fehlender Führung entstehen bei bestimmten Abschnitten Schwierigkeiten im Verständnis und bei der Befüllung der Datenfelder.

7.11.4.1.2 Anforderung 4.1.2: Binnendifferenzierung im Formular

Nr. / ID	1003 / 1021 / 1024	Priorität	2	Umsetzbarkeit	3
-----------------	--------------------	------------------	---	----------------------	---

Beschreibung Soll-Zustand

Verschiedene Bereiche sollten durch das Setzen von Abschnitten klarer semantisch voneinander abgegrenzt und evtl. auch mit Titel versehen werden. Hierzu können CSS-Formatierungen verwendet werden. Eine neue inhaltliche Gliederung wäre auf zwei Arten möglich: Entweder durch die „Informationstiefe“ (z.B. Datensatz u. Veröffentlichende Stelle / Ressourcenbeschreibung, Ansprechpersonen, Schlagwörter / Abdeckung, Raumbezug u.ä.) oder durch die Verbindlichkeit der Datenfelder nach DCAT-AP.de 2.0 durch „Muss – Soll – Kann“-Abschnitte. Farbliche Codierung könnte zur optischen Differenzierung eingesetzt werden.

Vergleich mit bestehender Lösung

Die aktuelle Struktur bündelt nur mangelhaft zusammengehörige Datenfelder. Zugleich werden die verschiedenen Gliederungspunkte in visueller Hinsicht nicht ausreichend voneinander getrennt. „Muss-Soll-Kann“-Felder werden nicht mit absteigender Wertung sortiert, lediglich „Muss“-Felder werden mittels Sternchen markiert.

7.11.4.1.3 Anforderung 4.1.3: Feedback beim Ausfüllen / Gamification

Nr. / ID	1012	Priorität	1	Umsetzbarkeit	3
-----------------	------	------------------	---	----------------------	---

Beschreibung Soll-Zustand

Die User sollten zum einen motiviert werden, ihre Datensätze mit möglichst vielen Metadaten zu versehen und zum anderen Sicherheit bei der Befüllung der Datenfelder vermittelt bekommen. Hierfür könnten z.B. „Easy“ und „Pro“-Ansichten, ein Fortschrittsbalken, bzw. eine Anzeige für die Qualität der Metadaten verwendet werden. Direktes Feedback beim Ausfüllen eines Datenfelds könnte ebenfalls hilfreich sein (z.B. grüner Haken). Auch könnte nach Ausfüllen des Formulars automatisiert eine Dankesmail versandt werden. Diese könnte auch dazu dienen, mittels einer Verteileradresse interne oder externe Kolleg:innen über das Bereitstellen eines Datensatzes zu informieren.

Vergleich mit bestehender Lösung

Das aktuelle Formular bietet wenig Anhaltspunkte dafür, inwieweit man den Anforderungen von GovData und den qualitativen Erwartungen an Open Data im Allgemeinen gerecht wird. Feedback oder ein Anreizsystem für qualitativ hochwertige Metadaten fehlen.

7.11.4.1.4 Anforderung 4.1.4: Import von CSV-Dateien

Nr. / ID	1010	Priorität	3	Umsetzbarkeit	3
-----------------	------	------------------	---	----------------------	---

Beschreibung Soll-Zustand

Durch die Möglichkeit zum Import von CSV-Dateien wird es für Institutionen mit regelmäßiger manueller Datenbereitstellung deutlich einfacher, ihre regelmäßig gleichlautenden Daten (z.B. zu „Veröffentlichende Stelle“) ohne großen Aufwand bereitzustellen. Weiterhin können Institutionen mit nachgeordneten Stellen oder in anderen Kollaborationen auf diesem Wege ihre Metadaten im Vorfeld gemeinsam bearbeiten, sammeln und anschließend über eine zentrale Stelle bereitstellen. Damit wäre der interne Organisationsaufwand für diese datenbereitstellende Institutionen deutlich verringert. Die User sollten mit einem fest definierten Schema CSV-Dateien in das Formular importieren können. Hierzu sollte innerhalb des Formulars ein Upload-Button erscheinen. Nach dem erfolgreichen Upload werden Metadatensätze als „Entwurf“ abgespeichert.

Weitere Informationen zur Umsetzbarkeit finden sich in Kapitel 7.11.6.

Vergleich mit bestehender Lösung

Bisher müssen datenbereitstellende Institutionen die Nutzung des von GovData zentral bereitgestellten Accounts mit ihren nachgeordneten Stellen koordinieren. Wie dies in Praxis umgesetzt wird (geteilte Zugangsdaten, eine Stelle zur Dateneintragung) steht ihnen frei, in jedem Fall bedeutet dies erhöhten Organisationsaufwand und damit eine Hürde für die Bereitstellung von Daten auf GovData. Es ist möglich, das Formular zwischenzuspeichern.

