

BUNDESAMT FÜR KARTOGRAPHIE UND GEODÄSIE

# Geokodierungsdienst der AdV

---

## Technisches Konzept

AG Geokodierungsdienste der AdV

**10.07.2015**

**Version 1.0**

## Änderungsverzeichnis

Versionsnummer	Datum	Änderung	Ersteller
<b>0.1</b>	07.03.2014	Dokument erstellt	Endrullis
<b>0.2</b>	13.03.2014	Ergänzung Kap. 5.1	Rath
<b>0.3</b>	14.03.2014	Einzelne Ergänzungen, Hinweise	Bork, Caffier, Rath, Thalheim, Strehmel
<b>0.4</b>	24.03.2014	Ergänzungen unter 4.4.1 Ergänzung Kap 5.2 Umsetzung einzelner Hinweise	Endrullis
<b>0.5</b>	31.03.2014	Einzelne Ergänzungen, Ergänzung 4.4.2	Rath, Endrullis Thalheim
<b>0.52</b>	11.06.2014	Austausch KVP durch OpenSearch GeoTemporal Service in 4.4.2.2 und im gesamten Dokument	Thalheim
<b>0.53</b>	06.08.2014	Kommentar DOG	Thalheim
<b>0.6</b>	12.08.2014	Fortschreibung Geokodierung für Flurstücke	Körkel
<b>0.61</b>	15.08.2014	Finale Abstimmung Anmerkungen DOG in der AG	Endrullis
<b>0.7</b>	04.02.2015	Kapitel 5 Geokodierung für Flurstücke komplett überarbeitet	Körkel
<b>0.8</b>	04.06.2015	Kapitel 5 gestrafft und aktualisiert	Endrullis
<b>0.9</b>	12.06.2015	Bilder in Kapitel 5 sowie Abschnitt Lösungen	Körkel
<b>0.91</b>	19.06.2015	Endredaktion aus Sicht BKG	Endrullis
<b>1.0</b>	10.07.2015	Angaben aus dem Änderungsmodus übernommen	Endrullis

## Arbeitsgruppe

Dr. Manfred Endrullis (Leitung)	Bundesamt für Kartographie und Geodäsie
Stephan Bork	Landesamt für Vermessung und Geoinformation Schleswig-Holstein
André Caffier	Ministerium für Inneres und Kommunales, Nordrhein-Westfalen
Christoph Rath	Information und Technik Nordrhein-Westfalen (IT.NRW)
Ralf Strehmel	Landesbetrieb Landesvermessung und Geobasisinformation Brandenburg (LGB)
Dirk Thalheim	Bundesamt für Kartographie und Geodäsie
Stefan Körkel	Bundesamt für Kartographie und Geodäsie

## Inhaltsverzeichnis

Änderungsverzeichnis .....	2
Arbeitsgruppe.....	2
Abbildungsverzeichnis.....	5
Tabellenverzeichnis .....	5
Abkürzungsverzeichnis .....	6
1 Zusammenfassung .....	7
2 Funktionale Dienstetypen.....	8
2.1 Ortssuchdienst.....	8
2.1.1 Beschreibung des Ortssuchdienstes .....	8
2.1.2 Hinweise zur Client-Server-Implementierung des Ortssuchdienstes.....	9
2.1.3 Anmerkung zur Verwendung des Begriffs „Gazetteerdienst“ .....	10
2.2 Geokodierungsdienst.....	10
2.2.1 Beschreibung des Geokodierungsdienstes.....	10
2.2.2 Varianten der Client-Server-Kommunikation.....	11
2.2.3 Geokodierungsanwendung.....	13
2.3 Reverser Geokodierungsdienst.....	14
3 Inhaltliches Stufenkonzept.....	15
4 Geokodierungsdienst für Adressen, Geonamen, POI.....	16
4.1 Vergleichende Analyse vorhandener Lösungsansätze.....	16
4.2 Verwendung von Standards .....	16
4.2.1 Verwendung standardisierter Dienste zur Geokodierung.....	16
4.2.2 Anmerkungen zum DOG-Profil.....	16
4.3 Anforderungen und Weiterentwicklungsbedarf.....	17
4.4 Schnittstellenbeschreibung Geokodierungsdienst.....	19
4.4.1 Datenmodell.....	19
4.4.2 Schnittstellen .....	21
4.5 Implementierung des Dienstes .....	24
5 Geokodierungsdienst für Flurstücke.....	25
5.1 Erforderliche Flurstücksinformationen.....	25
5.1.1 Attributive Eigenschaften.....	25
5.1.2 Geometrische Informationen .....	25
5.2 Grundlagen in den Bundesländern Ende 2014 .....	26
5.2.1 Ausgangslage bei den Daten.....	26

5.2.2	Ausgangslage bei den Diensten .....	26
5.2.3	Langfristige Entwicklung .....	27
5.3	Verwendung von Standards.....	28
5.4	Lösungsansätze für die Gesamtarchitektur.....	28
5.4.1	Geokodierungsdienst mit zentraler Datenhaltung.....	28
5.4.2	Geokodierungsdienst als reine Dienstekaskade.....	29
5.4.3	Geokodierungsdienst mit Dienstekaskade und zentralem Cache .....	30
5.5	Vorstellung geeigneter Lösungsansätze .....	31
5.5.1	Kurzfristige Lösung.....	31
5.5.2	Caching-Anwendung .....	32
5.5.3	NAS-Import .....	33
5.5.4	Mittel- und langfristige Lösung.....	33
6	Anlagenverzeichnis.....	35
7	Literaturverzeichnis .....	36

## Abbildungsverzeichnis

Abbildung 1: Ortssuche. Beispiel für die Kommunikation bei einer Vorschlagssuche.....	9
Abbildung 2: Ortssuche. Beispiel für die Kommunikation bei der Ergebnissuche.....	9
Abbildung 3: Beispiel für eine Vorschlagssuche in einer Anwendung .....	9
Abbildung 4: Beispiel für die Kommunikation mit einem Geokodierungsdienst.....	11
Abbildung 5: Geokodierung von Massendaten über eine Geokodierungsanwendung.....	14
Abbildung 6: GeoWebCluster des BKG an den Standorten Frankfurt a. M. und Leipzig. ....	24
Abbildung 7: Übersicht über die Bundesländer mit einem Flurstücksdienst.....	26
Abbildung 8: Geokodierungsdienst mit zentraler Datenhaltung.....	29
Abbildung 9: Geokodierungsdienst als Dienstekaskade.....	30
Abbildung 10: Geokodierungsdienst mit Dienstekaskade und zentralem Cache .....	31
Abbildung 11: Gesamtarchitektur der kurz- und mittelfristigen Lösung.....	32
Abbildung 12: Mittel- und langfristige Lösung.....	34

## Tabellenverzeichnis

Tabelle 1: Gemeinsame Anforderungen an den Geokodierungsdienst.....	17
Tabelle 2: Anforderungen an Ortssuche und Geokodierung.....	18
Tabelle 3: Anforderungen an die Reverse Geokodierung. ....	18
Tabelle 4: Attribute des Datentyps <i>Vorschlag</i> .....	19
Tabelle 5: Attribute des Datentyps <i>Ortsangabe</i> .....	20

## Abkürzungsverzeichnis

AdV	Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland
AG	Arbeitsgruppe
CSV	Comma-Separated Values (ein Dateiformat für strukturierte Daten in einfachen Textdateien)
DOG	Deutschland Online Gazetteer
EGovG	Gesetz zur Förderung der elektronischen Verwaltung (E-Government-Gesetz)
GDI-DE	Geodateninfrastruktur Deutschland
GN-DE	Geographische Namen Deutschland
GML	Geographic Markup Language
LG GDI-DE	Lenkungsgremium Geodateninfrastruktur Deutschland
OGC	Open Geospatial Consortium
OWS	OGC Web Service oder Open Web Service
POI	Points-of-Interest
URL	Uniform Resource Locator
WFS	Web Feature Service
WFS-G	Gazetteer Service - Application Profile of the Web Feature Service
WPS	Web Processing Service

## 1 Zusammenfassung

Das vorliegende Dokument dient der technischen Konzeption und Bereitstellung eines Geokodierungsdienstes der AdV. Ein Geokodierungsdienst ist ein Webdienst, der attributiv beschriebenen Objekten (geographischen Identifikatoren) eine räumliche Lagebeschreibung in Form einer Koordinate zuweist und die Objekte damit georäumlichen Analysen und Bearbeitungen zugänglich macht. Er erweitert die Funktionalität von Anwendungen, in die ein Dienst eingebunden wird und die den jeweiligen fachlichen Kontext abbilden und die passende Nutzeroberfläche bereitstellen.

Der Bedarf an einem solchen Dienst ist vielseitig. Unter anderem verlangt das E-Government-Gesetz (EGovG) im § 14 zur Georeferenzierung von Registern einen entsprechenden Dienst für Adressen, Flurstücke und in Rechtsvorschriften definierte Gebiete. Das Lenkungs-gremium Geodateninfrastruktur Deutschland (LG GDI-DE) hat mit einer Anfrage an die AdV vom 14.06.2013 um die Bereitstellung eines solchen Dienstes auch für Geonamen, Points-of-Interest (POI) und Kilometrierungssysteme ersucht. Die der AdV übergebenen konzeptionellen Ansätze [1] bis [3] des LG GDI-DE wurden hier berücksichtigt.

Der erarbeitete Lösungsvorschlag beschreibt die funktionalen Dienstetypen *Ortssuche* (mit Vorschlags- und Ergebnissuche), *Geokodierung* und *Reverse Geokodierung* konzeptionell und schlägt eine integrierte Implementierung in einem Geokodierungsdienst vor. Er analysiert die Ausgangslage hinsichtlich der erforderlichen Geodaten und der bereits vorhandenen Lösungsansätze und unterbreitet die folgenden Vorschläge:

Der am Dienstleistungszentrum des BKG entwickelte und betriebene Webdienst soll als Geokodierungsdienst der AdV für Adressen, Geonamen und POI eingesetzt werden. Schon jetzt liegt eine einsatzbereite Lösung vor, die unter anderem im Bundesbereich und im Geoportal.de verwendet wird. Weiterentwicklungen zur Umsetzung aller Aspekte aus diesem Konzept sollen zeitnah erfolgen und noch 2014 abgeschlossen werden.

Für die Geokodierung von Flurstücken wird ein bundesweiter, am BKG betriebener Geokodierungsdienst aufgebaut. Der Dienst soll als Dienstekaskade umgesetzt werden, die vorhandene Dienste der Bundesländer einbezieht. Ein Prototyp dieses Dienstes wurde am BKG entwickelt und erfolgreich erprobt.

