

Geographisches Institut
Rheinische Friedrich-Wilhelms-Universität Bonn

Konzept einer effizienten Suche in Metadaten unter
Beachtung räumlicher Aspekte auf Basis des
Geodatenkatalog.de

Bachelorarbeit

Vorgelegt von: Janek-Lukas Voß
B. Sc. Geographie
Matrikelnummer: 2543098

Betreuer: Prof. Dr. Klaus Greve
Datum: 12.10.2015

Inhaltsverzeichnis

1. Einleitung.....	1
2. Grundlagen.....	2
2.1. INSPIRE-Bestimmungen.....	2
2.2. INSPIRE-Umsetzung in Deutschland.....	5
2.3. Stand der Umsetzung der INSPIRE-Richtlinie.....	7
2.4. Der Geodatenkatalog.de.....	9
2.5. Metadaten der GDI-DE.....	10
3. Ist-Zustand der Suche im Geodatenkatalog.de.....	12
3.1. Die Solr-Suche.....	12
3.2. Die Aufbereitung der Geodaten.....	13
3.3. Der Suchvorgang im Geodatenkatalog.de.....	15
3.4. Die Volltextsuche im Geodatenkatalog.de.....	17
3.5. Die Ortssuche im Geodatenkatalog.de.....	18
4. Ist-Zustand der Metadaten im Geodatenkatalog.de.....	20
4.1. Die Extraktion der Metadaten.....	20
4.2. Analysewerkzeuge.....	20
4.3. Analysevorgang.....	21
4.4. Ergebnis der Metadatenanalyse.....	22
4.4.1. Konformität.....	22
4.4.2. Räumliche Bewertung.....	24

5. Konzept einer Suche in Metadaten.....	26
5.1. Vorgehen und Bewertungskriterien.....	26
5.2. Optimierung der Datenqualität.....	27
5.3. Optimierung des Suchvorgangs.....	30
6. Fazit.....	36
7. Quellen.....	38
7.1. Gesetze und Richtlinien.....	38
7.2. Literatur.....	39
7.3. Internetquellen.....	41
8. Glossar [*]	43
9. Anhang.....	45

* Begriffe, die im Glossar erklärt werden, sind bei ihrer ersten Nennung im Text kursiv und unterstrichen gekennzeichnet.

Abbildungsverzeichnis

Abbildung 1: Organisationsstruktur der Geodateninfrastruktur Deutschland (GDI-DE).....	6
Abbildung 2: Anzahl der Metadatensätze für INSPIRE-relevante Geodatenätze und -dienste innerhalb des Geodatenkatalog.de anhand von Metadaten von 2009 bis 2014.....	7
Abbildung 3: Architektur des Geodatenkatalog.de.....	9
Abbildung 4: Das „Publish-Find-Bind“-Schema am Beispiel der GDI-DE.....	10
Abbildung 5: Die Daten-Dienste-Kopplung der „Geodatensuche“ im Geodatenkatalog.de.....	14
Abbildung 6: Bildschirmfoto der Suche des Geoportal.de und seiner Ergebnisausgabe mit einer beispielhaften Suchanfrage nach dem Stichwort „wasser“.....	15
Abbildung 7: Schema des Suchvorgangs im Geodatenkatalog.de und im Geoportal.de.....	16
Abbildung 8: Beispiel der genutzten Lagebeziehungen der angefragten Bounding Box (X) zu möglichen Ergebnis-Bounding Boxen (A,B,C) der Ortssuche.....	19
Abbildung 9: Anteil der untersuchten Parameter an der zufälligen Stichprobe aus der Geonetwork-Datenbank des Geodatenkatalog.de in Prozent (%).....	24

1. Einleitung

Räumliche Informationen sind für uns ein zentraler Bestandteil des Alltags. Einerseits können sie für Privatpersonen von Nutzen sein – zum Beispiel als Navigationsgerät im Auto, als Wetterkarte oder bei der Planung des nächsten Urlaubs, andererseits haben auch Wissenschaft, Wirtschaft und Politik ein großes Interesse an räumlichen Informationen. Ob als Gefahrenkarten, Bevölkerungskarten, Ressourcen- und Infrastrukturkarten oder Wahlergebniskarten - der Nutzen von Geodaten ist groß und vielfältig.

Geodaten sind in diesem Sinne „alle Daten mit direktem oder indirektem Raumbezug zu einem bestimmten Standort oder geographischen Gebiet“ (INSPIRE:4).

Für die Politik und insbesondere die Verwaltung kann ihr Einsatz ein Mittel zur Effizienzsteigerung, Entscheidungsfindung, Steuerung und Planung sein (DEUTSCHER LANDKREISTAG BERLIN:3-4). Zugleich profitiert auch die Wirtschaft von diesen Vorteilen, denn sie kann Geodaten gewinnbringend nutzen und mit ihrer Hilfe neue Entwicklungen vorantreiben (LAKES 2011:1).

Schließlich helfen Sie in der Wissenschaft bei Forschungsprojekten, beispielsweise zu aktuellen Fragestellungen zum Klimawandel, zur Migration und Urbanisierung (LAKES 2011:1).

Aus diesen Gründen ist es von entscheidender Bedeutung, dass diese Daten für Wirtschaft, Wissenschaft, Politik und Verwaltung zugänglich gemacht werden.

In Deutschland werden behördliche Geodaten von Bund und Ländern über den Geodatenkatalog.de bereitgestellt. Er repräsentiert somit eine zentrale Sammelstelle für Informationen über Geodaten in Deutschland (WALTHER 2011:272). Mithilfe der Suchfunktion auf der Internetseite des Geoportal.de werden die Daten dann für jeden Internetnutzer zugänglich gemacht (VERWALTUNGSVEREINBARUNG GDI-DE:Anlage 2:1). So können potentielle Interessenten die für sie relevanten Geodaten finden und nutzen. Eine besondere Bedeutung kommt hierbei der Suchfunktion zu, denn nur wenn gesuchte Geodaten auch (schnell und einfach) gefunden werden können, kann man sie überhaupt nutzen. Deshalb beschäftigt sich die Arbeit mit dem Konzept einer effizienten Suchfunktion für den Geodatenkatalog.de.

Im Folgenden wird zunächst die Basis für das Verständnis des Themas geschaffen, indem die INSPIRE-Richtlinie, ihre Umsetzung in Deutschland, der Geodatenkatalog.de und der Ausdruck Metadaten erläutert wird. Des weiteren wird die Suchfunktion in ihrem aktuellen Zustand beschrieben und die Daten im Geodatenkatalog.de analysiert. Abschließend wird daraus ein Konzept für eine effiziente Suche, unter Berücksichtigung räumlicher Aspekte entwickelt. Die Umsetzung der hier gewonnenen Erkenntnisse kann im besten Fall zur Verbesserung der Geodateninfrastruktur Deutschlands und den Geodatenkatalog.de führen.

2. Grundlagen

2.1. INSPIRE-Bestimmungen

Mit der europäischen Richtlinie 2007/2/EG des europäischen Parlaments und des Rates wurde am 14.03.2007 die Einrichtung einer „**IN**frastructure for **SP**atial **InfoR**mation in the **E**uropean Community (kurz: INSPIRE)“ beschlossen - einer europaweiten Infrastruktur für alle behördlichen Geodaten (INSPIRE:1). Sie gilt für alle Geodaten, die bei den Behörden vorhanden sind oder für sie bereitgehalten werden, sowie für Geodaten die von den Behörden genutzt werden.

Die Gründe für den Beschluss dieser Richtlinie sind vielfältigster Art. Die Kernidee dahinter ist, dass die Behörden in der Europäischen Union auf verschiedenen Gebieten, wie zum Beispiel Umweltpolitik, Wirtschaft und Ressourcenmanagement, ihre digitalen Informationen gemeinsam nutzen wollen.

Bis zum Beschluss der Richtlinie durch das Europäische Parlament im März 2007 gab es zahlreiche Probleme bei der Verfügbarkeit, Qualität, Organisation, Zugänglichkeit und gemeinsamen Nutzung von Geodaten (INSPIRE:1).

Um diese Probleme zu lösen sind alle Mitglieder der EU durch die INSPIRE-Richtlinie verpflichtet eine solche Infrastruktur bis zum Jahre 2021 national bereitzustellen. Diese muss den Anforderungen der Vollständigkeit, Interoperabilität, Einheitlichkeit und freien Verfügbarkeit genügen.