7.11.4.1.5 Anforderung 4.1.5: Einbindung mehrerer Accounts

Nr. / ID	1007	Priorität	3	Umsetzbarkeit	3
-----------------	------	------------------	---	----------------------	---

Beschreibung Soll-Zustand

Institutionen mit nachgeordneten Stellen oder in anderen Kollaborationen haben die Möglichkeit, im Rahmen eines zentralen institutionellen Nutzerkontos mehrere Sub-Accounts/Nutzer hinzuzufügen, um mehreren Dienststellen das eigenständige Arbeiten zu ermöglichen. Damit wäre der interne Organisationsaufwand für datenbereitstellende Institutionen deutlich verringert. Auch ließen sich mit den Accounts Vorlagen verbinden, in denen Teile des Formulars bereits für die Institution vorausgefüllt sind (siehe 7.11.4.4.2).

Weitere Informationen zur Umsetzbarkeit finden sich in Kapitel 7.11.6.

Vergleich mit bestehender Lösung

Siehe Kapitel 7.11.4.1.4.

7.11.4.1.6 Anforderung 4.1.6: Formatierungsoptionen im Feld „Beschreibung“

Nr. / ID	1020	Priorität	3	Umsetzbarkeit	3
-----------------	------	------------------	---	----------------------	---

Beschreibung Soll-Zustand

In Textfeldern wie z.B. „Beschreibung“ wären Formatierungsfunktionen via Markdown eine sinnvolle Ergänzung, um bestimmte Abschnitte optisch herauszustellen (z.B. für Datenschemata).

Weitere Informationen zur Umsetzbarkeit finden sich in Kapitel 7.11.6.

Vergleich mit bestehender Lösung

Das Textfeld „Beschreibung“ hat begrenzte Formatierungsmöglichkeiten über HTML-Tags.

7.11.4.2 Handlungsfeld: Bezeichnung und Erklärung

7.11.4.2.1 Anforderung 4.2.1: Bezeichnung einiger Datenfelder anpassen

Nr. / ID	1014	Priorität	2	Umsetzbarkeit	1
-----------------	------	------------------	---	----------------------	---

Beschreibung Soll-Zustand

Folgende Begriffe sollten angepasst werden:

Begriff	Vorschlag
„Ausgeblendet“	Umbenennen in: „Sichtbarkeit“
„Webseite“ unter „Ressourcen“	Umbenennen in: „URL-Adresse der Ressource“
„Ebene der geopolitischen Abdeckung URI“	Umbenennen in: „Politische Ebene“

Vergleich mit bestehender Lösung

In einigen Fällen sind die gewählten Ausdrücke für Nutzer:innen ohne Vorkenntnisse unverständlich.

7.11.4.2.2 Anforderung 4.2.2: Löschen / Ergänzen von bestimmten Datenfeldern

Nr. / ID	1015	Priorität	3	Umsetzbarkeit	3
-----------------	------	------------------	---	----------------------	---

Beschreibung Soll-Zustand

Ein besonderes Augenmerk liegt hier auf der Möglichkeit der Mehrfachnennung im Feld „Autor“, die für viele Forschungsinstitutionen erforderlich ist. Ebenso sollte das Feld „Aktualisierungsfrequenz“ ergänzt werden, damit eine für Echtzeitdaten zentrale Information eingepflegt werden kann.

Folgende Datenfelder sollten ergänzt / gelöscht werden:

Datenfeld	Vorschlag
Qualitätssicherungsprozess URI	Löschen
Beschreibung (des Metadatensatzes)	Möglicher vorausgefüllter und überschreibbarer Text innerhalb des Feldes: "Bitte beschreiben Sie ihren Datensatz hier so genau wie möglich. Die Beschreibung ist die "Anleitung" zur Nutzung des Datensatzes. Die Beschreibung sollte bereits detailliert erklären, was in der Datei zu finden ist, ohne sie öffnen zu müssen. Wünschenswert sind vor allem Angaben zum Aufbau des Datensatzes, den enthaltenen Spalten, der Formatierung u.ä. (hier ein Beispiel ⁶⁴). Handelt es sich um eine sehr umfangreiche Datei oder gibt es hinsichtlich der Aussagekraft der Daten etwas zu beachten, benennen Sie diesen Umstand. Auch können Sie Anwendungsbeispiele (wie z.B. Apps) geben."
Aktualisierungsfrequenz	Ergänzung als Dropdown-Menü mit aus dct:accrualPeriodicity vorgegebenen deutschen Begriffen (monatlich, wöchentlich, täglich etc.); evtl. im Frontend anzuzeigen; evtl. Benennung als "Aktualisierungsfrequenz der Daten innerhalb des Datensatzes", damit hier keine Missverständnisse entstehen bei Institutionen, die bspw. jedes Jahr einen neuen Datensatz anlegen.
Abdeckung Text	Umbenennen in "Bezeichnung der verwaltungsgeografischen Abdeckung"
Autor	Feld sollte mehrfach belegbar sein. Andernfalls können Forschungsinstitutionen, bei denen oft mehrere Autor:innen an einem Datensatzes beteiligt sind, keine korrekten Metadaten erfassen.

