



IT-Planungsrat

Digitale Zukunft gestalten

Stand der Umsetzung der Spezifikation DIN SPEC 91379

Bericht der Fachgruppe String.Latin+ an den IT-Planungsrat

Version 1.0
11. August 2022
Rico Apitz (Jinit[AG)

Fachgruppe String.Latin+
UAG Berichtspflicht
Berlin

Inhaltsverzeichnis

1	Zusammenfassung und Schlussfolgerungen	4
1.1	Statistik	4
1.2	Rücklauf	4
1.3	Interpretation der Ergebnisse	5
1.4	Schlussfolgerungen und Ausblick	6
2	Hintergrundinformationen	7
3	Ergebnisse der Abfrage	7
3.1	Erläuterungen zum Vorgehen	7
3.2	Umfang der Rückmeldungen	9
3.3	Metadaten	11
3.4	Aktueller Status der Konformität	12
3.5	Planung / Vorgehen zum Erreichen der Konformität	16
3.6	Risiken beim weiteren Vorgehen zum Erreichen der Konformität	20
3.7	Probleme bei der Umsetzung	24
3.8	Allgemeine Hinweise	32
4	Erfahrungen für vergleichbare Umfragen	33
5	Abkürzungsverzeichnis	35
6	Quellenverzeichnis	37

Abbildungsverzeichnis

Abbildung 1: Gründe für Streichungen von Rückmeldungen	8
Abbildung 2: Verteilung der Rückmeldungen zwischen Bund, Länder und Kommunen	9
Abbildung 3: Verteilung der Rückmeldungen zwischen obersten Bundesbehörden, einzelnen Ländern und Kommunen	10
Abbildung 4: Verteilung der gemeldeten IT-Verfahren nach Themenbereich	11
Abbildung 5: Aktueller Status der Konformität zur DIN SPEC 91379 im Bereich Erfassung	14
Abbildung 6: Aktueller Status der Konformität zur DIN SPEC 91379 im Bereich Speicherung	14
Abbildung 7: Aktueller Status der Konformität zur DIN SPEC 91379 im Bereich Übermittlung	15
Abbildung 8: Aktueller Status der Konformität zur DIN SPEC 91379 im Bereich Anzeige	15
Abbildung 9: Aktueller Status der Konformität zur DIN SPEC 91379 im Bereich Ausdruck	16
Abbildung 10: Geplante oder bereits erreichte Unterstützung von UTF-8	17
Abbildung 11: Geplante oder bereits erreichte Unterstützung von String.Latin 1.1	18
Abbildung 12: Geplante oder bereits erreichte Unterstützung von DIN SPEC 91379 (String.Latin+ 1.2)	19
Abbildung 13: Bewertung des Risikos zum Erreichen der Konformität aufgrund Abhängigkeiten von Produktherstellern	20
Abbildung 14: Bewertung des Risikos zum Erreichen der Konformität aufgrund Abhängigkeiten von Schnittstellenpartnern.....	21
Abbildung 15: Bewertung des Risikos zum Erreichen der Konformität aufgrund Abhängigkeiten von Standards	22
Abbildung 16: Bewertung des Risikos zum Erreichen der Konformität aufgrund Abhängigkeiten von widersprüchlichen rechtlichen Vorgaben Dritter	23
Abbildung 17: Bewertung sonstiger Risiken zum Erreichen der Konformität	23
Abbildung 18: Bewertung von Umsetzungsproblemen mit der Eingabe der Schriftzeichen	25
Abbildung 19: Bewertung von Umsetzungsproblemen mit der Barrierefreiheit.....	26
Abbildung 20: Bewertung von Umsetzungsproblemen durch Schriftarten	27
Abbildung 21: Bewertung von Umsetzungsproblemen durch Multi-Codepoint-Zeichen	28
Abbildung 22: Bewertung von Umsetzungsproblemen durch die Corona-Pandemie	29
Abbildung 23: Bewertung von Umsetzungsproblemen aufgrund eines Mangels an Ressourcen	30
Abbildung 24: Bewertung von sonstigen Umsetzungsproblemen	31

1 Zusammenfassung und Schlussfolgerungen

1.1 Statistik

Vom 13. Dezember 2021 bis zum 10. Juni 2022 haben die Mitglieder des IT-Planungsrats (IT-PLR) aus Bund, Länder und Kommunen (nur Landeshauptstädte) entsprechend der Entscheidung 2019/16 [itplr2019a] zum zweiten Mal nach zwei Jahren der Fachgruppe „String.Latin+“ (FGSL) zum Stand der Umsetzung der DIN SPEC 91379 für IT-Verfahren, die dem Bund-Länder übergreifenden Datenaustausch oder dem Datenaustausch mit Bürgern und Wirtschaft dienen, berichtet. Die Fachgruppe legt mit diesem Dokument dem IT-PLR die Ergebnisse seiner Abfrage vor.

Es wurden Angaben zu 2.342 IT-Verfahren gemacht. Davon wurden 1.168 Angaben aus der statistischen Auswertung gestrichen, weil sie redundant, fehlerhaft oder unbrauchbar unvollständig waren oder weil sich die Verfahren nicht im Anwendungsbereich der DIN SPEC oder noch in der Entwicklung befinden.

Von den verbleibenden 1.174 IT-Verfahren (vor zwei Jahren waren es 1.410 IT-Verfahren) vermelden 65%, dass sie durchgehend (d.h. für Erfassung, Speicherung, Übermittlung, Anzeige und Ausdruck) ihre Zeichen mit UTF-8 kodieren, d.h. dass sie Unicode-fähig sind. Vor zwei Jahren waren es nur 35% (durch die niedrigere Grundgesamtheit haben sich die absoluten Zahlen jedoch nicht fast verdoppelt, sie sind von 495 auf 764 gestiegen). 43% haben bereits String.Latin 1.1 implementiert (19% vor zwei Jahren) und für 36% war die Umstellung auf die DIN SPEC 91379 (String.Latin+ 1.2) bereits erfolgreich (13% vor zwei Jahren). Die vollständige Produktivsetzung von String.Latin+ 1.2 planen innerhalb der beschlossenen Frist bis zum November 2024 (Entscheidung 2019/53, siehe [itplr2019b]) weitere 14%. In Summe machen bereits umgestellte und rechtzeitig für eine Umstellung geplante IT-Verfahren genau die Hälfte aller Rückmeldungen aus (im Vergleich zu 42% vor zwei Jahren, die absoluten Zahlen sind jedoch leicht gesunken).

Wie vor zwei Jahren geben zwei Prozent an, dass sie länger als bis zum Jahr 2024 brauchen werden. Vier Prozent geben an, keine Umstellung auf String.Latin+ 1.2 zu planen (drei Prozent vor zwei Jahren), wobei dafür meistens nachvollziehbare Gründe genannt werden. Mit zusammen 44% gibt es einen sehr großen Anteil von Verantwortlichen für IT-Verfahren, die eine Umstellung auf die DIN SPEC 91379 noch nicht geplant haben oder die in der Umfrage keine konkreten Angaben gemacht haben.

Am weitesten ist mit 59% die Umsetzung der DIN SPEC im Bereich der Speicherung (40% vor zwei Jahren). Ungefähr gleichauf ist die Umsetzung bei Erfassung, Übermittlung und Anzeige bei ca. 49% (32% vor zwei Jahren). Den größten Nachholbedarf gibt es unverändert beim Ausdruck, für den aber immerhin 40% die Umstellung auf die DIN SPEC als abgeschlossen betrachten (26% vor zwei Jahren).

Als größte Risiken für Planung und Umsetzung der DIN SPEC werden Abhängigkeiten von Schnittstellenpartnern (gestiegen von 24 auf 32%) und von Produktherstellern (gesunken von 38 auf 30%) genannt.

Weniger als zehn Prozent der Befragten geben an, bei der Umsetzung Probleme mit der Eingabe, Barrierefreiheit, Schriftarten oder der Corona-Pandemie zu haben. Für die bereits vor zwei Jahren abgefragten Probleme gab es keine nennenswerte Veränderung. Am häufigsten – mit jedoch nur zwölf Prozent – wurde bei der aktuellen Umfrage Ressourcenmangel genannt. Allerdings werden für 60 bis 75% der IT-Verfahren keine Angaben zu diesen Problemfeldern gemacht (75 bis 76% vor zwei Jahren, die konkreten nein-Antworten sind gestiegen).

1.2 Rücklauf

Insgesamt haben elf Bundesministerien und andere oberste Bundesbehörden (von 17), 14 Länder (von 16) und drei Landeshauptstädte (von 13) ausgefüllte Fragebögen bereitgestellt (fünf Länder mehr als vor zwei

Jahren, die gleiche Anzahl aus Bund und Kommunen, wenn auch nicht dieselben Absender). Die Rückmeldungen waren jedoch nicht alle vollständig. In der zur Verfügung gestellten Zeit konnten nicht für alle Behörden und für alle IT-Verfahren Angaben gemeldet werden. Ein erheblicher Anteil der Behörden vor allem auf der Bundesebene hat die Aufforderung zur Berichterstellung erst Anfang März erhalten, sodass die zwei Monate länger angesetzte Frist zur Beantwortung kaum Wirkung zeigte.

Erneut hat es vor allem von Landeshauptstädten wenig Rückmeldungen gegeben. Möglicherweise gab es in der Berichtskette ausgehend von der Geschäftsstelle des IT-PLR über Landesvertreter Probleme bei der gezielten Weitergabe in Richtung der kommunalen IT-Verantwortlichen.

Nicht alle Empfänger haben nachverfolgt / sichergestellt, dass nachgeordnete Behörden dem Wunsch nach Berichterstattung nachkommen. Einige nachgeordnete Behörden von Bundesressorts sowie Landesressorts haben sich direkt an das Postfach der FGSL gewandt, andere haben nicht auf die einmalige Aufforderung reagiert.

1.3 Interpretation der Ergebnisse

Für 36% der IT-Verfahren wird angegeben, die DIN SPEC 91379 bereits vollständig zu unterstützen. Das ist gegenüber 13% vor zwei Jahren eine erhebliche Steigerung. Angesichts der Umsetzungsfrist bis November 2024 besteht trotzdem akuter Handlungsbedarf. Eine konkrete Planung, diese Frist einzuhalten, gibt es für die Hälfte der betrachteten IT-Verfahren noch nicht. Es ist zu befürchten, dass dieser Anteil unter den nicht zurückgemeldeten IT-Verfahren noch höher ist.

Das Gesamtbild aller Antworten inklusive der Freitexte legt allerdings die Vermutung nahe, dass viele IT-Verfahren UTF-8 implementieren und deshalb angenommen wird, dass alle Unicode-Zeichen unterstützt werden. Das würde bedeuten, dass auch alle Zeichen der DIN SPEC 91379 unterstützt werden und somit eine Konformität zum Standard erreicht wird. Allein die Nutzung von UTF-8 zur software-technischen Kodierung der Zeichen stellt jedoch nicht sicher, dass alle möglichen Zeichen und Zeichensequenzen korrekt erfasst, übermittelt, angezeigt und ausgedruckt werden können. Besondere Herausforderungen liegen in den Eingabemöglichkeiten für Anwender und für automatische Texterkennung (OCR), im korrekten Umgang mit Multi-Codepoint-Zeichen in Textoperationen, wie Suche und Sortierung, in Definition und Implementation von Schnittstellen, die keine normativen Schriftzeichen ausschließen, in der Auswahl und Umstellung auf eine geeignete Schriftart für Anzeige und Ausdruck sowie in der Umsetzung der Barrierefreiheit.

Die Unterstützung von UTF-8 ist eine notwendige Voraussetzung für die Implementation der DIN SPEC 91379, die auch einen sehr großen Anteil des Migrationsaufwands abdeckt, aber sie ist dafür nicht hinreichend.

Die korrekte Umstellung des Ausdrucks auf String.Latin+ 1.2 von 40% aller IT-Verfahren muss angezweifelt werden. Es gibt nur wenige und vergleichsweise exotische Schriftarten, die wirklich alle Zeichen der DIN SPEC 91379 unterstützen. Da die oft problematischen Zeichen mit mehreren Codepoints in der Praxis nicht häufig vorkommen, darf vermutet werden, dass nicht allen Verantwortlichen bewusst ist, dass allein die Nutzung von UTF-8 nicht korrekte Darstellung und Ausdruck der Zeichen sicherstellen.

Bevor eine zur DIN SPEC konforme Schriftart eingeführt werden kann, muss oft zuerst das Corporate Design der Organisation angepasst werden. Weit verbreitete Schriftarten wie Arial, Verdana und BundesSans erfüllen nicht die Anforderungen der DIN SPEC 91379. Durch Unterschiede in der Laufweite der Schriftarten wird es bei amtlichen Dokumenten mit vorgegebenem Layout nicht selten Probleme geben, weil die Texte dann nicht mehr in die vorgesehenen Flächen passen.

Die Barrierefreiheit wird als Herausforderung vermutlich unterschätzt. Screenreader und sonstige Spezial-Software haben die Aufgabe noch nicht gelöst, einerseits mit den vielen diakritischen Zeichen umzugehen, aber andererseits für Menschen mit Behinderung beherrschbar zu bleiben.

Die offenen Herausforderungen bezüglich Eingabehilfen, Schriftarten, Barrierefreiheit und weiterer Punkte werden weiterhin in der FGSL bearbeitet. Die Ergebnisse dieser Arbeiten fließen in eine andauernd fortgeschriebene kommentierte Fassung der DIN SPEC 91379 inkl. FAQ [fgsl2020] ein.

Wer weitere Hinweise auf geeignete Lösungen (Best Practices) hat, möge diese bitte an String.Latin.Umsetzung@init.de senden. Das Postfach wird im Auftrag der KoSIT betrieben. Auch für Nachfragen zu diesem Bericht kann die E-Mail-Adresse verwendet werden.

1.4 Schlussfolgerungen und Ausblick

Die Umfrage hat einen ungebrochenen Bedarf an Informationen und Unterstützung für IT-Verantwortliche aufgezeigt. Neben dem Willen zur Umstellung auf eine korrekte Unterstützung aller lateinischen Namen und einer Finanzierung der dafür erforderlichen Aufwände ist auch Know-how für den Erfolg der Migration notwendig.

Auch bei der zweiten Umfrage wurde einige Male angegeben, von der Anforderung, die DIN SPEC 91379 implementieren zu müssen, erst durch die Umfrage erfahren zu haben. Es scheint ein systematisches Defizit zu geben, verbindliche Beschlüsse des IT-Planungsrates bei IT-Verantwortlichen der öffentlichen Verwaltung bekannt zu machen.

Auch wenn sich eine deutlich positive Entwicklung abzeichnet, was die Unterstützung der DIN SPEC 91379 angeht, ist der Anteil der IT-Verfahren, für die noch keine Maßnahmen zur Erreichung einer Konformität geplant wurden, 2,5 Jahre vor der Umsetzungsfrist viel zu hoch. Diverse Beispiele auch sehr komplexer Verfahren und Verfahrenslandschaften zeigen jedoch, dass die gesetzte Frist erreichbar war – wenn rechtzeitig mit der Umsetzung begonnen worden wäre. Nun sind erhöhte Anstrengungen bei den Nachzüglern erforderlich, um das gesetzte Ziel der flächendeckenden Konformität bis zum 1. November 2024 noch zu erreichen.

Oft wurde in den Freitexten angegeben, dass die Umstellung auf String.Latin aktuell durch technische Dienstleister umgesetzt wird. Deshalb konnten häufig noch keine Angaben zu möglichen Problemen gemacht werden.

Sehr oft wurde auf externe Schnittstellen verwiesen, die unterstützt werden müssen und die aktuell nicht die Zeichen der DIN SPEC 91379 zulassen. Neben verwaltungsextern gesteuerten Schnittstellen handelte es sich allerdings auch oft um andere IT-Verfahren der öffentlichen Verwaltung, die ebenso zur Implementation des Standards verpflichtet sind.

Offensichtlich war es erneut eine Herausforderung für die IT-Verantwortlichen, die abgefragten Aspekte für ihre IT-Verfahren zu bewerten, denn die Anzahl nicht ausgefüllter und mit „unbekannt“ bewerteter Fragestellungen ist zwar zurückgegangen, war aber auch in der zweiten Umfrage noch immer sehr hoch. Von den IT-Verantwortlichen wurden Anleitungen und Best Practices nachgefragt, um schneller und effizienter den Standard implementieren zu können. Es gibt erhebliche Wissenslücken und Missverständnisse in Bezug auf Geltungsbereich, Konformitätsanforderungen und technische Erfordernisse und Abhängigkeiten. Da es nicht gelingen wird, die Umsetzungsprojekte mit personellen Kapazitäten zu verstärken, sollte das vorhandene Know-how zur Nachnutzung dokumentiert, bereitgestellt und bekannt gemacht werden.

Erfahrungen aus den beiden Umfragen, mit denen vergleichbare Aktivitäten in der Zukunft verbessert werden können, werden im Kapitel 4 auf Seite 33 vorgestellt.

Mit dem Ausgabedatum 08/2022 ist am 29. Juli 2022 der Nachfolger der DIN SPEC 91379 vom DIN herausgegeben worden: die Norm DIN 91379 [din2022]. Laufende Aktivitäten zur Implementation der DIN SPEC 91379 sollten angepasst werden, um bis zum 1. November 2024 direkt die DIN-Norm zu unterstützen. Mit der Weiterentwicklung des Standards sind lediglich drei normative lateinische Buchstaben hinzugekommen, die für eine Konformität unterstützt werden müssen. Dafür wird ein zusätzliches diakritisches Zeichen benötigt. Im nicht-normativen Teil gibt es zwei neue Nicht-Buchstaben, die zur Konformität nicht benötigt werden.

Wie in anderen Branchen / Ressorts der öffentlichen Verwaltung üblich, muss die DIN-Norm kostenpflichtig bezogen werden. Entsprechend des deutschen Vergaberechts dürfen Normen in Ausschreibungen als MUSS-Anforderungen eingefordert werden, ohne dass es spezieller Begründungen bedarf. Die DIN SPEC 91379 ist keine Norm. Ihre Einhaltung in Ausschreibungen von Soft- oder Hardware verbindlich zu fordern, kann aber mit der Entscheidung 2019/16 des IT-Planungsrates [itplr2019a] begründet werden.

Als nächstes wird die Weiterentwicklung der DIN-Norm zu einer europäischen CEN-Norm angestrebt. Damit würde ermöglicht werden, dass Schnittstellen zu EU-Systemen den Standard für die Verarbeitung von Namen vorgeben. Die Unterstützung durch Software-Produkte, auch Standard-Software von Marktführern, kann dadurch hoffentlich gesteigert werden.