In diesem Zusammenhang meint Interoperabilität die „mögliche Kombination und [...] Interaktion ohne wiederholtes manuelles Eingreifen und in der Weise, dass das Ergebnis kohärent ist und der Zusatznutzen der Datensätze und Datendienste erhöht wird“ (INSPIRE:5). Anders ausgedrückt bedeutet Interoperabilität die übergreifende Kombinierbarkeit der Geodaten in sämtlichen Fällen durch die Festlegung von Standards.

Hierbei kommt den so genannten Metadaten eine zentrale Rolle zu. Gemäß INSPIRE sind Metadaten „Informationen, die Geodatensätze und Geodatendienste beschreiben und es ermöglichen, diese zu ermitteln, in Verzeichnisse aufzunehmen und zu nutzen“ (INSPIRE:5).

Ein Geodatensatz dagegen ist eine „identifizierbare Sammlung von Geodaten“. Mithilfe von Geodatendiensten lassen sich Geodaten von Computeranwendungen verarbeiten (INSPIRE:5).

Also sind Metadaten Daten bzw. Informationen über Daten. Im Bezug auf Geodaten werden Metadaten mit Informationen wie Titel, geographischer Standort oder Besitzer befüllt, um die Suche nach den zugehörigen Geodaten zu ermöglichen.

Metadaten spielen also eine essentielle Rolle, wenn es um Geodaten geht, denn sie sind der „Schlüssel“ deren Auffindung. Metadaten haben dahingehend auch in den Richtlinien rund um INSPIRE ein großes Gewicht, da deren Pflege unumstößlich mit den Zielen Interoperabilität, Einheitlichkeit und Vollständigkeit in Verbindung steht (vgl. INSPIRE).

Die europäische Kommission erließ zu diesem Zweck eine weitere Verordnung, die auf bereits existierenden Standards für Geodaten basiert (vgl. Kap.2.5.). Inhalt der Verordnung ist eine detaillierte Auflistung aller erforderlichen Metadatenelemente, sowie deren Beschreibung (INSPIRE METADATEN). Des weiteren konkretisieren sich diese Standards in den Technischen

Spezifikationen, welche von der Kommission zu verschiedenen Annex-Themen erlassen werden (INSPIRE:9).

Annexthemen definieren eine thematische Unterteilung von Geodaten und Geodatensätzen. Je nach Thema der Geodaten werden diese entweder den Themen aus Annex I, Annex II oder Annex III zugeordnet. Adressdaten zum Beispiel sind dem Annexthema I zugeordnet, welches vereinfacht gesagt geographische Basisdaten in sich zusammenfasst. Hierzu zählen unter anderem Koordinatenreferenzsysteme, Verkehrsnetze und Verwaltungseinheiten. Annex II befasst sich dagegen mit geologischen Daten wie Höhe, Bodenbedeckung und Geologie. Unter Annex III fallen schließlich alle weiteren Geodaten, wie zum Beispiel Bodennutzung, demographische Daten oder Versorgungsdaten.

Insgesamt werden alle betroffenen Geodaten und Geodatensätze 34 verschiedenen Geodaten Themen zugeteilt, welche wiederum in besagte Annexthemen gegliedert sind (vgl. INSPIRE).

Es existieren hierbei ganz klare Vorgaben, was die Datenformate und die Darstellung der Geodatensätze angeht. Im Sinne einer europaweiten Interoperabilität müssen sie je nach Geodaten Thema bestimmte Attribute, Datentypen und Hierarchien aufweisen (vgl. INSPIRE INTEROPERABILITÄT).

Welche Daten nun konkret unter die Richtlinie fallen, wird schon in der INSPIRE-Richtlinie klar definiert. Folgende Bedingungen müssen erfüllt sein, bevor die Richtlinie überhaupt Anwendung findet:

1. Die Geodatensätze müssen sich auf den Bereich, in dem ein Mitgliedstaat Hoheitsbefugnisse hat und/oder ausübt beziehen
2. Die Geodatensätze müssen in elektronischer Form vorliegen
3. a) Die Geodatensätze müssen bei einer Behörde vorliegen, oder sind bei ihr eingegangen, oder werden von ihr verwaltet oder aktualisiert und fallen unter ihren öffentlichen Auftrag
b) Die Geodatensätze liegen bei Dritten vor, welche die Geodaten im Auftrag der betroffenen Behörde verwalten.
4. Die Geodatensätze betreffen eines oder mehrere der Annexthemen (I-III)

(verändert nach INSPIRE:5)

Für die Umsetzung der Richtlinie wurde ein Zeitplan (Anhang Zeitplan) mit Meilensteinen bzw. Deadlines aufgestellt. Er regelt den schrittweisen Aufbau der EU-Infrastruktur.

In einem ersten Schritt werden die vorliegenden Geodaten über webbasierte Dienste durch ihre Metadaten such- und findbar gemacht. Anschließend werden die Geodaten mit Hilfe von Darstellungsdiensten über die Weboberfläche freigegeben und somit für den Nutzer sichtbar.

Im letzten Schritt können die Geodaten über Downloaddienste heruntergeladen und über Transformationsdienste in das INSPIRE-Datenschema des jeweiligen Annex-Themas umgewandelt werden.

Diese Schritte werden zunächst auf nationaler Ebene umgesetzt und dann für eine europaweite Nutzung freigegeben (INSPIRE ROADMAP).

Der festgelegte Zeitplan sieht bis zum 21.10.2015 folgendes vor: Sämtliche von der Richtlinie betroffenen Geodaten müssen über Such-, Darstellungs- und Downloaddienste bereits zugänglich gemacht worden sein. Transformationsdienste müssen bei Geodaten, die nicht standardkonform sind, zusätzlich bereitgestellt werden. Des Weiteren müssen neu erhobene bzw. umstrukturierte Geodaten, welche unter das Annex-Thema I fallen, standardkonform sein, während die bereits existierenden Geodatenbestände dies noch nicht sein müssen (INSPIRE ROADMAP).

In diesem Zusammenhang verlangt die INSPIRE-Richtlinie eine regelmäßige Berichterstattung über den Umsetzungsstand. Hierin enthalten sind außerdem die Kosten und der Nutzen der Richtlinie für das jeweilige Land. Alle EU-Mitgliedsstaaten müssen demnach seit 2010 im Abstand von 3 Jahren (2010, 2013, 2016, 2019) einen Bericht bei der EU-Kommission abgeben, die mit der Koordination der Richtlinie auf EU-Ebene beauftragt wurde. Nicht nur die Staaten der Europäischen Union dokumentieren ihren individuellen Fortschritt, sondern auch die EU-Kommission selbst liefert Umsetzungsberichte im Jahre 2014 und 2020 (INSPIRE:10).

Als zusätzliches Werkzeug zur Überwachung der Umsetzung beschloss die Kommission 2009 die Einführung eines INSPIRE-Monitoring. Am Ende jedes Jahres sollen dazu festgelegte Indikatoren, wie zum Beispiel der Grad der Konformität von Geodatensätzen oder Metadaten, von den EU-Staaten an die Kommission übermittelt werden. So lässt sich übersichtlich und quantitativ der Stand der Umsetzung der INSPIRE-Richtlinie eines jeden EU-Mitgliedslandes feststellen (INSPIRE MONITORING).

2.2. INSPIRE-Umsetzung in Deutschland

Die Umsetzung der Richtlinie in nationales Recht geschah über das Geodatenzugangsgesetz (10.02.2009) und über zahlreiche Gesetze der Länder. Als zentrale Aufgabe ist hierin der Aufbau einer Geodateninfrastruktur (GDI) formuliert. Eine Geodateninfrastruktur „ist eine Infrastruktur bestehend aus Geodaten, Metadaten und Geodatendiensten, Netzdiensten und -technologien, Vereinbarungen über gemeinsame Nutzung, über Zugang und Verwendung sowie Koordinierungs- und Überwachungsmechanismen, -prozesse und -verfahren mit dem Ziel, Geodaten verschiedener Herkunft interoperabel verfügbar zu machen“ (GEODATENZUGANGSGESETZ).

In Deutschland wird die erforderliche Aufgabe des Aufbaus einer solchen GDI zur Umsetzung der INSPIRE-Richtlinie bis 2021 von Bund, Ländern und Kommunen übernommen. Für die Bereitstellung von Geodaten sind die geodatenhaltenden Stellen in Deutschland verantwortlich. Das bedeutet, dass alle Bundes- und Landesbehörden, sowie Kommunen, die über betroffene Geodaten(sätze) verfügen, diese digital sammeln und über internetbasierte Dienste bereitstellen müssen. Auch hier gelten die gleichen Einschränkungen bei der Betroffenheit, wie bei der INSPIRE-Richtlinie (vgl. Kap. 1.1) (GEODATENZUGANGSGESETZ).