Die Quellen für die Geokodierungsdienste stellt das BKG bei Bedarf anderen AdV-Mitgliedern kostenfrei zur Verfügung.

Auf Nutzungsbedingungen für die Dienste wird in diesem Konzept nicht eingegangen, da diese durch die TF PRM erarbeitet werden.

## 2 Funktionale Dienstetypen

Grundsätzlich lassen sich die folgenden Dienstetypen funktional unterscheiden, unabhängig von einer möglichen integrierten Implementierung.

### 2.1 Ortssuchdienst

#### 2.1.1 Beschreibung des Ortssuchdienstes

Ein Ortssuchdienst ist ein Geodatendienst, der den über geographische Identifikatoren gesuchten Objekten in Geoanwendungen eine Koordinate (Geokodierung) zuweist. Der Ortssuchdienst unterstützt die Suche durch dynamisch generierte *Vorschlagslisten* sowie durch eine nach der Trefferqualität sortierte *Ergebnisliste*. Es findet *keine persistente* Speicherung der Koordinate statt.

Die Anfrage an den Dienst (Request) und dessen Antwort (Response) beinhalten die nachfolgend dargestellten, teilweise optionalen Informationen.

#### **Request:**

- identifizierende Angaben zu einem räumlichen Objekt (postalische Adresse, Bezeichnung eines Flurstücks, geographischer Name, POI oder Kilometrierungsangabe) – *strukturiert* über einzelne Attribute beschrieben oder *unstrukturiert* in einem einzigen Suchattribut angegeben
- das Koordinatenreferenzsystem, in dem die Geokodierung erfolgen soll
- attributive und/oder räumliche Filterangaben zur Einschränkung der Suche
- funktionale Parameter zur Steuerung der *Vorschlagssuche* oder *Ergebnissuche*

(beispielsweise Einschränkung der Ergebnisliste auf den besten Treffer oder auf eine Mindest-Trefferqualität bzw. auf eine Kombination von beidem)

#### **Response Vorschlagssuche:**

- nach Trefferqualität sortierte Liste mit Suchvorschlägen (mit Highlighting der gefundenen Suchbegriffe)

#### **Response Ergebnissuche:**

- nach der Trefferqualität sortierte Liste vollständiger Geoobjekte mit ihren Attributen, einer Geometrie in dem angeforderten Koordinatenreferenzsystem und der Trefferqualität

(ggf. eingeschränkt auf den besten Treffer oder alle Treffer einer bestimmten Qualität oder einer Kombination von beidem)



Ein Beispiel für die Kommunikation im Fall der Vorschlagssuche zeigt Abbildung 1. Die Vorschlagsliste zeigt unter anderem ein serverseitig generiertes Highlighting der gefundenen Suchbegriffe.

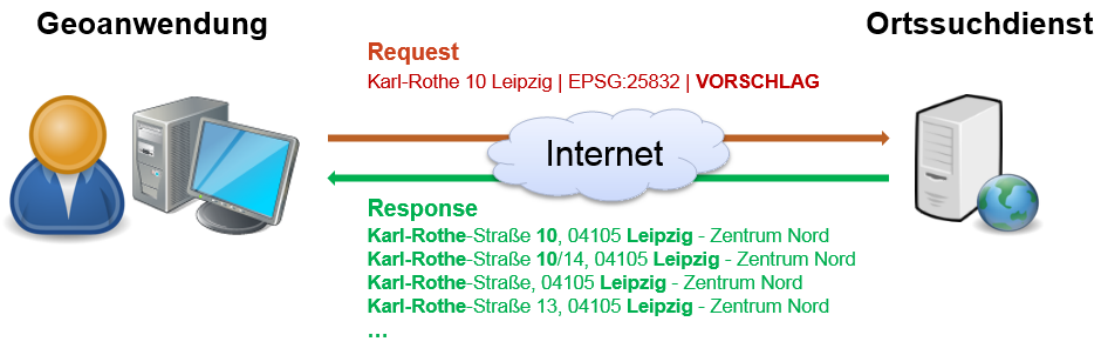


Abbildung 1: Ortssuche. Beispiel für die Kommunikation bei einer Vorschlagssuche.

An die Auswahl eines Vorschlags schließt sich die Ergebnissuche an, die vollständige Objekte mit allen Attributen einschließlich ihrer Lage zurück gibt, siehe Abbildung 2.

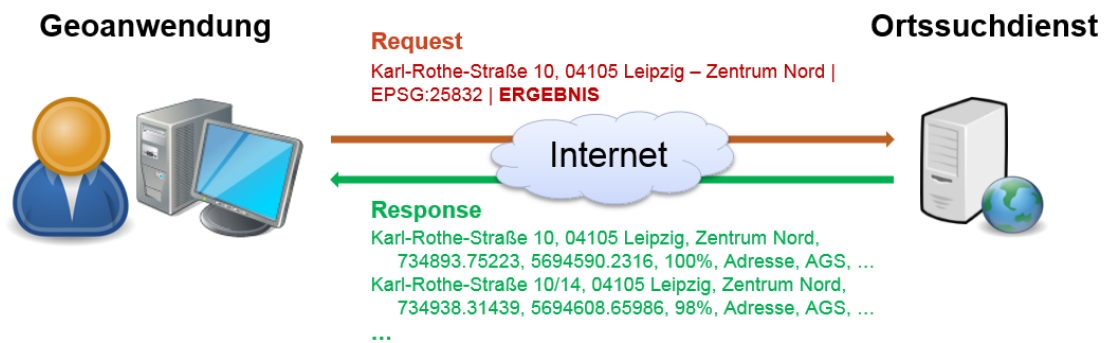


Abbildung 2: Ortssuche. Beispiel für die Kommunikation bei der Ergebnissuche.

### 2.1.2 Hinweise zur Client-Server-Implementierung des Ortssuchdiensts

Um die Funktionalität der Vorschlagssuche zu nutzen, muss auf Clientseite eine interaktive Eingabe verwirklicht werden, die bereits während der Eingabe eines Suchbegriffes Anfragen an den Dienst mit der Anforderung von Vorschlägen sendet. Die zurück gegebenen Vorschläge sind dem

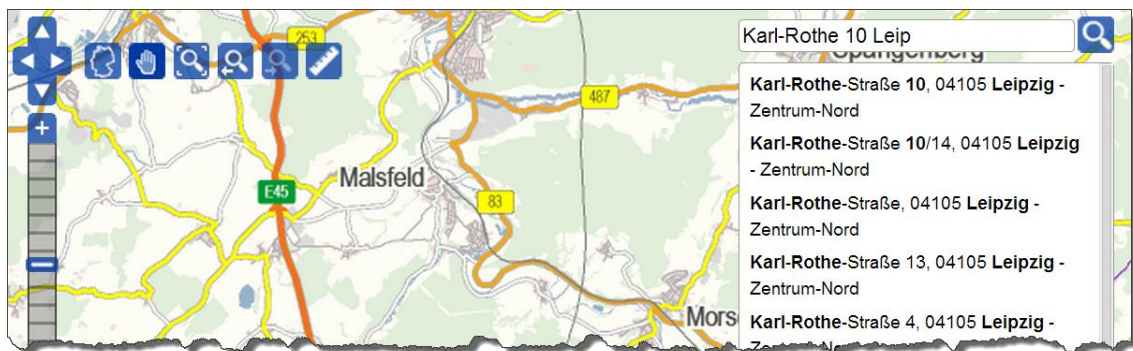


Abbildung 3: Beispiel für eine Vorschlagssuche in einer Anwendung

Benutzer in einer Auswahlliste anzubieten. Ein Beispiel hierfür zeigt **Fehler! Verweisquelle konnte nicht gefunden werden.**

Mit Fortsetzung der Eingabe des Suchbegriffes ist eine Aktualisierung der Auswahlliste durchzuführen. Vorschlagsanfragen sind in der Anwendung von Ergebnisanfragen zu unterscheiden. Hat der Nutzer beispielsweise aus der Liste der Vorschläge eine Auswahl getroffen, ist eine Ergebnisanfrage (die eigentliche Geokodierungsanfrage) an den Dienst zu richten.

### 2.1.3 Anmerkung zur Verwendung des Begriffs „Gazetteerdienst“

In den Unterlagen der GDI-DE wird alternativ zum Begriff *Ortssuchdienst* auch vom *Gazetteerdienst* gesprochen. Der Begriff *Gazetteerdienst* wird allerdings bereits im Zusammenhang mit dem *Gazetteer Service - Application Profile of the Web Feature Service Best Practice*. [4] des *Open Geo-spatial Consortium (OGC)* verwendet. Der hier beschriebene Ortssuchdienst hat jedoch nichts mit dieser Spezifikation gemeinsam. Um Verwechslungen zu vermeiden, sollte der Begriff *Gazetteerdienst* nur für das spezielle WFS-Profil des OGC verwendet werden.

## 2.2 Geokodierungsdienst

### 2.2.1 Beschreibung des Geokodierungsdienstes

Ein Geokodierungsdienst ist ein Geodatendienst, der einzelnen oder massenhaften Objekten mit geographischen Identifikatoren eine Koordinate (Geokodierung) zu deren persistenten Speicherung zuweist.

Der Request und der Response beinhalten die nachfolgend dargestellten, teilweise optionalen Informationen.

#### **Request:**

- identifizierende Angaben zu einem oder mehreren räumlichen Objekten (postalische Adressen, Bezeichnungen eines Flurstücks, geographische Namen, POI oder Kilometrierungsangaben) – jeweils strukturiert über einzelne Attribute beschrieben oder unstrukturiert in einem einzigen Suchattribut angegeben
- das Koordinatenreferenzsystem, in dem die Geokodierung erfolgen soll
- attributive und/oder räumliche Filterangaben zur Einschränkung der Suche(n)
- funktionale Parameter zur Steuerung der Rückgabeliste  
(z.B. Mindestqualität, maximale Anzahl zurück zu gebender Objekte)

#### **Response bei genau einem angefragten Objekt:**

- nach der Trefferqualität sortierte Liste vollständiger Geoobjekte mit ihren Attributen, einer Geometrie in dem angeforderten Koordinatenreferenzsystem und der Trefferqualität  
(ggf. eingeschränkt auf den besten Treffer oder auf eine Mindest-Trefferqualität bzw. auf eine Kombination von beidem)

### Response bei mehreren angefragten Objekten:

- eine Liste vollständiger Geoobjekte mit ihren Attributen, einer Geometrie in dem angeforderten Koordinatenreferenzsystem und der Trefferqualität für jedes angefragte Objekt

(ggf. für jedes angefragte Objekt eingeschränkt auf den besten Treffer oder auf eine Mindest-Trefferqualität bzw. auf eine Kombination von beidem)

Abbildung 4 zeigt ein Beispiel für die Kommunikation zwischen Anwendung und Dienst. In diesem Fall wurde die Anzahl der Antworten auf eine einzige eingeschränkt. Für die Geokodierung ganzer Register ist diese Einschränkung besonders geeignet.