2 Hintergrundinformationen

Mit der Entscheidung 2019/16 [itplr2019a] hat der IT-Planungsrat (IT-PLR) die DIN SPEC 91379 [din2019] für die Verarbeitung von Namen in IT-Verfahren der öffentlichen Verwaltung beschlossen. Alle IT-Verfahren, die dem Bund-Länder übergreifenden Datenaustausch oder dem Datenaustausch mit Bürgern und Wirtschaft dienen, müssen spätestens zum 1. November 2024 konform zur DIN SPEC 91379 sein (Entscheidung 2019/53 [itplr2019b]). Zur Nachverfolgung seiner Beschlusslage hat der IT-PLR festgelegt, dass jeweils zum 1. Februar 2020 und zum 1. Februar 2022 zum Stand der Umsetzung an die „Fachgruppe String.Latin+“ zu berichten ist.

3 Ergebnisse der Abfrage

3.1 Erläuterungen zum Vorgehen

Am 13. Dezember 2021 hat die Geschäftsstelle des IT-PLR dessen Mitglieder angeschrieben und nach zwei Jahren zum zweiten Mal um das Ausfüllen einer Vorlage zur Erfassung des Standes der Umsetzung der DIN SPEC 91379 gebeten. Zur Berichterstattung aufgefordert wurden:

- auf der Bundesebene: alle unmittelbaren und mittelbaren Bundesbehörden,
- auf der Landesebene: alle Landesministerien,
- auf kommunaler Ebene: nur die Landeshauptstädte (außer Berlin, Hamburg, Bremen).

Für jedes betroffene IT-Verfahren war eine Zeile auszufüllen. Die abgefragten Daten gliederten sich in die Bereiche:

- Metadaten
- Aktueller Status der Konformität
- Planung / Vorgehen zum Erreichen der Konformität
- Risiken beim weiteren Vorgehen zum Erreichen der Konformität
- Probleme bei der Umsetzung

Die ausgewerteten Daten dieses Berichts schließen alle bis zum 10. Juni 2022 im Postfach String.Latin.Umsetzung@init.de eingetroffenen Rückmeldungen mit ein.

Danach eingetroffene und noch weiter einlaufende Rückmeldungen werden von der FGSL weiter entgegen genommen und ausgewertet.

Insgesamt wurden Angaben zu 2.342 IT-Verfahren gemacht. Das sind 880 mehr als bei der Umfrage vor zwei Jahren. Für die statistische Auswertung wurden 1.168-mal Angaben über IT-Verfahren gestrichen, weil

- die IT-Verfahren keine Namen i.w.S. verarbeiten oder aus anderen Gründen nicht im Geltungsbereich des Beschlusses des IT-PLR liegen,
- offensichtlich Fehler / Widersprüche in den Angaben enthalten sind,
- die Angaben bis zur Unbrauchbarkeit unvollständig oder sogar leer sind,
- die IT-Verfahren spätestens im Jahr 2024 abgelöst oder eingestellt werden sollen,
- die IT-Verfahren sich noch in der Entwicklung befinden oder
- IT-Verfahren doppelt gemeldet wurden (wobei Produkte, die von verschiedenen Ländern oder Kommunen eingesetzt werden, auch separat erfasst werden – sie könnten sich in ihrer Version oder Konfiguration und damit in der Unterstützung von String.Latin+ 1.2 unterscheiden).

Welchen Anteil die verschiedenen Streichungsgründe an der Gesamtzahl aller Streichungen haben, zeigt die Abbildung 1. Den größten Block bilden Angaben über 770 IT-Verfahren eines Bundeslandes, die telefonisch übermittelt wurden und zu wenig Informationen für die Auswertung enthalten haben. Vor zwei Jahren sind nur 52-mal Angaben über IT-Verfahren aus den genannten Gründen gestrichen worden. Die Maßstäbe, wann ein Datensatz unvollständig oder widersprüchlich ist, wurden bei dieser zweiten Umfrage allerdings strenger ausgelegt.

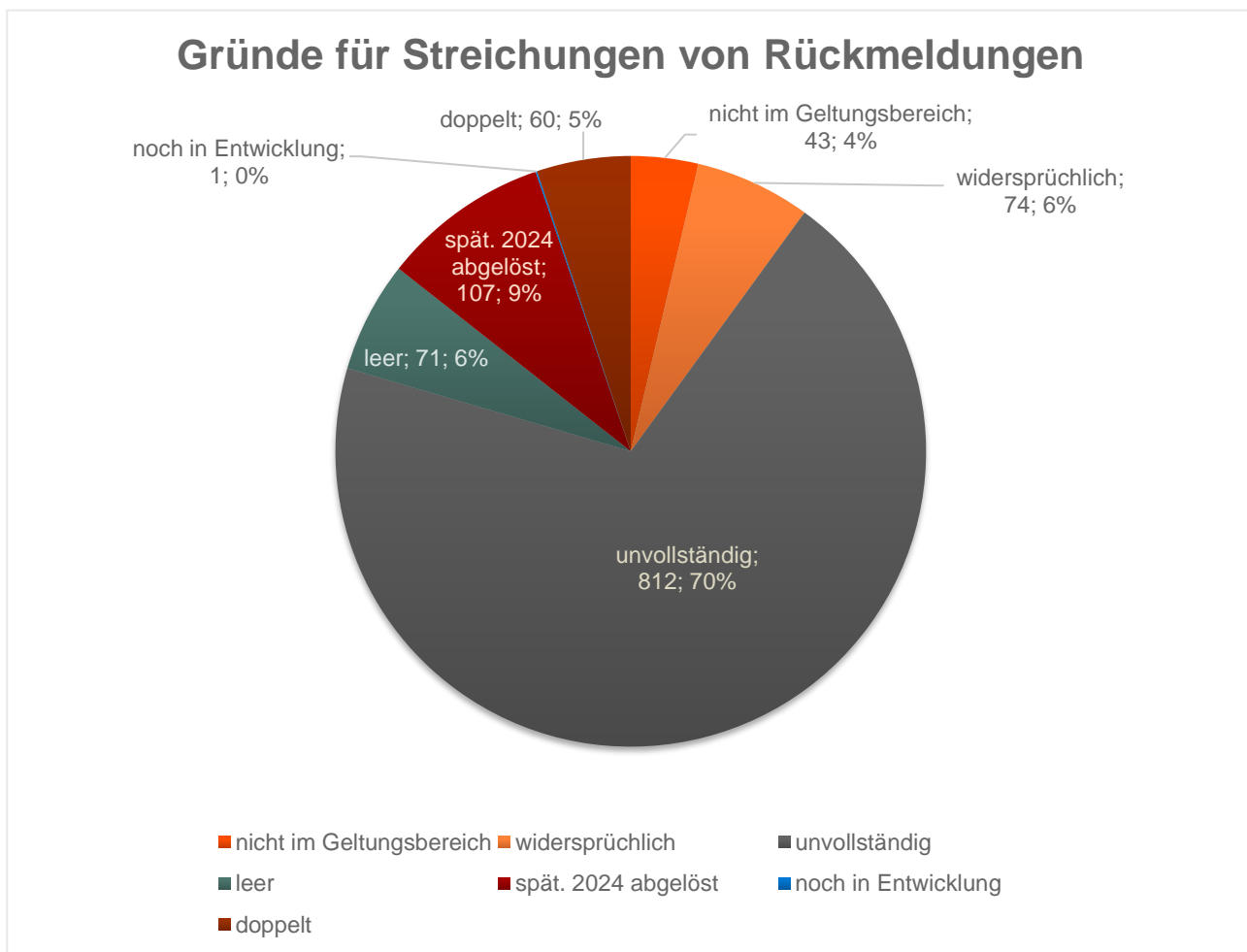


Abbildung 1: Gründe für Streichungen von Rückmeldungen

Die Excel-Vorlage sollte eigentlich unterbinden, dass bei Wertelisten Angaben außerhalb eines vorgegebenen Wertebereichs vorgenommen werden. Es ist trotzdem in vielen Fällen vorgekommen – vor allem deshalb, weil für 434 IT-Verfahren die Vorlage der ersten Umfrage zurückgesendet wurde. Speziell bei den Jahresangaben waren einige der gemeldeten Werte nicht mehr vorgesehen (z. B. eine zukünftige Produktivsetzung in den Jahren 2019 und 2020). Wenn sich die Angaben einem der vorgesehenen Werte eindeutig zuordnen ließen, wurden sie korrigiert, um die Auswertung nicht zu verzerren. Die falschen Jahresangaben wurden in der Auswertung separat betrachtet (nur weil das Datum der Produktivsetzung in der Vergangenheit liegt, ist nicht sichergestellt, dass die Konformität als bereits erreicht gelten kann). Änderungen an den ursprünglich gemeldeten Daten wurden intern markiert.

Insgesamt sind somit die Angaben über 1.174 IT-Verfahren der öffentlichen Verwaltung in Deutschland statistisch ausgewertet worden.

3.2 Umfang der Rückmeldungen

Insgesamt haben elf Bundesministerien und andere oberste Bundesbehörden (von 17), 14 Länder (von 16) und drei Landeshauptstädte (von 13) Berichte zur Umsetzung der DIN SPEC 91379 abgeliefert (fünf Länder mehr als vor zwei Jahren, die gleiche Anzahl aus Bund und Kommunen, wenn auch nicht dieselben Absender). Allerdings wurde oft darauf hingewiesen, dass in der zur Verfügung stehenden Zeit nicht alle nachgeordneten Behörden geantwortet haben und auch bei vorliegenden Antworten nicht immer alle IT-Verfahren untersucht werden konnten. Die Verlängerung der Antwortfrist im Vergleich zur vorherigen Umfrage um zwei Monate hat nur wenig Wirkung gezeigt.

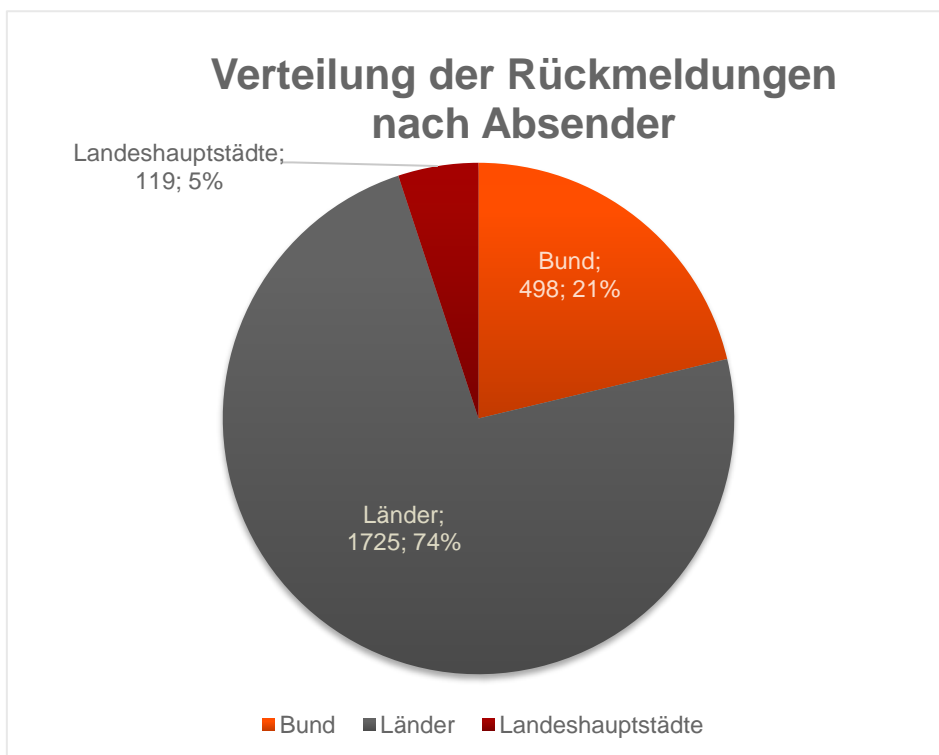


Abbildung 2: Verteilung der Rückmeldungen zwischen Bund, Länder und Kommunen

Im Vergleich zur Umfrage vor zwei Jahren sind aus allen Ebenen mehr IT-Verfahren gemeldet worden (auch wenn man 770 telefonisch gemeldete IT-Verfahren eines Bundeslandes abziehen würde, die wegen unvollständiger Angaben nicht ausgewertet wurden).

Nur von zwei Ländern gab es keine offizielle Reaktion:

- Bremen
- Nordrhein-Westfalen.

Jedoch haben fünf weitere Länder nur wenige IT-Verfahren (weniger als 30) zurückgemeldet, siehe Abbildung 3.

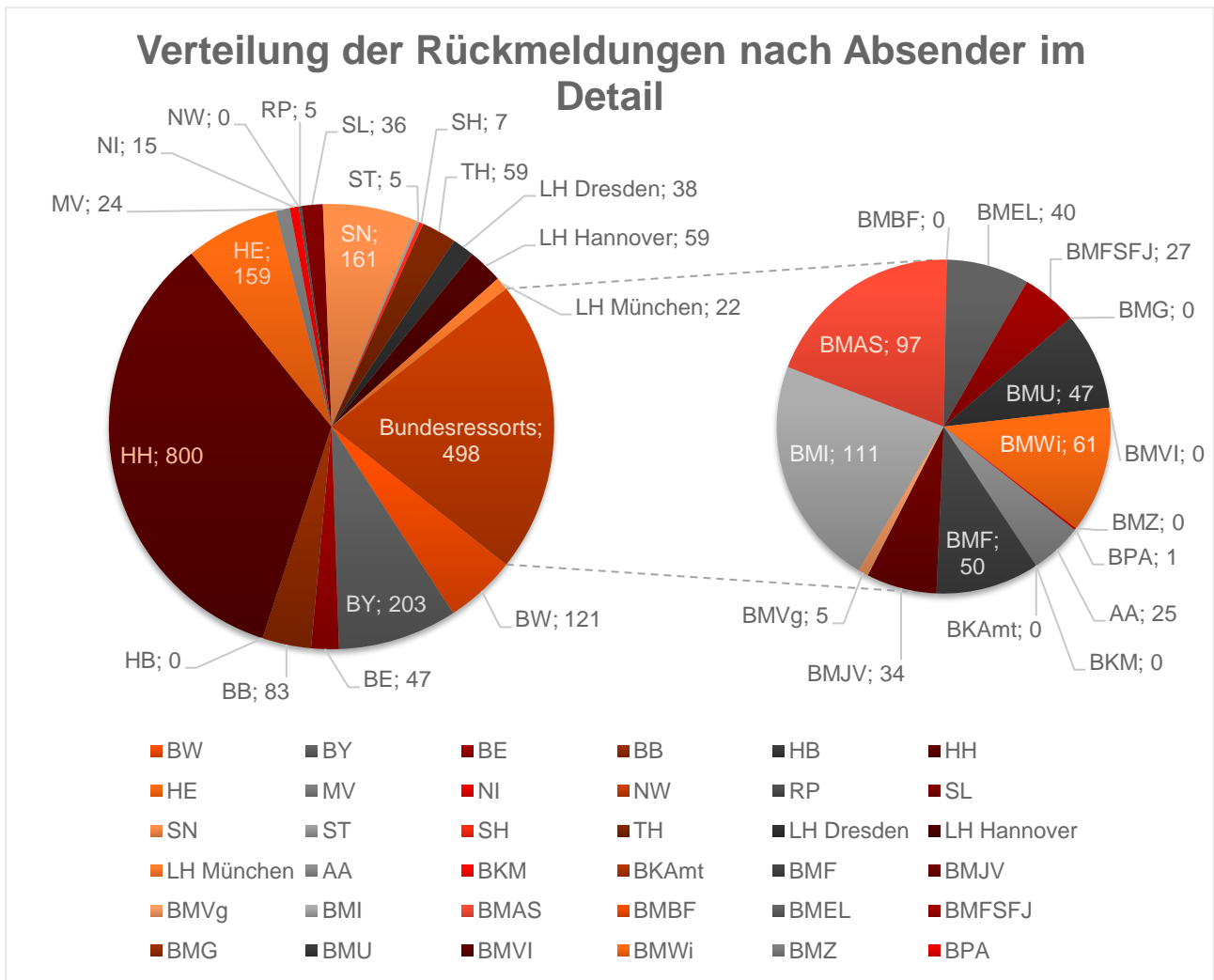


Abbildung 3: Verteilung der Rückmeldungen zwischen obersten Bundesbehörden, einzelnen Ländern und Kommunen
Auch von zehn Landeshauptstädten als Vertreter der Kommunen gab es bis zum 10. Juni 2022 keine gemeldeten IT-Verfahren:

- Potsdam / Brandenburg
- Stuttgart / Baden-Württemberg
- Wiesbaden / Hessen
- Schwerin / Mecklenburg-Vorpommern
- Düsseldorf / Nordrhein-Westfalen
- Mainz / Rheinland-Pfalz
- Saarbrücken / Saarland

- Magdeburg / Sachsen-Anhalt
- Kiel / Schleswig-Holstein
- Erfurt / Thüringen

Landeshauptstädte sollten über die Bundesländer angesprochen werden. Die hohe Zahl fehlender Rückmeldungen aus Kommunen lässt vermuten, dass in den Ländern nicht immer die richtigen Ansprechpartner bekannt sind. Nur aus Mainz gab es eine Rückmeldung, dass die Kapazitäten zur Teilnahme an der Umfrage aktuell nicht ausreichen.

3.3 Metadaten



Abbildung 4: Verteilung der gemeldeten IT-Verfahren nach Themenbereich

Die Angaben zum Themenbereich der IT-Verfahren dienen in erster Linie dem besseren Verständnis des Zwecks der Anwendungen. Es muss damit gerechnet werden, dass nicht alle Verfahren korrekt zugeordnet wurden. In überprüften (und korrigierten) Stichproben ist beispielsweise aufgefallen, dass viele polizeiliche IT-Verfahren nicht dem Themenbereich „Recht & Ordnung“ zugeordnet wurden, sondern „Sonstiges“. Die Anzahl nicht ausgefüllter Themenbereiche ist im Vergleich zur Umfrage vor zwei Jahren (noch 11%) vernachlässigbar gering (0%).

3.4 Aktueller Status der Konformität

In den Bereichen Erfassung, Speicherung, Übermittlung, Anzeige und Ausdruck wurde der aktuelle Status der Konformität zur DIN SPEC 91379 abgefragt. Die Abweichungen zwischen den Bereichen sind nicht gravierend. Den höchsten Umsetzungsstand hat wieder die Speicherung mit 59% (vor zwei Jahren 40%) erreicht und den niedrigsten Wert erreicht weiterhin der Ausdruck mit immerhin 41% (vor zwei Jahren 26%). Aufgrund der breiten Unterstützung von Unicode und UTF-8 durch aktuelle Datenbankprodukte hätte bei der Speicherung ein noch höherer Wert erwartet werden können. Vermutlich wird von den Fähigkeiten noch nicht überall Gebrauch gemacht, so lange nicht alle genutzten Software-Komponenten in den betrachteten IT-Verfahren Unicode-fähig sind.