Über die GDI-DE werden so sowohl „INSPIRE-betroffene“ Geodatensätze und ihre Metadaten, als auch davon ausgenommene Geodatensätze und ihre Metadaten verfügbar gemacht, denn nicht alle Geoinformationen sind „INSPIRE-relevant“ (z.B. Stadtplandienste) (vgl. Kap.2.1).

Im Sinne eines Open Government sollen die Daten in der Geodateninfrastruktur Deutschland (GDI-DE) auch für den Bürger zugänglich gemacht werden. Dabei bezeichnet Open Government eine generelle Öffnung von Staat und Verwaltung, um den Bürger stärker in politische Entscheidungen einzubeziehen und für Akzeptanz und Transparenz zu sorgen (VON LUCKE 2012:1-2).

Die Koordinierung dieses Prozesses übernimmt das dafür gegründete Lenkungs-gremium GDI-DE in Zusammenarbeit mit der Koordinierungsstelle GDI-DE und den Kontaktstellen von Bund, Ländern und Kommunen (vgl. Abb.1).

Das Lenkungs-gremium (LG) setzt sich aus Vertretern von Bund, Ländern und kommunalen Spitzenverbänden zusammen. Außerdem werden Wissenschaft und Wirtschaft eingebunden. Das LG ist die nationale Anlaufstelle für INSPIRE und steuert bzw. koordiniert seine Umsetzung. Es legt die Ziele und Grundsätze, sowie die Architektur und Organisation der GDI-DE fest. Zu seinen weiteren Aufgaben zählt die Finanzierung, die Nutzungsrechtregelung und die Berichterstattung an den übergeordneten IT-Planungsrat. Dieser ist das „zentrale Steuerungsgremium für die Informationstechnologie (IT) von Bund und Ländern“ (VERWALTUNGSVEREINBARUNG GDI:5-6).

Umgesetzt werden die Beschlüsse des LG von der Koordinierungsstelle GDI-DE. Dabei wird sie von den Kontaktstellen von Bund, Ländern und Kommunen unterstützt (vgl. Abb.1). Geführt wird sie vom Bundesamt für Kartographie und Geodäsie (BKG) in Frankfurt.

Diese Baumstruktur setzt sich mit den Kontaktstellen von Bund, Ländern und Kommunen fort. So sollen hier die technischen und strukturellen Maßnahmen in Zusammenarbeit mit der Koordinierungsstelle GDI-DE umgesetzt werden (VERWALTUNGSVEREINBARUNG GDI:8).

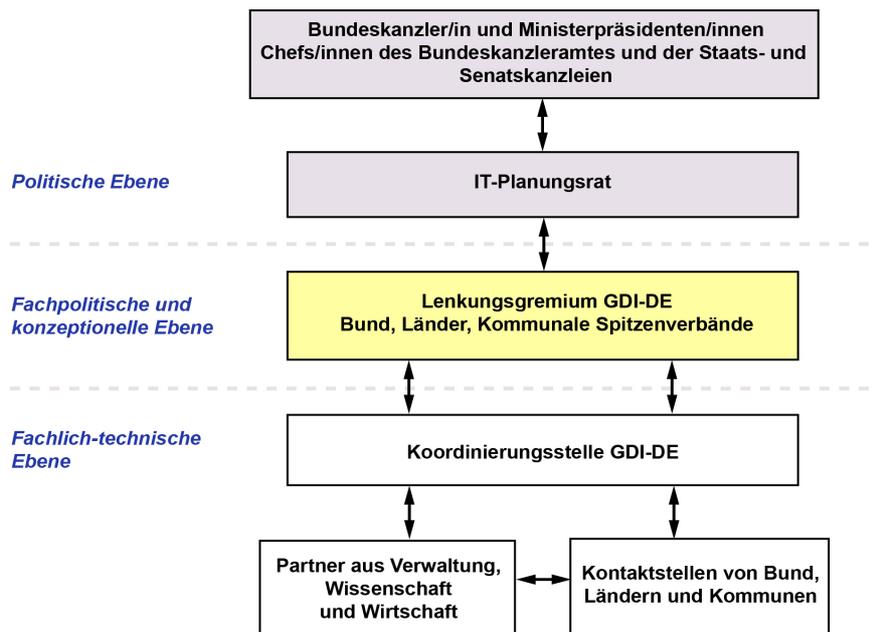


Abb.1. Organisationsstruktur der Geodateninfrastruktur Deutschland (GDI-DE), Abrufbar unter: http://www.geoportal.de/SharedDocs/Bilder/DE/GDI-DE/Organisationsstruktur_GDI-DE.jpg?__blob=poster, letzter Aufruf: 10.10.2015).

Die technische Architektur der GDI-DE besteht im Wesentlichen aus zwei Komponenten. Auf der einen Seite gibt es zentrale Komponenten, die das BKG betreibt. Auf der anderen Seite gibt es eine Vielzahl von dezentralen Diensten, die bei den geodatenhaltenden Stellen von Bund, Ländern und Kommunen vorliegen (AK METADATEN:7-8).

Die zentralen Komponenten beim BKG sind die GDI-Testsuite, die GDI-Registry, das Geoportal.de und der Geodatenkatalog.de (VERWALTUNGSVEREINBARUNG GDI: Anlage 2:1).

Die GDI-Testsuite ermöglicht es „Geodaten und Geodatendienste auf Konformität zu den Vorgaben von INSPIRE und GDI-DE zu prüfen“ (VERWALTUNGSVEREINBARUNG GDI:Anlage 2:1). Sie übernimmt somit eine Kontrollfunktion für alle von der Richtlinie betroffenen Behörden.

Die „GDI-Registry verwaltet Informationen, die vielfach verwendet werden und deren Einheitlichkeit sicherzustellen ist“ (VERWALTUNGSVEREINBARUNG GDI:Anlage 2:1). Hierzu zählen Monitoring-Register, Koordinatenreferenzsystem-Register und Schema-Register.

Das Geoportal.de dient der Suche und Anzeige von Diensten aller von den Behörden veröffentlichten Daten. Zudem gibt es dort allgemeine Informationen zur GDI-DE und der INSPIRE-Richtlinie.

Die Metadaten selbst liegen im Geodatenkatalog.de vor. Hier werden alle Metadaten innerhalb der GDI-DE gesammelt („*harvest*“) und bereitgestellt (VERWALTUNGSVEREINBARUNG GDI:Anlage 2:1).

Das Ziel der Bachelorarbeit ist die Suche in Metadaten, am Beispiel des Geodatenkatalog.de zu analysieren und gegebenenfalls Vorschläge zur Verbesserung zu machen. Hierbei wird auch räumlichen Aspekten eine Bedeutung zukommen.

2.3. Stand der Umsetzung der INSPIRE-Richtlinie

In Deutschland ist bei der Umsetzung der INSPIRE-Richtlinie ein eindeutiger Trend erkennbar. Absolut gesehen werden immer mehr INSPIRE-relevante Geodaten, Metadaten und Dienste über Geoportale veröffentlicht. Dies zeigen die Monitoringergebnisse der GDI-DE (GDI-DE MONITORING).

Anzahl der Metadatenätze für INSPIRE-relevanter Geodatenätze und -dienste

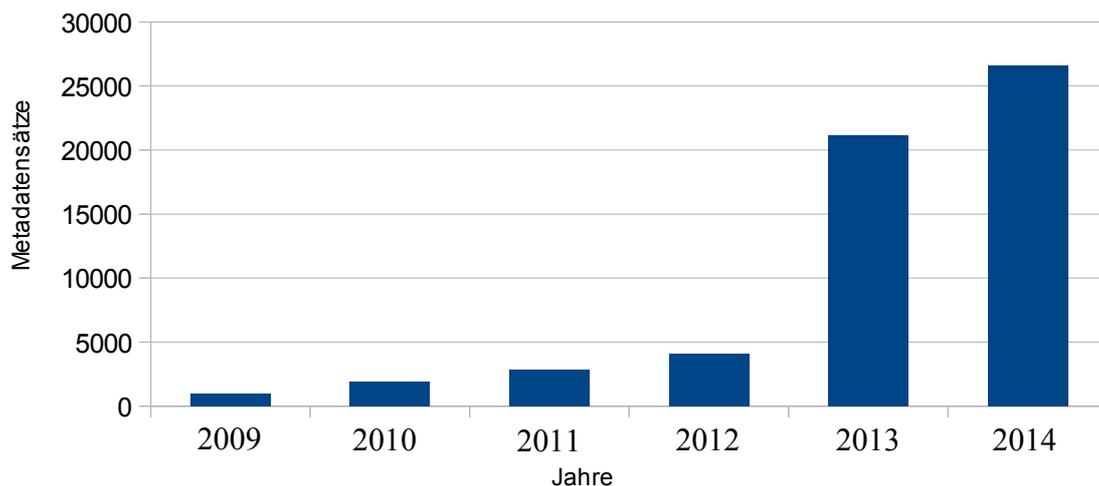


Abb.2: Anzahl der Metadatenätze für INSPIRE-relevante Geodatenätze und -dienste innerhalb des Geodatenkatalog.de anhand von Metadaten von 2009 bis 2014, eigene Darstellung auf Basis des Inspire-Monitoring Deutschland, Abrufbar unter: <http://www.geoportal.de/DE/GDI-DE/INSPIRE/Direktive/Monitoring-und-Reporting/monitoring-und-reporting.html?lang=de> (letzter Aufruf: 10.10.2015)