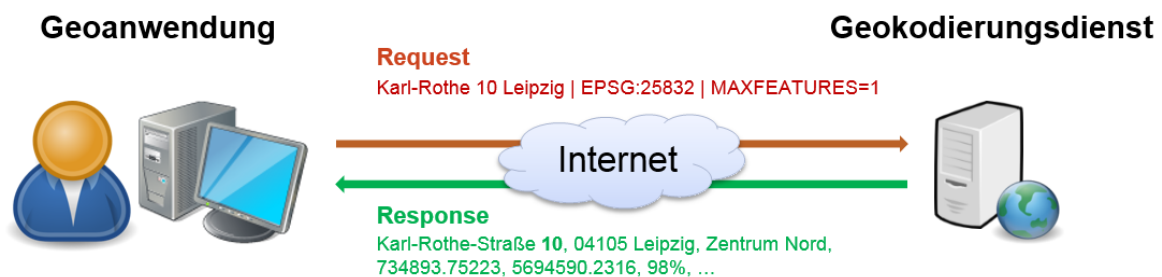


Abbildung 4: Beispiel für die Kommunikation mit einem Geokodierungsdienst.

## 2.2.2 Varianten der Client-Server-Kommunikation

Eine wesentliche Anwendung des Geokodierungsdienstes ist die Geokodierung ganzer Register, das heißt die Verarbeitung von Massendaten. Massendaten können dabei in einer „großen“ Anfrage oder in vielen „kleinen“ Anfragen an den Dienst verarbeitet werden. Beide Vorgehensweisen verlangen bestimmte Kommunikationsformen zwischen Clientanwendung und Serverdienst, die im Folgenden erläutert und diskutiert werden.

### 2.2.2.1 Asynchrone Geokodierung

Die asynchrone Geokodierung ist durch einen zeitlich versetzten Response des Servers ohne Blockieren des Kommunikationsprozesses seitens des Clients charakterisiert. Sie ist geeignet, große Datenmengen geschlossen an einen Server zur Bearbeitung zu übergeben. Hier der grundsätzliche Ablauf:

- Der Client sendet einen Datenstrom (eine Datei) mit allen zu georeferenzierenden Objekten an einen Server (Geokodierungsdienst).
- Aufgrund der Datenmenge und der daraus resultierenden Bearbeitungszeit wartet der Client nicht sofort auf das Geokodierungsergebnis. Er schließt die Kommunikation.
- Der Dienst prüft den Datenstrom, reagiert auf Fehler oder geokodiert die Datei bei Verarbeitbarkeit. Das Ergebnis wird auf dem Server für eine gewisse Zeit zum Download verwaltet.

- Der Client fragt den Geokodierungsdienst zyklisch an, ob dieser bereits fertig ist und lädt das Ergebnis nach Abschluss der Geokodierung.

Dieses Kommunikationsschema lässt sich prinzipiell mit Standards des OGC realisieren. Der relativ komplexe und aufwändig zu implementierende Web Processing Service (WPS) unterstützt eine asynchrone Kommunikation, in der ein Client zyklische Statusabfragen an den Server sendet (polling).

Die asynchrone Kommunikation hat den Vorteil, dass ein Client eine ganze Datei an den Server übergeben kann anstelle von vielen Einzelanfragen. Bei genauer Betrachtung besitzt diese Implementierung aber eine Reihe von Nachteilen auf Client- und Serverseite.

Nachteile:

- Eine Implementierung des WPS-Standards ist auf Client- und Serverseite aufwändig.
- Die Anwendung des Nutzers muss während der gesamten Bearbeitung durch den Server bestehen bleiben und das Ergebnis nach dessen Abschluss entgegen nehmen und verarbeiten. Bei sehr großer Objektanzahl kann dies einige Zeit in Anspruch nehmen. Für den Abbruchfall muss eine Technik für den Wiederaufsatz realisiert werden.
- Für die korrekte Zuordnung der Objekte aus der Anfrage und der Ergebnisse auf Clientseite sind besondere Maßnahmen zu ergreifen.
- Der Server erhält ganze Datenströme (Ressourcenbindung) bevor er diese auf ihre Korrektheit prüft und verarbeitet.
- Auf Serverseite muss das Geokodierungsergebnis bis zu seinem vollständigen Download durch den Client verwaltet werden.
- Die Datenverwaltung auf Serverseite erfordert besondere Maßnahmen unter den Aspekten des Datenschutzes.
- In einem skalierbaren und verteilten Serverumfeld sind Maßnahmen zur Kommunikation eines Clients mit einem dedizierten Server zu ergreifen.

### 2.2.2.2 Synchroner Geokodierung

Die synchrone Geokodierung ist durch einen sofortigen Response des Servers bei kurzzeitigem Blockieren des Kommunikationsprozesses (Warten auf die Antwort) seitens des Clients charakterisiert. Eine solche Kommunikation ist ideal für kleine Datenpakete an den Server, die schnell verarbeitet werden können. Hier der typische Ablauf:

- Der Client liest den Datenstrom (z.B. Datei), analysiert ihn und reagiert auf Fehler.
- Der Client sendet sequentiell Einzelanfragen (in ggf. mehreren parallelen Threads) an den Server und wartet sofort auf das schnell verfügbare Ergebnis.
- Der Client speichert die Ergebnisse sequentiell (z.B. in einer Datei).

Beim Senden von Einzelanfragen in einer synchronen Kommunikation muss der Client den Datenstrom analysieren und in Requests wandeln. Dadurch ergeben sich die folgenden Nachteile:

- Die Anwendung des Nutzers muss für den kompletten Verarbeitungsauftrag bestehen bleiben. Bei sehr großer Objektanzahl kann dies einige Zeit in Anspruch nehmen. Für den Abbruchfall muss eine Technik für den Wiederaufsatz realisiert werden.
- Eine nicht nur kurzzeitige Unterbrechung der Netzverbindung (abhängig von der Programmierung des Clients) führt zum Abbruch der Verarbeitung. Für den Abbruchfall sollte eine Technik für den Wiederaufsatz realisiert werden.

Verglichen mit den oben angeführten Nachteilen einer asynchronen Kommunikation ergeben sich dabei allerdings wesentliche Vorteile.

Vorteile:

- + Eine Implementierung kann auch auf Grundlage des weitaus besser etablierten und „schlankeren“ WFS-Standards (Web Feature Service) erfolgen.
- + Ressourcenbindung und Verwaltungsaufwand sind auf Serverseite deutlich geringer.
- + Datenschutzprobleme treten nicht auf, da keine Speicherung von Daten erfolgt.
- + Eine Kontrolle der Daten erfolgt anwendungsspezifisch schon vor ihrem Versand an den Server.
- + Die Zuordnung der Ergebnisse zu den Ausgangsdaten kann der Client einfach realisieren.

### **2.2.2.3 Schlussfolgerung**

Die synchrone Kommunikation erlaubt insgesamt eine weniger aufwändige Implementierung auf Client- und auf Serverseite. Der Web Feature Service kann die Grundlage für die Implementierung des Dienstes bilden. Die Geokodierung von Massendaten in dieser Kommunikationsform wird am BKG bereits seit längerem erfolgreich durchgeführt.

### **2.2.3 Geokodierungsanwendung**

Ein Geokodierungsdienst dient wie jeder Webdienst der Maschine – Maschine – Kommunikation. Ein Nutzer arbeitet dagegen mit einer Anwendung, die ihrerseits im Hintergrund Webdienste verwenden kann. Für die Geokodierung von Massendaten müssen den Nutzern Anwendungen zur Verfügung stehen, die spezifisch auf die jeweiligen Ausgangsdaten und die gewünschten Ergebnisdaten ausgerichtet sind. So können Ausgangsdaten in speziellen Datenbanken vorliegen oder in einfachen Excel-Tabellen. Die Georeferenzierungsergebnisse müssen in besondere Register zurück geschrieben werden oder es genügt dem erfahrenen Anwender beispielsweise eine CSV-Datei zur eigenen weiteren Verarbeitung.

Für einen optimalen Informationsfluss sind also spezielle Geokodierungsanwendungen erforderlich, die auf die jeweilige Infrastruktur und Datenstruktur ausgerichtet sind, vgl. Abbildung 5.



**Abbildung 5: Geokodierung von Massendaten über eine Geokodierungsanwendung.**

Die Anwendung gestattet eine nutzerorientierte Arbeit mit ganzen Dateien räumlicher Identifikatoren und eine Massengeokodierung kann über zeiteffiziente parallele synchrone Anfragen an einen Geokodierungsdienst erfolgen. Ferner kann eine graphische Oberfläche mit einer interaktiven Karte zur Klärung fehlgeschlagener Geokodierungen und zur interaktiven Zuordnung von Lageinformationen eine sinnvolle Komponente einer solchen Applikation darstellen.

### 2.3 Reverser Geokodierungsdienst

Ein reverser Geokodierungsdienst ist ein Geodatendienst, der ausgehend von einer Koordinate (Punkt) oder einem umfassenden Gebiet (Rechteck, Polygon) alle geographischen Identifikatoren mit einem bestimmten räumlichen Bezug (vollständig enthalten, angeschnitten) ermittelt.

Der Request und der Response beinhalten die nachfolgend dargestellt, teilweise optionalen Informationen.

#### Request:

- Suchgeometrie: Punkt mit Radius oder Rechteck oder Polygon
- Koordinatenreferenzsysteme der Suchgeometrie und der angeforderten Geoobjekte
- attributive Filterangaben zur Einschränkung der Suche
- räumlicher Operator (vollständig enthalten, angeschnitten)

#### Response:

- Liste vollständiger Geoobjekte (Adressen, Flurstücke, Geonamen, POI) mit ihren Attributen und einer Geometrie in dem angeforderten Koordinatenreferenzsystem

### 3 Inhaltliches Stufenkonzept

Die zunächst funktional unterschiedenen Dienstetypen sind inhaltlich nach unterstützten Objektarten zu gliedern. Als Anforderungen wurden durch das LG GDI-DE mit abnehmender Priorität von oben nach unten genannt:

- a) postalische Adressen
- b) geographische Namen
- c) Flurstücke
- d) Kilometrierungen
- e) POI's

Die Datenbestände a), b) und c) befinden sich in der direkten Zuständigkeit der AdV und sind für Geokodierungsdienste verfügbar.

Kilometrierungssysteme sind Gegenstand von Fachinformationssystemen und werden deshalb im Rahmen dieser Arbeitsgruppe nicht weiter betrachtet.