Die hohen Steigerungen der prozentualen Anteile im Vergleich zur vorherigen Umfrage sind auch darauf zurückzuführen, dass konsequenter unvollständige Rückmeldungen gestrichen wurden. Um in die Auswertung zu kommen, musste zum IT-Verfahren ausgefüllt worden sein: entweder der Zeitpunkt der vollständigen Produktivsetzung der Unterstützung der DIN SPEC 91379 (String.Latin+ 1.2) insgesamt oder mindestens vier der fünf Spalten aus diesem Bereich „Aktueller Status der Konformität zur DIN SPEC 91379“. Das führte dazu, dass es keine IT-Verfahren mehr in der Auswertung gibt, die in diesem Bereich auch nur teilweise nicht ausgefüllt wurden. Vor zwei Jahren machten nicht ausgefüllte Werte 13 bis 14% aller ausgewerteten IT-Verfahren aus.

Dass der Status als unbekannt gemeldet wurde, kam auf 14 bis 17% der Fälle. Vor zwei Jahren waren es jedoch noch 20 bis 24% (trotz der höheren Werte für „nicht ausgefüllt“).

Die Anteile der IT-Verfahren, deren Konformität zur DIN SPEC 91379 verneint oder als nicht erforderlich eingestuft wird, sind im Vergleich zu den Ergebnissen vor zwei Jahren relativ gleichgeblieben. Wenn einer der fünf Bereiche nicht erforderlich war (die Werte schwanken zwischen drei Prozent für die Speicherung und zehn Prozent für den Ausdruck), wurde dies meist nachvollziehbar fachlich begründet. Gab es Begründungen, die nicht plausibel waren, führte das meistens zu Korrekturen der Werte „nicht erforderlich“. War die Konformität zur DIN SPEC 91379 für das gesamte IT-Verfahren nicht erforderlich, wurde das IT-Verfahren als „nicht im Geltungsbereich“ aus der Statistik gestrichen.

Für ursprünglich 128 IT-Verfahren wurde einerseits angegeben, dass in allen fünf Bereichen die Konformität zur DIN SPEC 91379 bereits erreicht sei, und andererseits stand in der Spalte „Jahr der vollständigen Produktivsetzung der Unterstützung der DIN SPEC 91379 (String.Latin+ 1.2)¹“ nicht etwa „bereits erreicht“, sondern Daten in der Zukunft, „noch nicht geplant“, „unbekannt“ oder auch „nicht vorgesehen“. Das kann nicht beides richtig sein. Wenn ergänzende Angaben in den Freitexten den Sachverhalt aufgeklärt haben, wurden die entsprechenden Werte korrigiert. Schließlich wurden die Angaben zu 74 IT-Verfahren als „widersprüchlich“ aus der Auswertung gestrichen.

In den Freitexten erkennbare Ursachen für Widersprüche waren zum Beispiel:

¹ Siehe Abschnitt 3.5 „Planung / Vorgehen zum Erreichen der Konformität“

- Missverständnis, dass man die nicht-normativen Datentypen implementieren müsse, also die zulässigen Zeichen auf den Zeichenvorrat der DIN SPEC 91379 einschränken, um konform zu sein (weil man alle Zeichen unterstützt, wurde für die fünf einzelnen Bereiche trotzdem jeweils „ja“ ausgewählt)
- Irrtum, dass man für die fünf Bereiche mit „ja“ antworten dürfe, wenn die unterstützten Zeichen aktuell noch auf String.Latin 1.1 eingeschränkt werden (dann fehlen aber viele zusätzliche Zeichen der DIN SPEC 91379)

Es muss vermutet werden, dass für viele IT-Verfahren für die fünf Bereiche Erfassung, Speicherung, Übermittlung, Anzeige und Ausdruck deshalb fünfmal die Konformität zur DIN SPEC 91379 mit „ja“ bestätigt wurde, weil das IT-Verfahren Unicode/UTF-8 unterstützt (einige Freitexte unterstützen diese Vermutung). Dabei wird unterschätzt, dass Probleme mit Multi-Codepoint-Zeichen in allen Bereichen auftreten können, auch wenn UTF-8 unterstützt und die Anzahl der Unicode-Zeichen nicht aktiv eingeschränkt wird. Nur wenige Schriftarten unterstützen korrekt die Anzeige und den Ausdruck aller für die Konformität relevanten Zeichen. Allein aus der Implementation von UTF-8 heraus wurde für die Nutzer noch keine Möglichkeit geschaffen, spezielle Zeichen auch eingeben zu können. Für Schnittstellen zu externen Systemen ist nicht die theoretische Fähigkeit² zur Übertragung aller Zeichen ausschlaggebend, sondern dass wirklich bei fachlichem Bedarf³ alle normativen Zeichen der DIN SPEC 91379 erfolgreich empfangen und gesendet werden können. Nach der Eingabe und dem Empfang von Unicode-Zeichensequenzen müssen diese mit der Normalform C (NFC) normalisiert werden, sonst entstehen Interoperabilitätsprobleme. Das gilt beispielsweise auch für Suchanfragen.

Viele IT-Verfahren würden einer kritischen Überprüfung, ob wirklich alle normativen Zeichen der DIN SPEC 91379 erfolgreich erfasst, gespeichert, übermittelt, angezeigt und ausgedruckt werden können, vermutlich nicht standhalten. Aber wenn lückenlos UTF-8 unterstützt und die verwendbaren Zeichen nicht durch technische Barrieren eingeschränkt werden, lassen sich die meisten Namen im weiteren Sinne korrekt verarbeiten (grobe Schätzung: 99%). Die öffentliche Verwaltung Deutschlands ist jedoch durch rechtliche Vorgaben dazu verpflichtet, 100% der lateinischen Namen natürlicher Personen korrekt zu verarbeiten⁴. Eine Konformität zur DIN SPEC 91379 stellt das sicher.

² Wenn eine Schnittstelle UTF-8 unterstützt, kann sie theoretisch alle Unicode-Zeichen übertragen. Da man oft z. B. Steuerzeichen, Emojis, chinesische und klingonische Schriftzeichen ausschließen möchte, wird dann der Zeichensatz für Datentypen der Textfelder in der Schnittstellenspezifikation eingeschränkt. Ist die Einschränkung zu stark (z. B. „A..Z“, „a..z“, Ziffern und Satzzeichen), können nicht alle Unicode-Zeichen der DIN SPEC übertragen werden.

³ Es besteht beispielsweise kein fachlicher Bedarf alle normativen Zeichen der DIN SPEC 91379 in einem Datenfeld zu übertragen, wenn das Datenfeld gar keine Texte enthält, wenn die möglichen Werte des Datenfeldes fest vorgegeben sind oder auch dann, wenn ausschließlich Namen natürlicher Personen übertragen werden sollen. Es gibt normative Zeichen in der DIN SPEC, die für solche Namen ungeeignet sind, konkret die Schriftzeichengruppen der Nicht-Buchstaben N2, N3 und N4, z. B. Währungssymbole, Bruchzahlen und der Tabulator.

⁴ Zu den rechtlichen Vorgaben siehe DIN SPEC 91379, Kapitel „Einleitung“, Abschnitt „Ausgangslage“

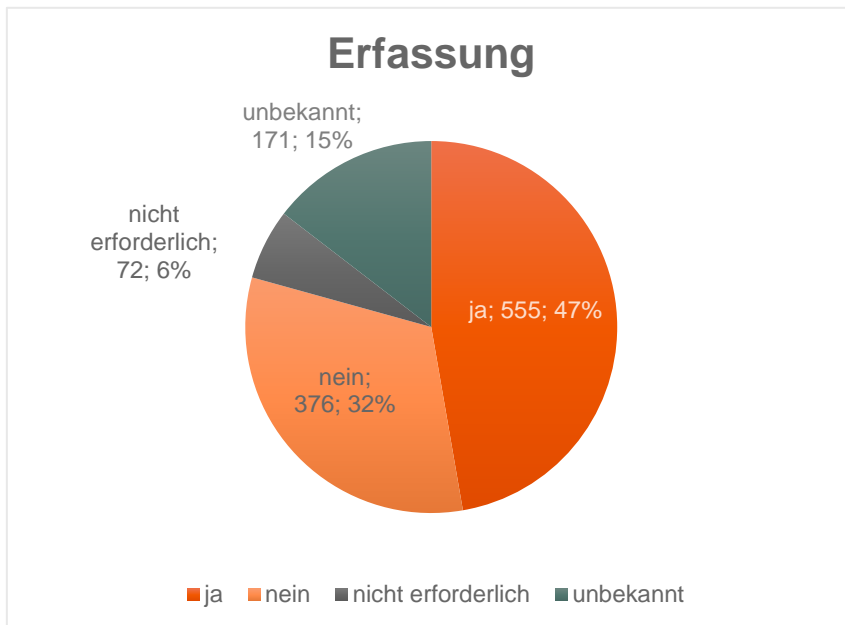


Abbildung 5: Aktueller Status der Konformität zur DIN SPEC 91379 im Bereich Erfassung

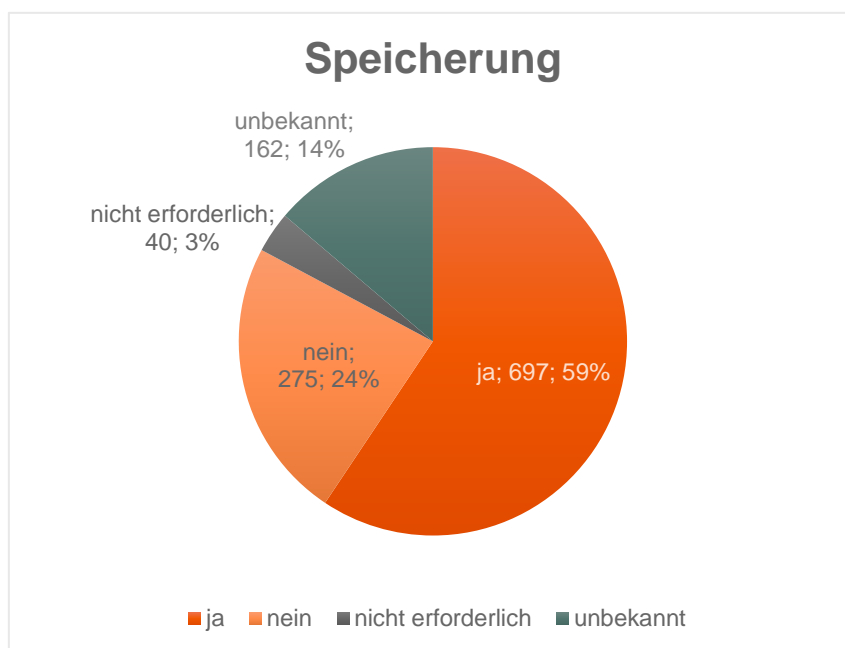


Abbildung 6: Aktueller Status der Konformität zur DIN SPEC 91379 im Bereich Speicherung

Zur Anforderung an IT-Verfahren, die Unicode-Zeichen der DIN SPEC 91379 speichern zu können, wurde u.a. zurückgemeldet, dass dies für viele Bestandssysteme nicht möglich sei. Diese laufen teilweise auf älteren Plattformen, die hardware-bedingt nicht oder nur mit einem unverhältnismäßig hohen Aufwand auf ein UTF-8-Encoding umgestellt werden können. Hier sei der Hinweis gestattet, dass Unicode-Zeichen eindeutig in ASCII-Zeichen hin- und wieder zurückkonvertiert werden können. Solange die Software Unicode-fähig ist, lässt sich das Speichern in Datenbanken mit 8-Bit- oder sogar 7-Bit-Zeichensätzen über beispielsweise „uencode“ realisieren, siehe Demonstration unter <https://www.dcode.fr/uu-encoding>.

Um Suchfunktionen in solchen Datenbanken zu ermöglichen, muss eine standardisierte Transformation von Unicode in den verwendeten Legacy-Zeichensatz angewendet werden. Für die gespeicherten Daten muss

eine Spalte mit dem Transformationsergebnis gespeichert werden und Suchanfragen müssen ebenfalls transformiert werden. Aufgrund des unausweichlichen Informationsverlustes solcher Transformationen werden möglicherweise mehr Treffer zurückgeliefert, als erwartet (eine Suche nach „Møller“ findet dann auch „Möller“). Damit muss die Software aber auch für übliche Suchanfragen umgehen können. Der digitale Anhang „2018-11-15_legacy-mappings.xlsx“ zur DIN SPEC 91379 enthält Vorschläge für Transformationen des Zeichenvorrats des Standards in bekannte Zeichensätze mit begrenztem Zeichenvorrat.

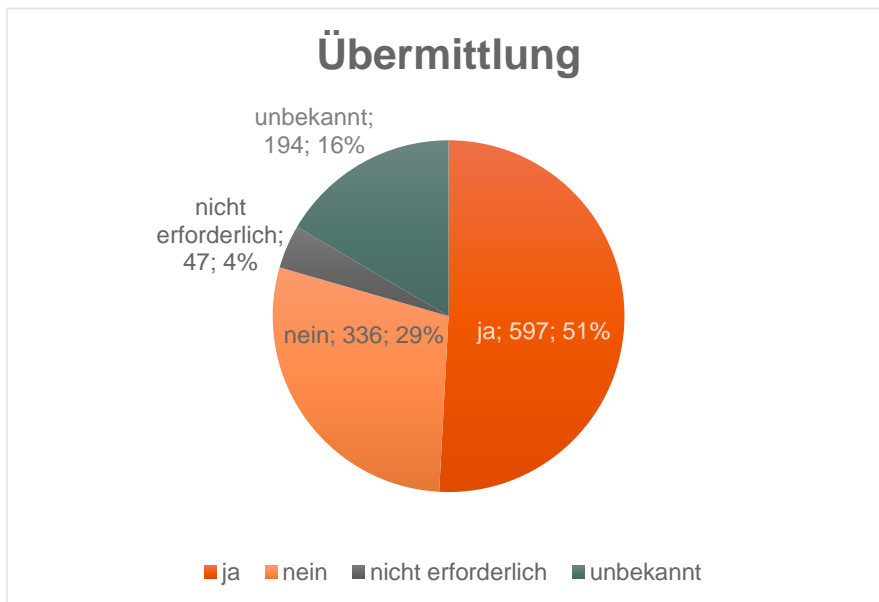


Abbildung 7: Aktueller Status der Konformität zur DIN SPEC 91379 im Bereich Übermittlung

In den Freitexten wurde darauf hingewiesen, dass die Koordination einer gleichzeitigen Umstellung aller Kommunikationspartner einen sehr hohen Aufwand erfordert. Auch Übergangslösungen, um zu überbrücken, dass einzelne Datenlieferanten und vor allem Datenempfänger noch nicht auf Unicode und die DIN SPEC 91379 umgestellt haben, erzeugen erhebliche Aufwände.

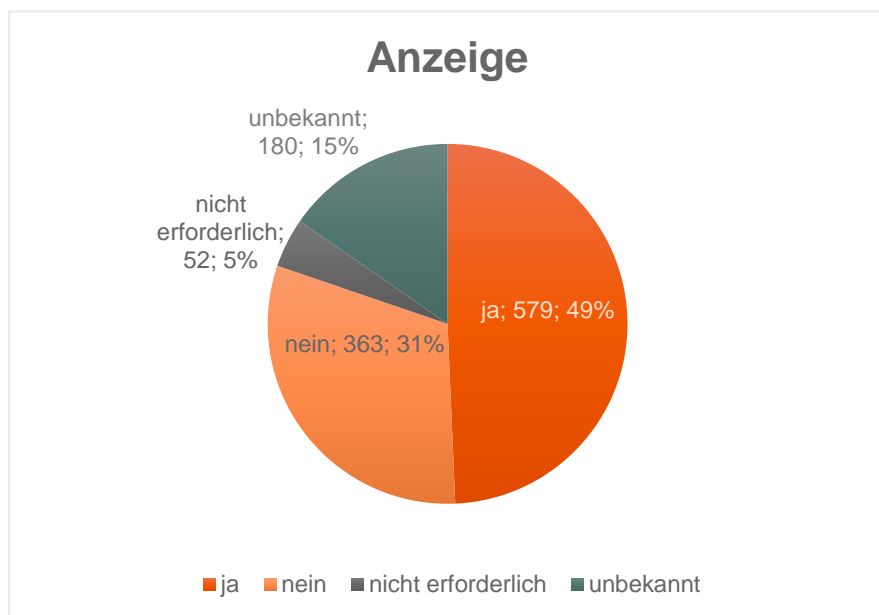


Abbildung 8: Aktueller Status der Konformität zur DIN SPEC 91379 im Bereich Anzeige

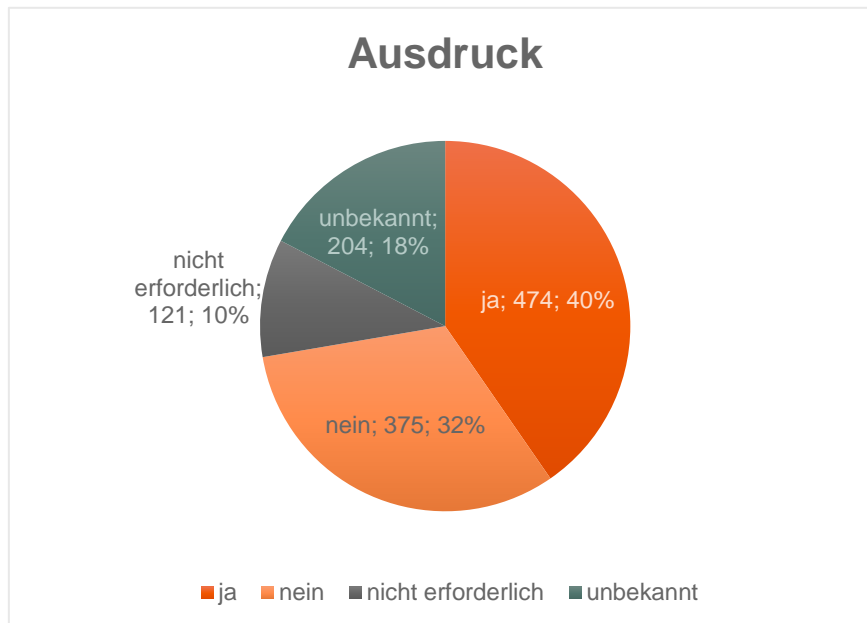


Abbildung 9: Aktueller Status der Konformität zur DIN SPEC 91379 im Bereich Ausdruck

Der Wert für den Ausdruck ist viel höher als man angesichts der nur wenigen Schriftarten, die wirklich alle Zeichen der DIN SPEC 91379 unterstützen, erwarten würde. Da die oft problematischen Zeichen mit mehreren Codepoints in der Praxis nicht häufig vorkommen, darf vermutet werden, dass nicht allen Verantwortlichen bewusst ist, dass allein die Nutzung von UTF-8 nicht korrekte Anzeige und Ausdruck der Zeichen sicherstellt.