Im letzten Jahr (2014) gab es bereits 26.593 INSPIRE-relevante Geodatenätze- und dienste innerhalb des Geodatenkatalog.de (vgl. Abb.2). Hinzu kommt, dass nicht klar ist, wie viele Geodaten(-sätze) und -dienste in Deutschland zwar von der INSPIRE-Richtlinie betroffen sind und damit veröffentlicht werden müssen, es jedoch noch nicht sind.

Fest steht, dass am Ende (2021) zu allen betroffenen behördlichen Geodaten Metadaten INSPIRE-konform erzeugt und zugänglich gemacht werden müssen (INSPIRE-ROADMAP).

Diese Entwicklung zeigt, dass eine Suche im Geodatenkatalog.de in Zukunft damit rechnen muss, mit steigenden Metadatenatzzahlen umzugehen.

Im „Midterm Evaluation Report“ der EU-Kommission von 2014 werden insgesamt 5 Kernprobleme bei der Umsetzung der INSPIRE-Richtlinie identifiziert: Fehlende oder unvollständige Geodaten, unvollständige Metadaten, Schwierigkeiten bei der Kombination verschiedener Geodatenätze, vielfache Unzugänglichkeit und verschiedene Barrieren beim Datenaustausch.

Deshalb schlägt die Kommission einige Anpassungen und Verbesserungen vor. Unter anderem soll es Trainingsprogramme für die betroffenen Behörden geben. Des weiteren müssen administrative Beschränkungen aufgehoben werden und die nationale und internationale Koordination soll

verbessert werden (EUROPEAN ENVIRONMENT AGENCY:5-6).

In Deutschland werden vom Lenkungsgremium GDI-DE zu diesem Zweck Workshops mit geodatenhaltenden Stellen abgehalten. Es wird Informationsmaterial (z.B. Infobroschüren) und ein Wiki der GDI-DE bereitgestellt. Schließlich wurde ein automatisiertes Monitoring mithilfe der GDI-Registry ab 2014 angestrebt (KOORDINIERUNGSSTELLE² GDI-DE:8-9) und schließlich auch umgesetzt (LVERMGEO).

Die Hauptprobleme in Deutschland sind finanzieller und personeller Art. Die Kosten für die Umsetzung einer deutschen Geodateninfrastruktur beliefen sich 2012 auf 800.000€ und werden bis zum Jahr 2016 auf insgesamt 2.308.000€ steigen (KOORDINIERUNGSSTELLE² GDI-DE:23).

Hinzu kommen Schwierigkeiten mit der hohen Komplexität der INSPIRE-Anforderungen. Die Nutzung wird durch heterogene Regelungen zu Lizenzen, Preisen und Datenschutz zusätzlich erschwert. Nicht zuletzt trägt hierzu die föderale Struktur Deutschlands bei. Insgesamt muss europaweit im Rahmen von INSPIRE noch viel geleistet werden. Vor allem die Konformitätsprobleme und Zugangsbeschränkungen stellen für alle Mitgliedsländer eine große Herausforderung dar. Auch die hohen Aufbaukosten der GDI-DE werden Bund und Länder in den nächsten Jahren noch stärker belasten (KOORDINIERUNGSSTELLE² GDI-DE:23-26).

Hinsichtlich der Suche in Metadaten im Geodatenkatalog.de könnten die genannten Schwierigkeiten unter anderem die Datenqualität der Metadaten beeinträchtigen. Nichtsdestoweniger ist die Datenqualität entscheidend für eine effiziente Suche (Rashid-Pour Anhang, Heimbürger Anhang).

2.4. Der Geodatenkatalog.de

Der Geodatenkatalog.de ist „ein zentraler Dienst, der Katalogdienste bzw. deren Daten zusammenführt, an andere Strukturen abgibt und Recherchen nach Geodaten und Diensten erlaubt“ (HÜBEN 2011:227). Die Geodatenbestände in Deutschland werden dadurch transparent und recherchierbar. Dies ermöglicht Nutzern im Idealfall intuitiv und komfortabel nach Daten zu suchen, diese in verständlicher Form zu erhalten und zu visualisieren (WALTHER 2011:271).

Um auch in Zukunft herstellerunabhängig und flexibel zu sein, findet die technische Umsetzung ausschließlich mit *OpenSource* Software statt (HÜBEN 2011:227).

Zunächst werden die dezentralen Metadaten der GDI-DE, welche in Katalogen bei den Bundesländern oder Institutionen (z.B. Deutsches Zentrum für Luft- und Raumfahrt) vorliegen, gesammelt („geharvestet“). (HÜBEN 2011:228). Die gesammelten Metadatenbestände liegen nun zentral in einer *Geonetwork*-Datenbank und können an INSPIRE ausgegeben werden.

Für die Suchfunktion im Geoportal.de müssen die Rohdaten zunächst bereinigt werden (vgl. Kap. 3.2.) (WALTHER 2011:272).

Es folgt eine Indexierung aller Metadaten in einer einzigen Tabelle. Die Bereinigung und Indexierung übernimmt die Applikation „Geodatenuche“. Schließlich werden die Ergebnisse einer Suche über das Geoportal.de an den Nutzer ausgegeben (vgl. Abb.3).

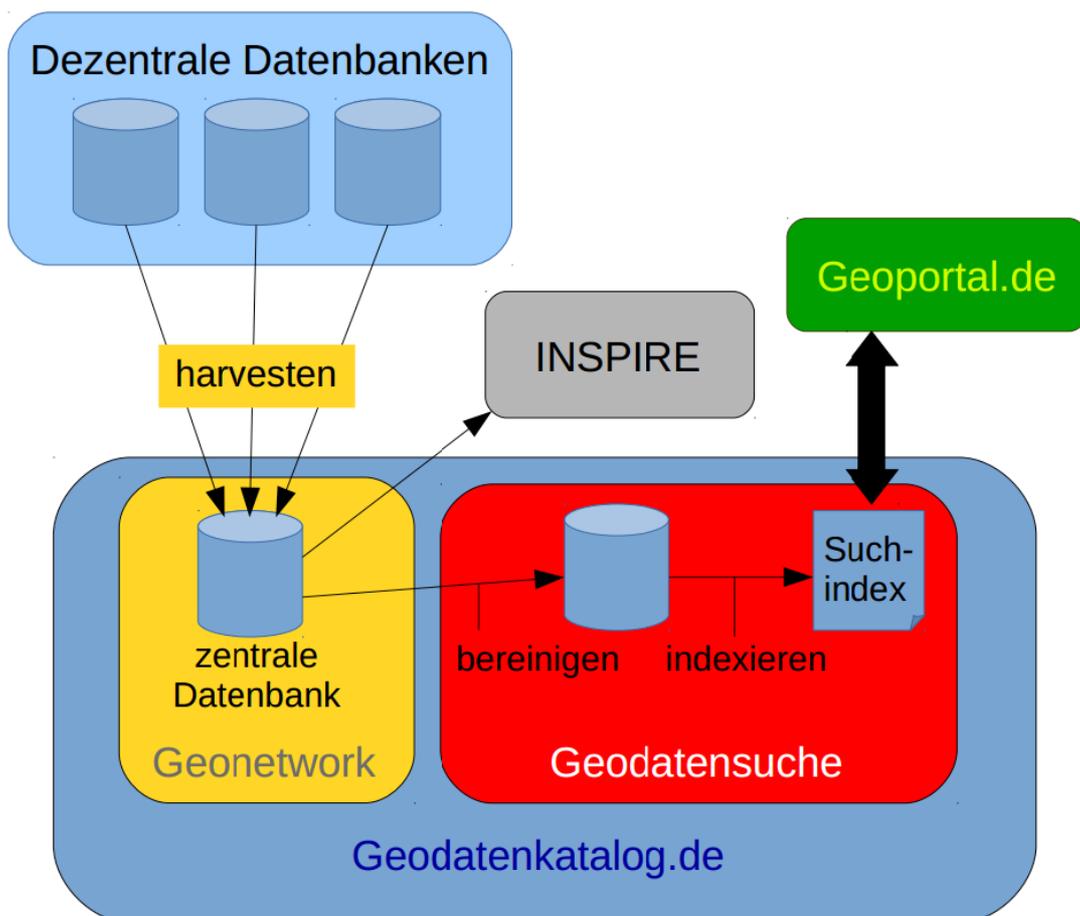


Abb.3: Architektur des Geodatenkatalog.de, eigene Darstellung.