Des Weiteren sollte das Formular dem DCAT-AP.de 2.0-Standard angepasst werden. Auch sollte es möglich sein, Dataservices / APIs anzugeben.

Weitere Informationen zur Umsetzbarkeit finden sich in Kapitel 7.11.6.

Vergleich mit bestehender Lösung

⁶⁴ <https://www.govdata.de/web/guest/suchen/-/details/cvd-sh-kreise-2022-08-19>

Bestimmte Datenfelder werfen bei den Datenbereitstellenden Zweifel hinsichtlich der Notwendigkeit der Befüllung oder einer korrekten Befüllung auf. In einigen Fällen fehlen Felder, die für eine korrekte Befüllung aus Sicht der Datenbereitstellenden nötig wären.

7.11.4.2.3 Anforderung 4.2.3: Unklarheiten beseitigen mit Mouseover/Infobutton

Nr. / ID	1005 / 1006	Priorität	2	Umsetzbarkeit	2
-----------------	-------------	------------------	---	----------------------	---

Beschreibung Soll-Zustand

Zu jedem einzelnen Feld sollten Erklärungen und eventuell ein Musterbeispiel direkt im Formular angezeigt werden können. Der BVA-Leitfaden zu Metadaten kann als Quelle für die FAQ Felder dienen.

Folgende Datenfelder sollten nach dem Feedback in den User-Interviews besser erklärt werden:

Datenfelder	Bemerkung
„Kalenderdatum“ und „Aktualisierungsdatum“ unter „Ressource“	Unklarheit, was genau an welcher Stelle erfragt wird und ob diese miteinander semantisch im Zusammenhang stehen; User wissen z.B. nicht, dass mit „Aktualisierungsdatum“ das Datum gemeint ist, wenn sich im Datensatz selbst etwas ändert und nicht das Datum, zu dem die Daten geupdatet werden
"Veröffentlichende Stelle" „Autor“ „Urheber" und „Ansprechpartner"	Unklarheit über das Verhältnis, zueinander und inwiefern hier Angaben evtl. mehrfach einzutragen sind
Kategorien	Trennschärfe der Kategorien klären; Unklarheit, ob Auswahl eher extensiv oder restriktiv gewünscht
Webseite bei Ressourcen	Unklar, ob Landingpage oder Unterseiten zu den Datensätzen gemeint sind
Verfügbarkeit	Kriterien für die Auswahl sollten genannt werden. Wann ist eine Ressource stabil.
By-Klausel	Unklarheit über die Bedeutung mit Beispiel beseitigen
Rechtsgrundlage	Unklarheit, ob bundesweite Regelungen oder nur landesspezifische gewünscht sind
Titel	Kriterien für einen guten Titel sollten genannt werden

Weitere Informationen zur Umsetzbarkeit finden sich in Kapitel 7.11.6.

Vergleich mit bestehender Lösung

Einige Datenfelder sind für Datenbereitstellende nicht intuitiv verständlich. Bisher dienen als Informationsmaterialien nur separate Dokumente, wie z.B. der BVA-Leitfaden.

7.11.4.3 Handlungsfeld: Hilfestellung bei der Eingabe

7.11.4.3.1 Anforderung 4.3.1: Verweise auf eindeutige Musterbeispiele

Nr. / ID	1022 / 1016	Priorität	2	Umsetzbarkeit	-
-----------------	-------------	------------------	---	----------------------	---

Beschreibung Soll-Zustand

Vielen Datenbereitstellenden wäre geholfen, wenn sie zum Überprüfen Ihrer Angaben eine Musterlösung als Vergleichsgrundlage hätten. Es wäre sinnvoll, wenn GovData ein idealtypisch befülltes Formular zur Verfügung stellen würde und dieses über Verlinkung im Formular für Datenbereitstellende einsehbar wäre.