3.5 Planung / Vorgehen zum Erreichen der Konformität

Die Verantwortlichen wurden gefragt, in welchem Jahr eine vollständige Produktivsetzung der Unterstützung von UTF-8 (also allgemein Unicode-Unterstützung mit oder ohne Einschränkung der Schriftzeichen), von String.Latin 1.1 und von der DIN SPEC 91379 (String.Latin+ 1.2) im gesamten IT-Verfahren erfolgen wird oder ob dies bereits erreicht wurde. In den Ausfüllhinweisen wurde deutlich gemacht, dass die Unterstützung von UTF-8 eine notwendige Bedingung ist, um auch String.Latin 1.1 vollständig unterstützen zu können. Aufgrund der Abwärtskompatibilität ist die Unterstützung von String.Latin 1.1 eine notwendige Bedingung, um auch String.Latin+ 1.2 (die DIN SPEC 91379) unterstützen zu können. Umgekehrt bedeuten eine Konformität zur DIN SPEC 91379 zwangsläufig die vollständige Implementation von String.Latin 1.1 und eine korrekte Unterstützung von UTF-8.

Vor zwei Jahren sind zu diesem Teil der Umfrage für 25 bis 27% der IT-Verfahren keine Angaben gemacht worden. Dieser Wert ist drastisch auf zwei bis acht Prozent gefallen. Die Ursache liegt aber sehr stark darin, dass bei der aktuellen Auswertung konsequenter IT-Verfahren mit unvollständigen Angaben aus der Statistik entfernt wurden.

Trotzdem darf festgestellt werden, dass die Anteile der zu den abgefragten Standards bereits konformen IT-Verfahren deutlich gestiegen sind. Auch wenn es mutmaßlich zu großen Teilen andere IT-Verfahren sind als vor zwei Jahren, sind die Anteile ohne Plan zum Erreichen der Konformität ähnlich hoch geblieben.

Das Einreichen von alten Fragebögen für die aktuelle Umfrage hat dazu geführt, dass häufig unplausible Angaben gemacht wurden. Anstatt „bereits erreicht“ eine Jahreszahl in der Zukunft zu nennen, ist nur für Jahre sinnvoll, die nicht in der Vergangenheit liegen. Zwei bis drei Prozent der Rückmeldungen aus dem Jahr 2022 haben die Jahre 2019, 2020 oder 2021 als geplante Termine für das Erreichen der Konformität zu UTF-8,

String.Latin 1.1 und/oder DIN SPEC 91379 angegeben. Mit einer gewissen Wahrscheinlichkeit haben diese IT-Verfahren die Konformität bereits erreicht, aber aus Gründen der Transparenz werden diese irregulären Werte separat ausgewiesen.

Für den unverändert geringen Teil der nicht für eine Konformität vorgesehenen IT-Verfahren wurden meistens plausible Begründungen im Freitext geliefert.

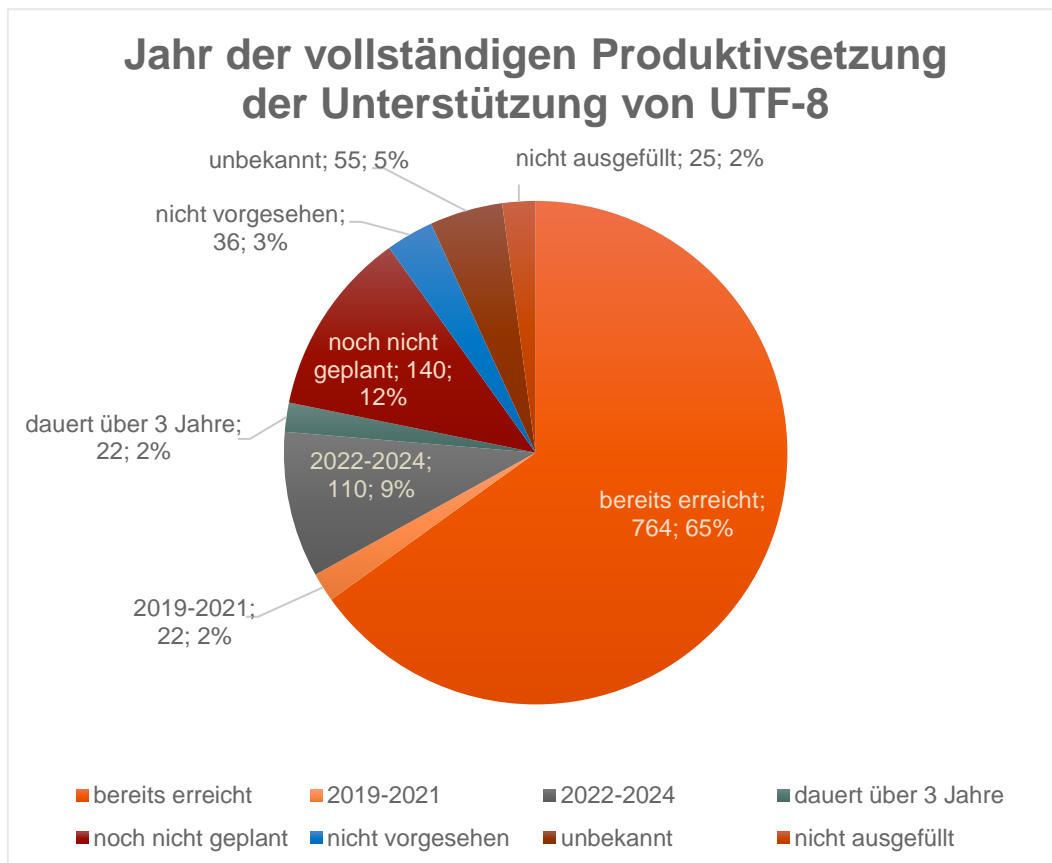


Abbildung 10: Geplante oder bereits erreichte Unterstützung von UTF-8

Allein die Nutzung von UTF-8 für die Kodierung von Zeichen stellt nicht sicher, dass alle normativen Zeichen und Zeichensequenzen der DIN SPEC 91379 korrekt erfasst, gespeichert, übermittelt, angezeigt und ausgedruckt werden können. Als notwendige Voraussetzung für eine Konformität zur DIN SPEC 91379 ist der hohe Grad der UTF-8-Konformität von mittlerweile 65% sehr zu begrüßen. Die Umstellung von 8-Bit Legacy Codepages, wie z. B. ISO-8859-1 oder CP1252, auf UTF-8 und Unicode ist ein sehr aufwändiger Teil auf dem Weg zur Konformität zur DIN SPEC 91379.

Für eine korrekte Verarbeitung der in den Standards String.Latin 1.1 und String.Latin+ 1.2 beschriebenen Schriftzeichen ist über die durchgängige Verwendung von UTF-8 hinaus erforderlich:

- Eingabemöglichkeit aller fachlich sinnvollen Zeichen mit einer deutschen Tastatur (die nicht alle Grundbuchstaben und diakritischen Zeichen enthält, erst recht nicht die diversen Nicht-Buchstaben)
- Normalisierung von Zeichensequenzen mit dem „Normalization Format C“ – NFC (es ist die übliche Normalform, Abweichungen führen aber zu Interoperabilitätsproblemen)
- korrekter Umgang mit Multi-Codepoint-Zeichen in Textoperationen, wie Suche und Sortierung
- Schnittstellen, deren Definition keine normativen Schriftzeichen ausschließt und deren Implementation keine Schriftzeichen konvertiert

- Schriftarten, die für Anzeige und Ausdruck alle Schriftzeichen des Standards korrekt darstellen
- Ggf. an das IT-Verfahren angeschlossene Hard- und Software-Komponenten müssen ebenfalls mit allen Zeichen korrekt umgehen, z. B. Lösungen zur automatischen Texterkennung (OCR) und Drucker.
- Wenn IT-Verfahren barrierefrei sein müssen (das gilt für die meisten), stellt eine Umstellung des Zeichensatzes auf viele neue Schriftzeichen eine besondere Herausforderung für den barrierefreien Zugang zu den Daten dar.

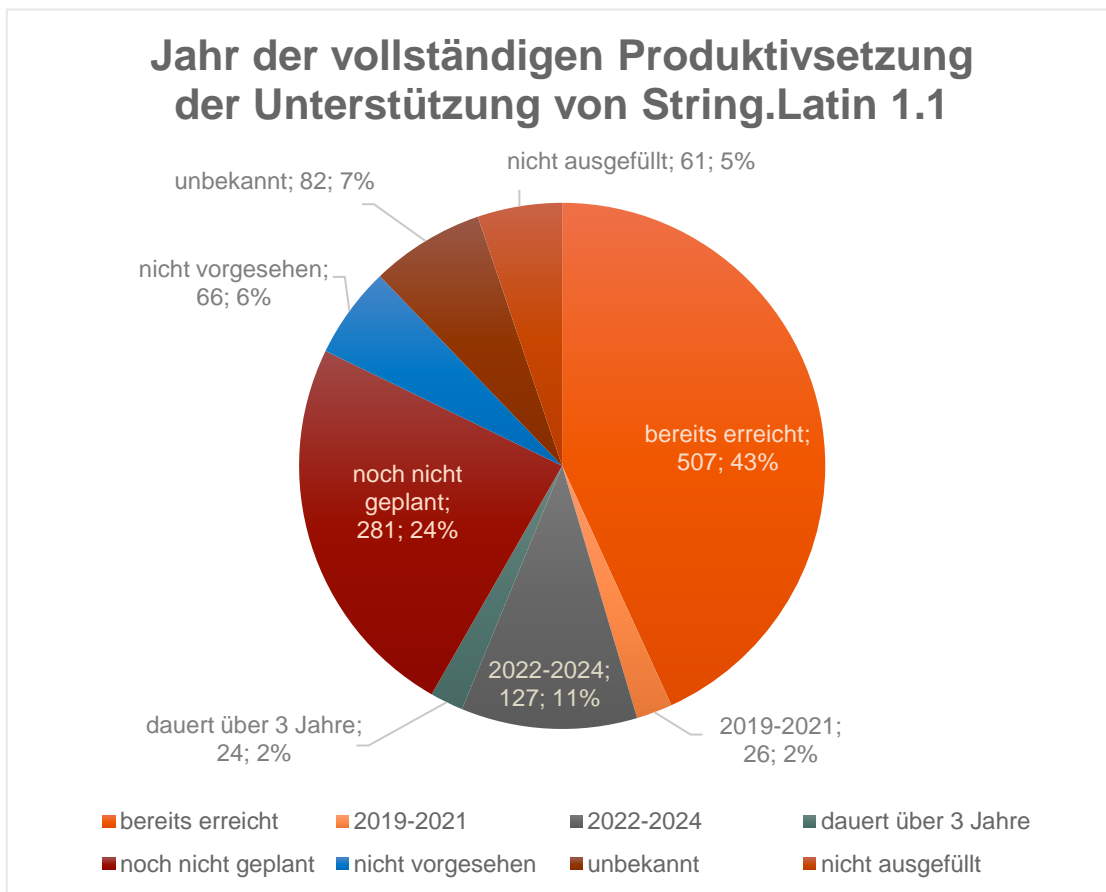


Abbildung 11: Geplante oder bereits erreichte Unterstützung von String.Latin 1.1

Der Anteil der zum Vorgänger der DIN SPEC 91379 (String.Latin+ 1.2) konformen IT-Verfahren ist in den letzten zwei Jahren von 19 auf 43% gestiegen.

Es sind drei IT-Verfahren verblieben, die zwar eine vollständige Unterstützung der DIN SPEC 91379 bereits als erreicht einordnen, bei der Produktivsetzung von String.Latin 1.1 jedoch „nicht vorgesehen“ oder „unbekannt“ angeben. Das ist eigentlich unlogisch, weil mit der Unterstützung der DIN SPEC 91379 auch eine Unterstützung von String.Latin 1.1 einhergeht. Noch deutlichere Widersprüche durch die Angabe von „nein“ oder Plandaten in der Zukunft für String.Latin 1.1 wurden entweder mithilfe von Informationen aus Freitexten korrigiert oder sie wurden als widersprüchlich gestrichen.

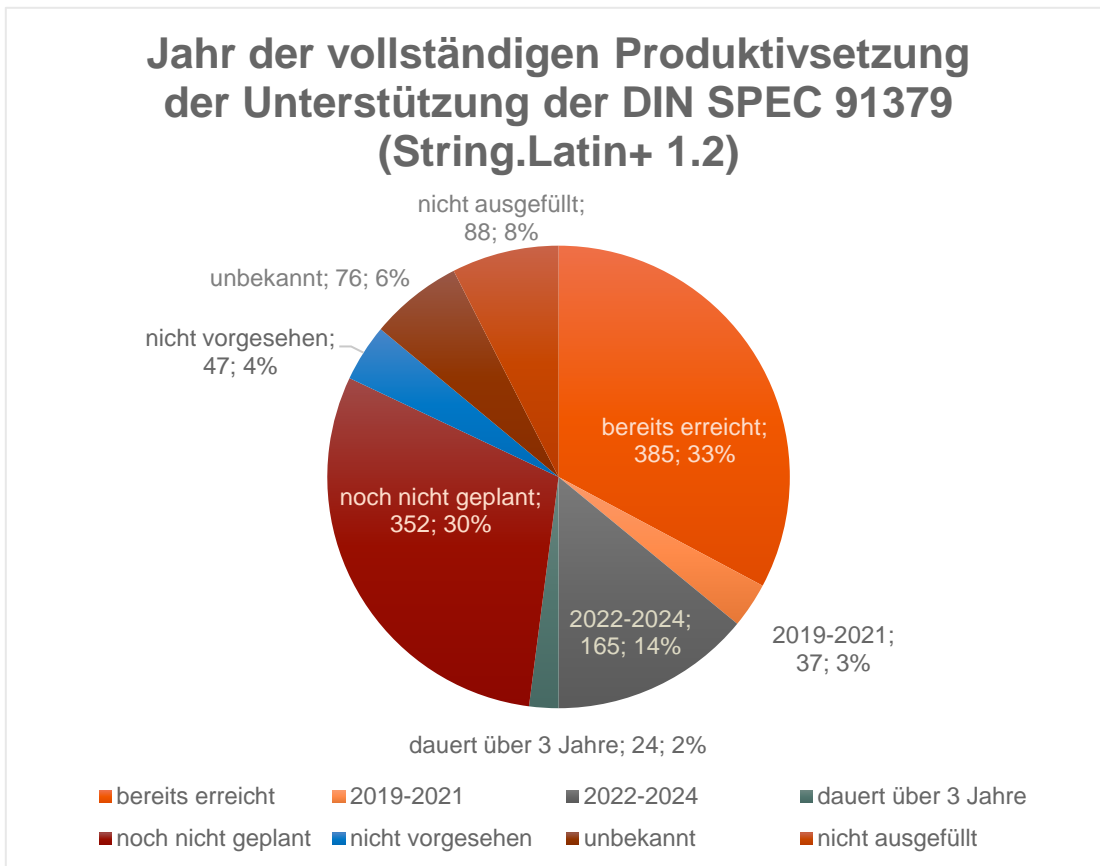


Abbildung 12: Geplante oder bereits erreichte Unterstützung von DIN SPEC 91379 (String.Latin+ 1.2)

Wenn für die Planung der Konformität zur DIN SPEC 91379 „nicht vorgesehen“ eingetragen wurde, gibt es häufiger den Fall, dass die Anwendung Unicode-fähig ist, aber im Grunde noch unklar ist, ob die Eingabe und der Ausdruck aller Zeichen wirklich gelingen würden. In diesem Fall sollte zuerst ermittelt werden, ob wirklich alle Zeichen (und Zeichensequenzen) der DIN SPEC unterstützt werden. Ist dies nicht der Fall, besteht Handlungsbedarf – auch wenn theoretisch alle Unicode-Zeichen unterstützt werden.

Eine häufig gegebene Erklärung dafür, dass eine Umsetzung der DIN SPEC 91379 nicht geplant wird, ist die baldige Ablösung / Einstellung des Systems. Wo dies im geänderten Formular der zweiten Umfrage angegeben wurde oder Freitexte eine entsprechende Information enthielten, wurden die IT-Verfahren als „spät. 2024 abgelöst“ aus der Statistik entfernt. Unklar bleibt jedoch, ob das ablösende Verfahren eine Unterstützung der DIN SPEC 91379 vorsieht.

Für 44% der ausgewerteten IT-Verfahren wurde die vollständige Produktivsetzung der DIN SPEC 91379 noch nicht geplant, ihr Termin ist unbekannt oder es wurden keine Angaben gemacht. Zur Umsetzung der Beschlüsse des IT-PLR besteht hier dringender Handlungsbedarf. Vor zwei Jahren waren es zwar noch 53%, aber die Vermutung liegt nahe, dass nicht gemeldete IT-Verfahren und IT-Verfahren mit unvollständigen oder widersprüchlichen Angaben ebenfalls zu den Problemfällen zu rechnen sind.

Dass der Anteil nicht ausgefüllter Werte seit der ersten Umfrage stark zurückgegangen ist, hat auch damit zu tun, dass unvollständige Zeilen aus der Wertung genommen wurden.

Der Anteil der bereits erreichten Konformität zur DIN SPEC 91379 ist in den letzten zwei Jahren erfreulich von 13 auf 33% angestiegen. Zusammen mit bereits geplanter Erreichung der Konformität werden allerdings nur 50% aller ausgewerteten IT-Verfahren erfasst. Angesichts einer Umsetzungsfrist bis November 2024 ist dieser Wert noch viel zu niedrig.