Points-of-Interest haben einen engen Bezug zu geographischen Namen, insbesondere lässt sich auch das gleiche Datenmodell anwenden. Einige Kategorien wurden bereits durch verschiedene AdV-Mitglieder, so auch durch das BKG, erhoben.

Für die Inhalte a), b) und e) können damit kurzfristig konkrete Lösungen erarbeitet werden, die auf vorhandenen Diensten basieren. Dies betrifft die Ortssuche, die Geokodierung und die Reverse Geokodierung. Diese Reihenfolge entspricht auch der zeitlichen Priorisierung wie sie seitens des LG GDI-DE gesehen wird. Dies wird im nächsten Kapitel weiter ausgeführt.

Auf Länderebene stehen die Flurstücke zur Verfügung. Ihre Bereitstellung für einen deutschlandweiten Geokodierungsdienst ist zunächst rechtlich und organisatorisch in der AdV zu klären. Die Lösungsmöglichkeiten für einen Geokodierungsdienst umreißt Kapitel 5 *Geokodierungsdienst für Flurstücke*.

## 4 Geokodierungsdienst für Adressen, Geonamen, POI

Die Bezeichnung *Geokodierungsdienst* steht hier für eine integrierte Implementierung der oben dargestellten Ortssuche, Geokodierung und Reversen Geokodierung.

### 4.1 Vergleichende Analyse vorhandener Lösungsansätze

Für den Geokodierungsdienst gibt es verschiedene Lösungsansätze, die bereits durch den LG GDI-DE vergleichend betrachtet worden sind. Aufgrund der besten Funktionalität konnte der Vergleich sofort auf die Lösungen ans NRW und BKG eingeschränkt werden. Die Analyse wurde durch die Arbeitsgruppe verfeinert (siehe Anlage Geokodierung\_Vergleich.xls) und durch Praxistests untersetzt. Im Ergebnis einigte sich die AG auf den Ansatz des BKG.

Es wird vorgeschlagen, den Geokodierungsdienst des BKG, installiert an dessen Dienstleistungszentrum, als Geokodierungsdienst der AdV für Adressen, Geonamen und POI einzusetzen. Schon jetzt liegt eine einsatzbereite Lösung vor, die unter anderem im Bundesbereich und im Geportal.de verwendet wird. Im laufenden Jahr 2014 werden darüber hinaus noch verschiedene Weiterentwicklungen vorgenommen werden, die sich aus den folgenden Abschnitten ergeben.

### 4.2 Verwendung von Standards

#### 4.2.1 Verwendung standardisierter Dienste zur Geokodierung

Der Geokodierungsdienst wird in zwei Schnittstellenversionen realisiert:

- a) standardisierter Web Feature Service gemäß AdV-WFS-Profil  
(mit einem Rückgabe-Datenstrom in GML)
- b) standardisierter OpenSearch GeoTemporal Service  
(unter anderem mit einem Rückgabe-Datenstrom in GeoJSON)

Mit der WFS-konformen Implementierung gemäß Profil der AdV wird eine etablierte standardisierte Schnittstelle des OGC unterstützt.

Das zusätzliche Angebot eines einfachen OpenSearch GeoTemporal Service und der Einsatz von GeoJSON entsprechen der Forderung nach größtmöglicher Einfachheit bei der Entwicklung von Webanwendungen.

#### 4.2.2 Anmerkungen zum DOG-Profil

Das durch viele Bundesländer bereits erfolgreich eingesetzte DOG-Profil HKFK Version 2.0.0 definiert eine Service-Schnittstelle zum Zugriff auf Haus- und Flurstückskoordinaten. Es basiert auf dem *Gazetteer Service - Application Profile of the Web Feature Service Best Practice* [4] (WFS-G) und legt Feature Types und deren Attribute in Bezug auf die Haus- und Flurstückskoordinaten Deutschlands fest.

Das Datenmodell und die funktionalen Eigenschaften des DOG sind auf eine strukturierte Suche über einzelne Attribute ausgelegt. Das Anwendungsszenario einer unstrukturierten Volltextsuche wird nicht abgebildet. Der Lösungsansatz im Abschnitt 4.4 trägt über die Fähigkeiten des



DOG hinaus auch der Nutzeranforderung unstrukturierter Ausgangsdaten bei Suche und Geokodierung Rechnung.

Unschärfe Suchen werden im DOG über die „Normalisierung“ von Namen und über den für den englischen Sprachraum entwickelten Soundex-Algorithmus durch spezielle Attribute und Funktionen ermöglicht. Der Lösungsansatz im Abschnitt 4.4 erweitert die Funktionalität um das Konzept der Rückgabe einer Objektliste mit Trefferqualitäten und entlastet den Anwender von einer durch ihn zu steuernden unscharfen Suche.

Mittlerweile existieren gute technologische Open-Source-Lösungen für unscharfe Volltextsuchen, die insgesamt noch mehr Möglichkeiten bieten, als sie im DOG-Profil aufgeführt sind. Die im DOG-Datenmodell enthaltenen Attribute für normalisierte und phonetische Werte werden hierdurch überflüssig. Für die unscharfe Suche ist darüber hinaus ein sehr einfaches Datenmodell von Vorteil. Es erleichtert die Suchanfragen und kann dennoch die Suchergebnisse vollständig abbilden. Auch die Integration eines WFS in bestehende GIS-Anwendungen wird so noch besser unterstützt. Dem trägt das Datenmodell unter Punkt 4.4.1 Rechnung.

Da sich die Anforderungen hinsichtlich unstrukturierter und fehlertoleranter Suchen bei Adressen und Flurstücken unterscheiden, wird das DOG-Profil bei der Konzeption eines Geokodierungsdienstes für Flurstücke weiter berücksichtigt werden.

### 4.3 Anforderungen und Weiterentwicklungsbedarf

In den folgenden Tabellen sind die wesentlichen zu erfüllenden Anforderungen an den Geokodierungsdienst zusammengestellt. In der Spalte *Realisierung* wird der erreichte Stand dargestellt.

Tabelle 1 enthält die für alle Dienstetypen relevanten Anforderungen bei der vorgesehenen Implementierung.

**Tabelle 1: Gemeinsame Anforderungen an den Geokodierungsdienst.**

Kategorie	Anforderung	Realisierung
Datengrundlage	postalische Adressen	Datensatz Hauskoordinaten der AdV
	geographische Namen	GN250 des BKG <i>oder</i> GN-DE basierend auf Basis-DLM
	POI	offen
Implementierung	offene Software	Eigenentwicklung des BKG auf Basis offener Software (Apache Solr, Tomcat, Java)
	Skalierbarkeit	durch Server-Virtualisierung am BKG gewährleistet
Dienstqualität	Hochverfügbarkeit	durch Standortredundanz und lokale Lastverteilung am BKG gewährleistet

	Performance (Antwortzeit und Datendurchsatz)	Anforderungen werden erfüllt (Millisekundenbereich und erforderlicher Datendurchsatz durch Skalierbarkeit)
Datenmodell	Adressen, Geonamen und POI in einem gemeinsamen Datenmodell	Adressen und Geonamen sind modelliert
Schnittstelle	Verwendung von Standards	WFS und OpenSearch GeoTemporal Service erfüllen dies
	Unterstützung verschiedener Koordinatenreferenzsysteme	vorhanden
	attributive Filterung	vorhanden

Tabelle 2 fasst die besonderen Anforderungen aus Sicht der Ortssuche und Geokodierung zusammen (vgl. 2.1.1 und 2.2.1).

**Tabelle 2: Anforderungen an Ortssuche und Geokodierung.**

Kategorie	Anforderung	Realisierung
Schnittstelle	unstrukturiertes Suchattribut	vorhanden
	strukturierte Suchattribute	vorhanden
	räumliche Filterung	vorhanden
	Vorschlagsliste	vorhanden
	Ergebnisliste mit Score	vorhanden
Funktionalität	Fehlertoleranz durch Normierung	vorhanden
	Fehlertoleranz durch Tokenbasierte Suche	vorhanden
	Fehlertoleranz durch Fuzzy-Suche	vorhanden

Tabelle 3 betrachtet die besonderen Anforderungen der Reversen Geokodierung (vgl. 2.3) und deren Umsetzungsstand.

**Tabelle 3: Anforderungen an die Reverse Geokodierung.**

Kategorie	Anforderung	Realisierung
-----------	-------------	--------------

Schnittstelle	Suchgeometrie Punkt (Suche des nächsten Objekts)	vorhanden
	Suchgeometrie Punkt mit Radius (Umkreissuche)	vorhanden
	Suchgeometrie Rechteck	vorhanden
	Suchgeometrie Polygon	genähert über einen GeoHash vorhanden
	Geometrische Operatoren	Within, Intersects

## 4.4 Schnittstellenbeschreibung Geokodierungsdienst

Nachfolgend werden das Datenmodell und die Schnittstellen für den Geokodierungsdienst beschrieben.

### 4.4.1 Datenmodell

Das nachfolgend beschriebene Datenmodell basiert auf einer flachen Struktur (einfache Objekttypen mit tabellarischen Attributen und ohne relationale Objektverknüpfungen), auf der besonders leistungsfähige Suchalgorithmen implementiert werden können. Dies entspricht dem am BKG bereits vorhandenen Lösungsansatz.

Eine Berücksichtigung und Implementierung der INSPIRE-Datenmodelle für Adressen und geographische Namen wird zum jetzigen Zeitpunkt für verfrüht erachtet. Die komplexen, relationalen INSPIRE-Datenmodelle, die nur über spezielle Erweiterungen vorhandener Geodatenserver unterstützt werden können und folglich ebenso auf Clientseite offene Fragen aufwerfen, befinden sich in einer noch anhaltenden Erörterung. Insbesondere angesichts des zeitnahen Bedarfs an Lösungen in Deutschland sollten INSPIRE-konforme Weiterentwicklungen zu einem späteren Zeitpunkt nach Konsolidierung der Datenmodelle erfolgen.

Der zur Umsetzung des EGovG rechtzeitig verfügbare, hocheffiziente Geokodierungsdienst besitzt in seiner flachen Datenstruktur die beiden grundlegenden Datentypen *Vorschlag* (Rückgabe bei Vorschlagsuchen) und *Ortsangabe* (Rückgabe bei Ergebnissuchen).

Die obligatorischen und optionalen Attribute der beiden Datentypen *Vorschlag* und *Ortsangabe* sind in den beiden folgenden Tabellen zusammengestellt.

**Tabelle 4: Attribute des Datentyps *Vorschlag*.**

Attribut	P/O <sup>1</sup>	Datentyp	Beschreibung
text	P	String	Volltext-Repräsentation des Objekts.