Insbesondere bei Lizenzen benötigen viele Neueinsteiger Aufklärung und ggf. kompetente Ansprechpersonen. Viele Organisationen fertigen einen internen Laufzettel an, um einheitliche und rechtsichere Angaben zu gewährleisten. Besonders für Organisationen, die hier ohne eigene Expertise sind, wäre eine musterhafte Beispiellösung von GovData empfehlenswert. Hierzu könnte eventuell Bezug auf die aktuell im BMI in Abstimmung befindliche Empfehlungen Bezug genommen werden.

Vergleich mit bestehender Lösung

Als Musterlösungen lassen sich bisher nur Metadatensätze auf GovData heranziehen. Lizenzempfehlungen wie z.B. unter <https://www.dcat-ap.de/def/licences> sind nicht verlinkt

7.11.4.3.2 Anforderung 4.3.2: Anlegen von Auto-Vervollständigungslisten und Auswahloptionen

Nr. / ID	1001 / 1004 / 1008	Priorität	2	Umsetzbarkeit	3
-----------------	--------------------	------------------	---	----------------------	---

Beschreibung Soll-Zustand

Bei Schlagwörtern kann man das Freitext-Problem umgehen, indem bei der Eingabe häufig frequentierte Begriffe mit einem Ranking als Auswahloptionen angezeigt werden. Durch die Standardisierung der Schlagwörter werden auch Variationen von Begriffen wie z.B. „OZG /Onlinezugangsgesetz/Online-Zugangsgesetz“ umgangen.

Bei den Datenfeldern „Geopolitische Abdeckung URI“ oder beim „Format“ für Ressourcen würden sich reine Auswahllisten anbieten. Dies könnte zukünftig ermöglichen, automatisierte Plausibilitätsprüfungen als Hilfestellungen zu integrieren.

Weitere Informationen zur Umsetzbarkeit finden sich in Kapitel 7.11.6.

Vergleich mit bestehender Lösung

Die Freitext-Eingabe sieht keine „Führung“ bei der Eingabe oder Überprüfung des Eingegebenen vor. Musterlösungen lassen sich bisher nur separiert, über den Output vom Frontend auf GovData, anschauen.

7.11.4.4 Handlungsfeld: Redundanz bei der Eingabe

7.11.4.4.1 Anforderung 4.4.1: Auto-Vervollständigung

Nr. / ID	1001 / 1004 / 1008	Priorität	2	Umsetzbarkeit	Wird mit 6.8.4.3.2 erledigt
-----------------	--------------------	------------------	---	----------------------	-----------------------------

Beschreibung Soll-Zustand

Innerhalb des Formulars werden bei Eingabefeldern mit Freitext Vorschläge zur automatischen Vervollständigung angeboten (siehe 7.11.4.3.2). Damit würde bei der Dateneingabe, insbesondere bei sich

wiederholenden Eingaben, Zeit gespart und Flüchtigkeitsfehler bei der Eintragung vermieden.

Vergleich mit bestehender Lösung

Das Formular ist statisch aufgebaut, jedes Feld muss händisch ausgefüllt werden, auch bei gedoppelter Information.

7.11.4.4.2 Anforderung 4.4.2: Mustervorlagen

Nr. / ID	1001 /1007	Priorität	2	Umsetzbarkeit	3
-----------------	------------	------------------	---	----------------------	---

Beschreibung Soll-Zustand

Institutionen können verschiedene Vorlagen zum Einpflegen ihrer Daten anlegen, in denen bestimmte Felder bereits vorausgefüllt sind, damit diese nicht für jeden Datensatz händisch eingegeben werden müssen (bspw. bei Adressdaten). Verschiedene Nutzerkonten einer Institution (s. Anforderung 4.1.5) können auf diese zugreifen und neue Vorlagen anlegen.

Die optionale Übernahme von Adressdaten, die außerhalb des Formulars an GovData übermittelt wurden, könnte die Arbeit ebenfalls erleichtern.

Weitere Informationen zur Umsetzbarkeit finden sich in Kapitel 7.11.6.

Vergleich mit bestehender Lösung

Siehe 7.11.4.4.1.

7.11.4.4.3 Anforderung 4.4.3: Automatische Übernahme

Nr. / ID	1009	Priorität	2	Umsetzbarkeit	3
-----------------	------	------------------	---	----------------------	---

Beschreibung Soll-Zustand

Die optionale Übernahme von Kontaktdaten von „Veröffentlichende Stelle“ zum „Ansprechpartner“ zum „Autor“ ist für viele Datenbereitstellende von zentraler Bedeutung.