Der Geodatenkatalog.de erlaubt so einen Gesamtüberblick über die Metadatenbestände der GDI-DE. Er erfüllt damit zusätzlich die Voraussetzung für das geforderte INSPIRE-Monitoring und eine Suche in den gesamten Metadatenbeständen. Des weiteren erlaubt er eine Überprüfung der (INSPIRE-)Konformität und Aktualität der Daten (WALTHER 2011:272).

Im Sinne des Publish-Find-Bind-Prinzips sollen Geodateninfrastrukturen ihre Geodaten über das Internet uneingeschränkt nutzbar machen. Sämtliche Kartenanwendungen im Internet können im Geodatenkatalog mithilfe ihrer Metadaten gefunden werden („find“). Dazu müssen sie ihre Geodaten veröffentlichen („publish“) und mit anderen geodatenhaltenden Stellen/Personen über einen Metadatenkatalog verknüpfen („bind“) (WALTHER 2011:273). Momentan ist dies bei den behördlichen Daten über die Länderkataloge und den Bundeskatalog umgesetzt.

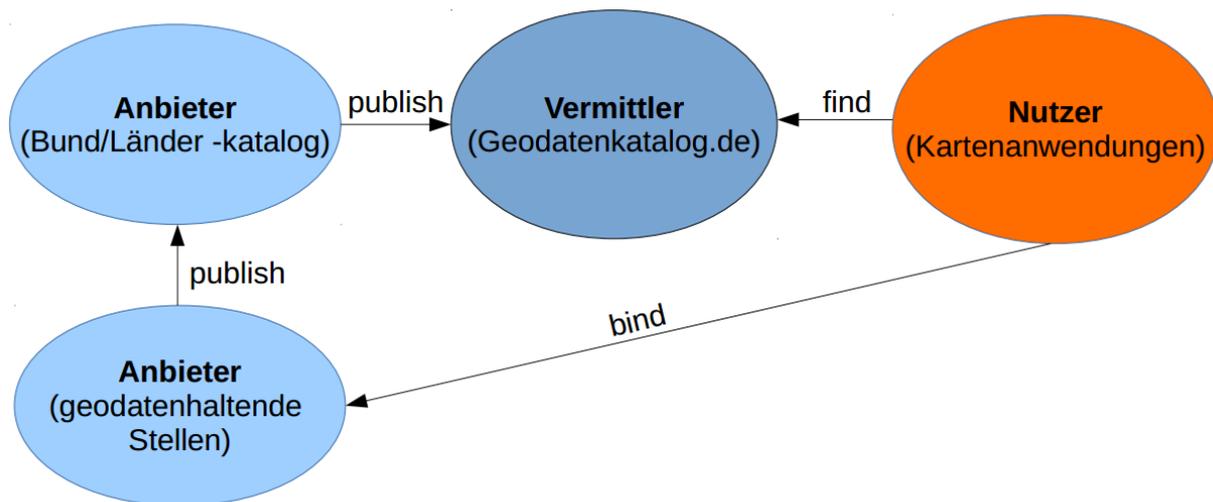


Abb.4: Das „Publish-Find-Bind“-Schema am Beispiel der GDI-DE, eigene Darstellung.

Die Umsetzung dieses Prinzips erfordert eine Standardisierung der Metadaten zur ihrer interoperablen Verwendung. Standardisierte Metadaten sind folglich die Grundlage für die Umsetzung des Prinzips „Publish-Find-Bind“ (WALTHER:273-274).

2.5. Metadaten der GDI-DE

Standardisierte Metadaten bei Geodaten bieten eine Reihe von Vorteilen: Sie verhindern überflüssige Datenerfassungen, zeigen Lücken in Datenbeständen auf, sichern die Qualität des Geodatensatzes, ermöglichen Vergleiche zwischen Geodatensätzen und sorgen für Interoperabilität und Transparenz des Geodatenbestandes (GDI-NI).

Metadaten können mit ihren Informationen auch den Vertrieb von Geodaten unterstützen. Im Geoportal gibt es für bestimmte Metadaten die Möglichkeit, sie direkt beim Anbieter per Mail zu bestellen (vgl. Abb.6).

Ein Anbieter von Metadaten möchte aber vor allem den Aufwand für die Datenabgabe minimieren und damit Kosten sparen. Ein Suchender hingegen erwartet, dass er einfach und gezielt das findet,

was er gesucht hat. Er weiß oft nicht was sich hinter dem Begriff „Metadaten“ verbirgt (KOORDINIERUNGSSTELLE GDI-DE¹:5).

Metadaten sind der Ausgangspunkt für die Beschreibung der Geodaten und Dienste. Mit Hilfe von Metadaten erhält man zunächst einmal Informationen zu den Daten selbst. Darüber hinaus kann sich der Nutzer in vielen Fällen, durch den Einsatz von weiterführenden (Hyper-)Links in den Metadaten, über Darstellungsdienste, Nutzungsbedingungen, Legenden und Datenmodelle informieren (KOORDINIERUNGSSTELLE GDI-DE¹:9-10).

Metadatenstandards werden in Deutschland durch die Vorgaben von INSPIRE, der GDI-DE und den ISO-Standards (International Standardization Organisation) beeinflusst. Die GDI-DE versucht dabei mit ihren Vorgaben eine gleichzeitige Realisierung von INSPIRE- und ISO-Standards zu fördern (INSPIRE INTEROPERABILITÄT) (AK METADATEN:9).

Schon seit 1946 entwickelt die ISO Standards, die weltweit gelten sollen. Die Organisation arbeitet hierbei unabhängig und nichtstaatlich an Spezifikationen für Produkte, Systeme und Dienste. Inzwischen sind so über 19500 verschiedene Standards entwickelt worden - so auch der Standard für die Metadaten geographischer Informationen (Nummer 19115) im Jahre 2003, der Standard für die Dienste geographischer Informationen (Nummer 19119) (2005) und der Standard für die XML-Kodierung geographischer Metadaten 19139 im Jahre 2007 (Iso).

Die technische Umsetzung der GDI-DE-Anforderungen wird vom Arbeitskreis Metadaten („AK Metadaten“) seit 2004 vorangetrieben. Der AK Metadaten ist eine Expertengruppe aus Vertretern der Wirtschaft und der Verwaltung. Er entwickelt Spezifikationen unter anderem auf Basis der ISO-Standards und arbeitet im Rahmen der GDI-DE (LENKUNGSGREMIUM GDI-DE:3).

Zu den Anforderungen der GDI-DE gehören unter anderem die Angabe einer ISO-Themenkategorie zusätzlich zu der Angabe eines INSPIRE-Annexthemas (z.B. INSPIRE-Annexthema: Gewässernetz, ISO-Themenkategorie: Binnengewässer), die Angabe eines Universal Unique Identifiers (UUID) und des Schlagwortes „inspireidentifiziert“ in den Metadaten, wenn der Metadatenatz von INSPIRE betroffen ist. Auch die Daten-Dienste-Kopplung, welche mithilfe von Metadaten der Dienste und Datensätze geschieht, wird hier beschrieben, um das Publish-Find-Bind-Konzept in der GDI-DE umzusetzen (AK METADATEN:10-11, 24).

Insgesamt haben INSPIRE-konforme Metadaten die folgende Struktur:

1. *Identifizierung* (Titel, Ressourcenverweis, Unique resource identifier...)
2. *Klassifizierung* (INSPIRE-Geodaten Thema)
3. *Schlüsselwort* (Schlüsselwort, u.U. Herkunft des Schlüsselwortes)
4. *Geographischer Standort* (Ausdehnung, mehrere Angaben möglich)
5. *Zeitangaben* (Erstellung, Veröffentlichung...)
6. *Qualität und Gültigkeit* (Herkunft der Daten, Auflösung)
7. *Übereinstimmung* (Inspire-Konformität, Durchführungsbestimmungen...)
8. *Zuständige Stelle* (Name, Adresse, Mail...)
9. *Zugangs- und Nutzungsbeschränkungen* (Beschränkung des öffentlichen Zugangs...)
10. *Metadaten über Metadaten* (Kontakt, Datum, Sprache...)