<sup>1</sup> P - Pflichtattribut / O - optionales Attribut

<b>score</b>	0	Double	Qualitative Bewertung des Objekts. Dieses Attribut ist abhängig von der Sucheingabe und wird für die Antwort an den Client generiert.
--------------	---	--------	---

**Tabelle 5: Attribute des Datentyps *Ortsangabe*.**

<b>Attribut</b>	<b>P/O</b>	<b>Datentyp</b>	<b>Beschreibung</b>
<b>id</b>	P	String	Die eindeutige ID (Zeichenfolge) des Objekts zur Identifizierung im Dienst.
<b>text</b>	P	String	Volltext-Repräsentation des Objekts.
<b>typ</b>	P	String	Klassifizierung des Objekts (Haus, Straße, Ort, Geoname, POI)
<b>score</b>	0	Double	Qualitative Bewertung des Objekts. Dieses Attribut ist abhängig von der Sucheingabe und wird für die Antwort an den Client generiert.
<b>geometry</b>	P	Point	Punktgeometrie des Objektes.
<b>bbox</b>	P	Polygon	BBOX-Geometrie des Objektes. Wird zum Zoom auf passenden Kartenausschnitt benötigt.
<b>ags</b>	0	String	Amtlicher Gemeindegchlüssel.
<b>rs</b>	0	String	Regionalschlüssel.
<b>bundesland</b>	0	String	Name des Bundeslandes, der das Objekt enthält.
<b>regbezirk</b>	0	String	Name des Regierungsbezirks, der das Objekt enthält.
<b>kreis</b>	0	String	Name des Kreises, der das Objekt enthält.
<b>verwgem</b>	0	String	Name der Verwaltungsgemeinschaft, der das Objekt enthält.
<b>gemeinde</b>	0	String	Name der Gemeinde, die das Objekt enthält
<b>plz</b>	0	String	Postleitzahl der Adresse.  Wird nur bei Objekten mit type=[ Haus   Strasse   Ort ] verwendet.
<b>ort</b>	0	String	Ortsname der Adresse.  Wird nur bei Objekten mit type=[ Haus   Strasse   Ort ] verwendet.

<b>ortsteil</b>	0	String	Ortsteilname der Adresse.  Wird nur bei Objekten mit type=[ Haus   Strasse   Ort ] verwendet.
<b>strasse</b>	0	String	Straßenname der Adresse.  Wird nur bei Objekten mit type=[ Haus   Strasse ] verwendet.
<b>haus</b>	0	String	Hausnummer mit Hausnummernzusatz der Adresse.  Wird nur bei Objekten mit type=Haus verwendet.
<b>name</b>	0	String	Geographischer Name des Objekts.  Wird nur bei Objekten mit type=[ Geoname   POI ] verwendet.
<b>kategorie</b>	0	String	Thematische Kategorie (Landschaft, Schule, ...) des Objekts.  Wird nur bei Objekten mit type=[ Geoname   POI ] verwendet.

#### 4.4.2 Schnittstellen

##### 4.4.2.1 Web Feature Service

Der Geokodierungsdienst implementiert einen Basic Web Feature Service in Version 1.1.0 und unterstützt damit eine etablierte standardisierte Schnittstelle des OGC. Hierbei werden die folgenden Operationen unterstützt:

- GetCapabilities
- DescribeFeatureType
- GetFeature

Während GetCapabilities und DescribeFeatureType nur Metadaten und beschreibende Informationen zum Dienst liefern, werden die eigentlichen Suchanfragen mit GetFeature durchgeführt. Mit Hilfe des FeatureTypes kann zwischen Vorschlagssuche (TYPENAME=Vorschlag) oder Ergebnissuche (TYPENAME=Ortsangabe) gewählt werden.

Die folgenden Filteroperatoren werden vom Dienst unterstützt:

- BBOX
- DWithin
- Intersects
- PropertyIsEqualTo
- PropertyIsLike
- And

- Or
- Not

Der Operator PropertyIsLike wird erweitert interpretiert, um die Vorteile der wortbasierten Volltextsuche auch in der WFS-Schnittstelle zu nutzen. Bei der wortbasierten Volltextsuche liegt bereits ein Treffer vor, wenn einzelne Wörter einer Suchanfrage im entsprechenden Attribut des Objekts vorkommen. Die Qualität der Übereinstimmung von Suchanfrage und dem jeweiligen Treffer beschreibt das Attribut *score*.

Die *Reverse Geokodierung* wird über die räumlichen Operatoren BBOX, DWithin und Intersects mit der erforderlichen Funktionalität implementiert.

#### 4.4.2.2 *OpenSearch GeoTemporal Service*

Der Geokodierungsdienst stellt weiterhin einen ebenfalls durch das Open Geospatial Consortium spezifizierte OpenSearch GeoTemporal Service bereit. „Bei OpenSearch handelt es sich um eine auf XML basierende Sammlung von Techniken, die es ermöglicht, Suchergebnisse von Suchmaschinen und Websites in einem standardisierten und maschinenlesbaren Format auszugeben.“<sup>2</sup> Diese Schnittstelle bietet eine einfache Implementierungsgrundlage und leichte Integrationsmöglichkeiten in Web-anwendungen.

Die Schnittstelle berücksichtigt zudem:

- die OpenSearch Suggestion Extension zur Generierung von Vorschlagslisten für die interaktive Unterstützung der Sucheingabe.
- die OpenSearch Geo and Time Extension des OGC zur Bereitstellung räumlicher Suchfunktionalitäten.

Folgende Operationen werden vom Dienst unterstützt:

- Beschreibung des Dienstes (OpenSearch Description Document)
- Generierung von Vorschlagslisten (suggest)
- Generierung von Ergebnislisten (search)
- Abfrage eines Objekts per ID (record)

Folgende Konformitätsklassen der OpenSearch Geo and Time Extension werden erfüllt:

- **Core (GeoSpatial Service)**
  - o Der Server generiert ein gültiges OpenSearch Description Dokument.
  - o Der Server definiert UTL-Templates für Ergebnisse im Atom-Format
  - o Der Server implementiert eine Bounding-Box Suche
- **Arbitrary Geometry Search**
  - o Der Server implementiert die Suche mit beliebiger Geometrie
- **Point and Radius Search**
  - o Der Server implementiert eine räumliche Suche mittels Punkt und Radius
- **Minimum spatial operator**
  - o Der Server implementiert als räumliche Operation “intersects”

---

<sup>2</sup> Quelle: Wikipedia

- ***Spatial relations***
  - Der Server implementiert weitere räumliche Operationen
- ***Get record by id.***
  - Der Server kann auf einen Datensatz mittels ID zugreifen.
- ***Search by name***
  - Der Server implementiert eine räumliche Suche unter Verwendung des Namen oder der Adresse eines Ortes.

Alle Anfragen an den Dienst erfolgen bei der OpenSearch Schnittstelle mit HTTP-GET. Über das OpenSearch Description Document werden parametrisierte URL-Templates bereitgestellt. Diese instruieren eine Clientanwendung, wie sich die Suchanfragen an den Dienst zusammensetzen. Ein URL-Template besteht aus der HTTP-GET URL inklusive Template-Parameter, die durch die Clientanwendung mit den entsprechenden Werten ersetzt werden.

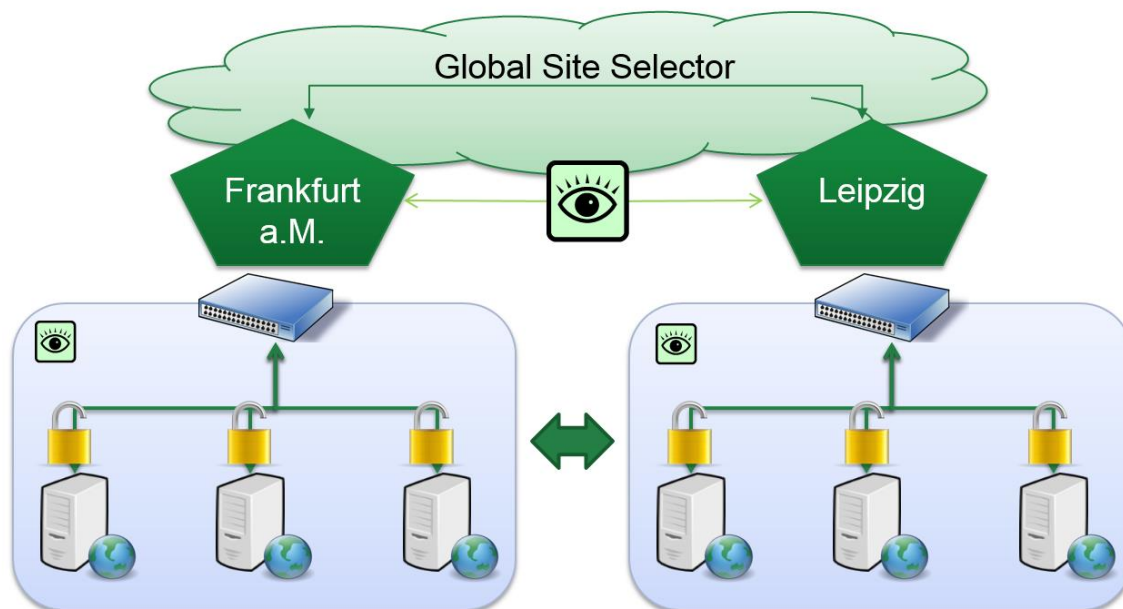
Im Unterschied zum WFS ist die Syntax für die Suchanfragen stark vereinfacht. Das Erfordernis eines komplexen XML-Filterausdrucks entfällt hier und die Parameter sind für die Suche von Ortsangaben optimiert. Der Suchstring wird über einen einzelnen Parameter angegeben. Zusätzliche Parameter haben funktionalen Einfluss auf die Suche, wie das unter Punkt 2.1.1 angeführte Highlighting gefundener Suchbegriffe.

Die Suchergebnisse werden standardmäßig im Atom 1.0-Format ausgegeben. Zudem kann über einen Format-Parameter das Ausgabeformat explizit gewählt werden. Folgende Formate werden vom Dienst unterstützt:

- application/atom+xml (Atom 1.0)
- application/gml+xml (Geographic Markup Language - GML)
- application/json (GeoJSON)
- application/x-suggestions+json (Suchvorschläge)

## 4.5 Implementierung des Dienstes

Die Implementierung des Geokodierungsdienstes erfolgt am Dienstleistungszentrum des BKG auf Grundlage der Open-Source-Software Apache Solr in einem standortredundanten und über lokale Lastverteilung und Servervirtualisierung skalierten GeoWebCluster.



**Abbildung 6: GeoWebCluster des BKG an den Standorten Frankfurt a. M. und Leipzig.**

Abbildung 6 zeigt die Struktur des GeoWebClusters mit seinen Komponenten für Lastverteilung, Sicherheitsmanagement, Synchronisierung und Monitoring.