Möglich wäre auch eine Vorbelegung des Feldes „Geopolitische Abdeckung“ auf Basis der Angabe aus „Veröffentlichende Stelle“.

Ebenso könnte das Feld „By Clauses“ bereits auf Basis der „Veröffentlichenden Stelle“ vorausgefüllt sein.

Weitere Informationen zur Umsetzbarkeit finden sich in Kapitel 7.11.6.

Vergleich mit bestehender Lösung

Siehe 7.11.4.4.1.

7.11.4.5 Handlungsfeld: Rahmenbedingungen und Sonstiges

7.11.4.5.1 Anforderung 4.5.1: Anpassungen UX

Nr. / ID	1024	Priorität	2	Umsetzbarkeit	2
-----------------	------	------------------	---	----------------------	---

Beschreibung Soll-Zustand

Der Button „Datensatz löschen“ ist zu zentral positioniert. Ein zusätzlicher „Share“-Button soll nach Veröffentlichung des Metadatensatzes zur Verfügung stehen.

Weitere Informationen zur Umsetzbarkeit finden sich in Kapitel 7.11.6.

Vergleich mit bestehender Lösung

Der Button „Datensatz löschen“ ist zentral unter jedem Dokument platziert. Eine Möglichkeit zum „Teilen“ der veröffentlichten Daten erscheint im unteren Bereich der Seite.

7.11.5 Nichtfunktionale Anforderungen

7.11.5.1 Allgemeine Anforderungen

- Das neue Formular wird weiterhin über das bestehende GovData Portal eingebettet;
- Das neue Formular soll Neueinsteigern einen leichten Zugang zu OpenData gewährleisten;
- Das neue Formular soll Power-Usern eine Zeitersparnis beim Ausfüllen ermöglichen;
- Das Formular soll motivieren mehr Open Data bereitzustellen und beispielsweise seine Datensätze in bevorzugten OpenData-Datentypen wie RDF/LOD bereitzustellen;
- Perspektivisch könnte eine Standalone-Variante des Formulars von kommunalen OpenData-Portalen nachgenutzt werden.

7.11.5.2 Gesetzliche Anforderungen

- Das neue Formular soll weitern den gesetzlichen Anforderungen wie Datenschutzrichtlinien entsprechen

7.11.5.3 Technische Anforderungen

- Das neue Formular soll das Erstellen von DCAT-AP.de 2.0-kompatiblen Metadaten ermöglichen

7.11.6 Hinweise zur Umsetzbarkeit durch den Dienstleister

Zu Anforderung 7.11.4.1.4: Import von CSV-Dateien

„Beim Import werden die Metadaten neu erstellt. Falls eine Aktualisierung mit dem Import einer CSV-Datei möglich sein soll, muss hierfür erst noch eine Export-Funktion (mit interner ID zur Identifikation) in GovData implementiert werden, da ansonsten die Metadaten nicht zu bestehenden Metadaten zugeordnet werden können.“

- Zu Anforderung 7.11.4.1.5: Einbindung mehrerer Accounts

„Initial muss ein Benutzerkonto für eine Organisation als Administrator festgelegt werden.“

„Eine Auswahl aus bereits registrierten Nutzern. Die eigentliche Implementierung hängt von den eingesetzten Anwendungen (Liferay/CKAN?) und dem Ort des Formulars (in CMS eingebettet oder Standalone-Anwendung) ab.“

- Zu Anforderung 7.11.4.1.6: Formatierungsoptionen im Feld „Beschreibung“

„Die Unterstützung weiterer Formatierungsfunktionen ist möglich, aber durch den Austausch der Metadaten sind im Text (wie aktuell auch schon mit HTML-Tags) Formatierungszeichen enthalten, die stören könnten.“

„Das betrifft neben dem Formular auch die Detailansicht der Metadaten, wo die neuen Formatierungszeichen interpretiert werden müssen.“

- Zu Anforderung 7.11.4.2.2: Löschen / Ergänzen von bestimmten Datenfeldern

„Wir entnehmen der Beschreibung, dass nicht nur einer, sondern bis zu drei Kontakte in der Rolle Autor eingegeben und gespeichert werden können sollen. Wir würden die Anzahl an möglichen Autoren nicht beschränken, analog zu den Ressourcen. Es ist nur eine Frage der Darstellung.

Allerdings müssen die zusätzlichen Daten auch irgendwo gespeichert werden und bei der Anzeige und im RDF berücksichtigt werden. Standardmäßig wird von CKAN (ckanext-dcat) nur die Speicherung von einem Kontakt unterstützt. Die Speicherung von mehreren Kontakten ist so bisher nicht vorgesehen und müsste deshalb erweitert oder sogar komplett geändert werden.“

- Zu Anforderung 7.11.4.2.3: Unklarheiten beseitigen mit Mouseover/Infobutton

„Ein Mouseover würden wir für komplexere Beschreibungen mit ggf. enthaltenen Formatierungen nicht empfehlen.