(vgl. INSPIRE METADATEN)

Dies ist jedoch nur ein kleiner Ausschnitt aus allen Elementen, die der ISO-Standard zu Metadaten von Geoinformationen definiert.

Man kann sich das Zusammenspiel von ISO, GDI-DE und INSPIRE wie folgt verdeutlichen: Es gibt einen sehr umfangreichen ISO-Standard zu Metadaten (ISO 19115) und einen ähnlich großen ISO-Standard zu Diensten (ISO 19119). Aus diesen Standards nimmt die INSPIRE-Richtlinie einige wenige Elemente, welche sie für entscheidend hält, heraus. Sie verändert einige der Anforderungen, fügt eigene Anforderungen hinzu und gibt sie EU-weit vor (vgl. INSPIRE). Die GDI-DE wiederum nimmt selbst einige ISO-Elemente aus den ISO-Standards und fügt sie ihren eigenen Anforderungen (z.B. Schlagwort „inspireidentifiziert“) hinzu (AK METADATEN:7,9). Die Bundesländer sind in der Verantwortung, gleichzeitig sowohl die INSPIRE-Vorgaben aufzunehmen, als auch die Anforderungen der GDI-DE umzusetzen. Dies ist durch die 16 Ländergesetze zur Geodateninfrastruktur in den Bundesländern umgesetzt worden (GEODATENZUGANGSGESETZ).

Um die Metadaten auf ihre Konformität zu überprüfen, stellt die GDI-DE die GDI-DE Testsuite bereit. Sie kann einerseits als Webanwendung (<http://testsuite.gdi-de.org/gdi/>), andererseits auch als lokale Anwendung und Schnittstelle genutzt werden (GDI-TESTSUITE).

Damit bietet die GDI-DE Testsuite die Möglichkeit Metadaten hinsichtlich ihrer INSPIRE-Konformität, Konformität zu GDI-DE-Konventionen und dem ISO-Standard zu validieren (AK METADATEN:31).

3. Ist-Zustand der Suche im Geodatenkatalog.de

3.1. Die Solr-Suche

Die Suche in Metadaten funktioniert im Geodatenkatalog.de mit der Kombination einer *php*-Suchapplikation und der Suchplattform „Solr“. Die Kenntnisse des Autors dieser Arbeit zu der Suche im Geodatenkatalog.de, stammen vor allem aus zahlreichen persönlichen Gesprächen mit dem damaligen Entwickler der Suche bei der WhereGroup GmbH & Co. KG.

Bevor die Suche so umgesetzt wurde, wie sie heute besteht, wurden die gesammelten Metadaten über das *CSW*-Protokoll ausgegeben. Diese Lösung existierte jedoch nur beim Vorgänger des Geoportal.de, dem GeoPortal.Bund (GEOPORTALBUND) (Heimbürger Anhang). Das CSW-Protokoll ist zwar bis heute der Standard, wenn es um den Austausch und die Katalogisierung von Metadaten geht, den Ansprüchen an eine moderne Suchapplikation wird es jedoch nicht gerecht. In Anbetracht der geringen Geschwindigkeit (Performance), den eingeschränkten Suchmöglichkeiten und dem Fehlen wichtiger Funktionen einer Suche (z.B. „autocomplete“ und „autosuggest“) wurde eine andere Metadatensuche im Geodatenkatalog.de gebraucht (Rashid-Pour Anhang).

Die Java-Suchplattform Solr ist geeignet den Anforderungen des Geodatenkatalog.de zu genügen und wurde deshalb mithilfe der *php*-Applikation „Geodatensuche“ in die Metadatensuche im Geodatenkatalog.de eingebunden. „Solr“ läuft als eigenständiger Suchserver unter einer Open-Source-Lizenz auf unterschiedlichen Plattformen wie Windows, Mac und Linux. Außerdem wird Solr kontinuierlich weiterentwickelt. Es kommen also immer neue Funktionen hinzu und Fehler und bestehenden Funktionen werden verbessert. Für die Geodatensuche sind vor allem die Ranking-Funktionen, die (räumlichen) Filterfunktionen und die einfache Bedienung von Solr von besonderer

Bedeutung (SOLR FEATURES) (Rashid-Pour Anhang).

Zahlreiche aktuelle und nennenswerte Internetseiten wie das Weiße Haus, AOL, Instagram, Netflix oder das deutsche Ebay nutzen Solr und sorgen so für eine aktive Gemeinschaft und rege Nachfrage (Solr WEBSITES).

Eng mit Solr verknüpft ist die Java-Bibliothek Lucene. Sie ist sozusagen der Motor von Solr, wenn Solr ein Auto wäre. Ein Auto kann nicht ohne seinen Motor fahren, jedoch kann man den Motor auch in ein anderes Auto einbauen (LUCENE). In diesem Sinn übernimmt Lucene die Indexierung nach den Vorgaben von Solr.

Solr besitzt zahlreiche Funktionen/Merkmale (features), welche über den reinen Indexierungsvorgang hinausgehen. Diese Funktionen setzen an verschiedenen Punkten des Suchvorgangs an (z.B. Indexierung, Indexaufbereitung, Suchergebnisausgabe und Ergebnisanalyse) und ermöglichen zahlreiche Anpassungsmöglichkeiten (vgl. Solr FEATURES).

Momentan wird die Solr-Version 3.3.0 aus dem Jahr 2011 im Geodatenkatalog.de genutzt (Heimbürger Anhang).

3.2. Die Aufbereitung der Geodaten

Sobald die Metadatenätze in der zentralen Geonetwork-Datenbank vorliegen, kann die Bereinigung der Metadaten beginnen. Die Aufbereitung der Metadaten ist entscheidend für eine performante Suche, denn Metadaten sind der „Schlüssel“ zu den dahinter liegenden Geodaten.

Die Suchapplikation „Geodatenuche“ ermöglicht die automatisierte Aufbereitung der Geodaten, bei der sie verschiedenste Aufgaben übernimmt. Zu den vergleichsweise einfachen Funktionen bei der Aufbereitung zählt eine Transformation aller Schriftzeichen der Metadaten in den Zeichensatz UTF-8, eine Beseitigung aller HTML, CSS und JavaScript-Elemente in den Metadaten und eine Berechnung der Fläche des Kartenausschnittes (Bounding Box) der Geodaten.

Die Angabe der Ausdehnung der Geodaten in einem so genannten geographischen Begrenzungsrechteck (Bounding Box) liegt zwar in den Metadaten vor (vgl. Kap.2.5.), kann jedoch fehlerhaft sein. Liegen die Koordinatenangaben der Bounding Box im Abschnitt „Ausdehnung“ der Metadaten außerhalb Europas, werden sie nicht in die bereinigte Datenbank übernommen.

Nicht-valide Werte bei den Längen- und Breitengraden im Wertebereich des WGS84 (z.B. -999) führen ebenso zur Nicht-Übernahme der Metadaten (Rashid-Pour Anhang).

Das WGS-84 (World Geodetic System 1984) ist ein geodätisches Datum und weltweites Bezugssystem. Es basiert auf dem WGS-84-Referenzellipsoid - einer mathematischen Annäherung an die Form der Erde. Eine Einteilung in Längen- und Breitengrade der Erde dient dabei der eindeutigen Positionsbestimmung. Später wurde das WGS-84 zur Grundlage für das GPS (Global Positioning System) (GPS).

In den Metadaten selbst gibt es keine Angabe des Koordinatenreferenzsystems, deshalb ist es nicht möglich vom WGS-84 abweichende Metadatenätze und -dienste auszusortieren, sofern sie Werte im validen Wertebereich des WGS-84 besitzen (-180° bis 180° für Längengrade, -90° bis 90° für Breitengrade) (Rashid-Pour Anhang).

Zu der Bereinigung der Metadaten kommt eine Kopplung der Dienste mit den dazugehörigen Daten

über die Metadaten der Dienste und der (*Layer*)-Daten. Die „Geodatenuche“ verwendet hierzu die Dienste-URL und die Layer-Ids der Geodaten, welche dem Dienst zugeordnet werden.