Die Konfiguration von Solr wird durch das BKG selbst durchgeführt, ebenso wie die Entwicklung der Webdienste mit der WFS- und OpenSearch-Schnittstelle. Die Dienste sind Java Servlets mit einer Implementierung unter Tomcat.



## 5 Geokodierungsdienst für Flurstücke

Die Bezeichnung *Geokodierungsdienst* steht für eine integrierte Implementierung der in Kapitel 2 dargestellten Ortssuche, Geokodierung und reversen Geokodierung, hier für Flurstücke. Die Funktionalität einer *schnellen Vorschlagssuche* wird allerdings nicht als erforderlich angesehen. Im Umgang mit Flurstücken darf von einer exakten Kenntnis der Attribute ausgegangen werden und die Eingabe der Parameter wird strukturiert und nicht in einer unstrukturierten Volltextzeile erfolgen.

### 5.1 Erforderliche Flurstücksinformationen

#### 5.1.1 Attributive Eigenschaften

Zu den wichtigsten attributiven Eigenschaften eines Flurstücks zählen die Flurstücksnummer sowie die Zugehörigkeit zu Land, Gemarkung und Flur. Diese Informationen sind für eine erfolgreiche Suche und Geokodierung erforderlich.

Mit der Einführung von ALKIS ist ein eindeutiges, deutschlandweites Kennzeichen zur Identifikation eines Flurstücks verfügbar. Es besitzt 20 Stellen und setzt sich in der folgenden Reihenfolge aus Nummern für die genannten Attribute zusammen (vgl. AdV 2008).

Aufbau des Flurstückskennzeichens:

- Land (2 Stellen)
- Gemarkung (4 Stellen)
- Flur (3 Stellen)
- Flurstück (9 Stellen)
  - Zähler (5 Stellen)
  - Nenner (4 Stellen)
- Flurstücksfolge (2 Stellen)

Attribute zur administrativen Zugehörigkeit in Textform und der Amtliche Gemeindeschlüssel können die Suche sinnvoll unterstützen:

- Name des Landes
- Name des Regierungsbezirks (optional)
- Name des Kreises
- Name der Verwaltungsgemeinschaft (optional)
- Name der Gemeinde
- AGS

#### 5.1.2 Geometrische Informationen

Für die Geokodierung ist eine einzelne repräsentative Koordinate (Flurstücksordinate) ausreichend. Das Umringspolygon wird für die reverse Geokodierung benötigt, kann aber auch zur Generierung der Flurstücksordinate erforderlich sein, wenn diese im Datenbestand nicht geführt wird. Zudem kann aus dem Umringspolygon eine Bounding-Box ermittelt werden, die für das Setzen des Kartenausschnitts in der Suche hilfreich ist.

Erforderliche geometrische Informationen sind demnach:

- Flurstücksordinate
- Umringspolygon

## 5.2 Grundlagen in den Bundesländern Ende 2014

Im Oktober 2014 hat der AK LK in Zusammenarbeit mit der AG Geokodierungsdienste eine Umfrage unter den Bundesländern durchgeführt. Ziel war die Erhebung des aktuellen Standes hinsichtlich der erforderlichen Flurstücksinformationen und deren Bereitstellung über Dienste oder Datensätze sowie die Ermittlung der strategischen Ausrichtung der Länder. Die wichtigsten Aussagen werden nachfolgend wiedergegeben.

### 5.2.1 Ausgangslage bei den Daten

Minimal erforderlich sind für die Geokodierung die Führung von Flurstückskennzeichen und Flurstücksordinate in einem Datenbestand. Diese beiden Informationen lagen **in 9 Bundesländern** (BW, BB, HH, HE, NW, RP, SL, ST, TH) vor.

### 5.2.2 Ausgangslage bei den Diensten

Ebenfalls **9 Bundesländer** (BW, BB, HH, HE, NI, NW, RP, SL, TH) verfügten bereits über einen Dienst, der für eine Geokodierung eingesetzt werden könnte. Sieben unterschiedliche Dienstspezifikationen wurden dabei vorgefunden, da auch die in drei Ländern eingesetzten WFS unterschiedliche Datenmodelle besaßen. Abbildung 7 gibt hierzu einen graphischen Überblick.

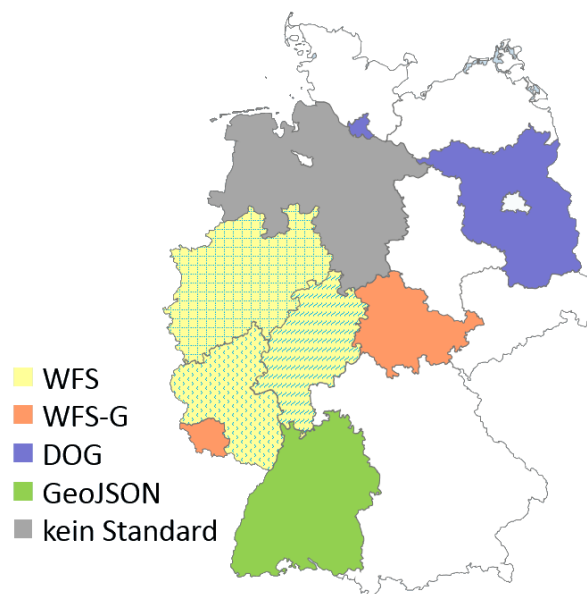


Abbildung 7: Übersicht über die Bundesländer mit einem Flurstücksdienst

Insgesamt wurden in den Antworten der Bundesländer sechs verschiedene Spezifikationen auf Basis der WFS-Schnittstelle angegeben:

- 1) OGC-Spezifikation WFS 1.1.0
  - a) mit eigenem Profil
  
- 2) OGC-Spezifikationen WFS 1.1.0/2.0.0 sowie „Gazetteer Service- Application Profile of the Web Feature Service Implementation Specification“ (WFS-G)
  - a) Deutschland Online Gazetteer-Profil Hauskoordinaten und Flurstückskoordinaten (DOG-Profil HKFK), 2.0.0
  - b) mit eigenem Profil
  
- 3) OGC-Spezifikationen WFS 1.1.0/2.0.0 sowie AdV-ALKIS-WFS-Produktspezifikation
  - a) gemäß Schemavariante „AAA-Modell-konform“
  - b) gemäß Schemavariante „NAS-konform“
  - c) gemäß Schemavariante „Vereinfachtes Datenaustauschschemata“

Eine genaue Untersuchung der vorhandenen Dienste zeigte jedoch, dass derzeit nicht alle WFS-G wirklich dieser Spezifikation entsprechen, sondern lediglich als WFS arbeiten.

Die Antworten zeigen grundsätzlich, **dass kurz- und mittelfristig nicht in allen Bundesländern einheitliche Dienste oder überhaupt Dienste für Flurstücke zur Verfügung stehen werden**. Für einen deutschlandweiten Geokodierungsdienst für Flurstücke sind derzeit **notwendig auch NAS-Datensätze vieler Bundesländer mit heranzuziehen**. Insbesondere erforderlich ist dabei die Objektart „AX\_Flurstueck“, mit der die wesentlichen Anforderungen bereits erfüllt werden.

### 5.2.3 Langfristige Entwicklung

Für jeden deutschlandweiten Geokodierungsdienst sind möglichst einheitliche Grundlagen in den Bundesländern hilfreich. Einheitliche Ausgangsdienste erlauben effiziente Lösungen, die in den nächsten Kapiteln diskutiert werden.

Um künftige einheitliche Grundlagen herauszuarbeiten, wurde die Frage nach langfristig eingesetzten Spezifikationen und Profilen gestellt. Das Ergebnis ist nachfolgend zusammengefasst, in Klammern die Anzahl der Bundesländer mit entsprechender Antwort.

- Ausrichtung nach konkreter Spezifikation und Schnittstelle:
  - DOG-Profil (2)
  - AdV-ALKIS-WFS (7)
  - AdV-ALKIS-Shape (5)
  - GeoJSON (1)
  
- Ausrichtung nach anderen Vorgaben:
  - alle geforderten Spezifikationen der AdV (4)
  - nur von Kunden verlangte Spezifikationen (1)
  - Fokus INSPIRE + wirtschaftliche/finanzierbare Spezifikationen (1)
  - nach der ALKIS-Einführung zu spezifizieren (1)

Auf die Schlussfolgerungen aus diesen Antworten wird im nächsten Punkt eingegangen.

### 5.3 Verwendung von Standards

Auch der Geokodierungsdienst für Flurstücke soll auf internationalen Standards basieren. Die Verwendung des WFS ist naheliegend.

Die Antworten der Bundesländer zeigen, dass ein WFS-G und das DOG-Profil keine erkennbare Rolle in der strategischen Entwicklung spielen. Das DOG-Profil würde zudem aufgrund der fehlenden polygonalen Beschreibung von Flurstücken die Reverse Geokodierung nicht unterstützen.

Erkennbar ist die vorrangige Ausrichtung der Bundesländer auf die Implementierung des AdV-ALKIS-WFS. Dieser erfüllt alle Anforderungen der Geokodierung in jeder seiner Schemavarianten.

Von den unter Punkt 5.2.2 bereits angeführten drei Schemavarianten des AdV-ALKIS-WFS bietet sich die einfachste Form „Vereinfachtes Datenaustauschschemata“ am meisten für eine Verwendung in einem Geokodierungsdienst an. Ein möglichst einfaches Datenmodell erleichtert die dynamische Verarbeitung von Daten bei Anfragen an den zentralen Dienst oder die Dienste der Bundesländer sowohl in einer Prozesskette als auch beim Nutzer des Dienstes selbst.