Wir würden einen „Infobutton“, ähnlich wie beim MQA-Dashboard vorsehen, mit dem bei Klick darauf der zusätzliche Infotext aus- und wieder eingeklappt werden kann.“

- Zu Anforderung 7.11.4.3.2: Anlegen von Auto-Vervollständigungslisten mit/und Auswahloptionen

„In der Spezifikation steht, dass es sich um ein Freitextfeld handelt. Aber falls Wertelisten existieren sollten, könnten deren Werte im Formular angeboten werden. Dabei sollte aber wahrscheinlich eine Mehrfachangabe möglich sein.“ (zu „Rechtsgrundlage“)

„Hier stehen mehrere Auswahllisten dahinter, z.B. „regionalKey“, „municipalityKey“, etc., von daher könnte es etwas unübersichtlich werden, und auch die Listen sehr lange, wie z.B. für den Regionalschlüssel.

Eventuell könnte hier auch analog zum Vorschlag für die Schlagwörter auch bei der Eingabe von Zeichen die dazu gefundenen Einträge angezeigt werden, aus denen ein Eintrag ausgewählt werden kann.“ (zu „Geopolitische Abdeckung URI“)

- Zu Anforderung 7.11.4.4.2: Mustervorlagen

„Hierfür benötigt es eine Eingabemaske, um diese Daten verwalten zu können. Die Daten müssen in einer Datenbank gespeichert werden. Hierzu könnte die bereits vorhandene DB und der Microservice „govdata-db“ genutzt und erweitert werden.“

„Falls mehrere Vorlagen erstellt wurden, muss eine Liste der Vorlagen angeboten werden, aus der die gewünschte Vorlage ausgewählt werden kann.“

- Zu Anforderung 7.11.4.4.3: Automatische Übernahme

„Man müsste das relativ flexibel halten und eine Übernahme von jedem Kontakt zu einem anderen Kontakt ermöglichen. Zum Beispiel könnte an jedem Kontakt ein „Button“ angezeigt werden und bei Klick auf den Button zu dem die Daten übernommen werden sollen, eine Liste der bereits vorhandenen Kontakte angezeigt werden, aus dem der Kontakt ausgewählt werden kann, von dem die Daten übernommen werden sollen.“ ; „Das könnte

mit der Anforderung „7.11.4.4.2 Mustervorlagen“ vielleicht in gewisser Weise ermöglicht werden. Aber falls das innerhalb eines Metadatensatzes öfter individuell möglich sein soll, wäre das möglich. Diesen Fall würden wir hier mit der Anforderung für eine Umsetzung berücksichtigen.“ (zu „Übernahme von Kontaktdaten“)

„Es ist fraglich, ob von dem Ort für die „Veröffentlichende Stelle“ immer die korrekte „Geopolitischen Abdeckung“ ermittelt werden kann. Dort müsste anschließend für den Benutzer eventuell noch eine Auswahl von den passendsten gefundenen Werten ermöglicht werden.“; „Das könnte mit der Anforderung „7.11.4.4.2 Mustervorlagen“ vielleicht in gewisser Weise ermöglicht werden.“ (zur Vorbelegung des Feldes „Geopolitische Abdeckung“ und „By Clauses“ auf Basis von „Veröffentlichende Stelle“)

- Zu Anforderung 7.11.4.5.1: Anpassungen UX

" Hier müsste zuerst die gewünschte Darstellung erarbeitet werden.“ (zu „Positionierung Button ‚Löschen‘“)

„Was genau soll nach dem Klick auf den „Share“-Button passieren?

Soll lediglich die URL zum Metadatensatz kopiert werden? Dieser Fall ist in der Einschätzung zur Umsetzbarkeit berücksichtigt.

Es existieren bisher Möglichkeiten zum Teilen des Metadatensatzes für E-Mail, Facebook, Twitter und Diaspora.