An den Dienst wird zu diesem Zweck eine *GetCapabilities*-Anfrage mithilfe der URL des Dienstes gestellt. Jeder Dienst besitzt eine URL mit der man ihn über das Internet über ein Protokoll (z.B. HTTP, FTP...) erreichen kann. Diese URL ist in den Metadaten zu dem Dienst angegeben und wird nun verwendet um durch eine *GetCapabilities*-Anfrage dem Dienst zugeordnete Layernamen und die Layer-Ids zu erfahren (beispielsweise Wasser=Layername 123=Layer-ID). Aus den Metadatensätzen der Geodaten, welche dem Dienst zugeordnet sind, werden nun die Layer-Ids (z.B. 123) extrahiert. Nun können den Layer-Ids der Metadaten (z.B. 123), durch die Informationen aus dem *GetCapabilities*-Aufruf des Dienstes Layernamen (z.B. Wasser) zugewiesen werden.

Ursprünglich enthalten die Metadaten der Layer nicht die Layernamen, sondern nur die Layer-IDs. Mit diesem Verfahren entsteht eine wechselseitige Kopplung zwischen Metadaten und Dienst. Schließlich werden die Layernamen der Layer und der zugehörige Dienst in die bereinigte Datenbank übertragen. Von dort aus wandern diese Informationen im nächsten Schritt in den Suchindex (Rashid-Pour Anhang) (vgl. Abb.5).

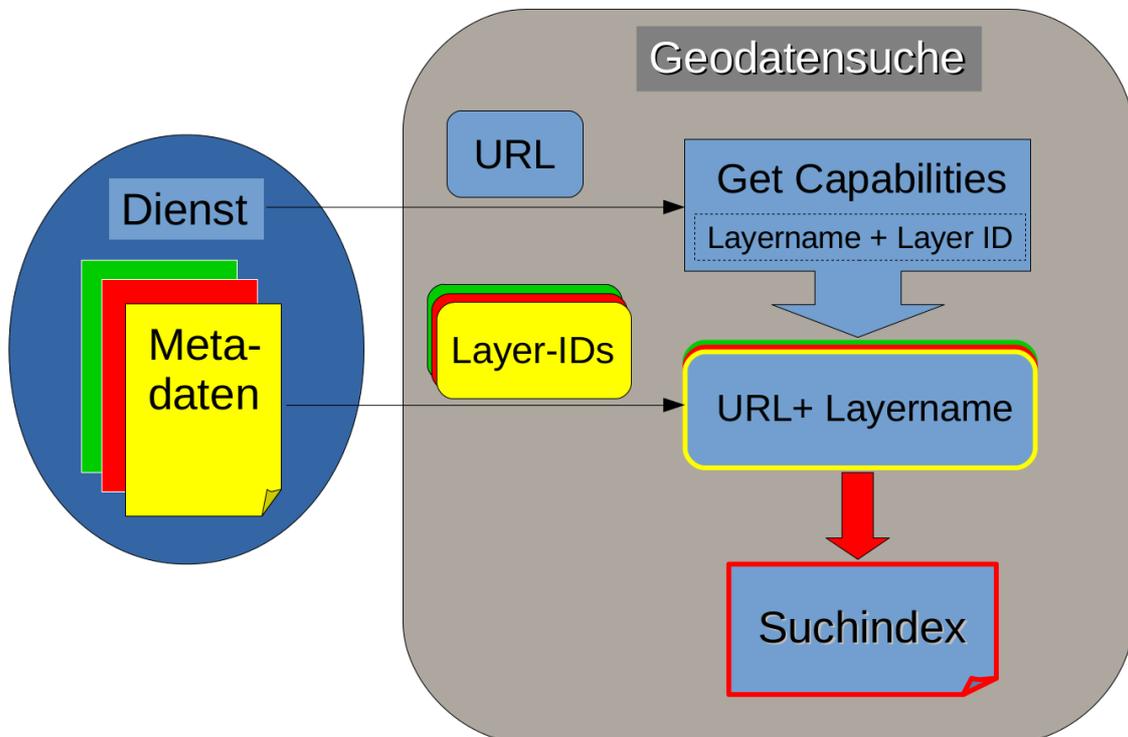


Abb.5: Die Daten-Dienste-Kopplung der „Geodatenuche“ im Geodatenkatalog.de, eigene Darstellung.

Mit der Daten-Dienste-Kopplung erhält man eine Unabhängigkeit der Daten vom Dienst. Einzelne Daten sind als Layer nun eindeutig einem Dienst zugeordnet. Sie verweisen so auf den Dienst und er verweist auf sie, obwohl Layer und Dienst im Geodatenkatalog.de durch verschiedene Metadatensätze repräsentiert sind. Wenn es sich bei dem Dienst um einen Darstellungsdienst handelt, können dadurch einzelne Layer dieses Dienstes im Geoportal.de ein- oder ausgeblendet werden (GEOPORTAL).

3.3. Der Suchvorgang im Geodatenkatalog.de

Liegen die aufbereiteten Daten erst einmal in der Datenbank, so kann daraus ein Index aufgebaut werden. Im Prinzip ist der Index eine einzige große Tabelle. In den Zeilen stehen die Metadaten der Geodaten und Dienste, während in den Spalten die Metadateneigenschaften (z.B. Bounding Box) abgetragen werden. Die Lucene-Komponente der Suche baut den Index auf und liefert die Ergebnisse an den Solr-Suchserver. Die Ergebnisse der Suche werden wiederum an die „Geodatenuche“ weitergeleitet und können schließlich im Geoportal angezeigt werden (vgl. Abb.6).

The screenshot shows the search interface of Geoportal.de. On the left, there are filter panels for 'Themengebiete' and 'Datentyp'. The main search area has two input fields: 'Was suchen Sie?' containing 'wasser' and 'Wo suchen Sie?'. Below these are example suggestions and a 'Suchen' button. The results are categorized into 'Karten (987)', 'Geodaten (7106)', and 'Webseite (55)'. The first result is 'WMS Leitungsnetz des Spremberger Wasser- und Abwasserzweckverbandes' with a date of 2015-07-01 and a rating of 3 stars. The second result is 'Wasser- und Abwasserzweckverband Werder Havelland (WMS)' with a date of 15-09-11 and a rating of 3 stars. The interface includes a 'Sortieren nach' dropdown set to 'Relevanz' and a 'Treffer 1 - 10 von 987' indicator.

Abb.6: Bildschirmfoto der Suche des Geoportal.de und seiner Ergebnisausgabe mit einer beispielhaften Suchanfrage nach dem Stichwort „wasser“, Abrufbar unter:

<http://www.geoportal.de/DE/Geoportal/Suche/suche.html?lang=de&what=wasser> (letzter Aufruf 10.10.2015)

Eine Suche läuft in folgenden Schritten ab:

1. Suchanfrage im Geoportal.de
2. Generieren eines Suchstrings (Suchanfrage) der Geodatenuche
3. Übermitteln des Suchstrings an den Solr-Suchserver
4. Anfrage des Solr-Suchservers an die Lucene-Bibliothek
5. Ausgabe der Ergebnisse an den Solr-Suchserver
6. Verarbeitung der Ergebnisse
7. Ausgabe der Ergebnisse an die „Geodatenuche“ mit Hilfe einer *Json*-Datei
8. Darstellung der Ergebnisse über die Seite Geoportal.de

(Rashid-Pour Anhang)