***Es wird deshalb vorgeschlagen, den zentralen AdV-Geokodierungsdienst für Flurstücke als AdV-ALKIS-WFS mit „Vereinfachtem Datenaustauschschemata“ zu implementieren.***

### 5.4 Lösungsansätze für die Gesamtarchitektur

#### 5.4.1 Geokodierungsdienst mit zentraler Datenhaltung

Ein Dienst, der auf einer zentralen Datenhaltung basiert, besitzt die Vorteile:

- hohe Performance
- hohe Verfügbarkeit
- Verwendbarkeit vieler Frameworks für OGC-Dienste

Nachteilig sind allerdings:

- sehr hoher Aufwand für die Datenpflege
- in der Regel geringere Aktualität der Daten als im Bundesland

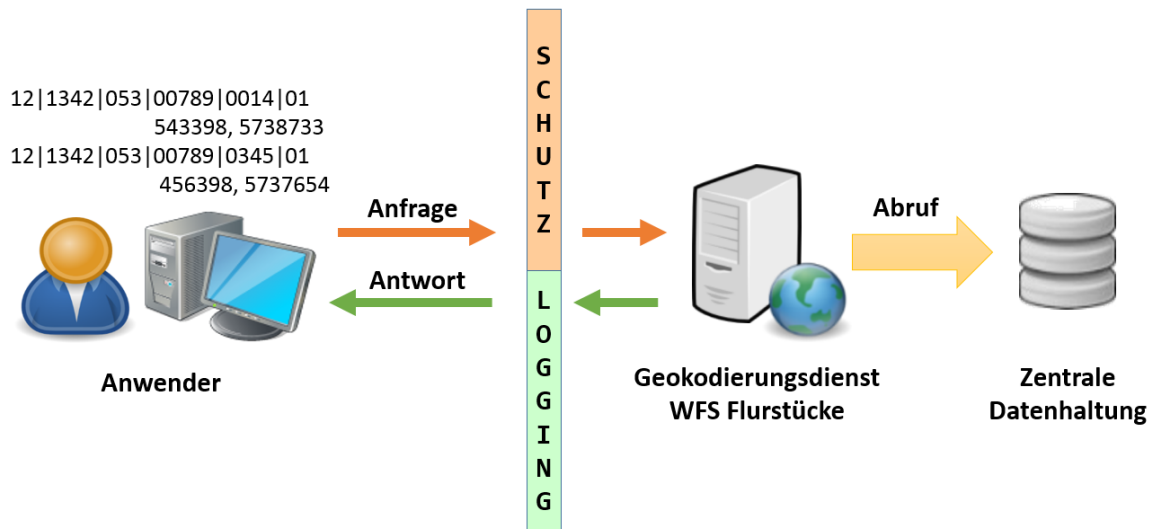
Grundvoraussetzung für eine solche Lösung ist ein effizientes Verfahren zur Datenaktualisierung auf Seiten der Bundesländer und des BKG.

Unumgänglich ist eine zentrale Datenhaltung, wenn der Dienst anders nicht realisierbar ist oder die erforderliche Performance und Verfügbarkeit des Dienstes nur so gewährleistet werden können. Das ist insbesondere der Fall, wenn

- nicht in allen Ländern geeignete Dienste vorhanden sind, also auch NAS-Datensätze mit eingesetzt werden müssen

- die vorhandenen Dienste nicht einheitlich sind und eine Harmonisierung on-the-fly aus Performancegründen ausscheidet
- eingebundene Dienste qualitative Mindestanforderungen nicht erfüllen

In Abbildung 8 ist die Gesamtarchitektur dieses Lösungsansatzes visualisiert.



**Abbildung 8: Geokodierungsdienst mit zentraler Datenhaltung**

### 5.4.2 Geokodierungsdienst als reine Dienstekaskade

Eine reine Dienstekaskade unter dynamischem Zugriff ("on-the-fly") auf die WFS-Dienste der Bundesländer hätte die Vorteile:

- Wegfall der zentralen Datenhaltung und dessen Aktualisierungsaufwands
- stets Zugriff auf dezentral gepflegte, aktuelle Daten der Bundesländer
- schlanke Gesamtarchitektur bei einheitlichen Ausgangsdiensten

Zu den wesentlichen Nachteilen zählen:

- eine verminderte Performanz
- eine verminderte Verfügbarkeit

Grundvoraussetzung für diese Lösung sind in allen Bundesländern verfügbare AdV-ALKIS-WFS mit „Vereinfachtem Datenaustauschschemata“. Eine On-the-fly-Transformation von Daten aus verschiedensten Ausgangsschemata in ein einheitliches Datenmodell des zentralen Geokodierungsdienstes ist aus Aufwands- und Performancegründen nicht möglich bzw. wäre kein effizienter Lösungsansatz.

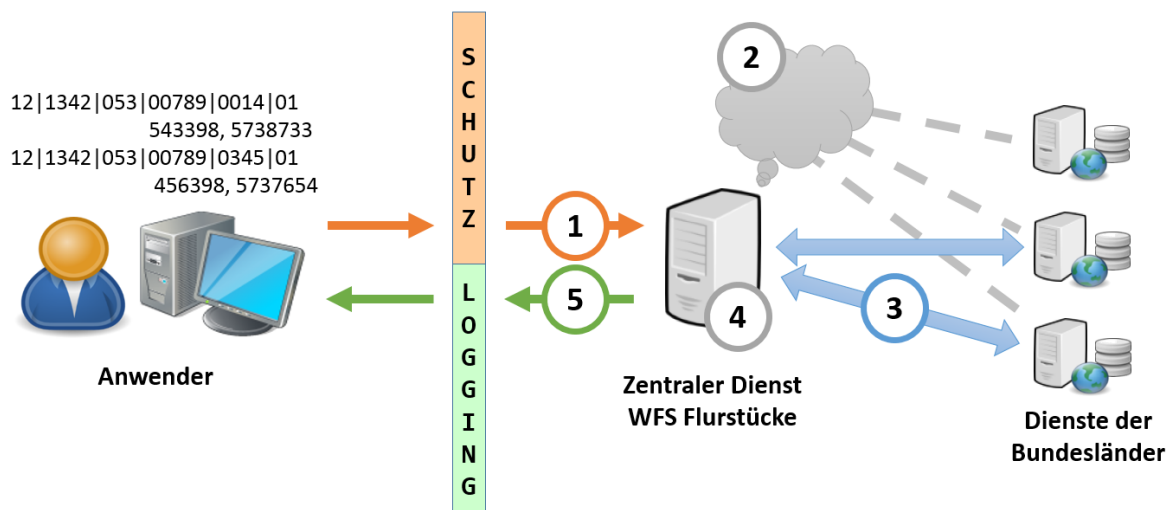
Realisierbar ist diese Dienstekaskade, weil die Performanceanforderungen an den Dienst durch den Wegfall einer schnellen Vorschlagssuche nicht außergewöhnlich hoch sind.

Ein weiterer wesentlicher Nachteil einer reinen Dienste-Kaskade ist, dass in allen Bundesländern ein entsprechender Dienst zur Verfügung stehen muss. Die zeitnahe Realisierung einer

reinen Dienste-Kaskade ist vor diesem Hintergrund erst in einigen Jahren möglich, da noch nicht alle Länder einen Geokodierungsdienst für Flurstücke im Einsatz haben.

Abbildung 9 veranschaulicht die Funktionsweise der Dienstekaskade:

- (1) Anfrage des zugangsberechtigten Anwenders an den zentralen Dienst
- (2) Zuordnung zu den Bundesländern über eine bundesweite Liste von Gemarkungen
- (3) spezielle Anfrage an die entsprechenden Dienste Bundesländer
- (4) Zusammenführung der Ergebnisse
- (5) Logging der abgerufenen Features und Weiterleitung an den Anwender



**Abbildung 9: Geokodierungsdienst als Dienstekaskade**

Der Dienst trifft anhand der Suchanfrage eine Vorauswahl, und bezieht somit nur nötige Dienste für die Suche ein. Hierdurch wird eine Weiterleitung der Suchanfrage an alle Länderdienste vermieden und die Antwortzeit minimiert.

### 5.4.3 Geokodierungsdienst mit Dienstekaskade und zentralem Cache

Aufgrund der inhomogenen Ausgangslage in den Bundesländern ist eine Übergangslösung erforderlich, um einen zentralen Dienst möglichst zeitnah einsetzen zu können. Wie in Abbildung 10 skizziert, stellt die Architektur dieses Ansatzes einen Kompromiss der beiden zuvor betrachteten Lösungen dar, d. h.:

- der zentrale Dienst ist ein AdV-ALKIS-WFS mit „Vereinfachtem Datenaustauschsche-ma“,
- vorhandene Landesdienste AdV-ALKIS-WFS mit „Vereinfachtem Datenaustauschsche-ma“ werden dynamisch in den zentralen Dienst eingebunden,
- andere XML-basierte Landesdienste werden über XML-Transformationen in einen zent-ralen Cache mit einheitlicher Datenstruktur überführt,
- NAS-Datensätze der Bundesländer ohne geeigneten Dienst werden über XML-Transformationen in einen zentralen Cache mit einheitlicher Datenstruktur überführt.

Zu den wesentlichen Vorteilen zählen dann:

- die Unterstützung von Bundesländern ohne verwendbaren WFS-Dienst,

- die Unterstützung aller aktuell verfügbaren WFS-Dienste,
- die Sicherstellung der erforderlichen Verfügbarkeit und Performanz des zentralen Dienstes durch Caching-Techniken, sowie
- die sukzessive Erweiterbarkeit der Dienste-Kaskade.

Damit werden alle wesentlichen Anforderungen an den Geokodierungsdienst für Flurstücke erfüllt. Zudem kommt das Konzept bereits dem Wunsch nach dezentraler Datenhaltung entgegen.