Die Buttons dazu befinden sich auf jeder Seite im rechten unteren Bereich.“

8 Literaturverzeichnis

- Automatic, WP Super Cache, WordPress, zuletzt abgerufen am 08.12.2022, <https://wordpress.org/plugins/wp-super-cache/>
- Bundesministerium des Innern, „Regierungsprogramm. Vernetzte und transparente Verwaltung“, Sep. 2010, Bundesministerium des Innern.
- Redaktion Organisationshandbuch, „2.4.3 Personalbedarfsermittlung (PBE) – PBE-Leitfaden“, Bundesministerium des Innern, für Bau und Heimat / Bundesverwaltungsamt, zuletzt abgerufen am 10.08.2022, https://www.orghandbuch.de/OHB/DE/OrganisationshandbuchNEU/2_Organisationsmanagement/2_4_Ressourcen/2_4_3_Leitfaden/leitfaden-node.html.
- Bundesverband Deutscher Unternehmensberatungen, „Studie: Honorare im Consulting steigen leicht an“, zuletzt abgerufen am 21.10.2022, <https://www.bdu.de/newsletter/ausgabe-12020/studie-honorare-im-consulting-steigen-leicht-an/>
- Bundesverwaltungsamt, „Leitfaden Metadaten“, 24. Aug. 2020, zuletzt abgerufen am 21.11.2022, https://www.bva.bund.de/SharedDocs/Downloads/DE/Behoerden/Beratung/Methoden/open_data_leitfaden_metadaten.pdf
- CKAN, ckanext-harvest, GitHub, zuletzt abgerufen am 25.11.2022, <https://github.com/ckan/ckanext-harvest>
- CKAN, ckanext-dcat, GitHub, zuletzt abgerufen am 25.11.2022, <https://github.com/ckan/ckanext-dcat>
- Datopian Tech, „CKAN v3“, Datopian Tech, 4. Feb. 2022, zuletzt abgerufen am 25.11.2022, <https://tech.datopian.com/ckan-v3/>
- Deutscher Bundestag, Drucksache 19/27442, 09. Mär. 2021, Deutscher Bundestag.
- DCAT-AP.de, „DCAT-AP.de Spezifikation 2.0“, 28. Feb. 2022, zuletzt abgerufen am 21.11.2022, <https://www.dcat-ap.de/def/dcatde/2.0/spec/>
- DCAT-AP.de, „DCAT-AP.de Konventionenhandbuch 2.0“, 28. Feb 2022, zuletzt abgerufen am 21.11.2022, <https://www.dcat-ap.de/def/dcatde/2.0/implRules/>
- DCAT-AP.de, „DCAT-AP.de Validator“, zuletzt abgerufen am 21.11.2022, <https://www.itb.ec.europa.eu/shacl/dcat-ap.de/upload>
- DCAT-AP.de, „Herzlich willkommen auf DCAT-AP.de“, zuletzt abgerufen am 21.11.2022, <https://www.dcat-ap.de/>
- DCAT-AP.de, „Liste der Datenbereitsteller“, zuletzt abgerufen am 21.11.2022, <https://www.dcat-ap.de/def/contributors/>
- DCAT-AP.de, „Vokabulare und Dokumente für DCAT-AP.de“, zuletzt abgerufen am 21.11.2022, <https://www.dcat-ap.de/def/>
- Die Techniker, „Wer hat die meisten Urlaubstage in Deutschland?“, zuletzt abgerufen am 24.08.2022, https://www.tk-lex.tk.de/web/guest/externalcontent?_leongshared_template=HAUFEDetail&_leongshared_externalcontentid=15_PORTLET_17979109
- Drafting Team Metadata and European Commission Joint Research Centre, INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119, European Commission Joint Research Centre, 6. Nov. 2013.
- GovData, „Anleitung zur Datenbereitstellung auf GovData“, GovData, zuletzt abgerufen am 21.11.2022,