3.6 Risiken beim weiteren Vorgehen zum Erreichen der Konformität

Die Anzahl der Angaben, dass Risiken unbekannt sind oder dass Felder mit Risiken nicht ausgefüllt wurden, ist im Vergleich zur Umfrage vor zwei Jahren zurückgegangen. Bei der Abhängigkeit von Produktherstellern ging der ursprünglich sehr hohe Zustimmungswert etwas zurück (von 38% auf 30% gesunken). Die anderen Risiken wurden zwei Jahre später häufiger als relevant eingeschätzt. Die Abhängigkeit von Schnittstellenpartnern wurde nun am häufigsten bejaht (von 24% auf 32% gestiegen). Es ist zu vermuten, dass in den letzten zwei Jahren Gespräche mit Herstellern und Schnittstellenpartnern zur DIN SPEC 91379 geführt wurden. Auch aus den Freitexten ist oft zu erkennen, dass die Software-Hersteller zunehmend eine Umstellung auf UTF-8 und einen stark erweiterten Zeichenvorrat vorbereitet haben, aber Schnittstellenvereinbarungen (direkte mit Partnern oder indirekte über Standards), fehlende Unterstützung durch die Hausschrift und Bedenken zur Barrierefreiheit verhindern, die Unterstützung der DIN SPEC in den produktiven Betrieb gehen zu lassen. In sehr vielen Fällen wird keine Möglichkeit gesehen, Einfluss auf den Software-Hersteller zu nehmen.

Aus den Freitexten zur Erläuterung der identifizierten Risiken lassen sich viele Missverständnisse erkennen. Es ist nicht überall bekannt, dass nur die normativen Zeichen der DIN SPEC 91379 unterstützt werden müssen und dass eine Verwendung zusätzlicher Zeichen die Konformität nicht verletzt. Es müssen nicht die vom Standard lediglich informativ zur Verfügung gestellten Datentypen zur Einschränkung der Zeichen zum Einsatz kommen. Es müssen auch nicht griechische und kyrillische Zeichen zugelassen werden, da sie nicht normativ sind.

Auch fehlende Grundkenntnisse kommen zum Vorschein, wenn bei einer Unterstützung von ISO Latin-1 (=ISO 8859-1) angenommen wird, damit Unicode zu unterstützen und somit konform zur DIN SPEC 91379 zu sein. ISO Latin-1 scheint zwar eine Teilmenge der DIN SPEC zu sein – tatsächlich fehlt ein Zeichen, der bedingte Trennstrich SHY (soft hyphen) – aber die DIN SPEC 91379 ist ein Mindeststandard. Es müssen alle ihre normativen Zeichen unterstützt werden. 8-Bit-Zeichensätze wie ISO Latin-1 reichen nicht aus, um alle normativen Zeichen des Standards verarbeiten zu können.

Notwendige Abstimmungen mit allen Beteiligten (Software-Hersteller, Schnittstellenpartner, Herausgeber von Standards, Betreiber von IT-Verfahren), um eine Umstellung des Zeichenvorrats zu erreichen, werden oft als sehr aufwändig und damit als großes Hemmnis für das Erreichen der Konformität beschrieben.

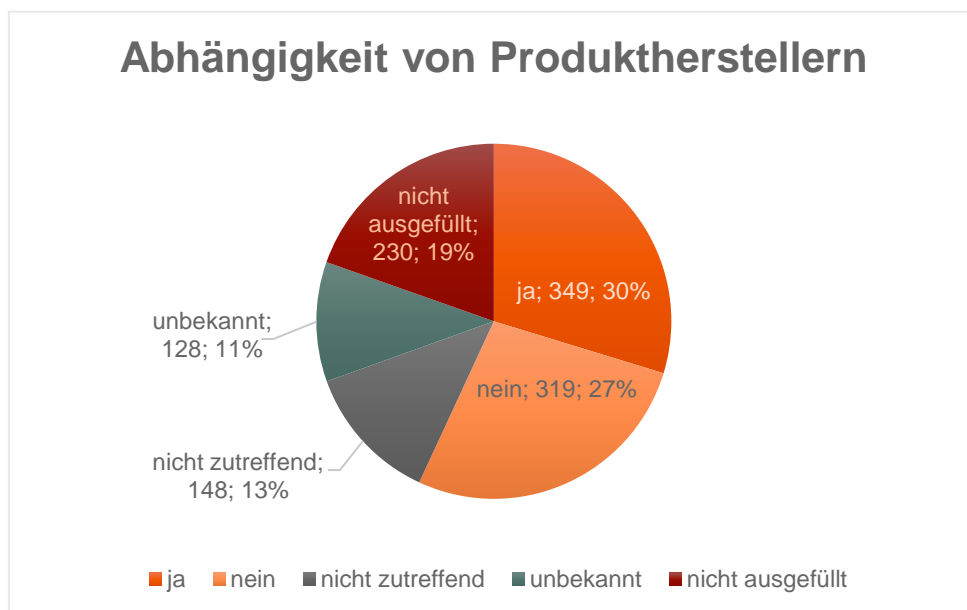


Abbildung 13: Bewertung des Risikos zum Erreichen der Konformität aufgrund Abhängigkeiten von Produktherstellern

Es wurden diverse konkrete Beispiele für Abhängigkeiten von Produktherstellern genannt, u.a. Microsoft, Adobe, SAP, diverse IT-Dienstleister des Bundes, der Länder und Kommunen wie z. B. das ITZBund für GSB, FMS und fachspezifische IT-Verfahren, diverse privatwirtschaftliche IT-Dienstleister wie z. B. die Bundesdruckerei, Länderverbünde wie die LAG GADSYS (Länderarbeitsgruppe Gemeinsame Abfall-DV-Systeme), Liferay, aber auch allgemeine Abhängigkeiten von Herstellern eingekaufter Software. Nicht selten gibt es einen Zeitplan zur rechtzeitigen Umstellung der Software. Die Abhängigkeit vom Hersteller zur Erreichung der Ziele besteht selbstverständlich trotzdem.

Bemängelt wurde, dass weit verbreitete Standardprodukte von beispielsweise SAP und Microsoft die DIN SPEC 91379 aus dem Jahr 2019 nicht explizit unterstützen. Mit einer korrekten Implementation von Unicode und UTF-8 ist es allerdings möglich, diese Produkte konform zum Standard einzusetzen. Allerdings ist die Eingabe der Zeichen oft nicht komfortabel gelöst und es müssen für die Anwendungen geeignete Schriftarten ausgewählt werden.

Auch Datenbanken von Oracle und PostgreSQL wurden als Abhängigkeit genannt, weil nur bestimmte Zeichensätze, wie das „AL32UTF8 character set“ verwendet werden können. Korrekt konfiguriert und mit passender Software angesprochen stellen aktuelle Datenbanken – weder von Oracle noch von anderen Herstellern – kein Hemmnis dar, um Unicode-Zeichen inkl. Multi-Codepoint-Zeichen zu unterstützen. Wo allerdings noch dBase4 eingesetzt wird, dürfte eine Umstellung auf UTF-8 und die DIN SPEC 91379 nicht ohne aufwändigen Austausch der Datenbank möglich sein.

Speziell für eine OCR-Lösung (optische Zeichenerkennung) wird auf die Abhängigkeit vom Hersteller verwiesen, inwieweit der Zeichenvorrat der DIN SPEC 91379 unterstützt wird.

Eine sehr typische und in ihrer Aussage sehr häufige Rückmeldung lautet: „Umstellungen muss der Verfahrenshersteller vornehmen. Wir sind "nur" Anwender.“ Solange die Anwender die Umstellungen nicht verlangen, wird der Verfahrenshersteller nicht aktiv werden. Die Beschlüsse des IT-Planungsrates sind nicht für Software-Hersteller bindend, sondern für die öffentliche Verwaltung, die Software-Systeme zur Erledigung ihrer Aufgaben beschafft.

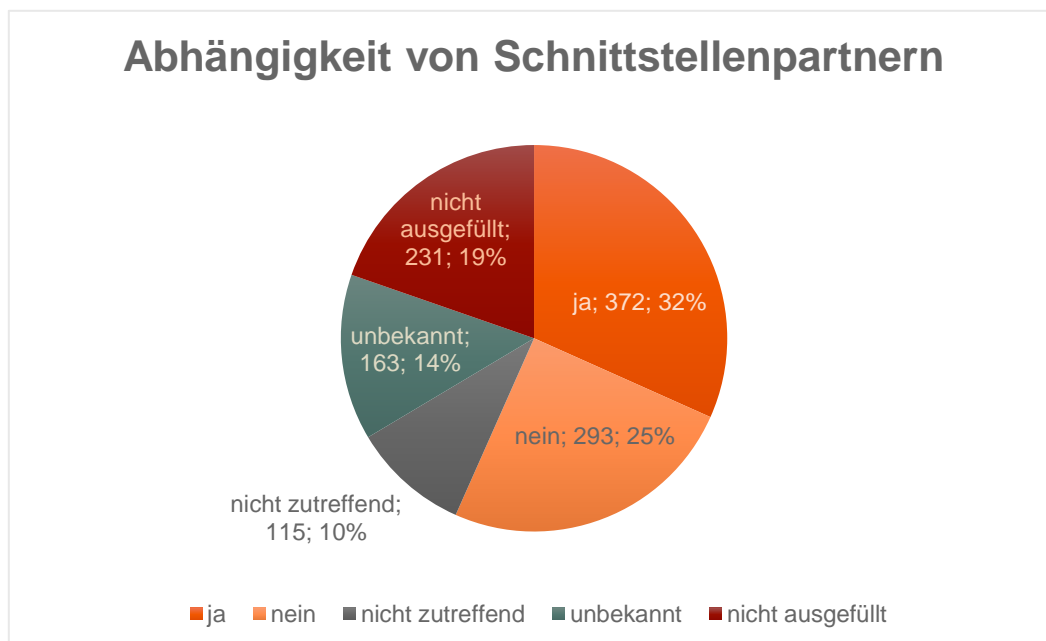


Abbildung 14: Bewertung des Risikos zum Erreichen der Konformität aufgrund Abhängigkeiten von Schnittstellenpartnern

Es werden diverse Beispiele für Abhängigkeiten von Schnittstellenpartnern genannt, u.a. die europäische Kommission (EESSI), KONSENS (RMS-KMV), ELSTER, BZSt-Online-Portal, Landesfinanzverwaltung, Auskunftssysteme in der Kreditwirtschaft, Destatis (diverse Statistiken), BfJ (BZR), BKA (INPOL), KBA (diverse Register), BVA (Visa, NWR), EGVP, Krankenkassen (KV-Meldung), DHL, österreichische und amerikanische Kommunikationspartner.

Zum Umgang mit dem Risiko wurde berichtet, dass bereits bei den Umstellungen auf UTF-8 und String.Latin 1.1 Adapter zu Kommunikationspartnern etabliert wurden. Das sei auch bei der Umstellung auf String.Latin+ 1.2 vorgesehen.

Es wurde auch erwähnt, dass Kommunikationspartner die Unterstützung von griechischen und kyrillischen Zeichen ablehnen. Hierbei ist zu beachten, dass diese Zeichen nicht zum normativen Teil der DIN SPEC 91379 gehören. Sie nicht zu unterstützen, verletzt nicht die Konformität zum Standard.

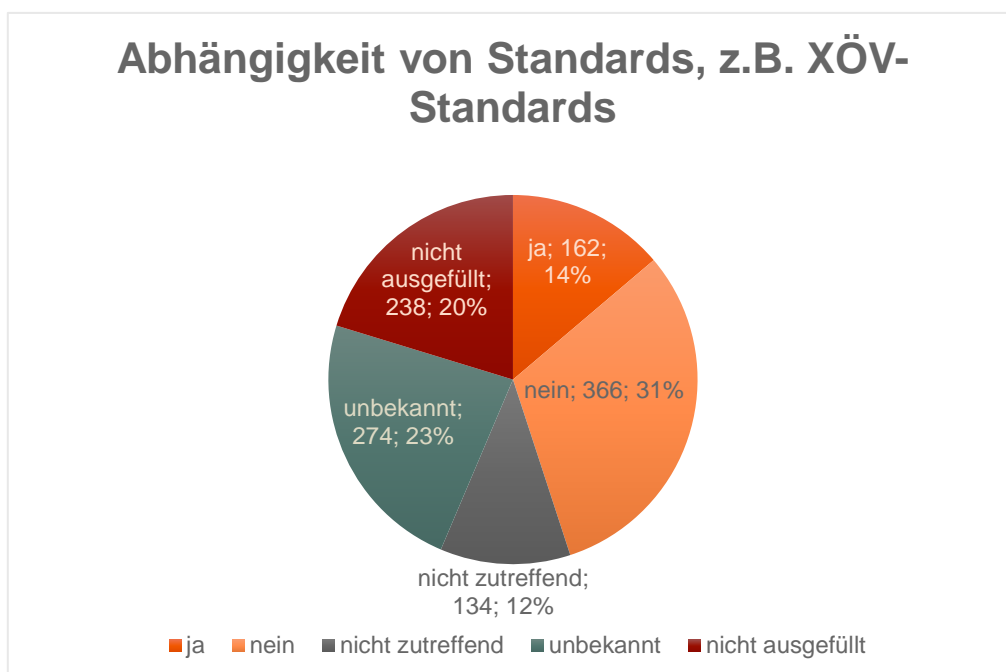


Abbildung 15: Bewertung des Risikos zum Erreichen der Konformität aufgrund Abhängigkeiten von Standards

Konkret wurden u.a. XMeld, XPersonenstand, XAusländer, XJustiz, XPolizei, XGewerbe, XFamilie, XRechnung und xdomea als Standards genannt, von denen sehr hohe Abhängigkeiten bestehen.

Allerdings kann für einige der XÖV-Standards bereits ein Datum benannt werden, ab wann sie die DIN SPEC 91379 integrieren, z. B. der 31.10.2023 für XJustiz. Das scheint allerdings vielen Verantwortlichen für betroffene IT-Verfahren nicht bekannt zu sein.

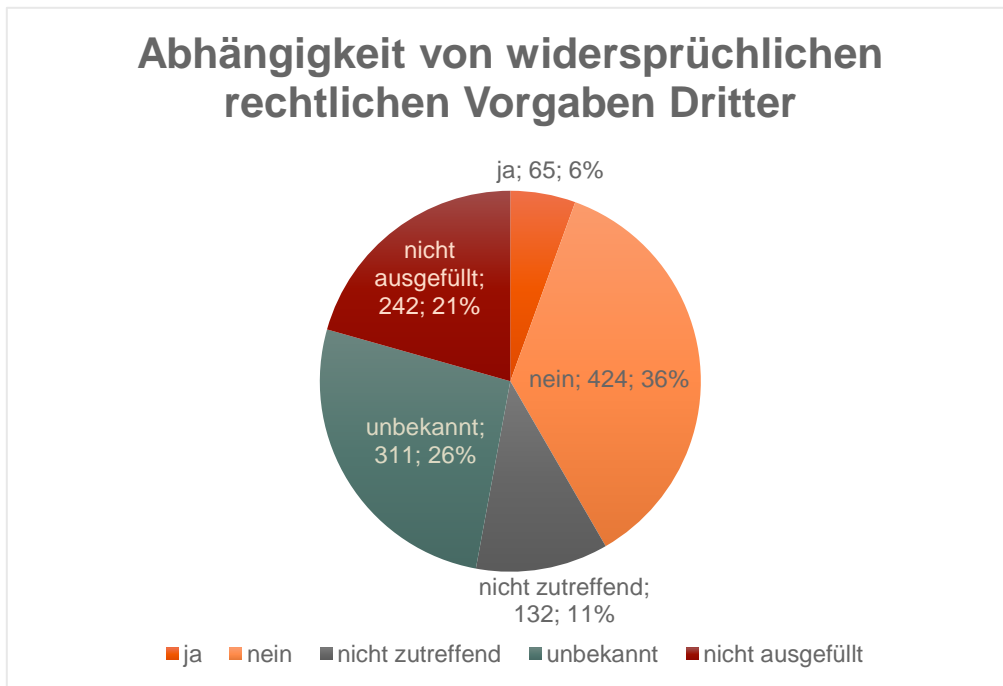


Abbildung 16: Bewertung des Risikos zum Erreichen der Konformität aufgrund Abhängigkeiten von widersprüchlichen rechtlichen Vorgaben Dritter

Als Beispiele für widersprüchliche rechtliche Vorgaben Dritter wurden u.a. die Datenerfassungs- und -übermittlungsverordnung (DEÜV) genannt, die kein String.Latin erlaube, das DFÜ-Abkommen für die Kommunikation mit Geld- und Kreditinstituten und die Beschlusslage des GKV-Spitzenverbandes (auf der Grundlage § 199a SGB V)⁵.

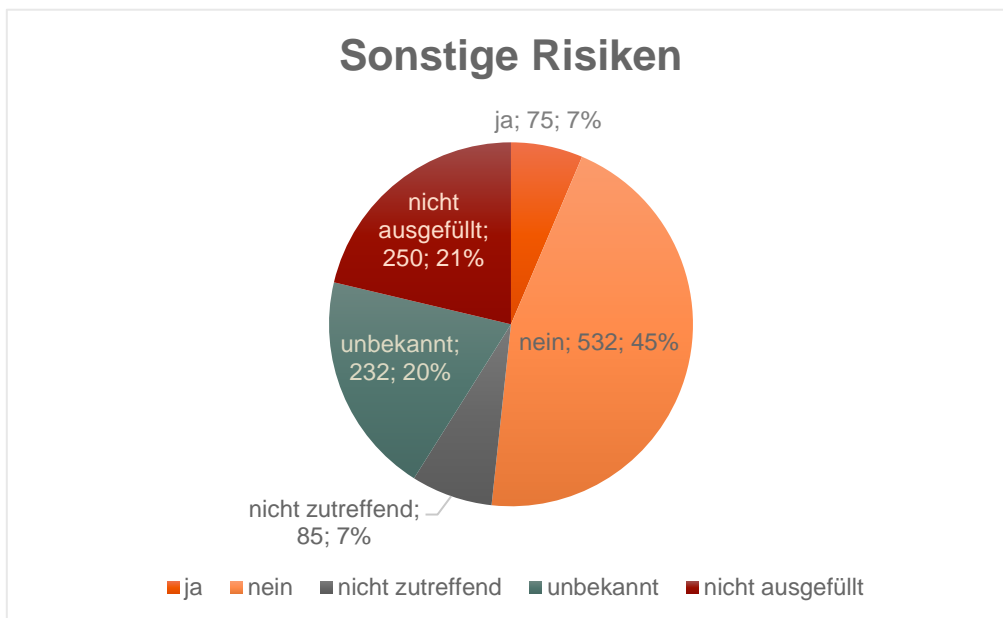


Abbildung 17: Bewertung sonstiger Risiken zum Erreichen der Konformität

⁵ https://www.gkv-datenaustausch.de/studenten_meldeverfahren/studenten_meldeverfahren.jsp

Über die vier konkret genannten Risiken hinaus werden scheinbar kaum weitere Risiken gesehen. Von den sieben Prozent, die sonstige Risiken identifiziert haben, wurden u.a. genannt:

- technische Risiken bei der Migration der in der Datenbank gespeicherten, nicht unicode-konformen Daten in das gewünschte Format. Hierbei könnten Inkonsistenzen entstehen und die Integrität der Daten verletzen. Daher sind ausführliche Tests und Planungen notwendig.
- Unsicherheit bezüglich der Qualität der in den Systemen der Schnittstellenpartner implementierten Suchalgorithmen und den daraus resultierenden Treffermengen
- wegen fachlicher Begriffe aus dem medizinischen Bereich und globalen Firmennamen sei eine Einschränkung von UTF-8 auf den Zeichenvorrat der DIN SPEC 91379 nicht sinnvoll
- fraglich, ob die Konformität in einer seit 2007 existenten Datenbank implementiert werden kann
- nicht alle eingesetzten Drucker sind unicode-fähig
- fehlende Haushaltsmittel

Oft wurden im Freitextfeld die sonstigen Risiken nicht näher ausgeführt. Es werden vermutlich noch überraschende neue Gefahren vermutet.