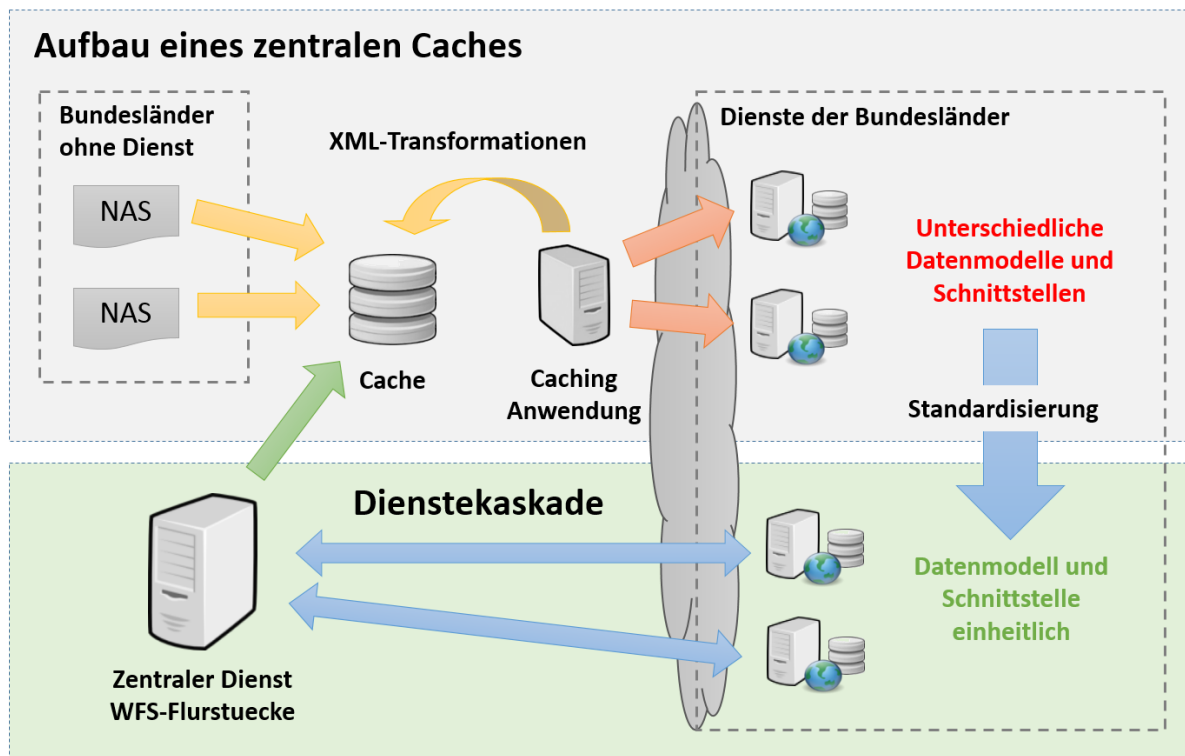
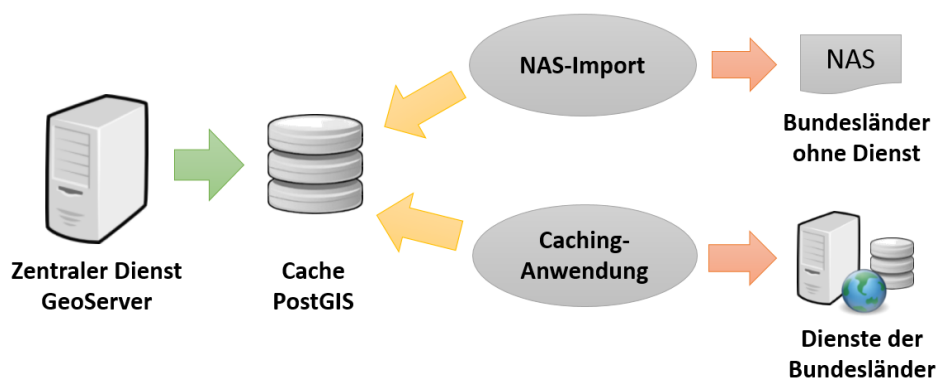


Abbildung 10: Geokodierungsdienst mit Dienstekaskade und zentralem Cache

## 5.5 Vorstellung geeigneter Lösungsansätze

### 5.5.1 Kurzfristige Lösung

Kurzfristig wäre eine Lösung als Geokodierungsdienst mit Dienstekaskade und zentralem Cache mithilfe von geeigneten Open-Source-Lösungen beim BKG verfügbar, wie in Abbildung 11 dargestellt.



**Abbildung 11: Gesamtarchitektur der kurz- und mittelfristigen Lösung**

Die Realisierung eines Caches könnte über eine PostgreSQL-Datenbank mit PostGIS-Erweiterung erfolgen.

Als WFS-Server kann zunächst die Open-Source-Software GeoServer dienen, welche die Einbindung von PostGIS erlaubt. Der GeoServer stellt bereits den vollen Funktionsumfang eines WFS-Servers zur Verfügung. Dieser kann entsprechend dem AdV-ALKIS-WFS Profil mit „Vereinfachtem Austauschschema“ konfiguriert und kann als zentraler Dienst eingesetzt werden.

### 5.5.2 Caching-Anwendung

Für den Import von verwendbaren WFS-Diensten befindet sich beim BKG eine Anwendung auf Basis von Java in der Entwicklung, die verwendet werden könnte. Dabei kommen diverse Open-Source-Bibliotheken zum Einsatz:

- Java Database Connectivity (JDBC) für die Schnittstelle zur Datenbank
- Simple API for XML (SAX) für die Interpretation der XML-Daten
- Document Object Model and XML (JDOM) zur Konfiguration des gesamten Frameworks
- Apache HttpClient für die Kommunikation mit den WFS-Diensten
- Java Concurrency zur Parallelisierung der Anfragen

Das Framework erlaubt es, die Daten der verschiedenen Dienste automatisiert abzurufen und in ein einheitliches Datenmodell zu überführen. Die Konfiguration des Frameworks umfasst die WFS-Dienste, entsprechende Abfragen, die Transformation der XML-Daten und den Datenbankzugriff.

Als verwendbar werden alle Dienste betrachtet, die folgende Voraussetzungen erfüllen:

1. Der Dienst ist ein WFS 1.0.0 oder WFS 1.1.0.
2. Das Datenmodell enthält Flurstückskennzeichen (FKZ) und Umringspolygon (GML).
3. Die maximale Feature-Anzahl ist größer als die Anzahl aller Gemarkungen im Bundesland und größer als die Anzahl aller Flurstücke der umfassendsten Gemarkung (Flur).

Dabei spielt es keine Rolle, in welcher Form das FKZ vorliegt. Entscheidend ist lediglich, dass alle Informationen vorhanden sind, da sie ohnehin mittels Transformationen vereinheitlicht werden.



Die dritte Voraussetzung hat ihren Ursprung in der hierarchischen Vorgehensweise der Anwendung, welche sich wie folgt gliedert:

- A. Abruf aller n Gemarkungen (1 Abfrage)
- B. Paralleles Abrufen der Flurstücke mit Filter auf der Gemarkung (n Abfragen)

Insgesamt werden so  $1+n$  Abfragen an den Dienst gesendet. Liegen im Bundesland ebenfalls Fluren vor, ist ein Zwischenschritt über diese mit Filter auf der Gemarkung möglich.

Erste Tests mit dem Framework verliefen positiv. So konnten bereits Daten von Diensten aus drei Ländern in den Cache geladen werden, wovon eines der Länder bereits flächendeckend abgefragt wurde. Der Abruf konnte hier durch Parallelisierung auf unter eine Stunde pro Mio. Flurstücke gesenkt werden. Ein nächtlicher Austausch garantiert, dass die Dienste während des Tages nicht unnötig mit Anfragen belastet werden. Ein Abruf der Daten sollte sich nach den Aktualisierungszyklen des ursprünglichen Dienstes richten.

### 5.5.3 NAS-Import

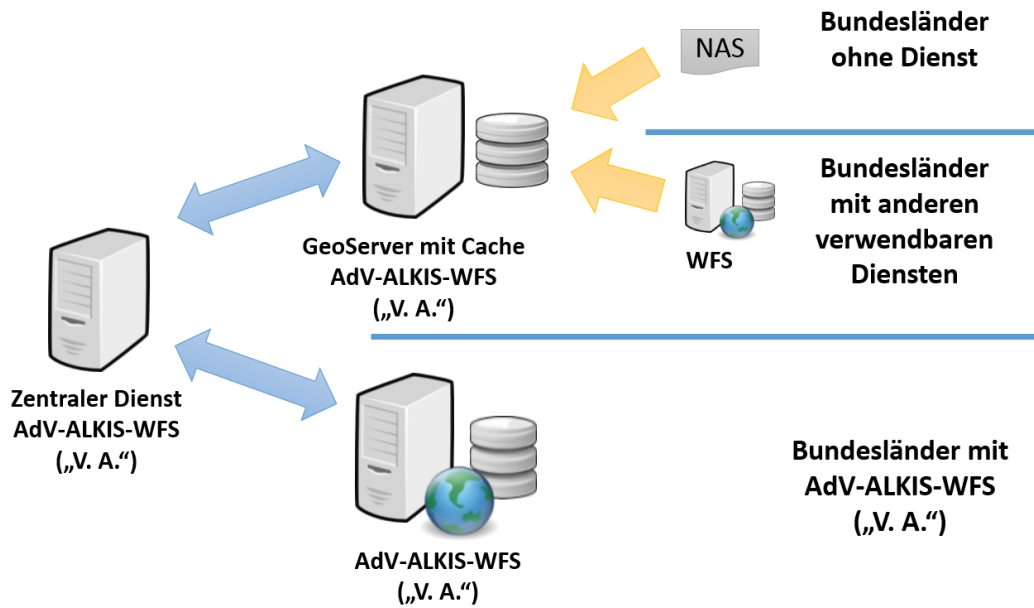
Über den Objekttyp „AX\_Flurstueck“ der NAS können Daten in den Cache übernommen werden. Mithilfe von Testdatensätzen aus Berlin, Hessen und Baden-Württemberg ist dieses Vorgehen bereits erfolgreich beim BKG getestet wurde.

Bei der Implementierung kommen ebenfalls JDBC und SAX zum Einsatz. Die Geometrien (GML) werden dabei in Extended Well-Known Text (EWKT) umgewandelt, den PostGIS interpretieren kann. Damit bleiben die Kreisbögen im zentralen Cache erhalten.

### 5.5.4 Mittel- und langfristige Lösung

Langfristig sollen die AdV-ALKIS-konformen WFS-Dienste nach dem „Vereinfachten Datenaustauschschemata“ („V. A.“) in eine Dienstekaskade „on-the-fly“ eingebunden werden. Um eine sukzessive Erweiterung der Dienste zu gewährleisten, wird vor den bspw. GeoServer ein weiterer Dienst geschaltet, der dann als zentraler Dienst arbeitet.

Der GeoServer mit zentralem Cache wäre in dieser Lösung gleichwertig zu den Diensten der Bundesländer einzuordnen. Er würde einen Dienst für Länder zur Verfügung stellen, die über keinen konformen Dienst verfügen. Abbildung 12 veranschaulicht diesen möglichen Lösungsschritt.



**Abbildung 12: Mittel- und langfristige Lösung**

Der zentrale Dienst übernimmt die eigentliche Aufgabe in der Dienstekaskade, die notwendigen Dienste zu identifizieren und die Anfrage weiterzuleiten. Falls erforderlich müssen die Ergebnisse anschließend miteinander verschnitten werden. Solange alle Dienste einem einheitlichen Profil entsprechen, sind dann keine XML-Transformationen notwendig und die Auswirkung auf die Performanz und den Implementierungsaufwand bleibt gering.

## **6 Anlagenverzeichnis**

Geokodierung\_Vergleich.xls. Ein Vergleich der Lösungsansätze zur Geokodierung aus NRW und dem BKG.

## 7 Literaturverzeichnis

- [1] LG GDI-DE. (02.05.2013). *Bericht der AG Routing und Geokodierung (Langfassung)*.  
LG GDI-DE. (02.05.2013). *Kurzbericht der AG Routing und Geokodierung*.
- [2] LG GDI-DE. (02.05.2013). Bericht der AG Routing und Geokodierung (Langfassung).
- [3] LG GDI-DE. (02.05.2013). Anforderungen und Weiterentwicklungsaufwand.
- [4] Open Geospatial Consortium. (17.02.2012). Gazetteer Service - Application Profile of the Web Feature Service Best Practice.