- <https://www.govdata.de/web/guest/datenbereitstellungaufgovdata>
- GovData, „Das Portal“, GovData, zuletzt abgerufen am 20.09.2022, <https://www.govdata.de/web/guest/hilfe>.
- GovData, „Schnittstellen. SPARQL Assistent“, GovData, zuletzt abgerufen am 16.09.2022, <https://www.govdata.de/web/guest/sparql-assistent>
- IT-Planungsrat, „Die Nationale E-Government-Strategie“, IT-Planungsrat, zuletzt abgerufen am 08.12.2022, <https://www.it-planungsrat.de/der-it-planungsrat/nationale-e-government-strategie>.
- IT-Planungsrat, „Steuerungsprojekt „Open Government““, IT-Planungsrat Beschluss 2012/43, 25. Okt. 2012, zuletzt abgerufen am 08.12.2022, <https://www.it-planungsrat.de/beschluss/beschluss-2012-43>
- IT-Planungsrat, „Standardisierungsbedarf zur Metadatenstruktur für offene Verwaltungsdaten“, IT-Planungsrat Beschluss 2018/11, 16. Apr. 2018, zuletzt abgerufen am 21.11.2022, <https://www.it-planungsrat.de/beschluss/beschluss-2018-11>
- IT-Planungsrat, „Standardisierungsbedarf ‚Metadatenstruktur für offene Verwaltungsdaten‘“, IT-Planungsrat Beschluss 2018/30, 28. Jun. 2018, zuletzt abgerufen am 28.11.2022, <https://www.it-planungsrat.de/beschluss/beschluss-2018-30>
- IT-Planungsrat, „Änderung des E-Government-Gesetzes und Datennutzung auf GovData“, IT-Planungsrat Beschluss 2021/18, 17. Mär. 2021, zuletzt abgerufen am 21.09.2022, <https://www.it-planungsrat.de/beschluss/beschluss-2021-18>.
- IT-Planungsrat, „Änderung des E-Government-Gesetzes und Datennutzung auf GovData“, IT-Planungsrat Beschluss 2021/48, 29. Okt. 2021, zuletzt abgerufen am 21.09.2022, <https://www.it-planungsrat.de/beschluss/beschluss-2021-48>.
- J. Klessmann, P. Denker, I. Schieferdecker, S. E. Schulz et al., „Open Government Data Deutschland. Eine Studie zu Open Government in Deutschland im Auftrag des Bundesministerium des Innern“, 2012, Fraunhoferinstitut für offene Kommunikationssysteme FOKUS.
- M. Smithson (2003): *Confidence Intervals* (No. 140). Sage, ISBN: 9780761924999.
- Michael Simpson, contact-form-7-to-database-extension, WordPress, zuletzt abgerufen am 08.12.2022, <https://wordpress.org/plugins/contact-form-7-to-database-extension/>
- N. Shemtov und I. Walden (2013): *Free and Open Source Software: Policy, Law And Practice*, Oxford University Press, ISBN: 9780199680498.
- Nocean, HoneyPot for Contact Form 7, WordPress, zuletzt abgerufen am 08.12.2022, <https://wordpress.org/plugins/contact-form-7-honeypot/>
- Open Source Initiative, „The Open Source Definition“, Open Source Initiative, zuletzt abgerufen am 16.11.2022, <https://opensource.org/osd>.
- P. Wagner-Schelewsky und L. Hering (2019): *Online-Befragung*, in: N. Baur und J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung*, https://doi.org/10.1007/978-3-658-21308-4_49.
- Pluginkollektiv, Antispam Bee, WordPress, zuletzt abgerufen am 08.12.2022, <https://wordpress.org/plugins/antispam-bee/>
- Redaktions-Team GovData.de, „Verwaltungsvereinbarung GovData tritt in Kraft“, GovData, zuletzt abgerufen am 21.09.2022, <https://www.govdata.de/neues/-/blogs/verwaltungsvereinbarung-govdata-tritt-in-kraft>.
- R. Rehm, „Wie mache ich eigentlich Open Source?“, 2012 Java Script Magazin, zuletzt abgerufen am 14. 11 2022, <http://www.magjs.de/2012-01/rehm/rehm.html>.

- Statistisches Bundesamt, „Krankheitstage je Beschäftigten pro Jahr“, 2022, Statistisches Bundesamt, zuletzt abgerufen am 24. 08. 2022, https://www.destatis.de/DE/Themen/Arbeit/Arbeitsmarkt/Qualitaet-Arbeit/Dimension-2/_Interaktiv/2-4-krankenstand.html
- T. Costa, “Maintaining Open Source Projects”, 2016, thoughtbot, zuletzt abgerufen am 08.12.2022, <https://books.thoughtbot.com/assets/maintaining-open-source-projects.pdf>.
- Takayuki Miyoshi, Contact Form 7, WordPress, zuletzt abgerufen am 08.12.2022, <https://wordpress.org/plugins/contact-form-7/>
- Team Yoast, Yoast SEO, WordPress, zuletzt abgerufen am 08.12.2022, <https://wordpress.org/plugins/wordpress-seo/>
- The bbPress Contributors, bbPress, WordPress, zuletzt abgerufen am 08.12.2022, <https://wordpress.org/plugins/bbpress/>
- Deutscher Städte- und Gemeindebund, „Wettbewerb "Apps für Deutschland"“, 18. Okt. 2011, DStGB, zuletzt abgerufen am 08.12.2022, <https://www.dstgb.de/aktuelles/archiv/archiv-2011/wettbewerb-apps-fuer-deutschland/>.
- Wikipedia, „Bugtracker“, Wikipedia, zuletzt abgerufen am 14.11.2022), <https://de.wikipedia.org/wiki/Bugtracker>.