Zu der Befürchtung, die DIN SPEC könnte bereits vorhandene Funktionalität beschneiden, sei darauf hingewiesen, dass sie ein Mindeststandard ist. Die Konformität zur DIN SPEC wird nicht verletzt, indem zusätzliche Zeichen unterstützt werden.

3.7 Probleme bei der Umsetzung

Die Verantwortlichen wurden gefragt, was ihnen bei der Umsetzung von UTF-8, String.Latin 1.1 oder DIN SPEC 91379 (String.Latin+ 1.2) Probleme bereitet hat.

Für den Umgang mit typischen Problemen, wie der Eingabe von Buchstaben mit diakritischen Zeichen oder der Auswahl einer geeigneten Schriftart, sei auf die kommentierte Fassung der DIN SPEC „String.Latin+ 1.2“ [fgsl2020] verwiesen, die von der Fachgruppe String.Latin+ (FGSL) herausgegeben wird. Das Dokument enthält neben zusätzlichen Erläuterungen auch häufig gestellte Fragen und deren Antworten (FAQ). Es kann ohne Einschränkungen heruntergeladen werden.

Die bereits vor zwei Jahren abgefragten Probleme wurden signifikant seltener nicht ausgefüllt (20 - 21%) als die in der aktuellen Abfrage ergänzten Problemfälle (alle vier bei 36%). Das liegt daran, dass fast ein Viertel der ausgewerteten Rückmeldungen mit den alten Fragebögen erfolgte, bei denen die zusätzlichen Probleme fehlten.

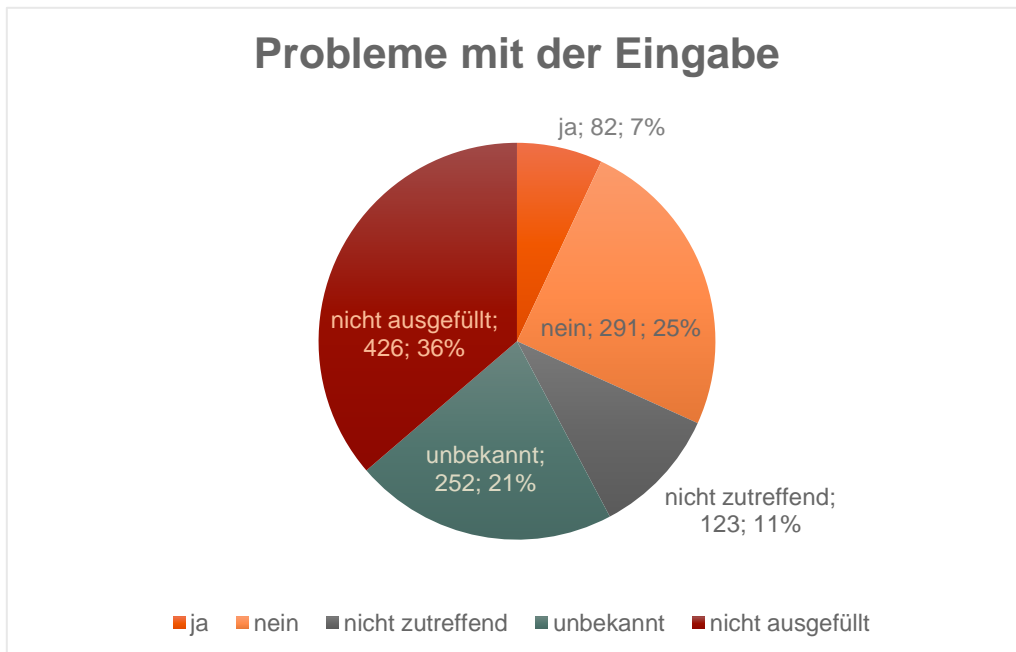


Abbildung 18: Bewertung von Umsetzungsproblemen mit der Eingabe der Schriftzeichen

Da viele IT-Verfahren keine Oberfläche zum Eingeben von Daten haben, weil sie lediglich über Programmierschnittstellen oder replizierte Datenbanken mit Daten gespeist werden, ist der Wert plausibel, dass für 11% der IT-Verfahren eine Eingabe nicht zutreffend ist.

Erstaunlich ist, dass den 7% der IT-Verfahren mit Herausforderungen bei der Dateneingabe 25% gegenüberstehen, die keine Probleme mit dem Eingeben der Zeichen der DIN SPEC 91379 haben. Aus Kommentaren in Freitexten kann man allerdings entnehmen, dass in der Praxis vieler Verfahren nur sehr selten spezielle Zeichen zum Einsatz kommen. Weil es keinen Bedarf gibt, Grundbuchstaben mit exotischen diakritischen Zeichen zu kombinieren, treten bei der Eingabe keine nennenswerten Probleme auf. In der kommentierten Fassung der DIN SPEC „String.Latin+ 1.2“ [fgsl2020] wird unter der Frage „Wie kann die Eingabe der vielen Zeichen der DIN SPEC implementiert werden?“ erläutert, wie man mit kostenfreien Mitteln der Betriebssysteme Windows und macOS spezielle Schriftzeichen eingeben kann. Die Windows-Standardmittel wurden allerdings als unkomfortabel bezeichnet und eigene Implementierungen als aufwändig eingeschätzt. In der Fachgruppe String.Latin+ (FGSL) wurden spezielle Entwicklungen zur Eingabe des umfangreichen Zeichenvorrats vorgestellt, die nachgenutzt werden können.

Die sichere optische Unterscheidung der diakritischen Zeichen ist nicht trivial. IT-Lösungen für die Eingabe der 646 lateinischen Buchstaben der DIN SPEC 91379 sollten das Kombinieren der 35 Grundbuchstaben mit 19 diakritischen Zeichen intelligent unterstützen. Der digitale Anhang „2018-11-15_din-spec-91379.xlsx“ der DIN SPEC 91379 enthält zu den Buchstaben zusätzliche Informationen, in welchen Ländern mit lateinischer Schrift sie beheimatet sind (Spalte „countries“) und durch welche Transliterationen aus nicht-lateinischen Schriften sie entstehen können (Spalte „transliteration“). Diese Informationen bei der Eingabe zu beachten, hilft Fehler zu vermeiden.

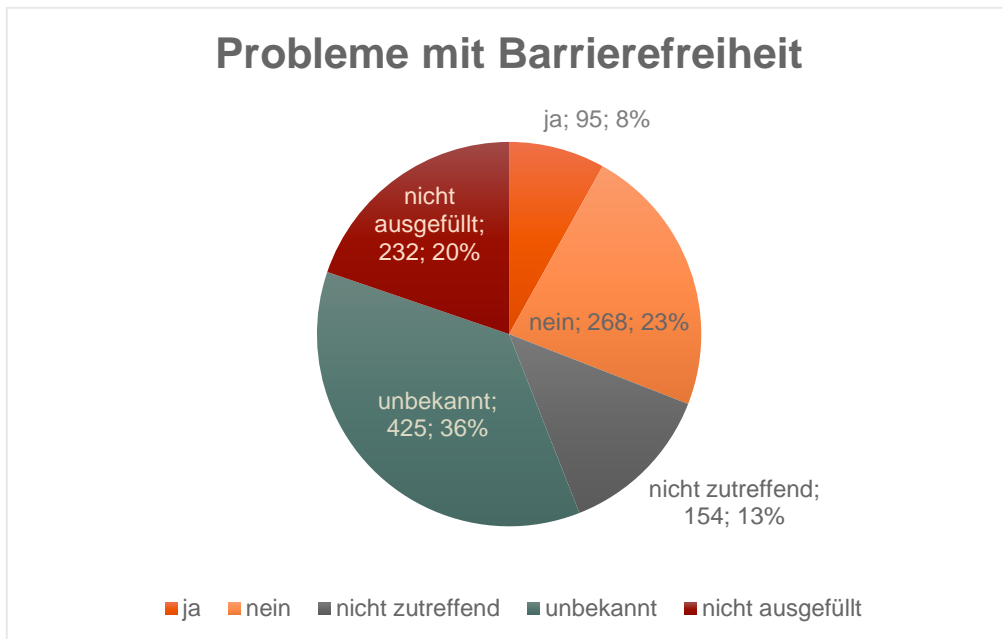


Abbildung 19: Bewertung von Umsetzungsproblemen mit der Barrierefreiheit

Nach Problemen mit der Barrierefreiheit wurde auch schon vor zwei Jahren gefragt. Der Anteil der ja-Antworten ist leicht zurückgegangen (von 9% auf 8%) und der Anteil der nein-Antworten ist stark gestiegen (von 15% auf 23%).

Möglicherweise stellt die Barrierefreiheit ein unterschätztes Problem dar. Obwohl nur acht Prozent der Antworten im Bereich Barrierefreiheit ein Problem sehen, scheint hier Handlungsbedarf zu bestehen. Damit Menschen mit z. B. eingeschränktem Sehvermögen oder blinde Menschen auch weiterhin die IT-Verfahren der öffentlichen Verwaltung nutzen können, müssen standardisierte Lösungen für Screenreader und sonstige Spezial-Software⁶ entstehen, die einerseits die vielen diakritischen Zeichen vorlesen, buchstabieren oder geeignet repräsentieren, aber andererseits für Menschen mit Behinderung beherrschbar bleiben. Ohne Unterstützung ist z. B. die Unterscheidung ähnlicher Diakritika für Menschen mit eingeschränktem Sehvermögen schwierig.

In einer Rückmeldung wird sogar vermutet, dass die DIN SPEC 91379 nicht barrierefrei umsetzbar sei, zum einen da die Braille-Zeile nicht so viele Zeichen (nur maximal 256) unterstützen würde und zum anderen weil das Verstehen der Ausgabe eines dafür konfigurierten Screenreaders das Auswendiglernen von hunderten ausgesprochenen Sonderzeichen erfordern würde.

In der 8-Punkte-Braille-Schrift können zwar nur maximal 256 Zeichen unterschieden werden (in der 6-Punkte-Braille-Schrift sind es sogar nur maximal 64 Zeichen), aber es wäre möglich, Zeichensequenzen mit einem vergleichbaren Mechanismus wie in Unicode in Grundbuchstaben und diakritische Zeichen zu zerlegen. Durch das Kombinieren von Grundbuchstaben und diakritischen Zeichen lassen sich die zusätzlichen Buchstaben bilden. Dafür, wie die meisten der 646 Buchstaben der DIN SPEC 91379 auf der Braille-Zeile dargestellt werden sollen, gibt es derzeit aber noch keine standardisierte Festlegung.

⁶ Grundsätzlich unterstützten die Screenreader verschiedene Nutzergruppen, die blind und unterschiedlich stark sehbeeinträchtigt sind. Die Braille-Zeile dient in der Regel ergänzend als Unterstützung für Menschen, die blind sind.

Um die besonderen Buchstaben der DIN SPEC 91379 mit Braille-Zeile oder Screenreader zu unterscheiden, müssen nicht alle 646 lateinischen Buchstaben auswendig gelernt werden. Diese bestehen aus den bekannten 27 Grundbuchstaben der deutschen Schrift (A bis Z und das Eszett), aus nur fünf weiteren Grundbuchstaben anderer lateinischer Schriften und aus drei Ligaturen – das sind Verbindungen aus bekannten Grundbuchstaben. Vier der acht zusätzlichen Grundbuchstaben sind bereits Bestandteil von 8-Bit-Zeichensätzen wie ISO 8859-15. Die vielen verschiedenen Buchstaben entstehen durch Kombinationen der 35 Grundbuchstaben mit 19 diakritischen Zeichen. Teil der deutschen Schrift ist beispielsweise das diakritische Zeichen „DIAERESIS“ (zwei Punkte über dem Grundbuchstaben), das kombiniert mit a, o und u die Umlaute ä, ö und ü ergibt. Über Fremdwörter und Namen ist im Deutschen auch das „ACUTE“ (schräger Strich nach rechts oben über dem Grundbuchstaben) weit verbreitet, z. B. in „René“.

Die sichere optische Unterscheidung der diakritischen Zeichen ist auch für nicht sehbehinderte Menschen herausfordernd, siehe vorheriger Abschnitt zu Problemen mit der Eingabe. Von einer standardisierten gut verständlichen Beschreibung des Aussehens der diakritischen Zeichen (z. B. „zwei Punkte über dem Grundbuchstaben“ anstelle von „DIAERESIS“) für die Verwendung durch Screenreader könnten alle Nutzer profitieren. Individuelle Konfigurationen der Screenreader, die sich unterscheiden, wären nicht zielführend. Wenn Menschen mit Behinderung den Arbeitgeber wechseln, müssten sie die Benennung der diakritischen Zeichen neu lernen.

Die barrierefreie Eingabe der vielen Zeichen der DIN SPEC 91379 ist noch herausfordernder als für nicht sehbehinderte Menschen. Die Standardmittel von Windows sind ungeeignet, da alle zur Auswahl stehenden Zeichen vorgelesen werden müssten. Für eine barrierefreie Lösung müsste auch hier eine Möglichkeit geschaffen werden, zuerst den Grundbuchstaben anzugeben und dann die noch sinnvoll zur Verfügung stehenden diakritischen Zeichen zur Auswahl zu stellen. Für eine verlustfreie Übernahme von Namen aus digitalen Dokumenten sind technische Lösungen möglich. Dabei muss auf die Anwendung der korrekten Normalform (NFC) geachtet werden.

In den Freitexten wurde sehr oft angegeben, dass man noch keine Kapazitäten hatte, sich mit Fragen der Barrierefreiheit im Kontext von String.Latin zu beschäftigen. Das erklärt die sehr hohen Anteile der Werte „unbekannt“ (36% – nahezu unverändert zu 35% vor zwei Jahren) und „nicht ausgefüllt“ (auch durch das konsequentere Löschen stark unvollständiger Rückmeldungen gesunken von 38% auf jetzt noch 20%).

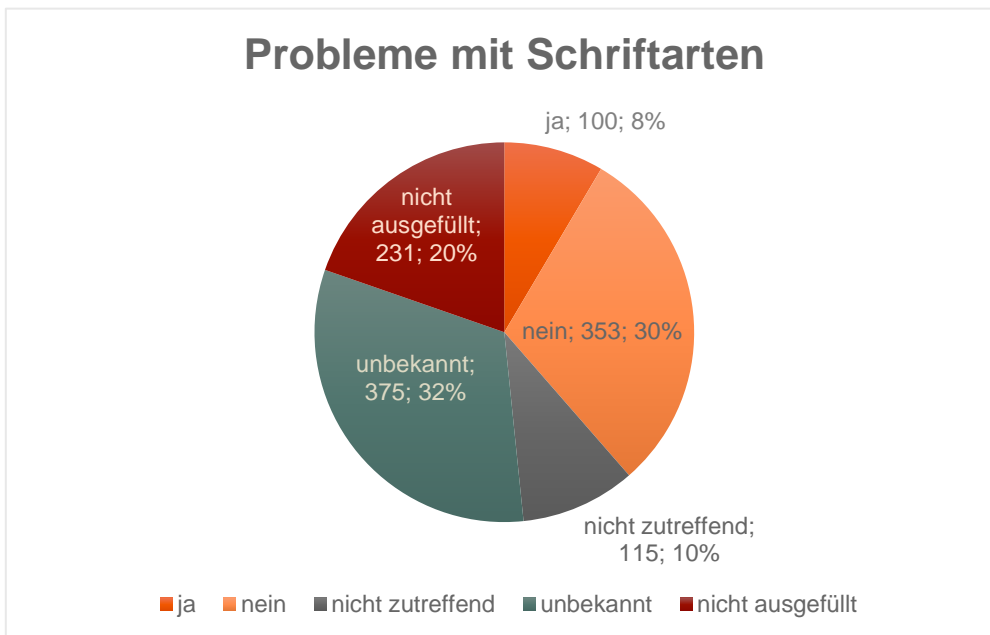


Abbildung 20: Bewertung von Umsetzungsproblemen durch Schriftarten

Die Antworten zu Problemen mit Schriftarten haben sich ähnlich entwickelt, wie die Antworten zu Problemen mit Barrierefreiheit. Die ja-Antworten sind auf niedrigem Niveau nahezu gleich geblieben (von 9% auf 8% gefallen). Die nein-Antworten haben stark zugenommen (fast verdoppelt von 16% auf 30%). Der Anteil der Antwort „unbekannt“ blieb nahezu gleich (von 35% auf 32% gefallen), während die Fragestellung seltener nicht ausgefüllt wurde (statt 38% jetzt noch 20%).

Es ist zu vermuten, dass Probleme auch deshalb eher selten auftreten, weil die weit verbreiteten Schriftarten alle Zeichen unterstützen, die in Deutschland sehr häufig vorkommen. Von Schriftarten wie Arial, Verdana und BundesSans werden einige normative Schriftzeichen der DIN SPEC 91379 nicht unterstützt. Diese Zeichen werden aber nur selten benötigt. So lange vorgelagerte IT-Verfahren noch keine speziellen Zeichen anliefern, treten bei Anzeige und Ausdruck keine Probleme auf. Die Komplexität und die Tücken dieses Themas werden sehr wahrscheinlich noch unterschätzt.

Als konkrete Probleme werden genannt, dass erst das Corporate Design umgestellt werden müsste, bevor neue Schriftarten verwendet werden dürfen, die alle Zeichen der DIN SPEC 91379 darstellen können. Eine solche Umstellung findet in der Regel nicht statt und Ausnahmen vom Corporate Design werden nicht genehmigt. Die Unterstützung der DIN SPEC 91379 sollte zukünftig bei der Bewertung und Auswahl von Schriftarten eine wesentliche Rolle spielen.

Bei bestimmten amtlichen Dokumenten mit vorgegebenem Layout werden Probleme vermutet, weil der Wechsel der Schriftart zu einer breiteren Laufweite der Schrift führen kann, die dann nicht mehr in die bestehenden Formulare passt. Der Aufwand, alle Formulare wegen dem Wechsel der Schriftart umstellen zu müssen, wird aus wirtschaftlichen Erwägungen gescheut. Die nichtproportionale Schriftart UnicodeDoc, die zumindest alle Zeichen des Standards String.Latin 1.1 korrekt darstellen kann, wurde aufgrund der konstanten Breite aller Zeichen (wie eine Schreibmaschinenschrift) für die Bescheiderstellung als völlig ungeeignet bewertet.

Es wurde vermutet, dass keine Schriftart existiert, die alle Schriftzeichen der DIN SPEC 91379 unterstützt. Das ist nicht richtig. In der kommentierten Fassung des Standards werden geeignete Schriftarten benannt, siehe Frage „Welche Schriftarten stellen alle Unicode-Zeichen der DIN SPEC 91379 korrekt dar?“ in „String.Latin+ 1.2“ [fgsl2020].

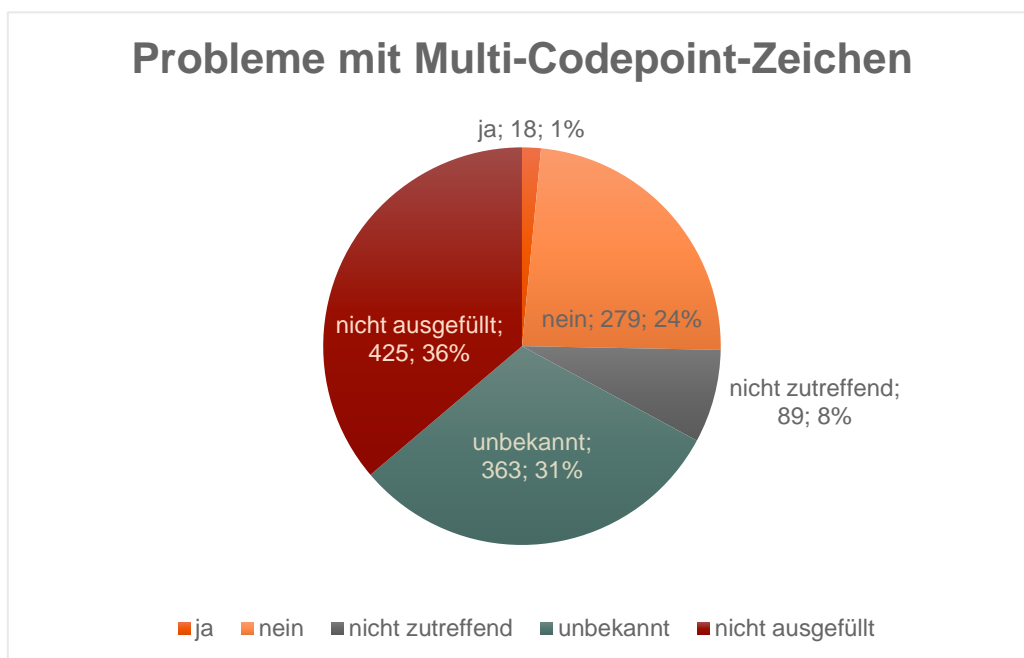


Abbildung 21: Bewertung von Umsetzungsproblemen durch Multi-Codepoint-Zeichen

Bei dem technischen Problem der Multi-Codepoint-Zeichen⁷ ist die Summe aus dem Wert „unbekannt“ und nicht ausgefüllter Tabellenzellen mit Abstand am höchsten (67%) und die Auswahl des Wertes „ja“ mit Abstand am niedrigsten (nur 1%).

Offen bleibt, ob das Problem deshalb so selten auftritt, weil moderne Datenbanken, moderne Entwicklungs-Frameworks und sehr gut qualifizierte Software-Entwickler zum Einsatz kommen, oder ob die Multi-Codepoint-Zeichen in der Praxis so selten vorkommen, dass es keine Probleme gibt, oder ob die Multi-Codepoint-Zeichen von den Eingabemöglichkeiten und Schnittstellen gar nicht erst zugelassen werden, um die ggf. dramatischen Probleme zu umgehen.

Für ein IT-Verfahren wurde informiert, dass es 1997 entwickelt und eingeführt wurde. Auch wenn weniger alte Datenbanken und Source Code befähigt werden müssen, mit Multi-Codepoint-Zeichen umzugehen, ist mit sehr großen Problemen mit Stringlängen, Such- und Sortieralgorithmen zu rechnen. Auch Schriftarten haben oft ausschließlich mit den Multi-Codepoint-Zeichen Darstellungsprobleme, weil Grundbuchstabe und die Zeichen der weiteren Codepoints ad-hoc kombiniert werden, was zu ungeeigneten Darstellungen führen kann.

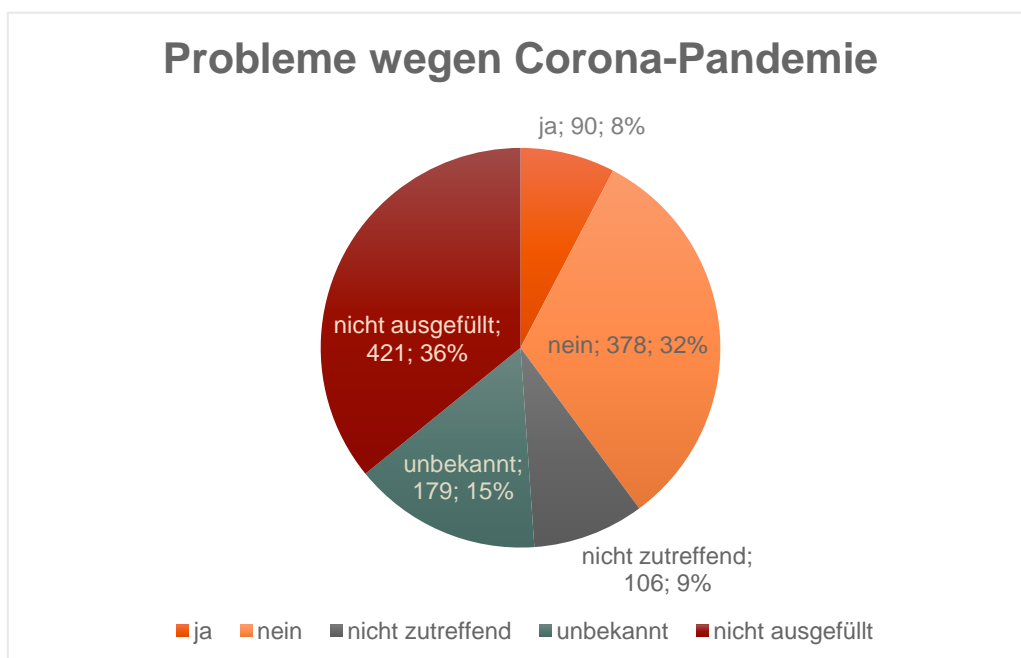


Abbildung 22: Bewertung von Umsetzungsproblemen durch die Corona-Pandemie

⁷ Multi-Codepoint-Zeichen sind Zeichensequenzen, die aus mehreren Single-Codepoint-Zeichen zusammengesetzt werden. In der DIN SPEC 91379 gibt es Multi-Codepoint-Zeichen, die aus bis zu drei Codepoints bestehen. Sie beanspruchen dementsprechend den doppelten oder dreifachen Speicherplatz eines Single-Codepoint-Zeichens und verändern auf der technischen Ebene die Länge von Strings. Mit diesem Mechanismus unterstützt Unicode sehr flexibel auch die Verarbeitung von Schriftzeichen, denen nicht explizit ein Codepoint im Standard zugewiesen wurde. Für den Buchstaben ċ gibt es beispielsweise keinen Codepoint im Unicode-Standard. Deshalb muss er aus den vorhandenen Zeichen c (Grundbuchstabe mit dem Codepoint 0063) und ˇ (kombinierendes diakritisches Zeichen mit dem Codepoint 0308) zusammengesetzt werden.

Vergleichsweise selten werden Probleme bei der Einführung von UTF-8, String.Latin 1.1 oder DIN SPEC 91379 auf die Corona-Pandemie zurückgeführt. Bei den konkret benannten Problemen liegt der nein-Wert mit 32% am höchsten.

Bei 8% der ausgewerteten IT-Verfahren haben Corona-bedingte Verschiebungen der Prioritäten und Ressourcen sowie längere Ausfällen von Personal die Umstellung auf neue Zeichensätze behindert und verzögert.

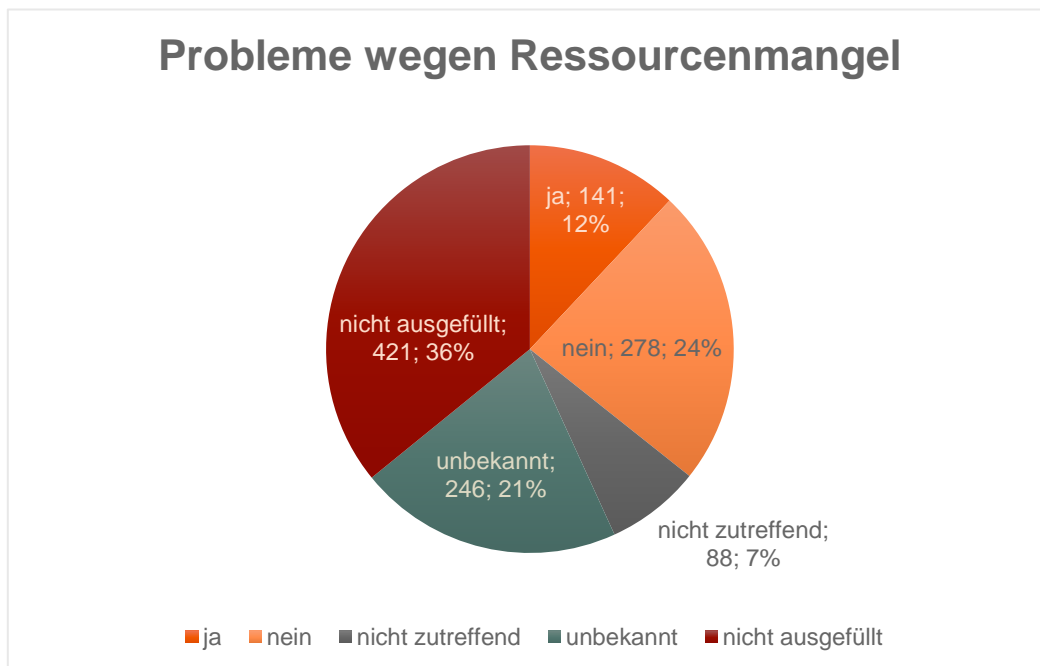


Abbildung 23: Bewertung von Umsetzungsproblemen aufgrund eines Mangels an Ressourcen

Das am häufigsten auftretende Problem bei der Einführung moderner Zeichensätze ist der Ressourcenmangel (12% gegenüber 1 bis 8% bei den anderen Problemen). Projekte, die allein der Umstellung von Zeichensätzen dienen, werden als erste depriorisiert, wenn gesetzliche Vorgaben und IT-Sicherheit gewährleistet werden müssen, wenn es ungeplante Bedarfe gibt, auf externe Krisen reagiert werden muss oder Mittel / Personal pauschal gekürzt werden.

Dass diese Projekte bei Umstellungen großer Datenbanken von 8-Bit-Zeichensätzen auf UTF-8 meistens mehrere Jahre dauern, bietet sehr viel Angriffsfläche. Während der langen Laufzeit gibt es immer wieder Verzögerungen, weil Ressourcen für konkurrierende Vorhaben abgezogen werden. Ohne eigenes Projekt steht die Umstellung auf String.Latin stets in Konkurrenz mit der Umsetzung fachlicher Anforderungen und technischer Anpassungen.

Bei dieser Problemlage ist es schwer, außer offensichtlichen Banalitäten hilfreiche Hinweise zur Vermeidung der Schwierigkeiten zu geben. Äußerst wichtig ist, dass bei ohnehin anstehenden Ablösungen von IT-Verfahren bei der Beschaffung oder Neuentwicklung der Nachfolger von Anfang an auf eine UTF-8-Fähigkeit und auf die Unterstützung des Zeichenvorrats der DIN SPEC 91379 geachtet wird. Dann müssen nicht später zusätzliche Migrationsprojekte aufgesetzt und mit Ressourcen versorgt werden.

In einigen Freitexten wurde angegeben, dass IT-Verfahren im Regelfall aufgrund personeller und finanzieller Engpässe nur dann an allgemeine Architekturvorgaben angepasst werden, wenn ohnehin wegen fachlicher oder technischer Änderungen Projekte angestoßen werden. Alternativ werden die Vorgaben beachtet, sobald ein IT-Verfahren durch seinen Lebenszyklus erneuert wird.

An Betreiber zentraler Register mit sehr vielen Schnittstellen vor allem zu kommunalen IT-Verfahren und an Hersteller weit verbreiteter Software-Komponenten muss appelliert werden, dass von ihnen die Umstellung sehr vieler anderer Systeme abhängt. Das muss beim Priorisieren dieser Migrationsprojekte beachtet werden.

Verzögerungen werden auch darauf zurückgeführt, dass es nicht gelingt, nach Abgängen im Zuge des demographischen Wandels IT-Personal zu gewinnen.

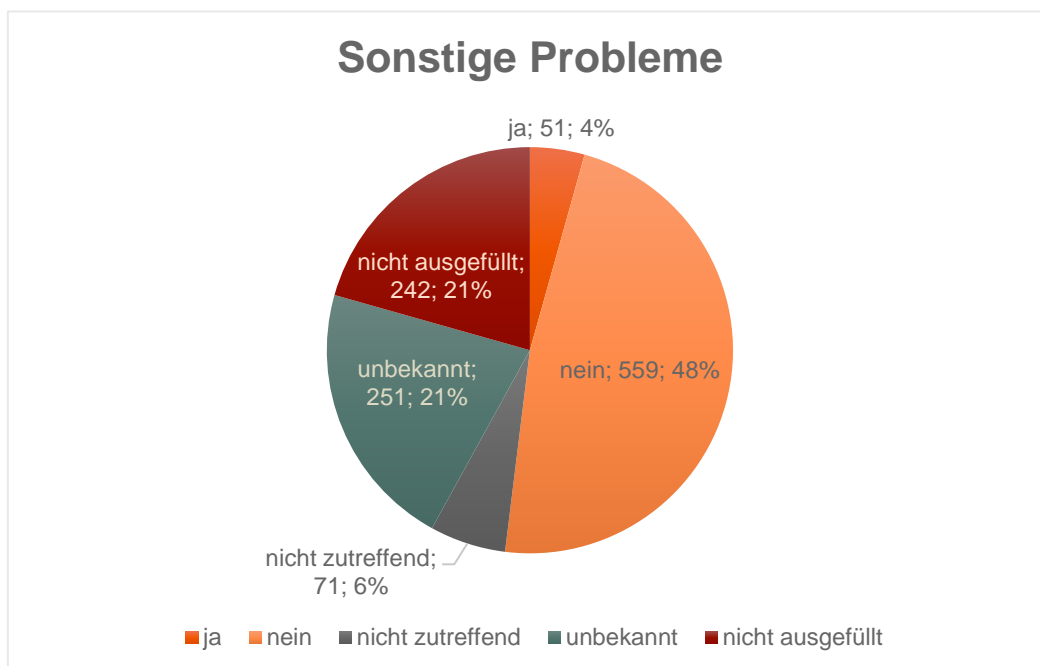


Abbildung 24: Bewertung von sonstigen Umsetzungsproblemen

Es ist erfreulich, dass neben den sechs explizit abgefragten Problemen kaum weitere Umsetzungsprobleme aufgetreten sind. Von den vier Prozent, die sonstige Probleme hatten oder haben, wurden genannt:

- Testfälle und Testdaten wurden vermisst, um die Konformität zur DIN SPEC 91379 überprüfen zu können
- Bei der Digitalisierung von Akten mit automatisierter Schrifterkennung (OCR) können Zeichen mit Diakritika nicht immer korrekt erkannt werden, was z. B. (Firmen-)Namen verfälschen kann.
- Der Datenpool unterstützt bereits UTF-8 und kann alle Unicode-Zeichen speichern. Es fehlt jedoch eine NFC-Normalisierung. Diese soll noch beim Eingabeprozess ergänzt werden, um bei der Speicherung und der Übertragung die Verwendung NFC-normalisierter Daten sicherzustellen.
- Office-Anwendungen enthalten nicht alle Zeichen des Datentyps C der DIN SPEC 91379
- Es ist zum jetzigen Zeitpunkt noch offen, ob das jetzige IT-Verfahren weiterverwendet wird oder ein neues Produkt eines anderen Herstellers eingeführt wird.
- keine latinchars.xml für String.Latin+ auffindbar
- die bei den Risiken genannten Abhängigkeiten verlangsamten den Fortschritt

Es ist unwahrscheinlich, dass Office-Anwendungen nicht alle normativen Zeichen der DIN SPEC 91379 und damit alle Schriftzeichen des Datentyps C unterstützen. Vermutlich können die konfigurierten Schriftarten nicht alle Zeichen korrekt darstellen.

Die Datei latinchars.xml ist Teil der digitalen Anhänge der DIN SPEC 91379. Sie kann direkt beim Beuth-Verlag kostenfrei bezogen werden, siehe [din2019]. Auch die kommentierte Fassung des Standards enthält alle digitalen Anhänge, siehe [kosit2022].

Bei der Weiterentwicklung des Standards zur Norm DIN 91379 wurde die NFC-Normalisierung bei den Konformitätsanforderungen explizit ergänzt. Der digitale Anhang latinchars.xml wurde für die neue Norm in chars.xml umbenannt, da er nicht nur lateinische Schriftzeichen enthält.

3.8 Allgemeine Hinweise

Nur selten wurde die (optionale) Möglichkeit genutzt, mit dem Fragebogen allgemeine Anmerkungen mitzugeben, um die Weiterentwicklung des Standards zu unterstützen, um Hinweise auf Lösungen / Best Practices zu geben oder um zukünftige Umfragen zu verbessern. Häufiger wurden mit den E-Mails zur Rücksendung der ausgefüllten Fragebögen noch weitere Hinweise gegeben oder eigener Bedarf an Anleitungen und Best Practices geäußert.

Es wurde der Mehrwert hinterfragt, warum nach einer Umstellung auf UTF-8 noch an der Konformität zur DIN SPEC 91379 gearbeitet werden soll. Allein den Beschluss des IT-Planungsrates bekannt zu machen, reicht also nicht aus. Es muss auch erläutert werden, dass Eingabemöglichkeiten für alle Zeichen sichergestellt werden müssen, dass es nicht selbstverständlich ist, dass Software die Normalform C (NFC) für die Repräsentation von Zeichensequenzen erzwingt und mit Multi-Codepoint-Zeichen korrekt umgeht, dass ggf. Zeichenbeschränkungen von Schnittstellen erweitert werden müssen und dass nur wenige Schriftarten alle Zeichen korrekt anzeigen und ausdrucken können. Selbst UTF-8-fähige Drucker sowie Hard- und Software zur Texterkennung sind auch nicht immer in der Lage, alle Zeichen der DIN SPEC 91379 korrekt zu verarbeiten. Nur wenn die gesamte Kette aller Komponenten überprüft und bei Bedarf angepasst wurde, sind die IT-Verfahren in der Lage, die gesetzlichen Vorschriften zur korrekten Verarbeitung von Namen natürlicher Personen umzusetzen.

Ein Produkthersteller hat die Unterstützung von UTF-8 und DIN SPEC 91379 abgelehnt mit der Begründung, dass UTF-8 Zeichen enthalte, die für eine Fachanwendung ungeeignet sind, beispielsweise die Phantasiesprache „Klingonisch“. Das Zulassen solcher Zeichen führe zu Problemen bei der Suche. Die 734 normativen Schriftzeichen der DIN SPEC 91379 sind eine vergleichsweise sehr kleine Teilmenge von Unicode, dem Zeichensatz, der von UTF-8 komplett unterstützt wird. Ein konformes IT-Verfahren muss diese 734 Schriftzeichen korrekt verarbeiten. Es dürfen auch mehr Zeichen verarbeitet werden, aber die klingonischen Zeichen sollten in jedem Fall ausgeschlossen werden. Das ist technisch leicht umsetzbar. Zu den digitalen Anhängen des Standards gehört eine XSD-Datei, die u.a. einen Datentyp definiert, der die zulässigen Zeichen auf den normativen Teil der DIN SPEC 91379 einschränkt („datatypeC“ in din-91379-datatypes.xsd).

Es wurde die Wahrnehmung mitgeteilt, dass die DIN SPEC 91379 widersprüchliche Aussagen / Empfehlungen enthalte hinsichtlich der Erweiterbarkeit der beschriebenen Zeichen oder einer Notwendigkeit, die unterstützten Zeichen auf die beschriebenen zu beschränken. Die Aussagen des Standards werden offensichtlich häufiger missverstanden. Im Abschnitt 4.1 „Konformität von IT-Verfahren“ heißt es: „Jedes konforme IT-Verfahren MUSS die in Tabelle 3 bis Tabelle 8 genannten Schriftzeichen bei allen für Namen (im weiteren Sinne) bestimmten Datenfeldern in vollem Umfang verarbeiten können. Dies umfasst insbesondere die Erfassung, Speicherung, Übermittlung, Anzeige und den Ausdruck aller Schriftzeichen.“ Es wird nicht gefordert, den unterstützten Zeichenvorrat auf die genannten Schriftzeichen zu reduzieren. Als informativer Anhang werden technische Datentypen mit dem Standard ausgeliefert. Diese unterstützen den Anwendungsfall, dass aus fachlichen Gründen der unterstützte Zeichenvorrat auf die Inhalte der DIN SPEC 91379 reduziert werden soll. Die Anwendung dieser Datentypen ist nicht normativ und somit keine Voraussetzung für die Konformität zum Standard. Dieser Zusammenhang wird im Kapitel 1 „Anwendungsbereich“ des Standards erläutert: „Durch den normativen Teil wird eine Mindestmenge von Schriftzeichen bestimmt. Zur DIN

SPEC konforme IT-Verfahren dürfen darüber hinausgehende Schriftzeichen unterstützen. In Schnittstellenvereinbarungen ist es jedoch erforderlich, den jeweiligen Zeichenvorrat abschließend festzulegen. Die DIN SPEC stellt in Anhang B Datentypen für fachlich bestimmte Anwendungsfälle zur Verfügung.“

Dass es einerseits zulässig ist, mehr Zeichen zu unterstützen, als in der DIN SPEC 91379 aufgelistet werden, und dass es andererseits eine fachlich sinnvolle Lösung sein kann, die zulässigen Zeichen auf den Umfang des Standards zu begrenzen, ist nicht allen IT-Verantwortlichen klar geworden. Erläuterungen dazu befinden sich in der Einleitung der DIN SPEC 91379 im Abschnitt „Abwägung von Alternativen“ [din2019].

Es wurde angeregt, in den Dokumenten des Standards klarer zu beschreiben, dass allein eine Öffnung für alle Unicode-Zeichen nicht automatisch bedeutet, dass die DIN SPEC 91379 korrekt implementiert wurde, siehe auch die Ausführungen unter Abbildung 10 „Geplante oder bereits erreichte Unterstützung von UTF-8“.

Es wurde kritisiert, dass nicht klar definiert sei, dass Schnittstellen mit Kommunikationspartnern im Ausland außerhalb des Geltungsbereichs der DIN SPEC 91379 liegen. Hierzu sei darauf hingewiesen, dass der Standard selbst keine Aussagen über seinen Geltungsbereich trifft. Die Entscheidung 2019/16 des IT-Planungsrates [itplr2019a] hat den Geltungsbereich für die deutsche Verwaltung festgelegt auf „IT-Verfahren, die dem Bund-Länder übergreifenden Datenaustausch oder dem Datenaustausch mit Bürgern und Wirtschaft dienen“. Wenn solche IT-Verfahren zusätzlich Schnittstellen mit Kommunikationspartnern im Ausland haben, nimmt sie das nicht aus dem Geltungsbereich heraus. Dann sind an diesen Schnittstellen Konvertierungen notwendig.

Insgesamt sind sehr viele der Rückmeldungen in Freitexten von Missverständnissen und falschen Annahmen geprägt. Es gibt nach wie vor einen hohen Informationsbedarf bezüglich der Umstellung von IT-Verfahren auf UTF-8 und auf Unterstützung der Unicode-Teilmenge des Standards String.Latin+ 1.2.

4 Erfahrungen für vergleichbare Umfragen

Es gab in Relation zur Gesamtzahl der Rückmeldungen nur sehr wenige Nachfragen an das Postfach String.Latin.Umsetzung@init.de zur Verwendung des Fragebogens und auch Verständnisfragen. Die Möglichkeit, Feedback zum Fragebogen zu geben, wurde nur einmal genutzt.

Es hat sich bewährt, die Dateien nicht per E-Mail zu versenden, sondern für sie eine Download-Möglichkeit zu schaffen. Dadurch konnte ausgeschlossen werden, dass E-Mails in der Weiterleitung der Umfrage aus Sicherheitsgründen abgewiesen werden.

Es wurden Bedenken geäußert, Umfrageergebnisse und Best Practices an ein E-Mail-Postfach einer Firma zu senden. Es hätte entweder deutlicher gemacht werden sollen, dass das Postfach im Auftrag einer Behörde betrieben wird, oder die Rückmeldungen hätten per Datei-Upload auf einer Webseite der Behörde erfolgen sollen – ggf. alternativ. Eine automatische Weiterleitung der an eine Behörde gerichteten E-Mails an einen privaten Dienstleister war aus rechtlichen Gründen keine Option. Der Zugriff auf ein Behördenpostfach durch einen privaten Dienstleister ist nur mit hohem Aufwand zur Absicherung der Infrastruktur möglich.

Dass die Dateien in der zweiten Umfrage nur im Excel-Format und nicht wie in der ersten Umfrage zusätzlich im OpenDocument-Format bereitgestellt wurden, hat keine erkennbaren Probleme verursacht.

Das Excel-Format hat nicht verhindern können, dass in den Feldern mit vorgegebenen Wertelisten andere Werte eingetragen wurden. Das erschwert die Auswertung und verschlechtert die Qualität der Rückmeldungen. Ein Tool zur webbasierten Online-Umfrage wäre deutlich aufwändiger, würde aber die Qualität der Rückmeldungen verbessern.

Von der ersten zur zweiten Umfrage wurde das Format der Excel-Datei um zusätzliche Spalten ergänzt. Es wurde nach dem Namen des Bundesressorts/Bundeslands/Kommune gefragt, es wurde gefragt, ob eine Abschaltung des IT-Verfahrens geplant ist, es wurde gezielt nach möglichen zusätzlichen Problemen bei der

Umsetzung gefragt (Eingabe, Multi-Codepoint-Zeichen, Corona-Pandemie und Ressourcenmangel) und es wurden in bestehenden Wertelisten zusätzliche Werte zugelassen, wie z. B. „unbekannt“ für „Sonstige Probleme“. Diese Änderungen haben die inhaltliche Auswertung der zweiten Umfrage verbessert. Beim Kopieren der alten Umfragewerte in die neue Vorlage wurden allerdings sehr oft Fehler gemacht, die aufwändig korrigiert werden mussten. Da fast ein Viertel der Rückmeldungen im Format der ersten Umfrage erfolgte, blieben viele der neuen Werte unbeantwortet. Es ist von Vorteil, das Format einer mehrfach durchgeführten Umfrage nicht zu ändern.

Dass die Ergebnisse der Umfrage vor zwei Jahren sehr oft unverändert als aktuelle Antworten weitergeleitet wurden, offenbart eine schlechte Qualität der Antworten. Es ergibt keinen Sinn, im Jahr 2022 zu antworten, dass man plant, die Konformität zur DIN SPEC 91379 im Jahr 2020 umzusetzen. Immerhin drei Prozent der Rückmeldungen bestand aus falschen Werten, die in der aktuellen Vorlage gar nicht mehr zur Auswahl standen. Ob wirklich alle IT-Verfahren wie geplant umgestellt wurden, ist fraglich. Diese Antworten wurden in den Ergebnissen deshalb separat ausgewiesen.

Wenn eine Umfrage mehrfach durchgeführt wird, sollte die Vorlage versioniert werden, auch wenn die Vorlage nicht verändert wird. Die unterschiedlichen Versionsnummern der Vorlagen der ersten und zweiten Umfrage haben den Umgang mit Zulieferungen im alten Format erleichtert und sie haben eine statistische Auswertung ermöglicht, wie oft mutmaßlich einfach die alten Angaben unverändert (ohne Aktualisierung) ausgeliefert wurden.

Aus technischen Gründen der Auswertung und um beim Übertragen aus den einzelnen Excel-Dateien in eine Gesamtliste keine Rückmeldungen zu übersehen, sollte stets ein Hinweis gegeben werden, möglichst keine Leerzeilen zwischen den Einträgen zu lassen. Ein solcher Hinweis hat bei der zweiten Umfrage zu erheblich weniger Rückmeldungen mit Leerzeilen geführt.

Es sollte vorgesehen werden, auch Angaben zu IT-Verfahren abzufragen, die noch nicht produktiv sind, aber bereits geplant werden oder sich in Entwicklung befinden. Ggf. werden diese IT-Verfahren aktuell laufende Systeme ersetzen. Die Fragen müssen für diesen Fall etwas anders formuliert werden. Das Excel-Format ist für Fallunterscheidungen nicht gut geeignet. In einer webbasierten Umfrage wäre das leichter umsetzbar.

Es hat sich bewährt, ausgefüllte Beispiele mit der Vorlage des Fragebogens bereitzustellen. Es gab nur wenige Rückfragen (keine zur Verwendung des Fragebogens) und für die Felder mit Freitexten wurde die Erwartungshaltung deutlich und die Antworten wurden harmonisiert. Eine Orientierung vieler Antworten an den Beispielen war deutlich erkennbar.

Um widersprüchliche Antworten zu reduzieren, ist es überlegenswert, Beispiele mitzuliefern, welche Antwortkonstellationen nicht sinnvoll wären. Ein webbasiertes Umfrage-Tool könnte direkt beim Ausfüllen auf Widersprüche hinweisen und die Qualität der Antworten verbessern.

Grundsätzlich gilt, dass Erläuterungen und Beispiele möglichst niedrigschwellig sein sollten, also nicht sehr umfangreich, damit sie auch beachtet werden.

In der Excel-Datei zum Eintragen der Informationen über IT-Verfahren war die linke Spalte mit den Namen der IT-Verfahren fixiert, sodass immer zu sehen war, welche Zeile gerade bearbeitet wird. Wenn in einer Excel-Datei sehr viele Zeilen eingetragen wurden, führte das Scrollen nach unten allerdings dazu, dass man die Titel der auszufüllenden Spalten nicht mehr lesen konnte. Es hätte beides fixiert werden sollen: die linke Spalte mit den Namen der einzutragenden Objekte und die Zeile mit den Titeln aller Spalten.

Dort, wo anstatt der zwei Monate der ersten Umfrage beim zweiten Durchlauf vier Monate zur Verfügung standen, hat sich die Anzahl der Rückmeldungen signifikant erhöht. Es wurde beim zweiten Mal extra ein Hinweis gegeben, dass nachgeordneten Behörden eine ausreichende Frist gewährt werden sollte, weil das Beantworten der Fragen nicht in wenigen Tagen möglich ist, sondern ggf. umfangreiche Recherchen erfordert.

Dass ausgehend von der Geschäftsstelle des IT-Planungsrats über die direkten Mitglieder wirklich alle relevanten Verantwortlichen für IT-Verfahren in Bund, Länder und Kommunen erreicht werden, ist durch die bestehenden Strukturen noch nicht sichergestellt.

5 Abkürzungsverzeichnis

AA	Auswärtiges Amt
BB	Brandenburg
BE	Berlin
BfJ	Bundesamt für Justiz
BKA	Bundeskriminalamt
BKAmt	Bundeskanzleramt
BKM	Beauftragter der Bundesregierung für Kultur und Medien
BMAS	Bundesministerium für Arbeit und Soziales
BMBF	Bundesministerium für Bildung und Forschung
BMEL	Bundesministerium für Ernährung und Landwirtschaft
BMF	Bundesministerium der Finanzen
BMFSFJ	Bundesministerium für Familie, Senioren, Frauen und Jugend
BMG	Bundesministerium für Gesundheit
BMI	Bundesministerium des Innern, für Bau und Heimat
BMJV	Bundesministerium der Justiz und für Verbraucherschutz
BMU	Bundesministerium für Umwelt, Naturschutz und nukleare Sicherheit
BMVg	Bundesministerium der Verteidigung
BMVI	Bundesministerium für Verkehr und digitale Infrastruktur
BMWi	Bundesministerium für Wirtschaft und Energie
BMZ	Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung
BPA	Presse- und Informationsamt der Bundesregierung
BVA	Bundesverwaltungsamt
BW	Baden-Württemberg
BY	Bayern
BZR	Bundeszentralregister
BZSt	Bundeszentralamt für Steuern
Destatis	Deutsches Statistik-Informationssystem, Internetauftritt des Statistischen Bundesamtes (StBA)
DFÜ	Datenfernübertragung
DIN	Deutsches Institut für Normung e. V.
EESSI	Electronic Exchange of Social Security Information

EGVP	Elektronisches Gerichts- und Verwaltungspostfach
ELSTER	Elektronische Steuererklärung
FAQ	Frequently Asked Questions (Häufig gestellte Fragen und Antworten)
FGSL	Fachgruppe String.Latin+
FMS	Formularmanagementsystem
GKV	Gesetzliche Krankenversicherung
GSB	Government Site Builder
HB	Bremen
HE	Hessen
HH	Hamburg
INPOL	Informationssystem der Polizei
IT-PLR	IT-Planungsrat
ITZBund	Informationstechnikzentrum Bund
i.w.S.	im weiteren Sinne
KBA	Kraftfahrt-Bundesamt
KMS	KONSENS-Mitteilungsverfahren
KONSENS	Koordinierte neue Software-Entwicklung der Steuerverwaltung
KoSIT	Koordinierungsstelle für IT-Standards
KV	Krankenversicherung
LH	Landeshauptstadt
MV	Mecklenburg-Vorpommern
NI	Niedersachsen
NW	Nordrhein-Westfalen
NWR	Nationales Waffenregister
OCR	Optical Character Recognition (optische Zeichenerkennung, Texterkennung)
RMS	Risikomanagementsystem
RP	Rheinland-Pfalz
SGB	Sozialgesetzbuch
SH	Schleswig-Holstein
SL	Saarland
SN	Sachsen
ST	Sachsen-Anhalt
TH	Thüringen
UCS	Universal Coded Character Set, der Unicode-Zeichensatz

UTF	UCS Transformation Format
XML	Extensible Markup Language
XSD	XML Schema Definition

6 Quellenverzeichnis

- [din2019] DIN SPEC 91379: „Zeichen in Unicode für die elektronische Verarbeitung von Namen und den Datenaustausch in Europa“; März 2019; kostenfrei beziehbar vom Beuth-Verlag unter <https://www.beuth.de/de/technische-regel/din-spec-91379/301228458>
- [din2022] DIN 91379:2022-08 „Zeichen und definierte Zeichensequenzen in Unicode für die elektronische Verarbeitung von Namen und den Datenaustausch in Europa“, August 2022, kostenpflichtig (177,80 EUR für den digitalen Download) beziehbar vom Beuth-Verlag unter <https://www.beuth.de/de/norm/din-91379/353496133>
- [fgsl2020] Fachgruppe String.Latin+ (FGSL): „String.Latin+ 1.2 – Zeichen in Unicode für die elektronische Verarbeitung von Namen und den Datenaustausch in Europa“; kommentierte und erweiterte Fassung der DIN SPEC 91379, inklusive einer umfangreichen Liste häufig gestellter Fragen (FAQ); 15. Januar 2020; <https://www.xoev.de/downloads-2316#StringLatin>
- [itplr2014] IT-Planungsrat (IT-PLR): „Entscheidung 2014/04 - Einheitlicher Zeichensatz für Datenübermittlung und Registerführung“; 13. Sitzung des IT-Planungsrats vom 12. März 2014; <https://www.it-planungsrat.de/beschluss/beschluss-2014-04>
- [itplr2019a] IT-Planungsrat (IT-PLR): „Entscheidung 2019/16 - DIN SPEC“; 28. Sitzung des IT-Planungsrats vom 12. März 2019; <https://www.it-planungsrat.de/beschluss/beschluss-2019-16>
- [itplr2019b] IT-Planungsrat (IT-PLR): „Entscheidung 2019/53 - DIN SPEC 91379“; 30. Sitzung des IT-Planungsrats vom 23. Oktober 2019; <https://www.it-planungsrat.de/beschluss/beschluss-2019-53>
- [kosit2012] Koordinierungsstelle für IT-Standards (KoSIT): „Lateinische Zeichen in Unicode“; Version 1.1.1; 27. Januar 2012; http://xoev.de/latinchars/1_1/latinchars.pdf
- [kosit2022] Koordinierungsstelle für IT-Standards (KoSIT): „String.Latin+ – Lateinische Zeichen in Unicode“; https://www.xoev.de/die_standards/lateinische_zeichen_in_unicode-4813