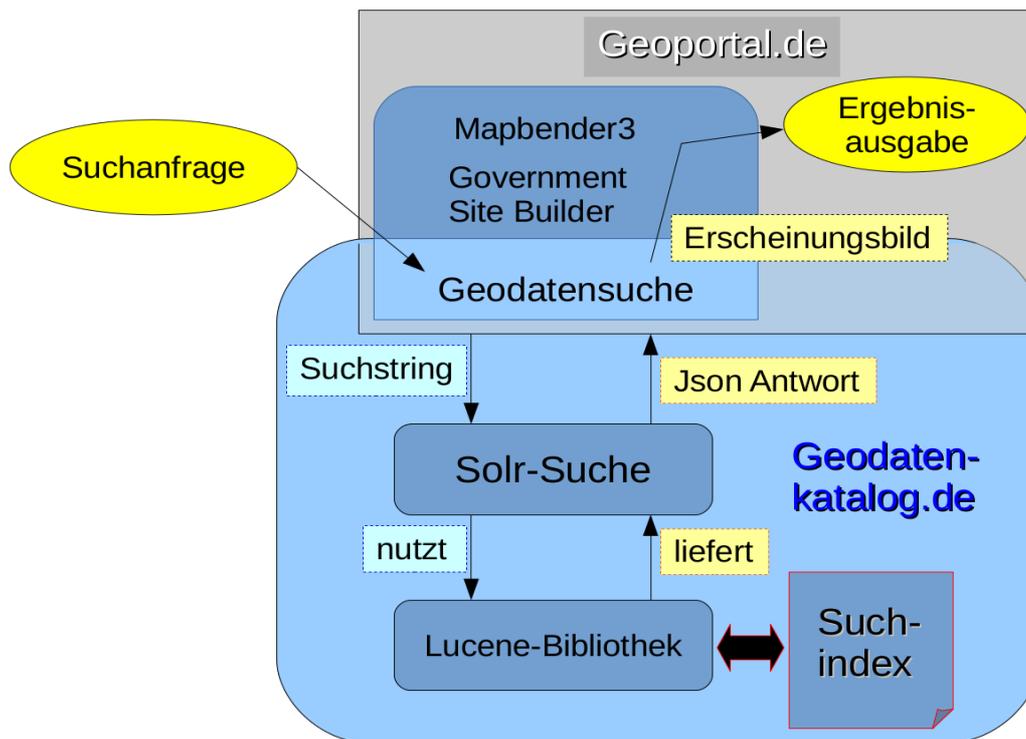


Abb.7: Schema des Suchvorgangs im Geodatenkatalog.de und im Geoportal.de, eigene Darstellung.

Das Geoportal.de besteht aus den 3 Komponenten: Mapbender3, dem Government Site Builder (GSB) und der „Geodaten suche“. Die Grundlage für die Webseite des Geoportal.de ist das Content-Management-System GSB für Webanwendungen der deutschen Verwaltung. Es bildet den Rahmen für alle Inhalte der Webseite. Sämtliche Texte, Fotos, Videos usw. können mit seiner Hilfe der Seite hinzugefügt werden (BVA).

Mapbender3 ist ebenso ein Content-Management-System wie GSB, jedoch ist es speziell für Kartenanwendungen konzipiert. Mit Mapbender3 kann man Kartenanwendungen auf Webseiten anzeigen, diese konfigurieren und den Zugriff darauf kontrollieren (MAPBENDER3).

Im Hintergrund der Webseite wirkt die „Geodaten suche“, welche die Suchanfrage entgegen nimmt und die Suchergebnisse aus dem Geodatenkatalog.de ausgibt. Sie ist gewissermaßen das Bindeglied zwischen dem Geoportal.de und dem Geodatenkatalog.de (vgl. Abb.7) (Rashid-Pour Anhang).

Alle Suchergebnisse werden mit der Applikation „Geodaten suche“ im Geoportal.de angezeigt. Das schließt die gefilterte Darstellung der Ergebnisse nach den folgenden Filteroptionen mit ein (vgl. Abb.6):

- *Themengebiete* (z.B. Wetter/Klima, Geologie/Bodenkunde, Wasser usw.)
- *Datentyp* (z.B. Geodatenatz, Kartendienst, Download-Dienst usw.)
- *Bezug* (Bestellen, Download)
- *Datum* (von „Datum“ bis „Datum“)
- *Maßstab* (von „Maßstab“ bis „Maßstab“)
- *Bewertung* (von 1 bis 5 Sterne)
- *INSPIRE* (INSPIRE(-identifiziert) (vgl. Kap.2.5.)

Solche Filterfunktionen für die Metadatenuche im Geodatenkatalog.de beispielsweise hätten mit einem CSW-Protokoll nicht umgesetzt werden können (Rashid-Pour Anhang) (CSW-FEATURES).

3.4. Die Volltextsuche des Geodatenkatalog.de

Für die Volltextsuche werden beim Indexieren drei neue Spalten im Index angelegt. In eine Spalte wird der Titel (title) des Geodatensatzes bzw. -dienstes abgelegt. In die zweite Spalte fallen alternativer Titel (alternatetitle), Kurzbeschreibung (abstract) und Schlüsselwörter (keywords) der Metadaten. Die dritte Spalte, welche für die Volltextsuche im Index angelegt wird, enthält die Informationen aus den Feldern Themenkategorie (topicCategory) und Anbieter (provider).

Alle anderen Informationen aus den Metadaten werden bei der Suche bisher nicht beachtet (Rashid-Pour Anhang).

Mithilfe der Administrationsfunktionen der „Geodatenuche“ kann der Betreiber des Geodatenkatalog.de, also das BKG, die Gewichtungen der Spalten bei der Volltextsuche frei festlegen. Bei der aktuellen Suche sind die 3 Spalten wie folgt gewichtet:

- 1 Spalte (title) > 2 Spalte (alternatetitle, abstract, keywords) > 3 Spalte (topicCategory, provider)

Sobald ein Nutzer des Geodatenkatalog.de eine Sachsuche über das Geoportal.de startet, werden nach dem bereits beschriebenen Suchschema (Abb.7) die drei Indexfelder nach Entsprechungen durchkämmt.

Hierbei sind folgende Suchergebnisse zugelassen:

- text (exakte Entsprechungen des Wortes „text“ z.B. Wasser)
- text* (Wörter, die mit „text“ beginnen, sich danach aber fortsetzen, z.B. Wasserwege)
- *text (Wörter, die mit anderen Buchstaben beginnen und danach mit „text“ fortfahren, z.B. Abwasser)
- *text* (Wörter, die „text“ als mittleres Element enthalten, z.B. Trinkwassernetz)

Solr gewichtet diese Treffer bei der Suche noch einmal intern. Je öfter ein Wort in einer Spalte (z.B. in der Kurzbeschreibung) steht, desto höher wird es gewichtet. Außerdem gewichtet Solr, je nach Übereinstimmung mit dem gesuchten Begriff. Die Höhe der Gewichtung verhält sich hierbei wie folgt:

text > text* > *text > *text*

(Rashid-Pour Anhang)

3.5. Die Ortssuche im Geodatenkatalog.de

Ganz ähnlich funktioniert auch die Ortssuche im Geodatenkatalog.de. Auch hier werden die Ergebnisse aus Indexspalten extrahiert, die für die Ortssuche vorher von Lucene angelegt wurden - auch hier findet eine Gewichtung der Ergebnisse statt.

In der bereinigten Datenbank im Geodatenkatalog.de sind die Metadaten der Dienste und Geodatenätze hinterlegt. Im Abschnitt „Ausdehnung“ (extent) sind die Begrenzungsrechtecke (bounding box) der Metadaten angegeben (vgl. Kap. 2.5.). Die Ortssuche findet anhand der ersten dort eingetragenen Bounding Box statt. Wenn im „Ausdehnung“-Abschnitt mehrere Bounding Boxen eingetragen wurden, wird nur die erste Bounding Box aus den Metadaten in den Lucene-Index abgelegt (Rashid-Pour Anhang).

Über die Ortssuche angefragte Kartenausschnitte in Form eines Begrenzungsrechtecks können in verschiedenen Lagebeziehungen zu den Kartenausschnitten in den Metadaten des Geodatenkatalog.de stehen.

Fragt man zum Beispiel die Bounding Box „X“ über das Geoportal.de an, so generiert die „Geodatenuche“ einen Suchstring an Solr (vgl. Abb.7).

Die Anfrage wird im Geoportal.de über ein separates Suchfeld verarbeitet, wobei hier durch eine automatische Vervollständigungsfunktion (autosuggest) einen Ort vorgibt. Hinter den vorgegeben Ortsangaben (z.B. 53113 Bonn) stehen die Datensätze der Adressdaten des Bundes und Geographische Namensbestände (z.B. Landschaft Odenwald). Der Nutzer kann alternativ ein Gebiet über Koordinatenangaben oder einen Kartenausschnitt manuell bestimmen (vgl. Abb. 6) (DATENGRUNDLAGE GEOPORTAL).

Mit diesen Funktionen kann ein Nutzer den Suchbereich festlegen, in dem er vorwiegend Suchergebnisse erhalten möchte.

Im Folgenden entscheidet nun die relative Lage des angefragten Suchbereichs bzw. der Bounding Box zu dem Begrenzungsrechtecken der Metadatenätze des Geodatenkatalog.de, welche im Index hinterlegt sind.

Die folgenden Lagebeziehungen der Bounding Boxen gehen in die Gewichtung der Suchergebnisse ein (vgl. Abb.8):

1. Die angefragte Bounding Box (X) enthält („contains“) eine Bounding Box (A)
2. Die angefragte Bounding Box (X) liegt innerhalb („within“) einer Bounding Box (B)
3. Die angefragte Bounding Box (X) überlappt („intersect“) mit einer Bounding Box (C)

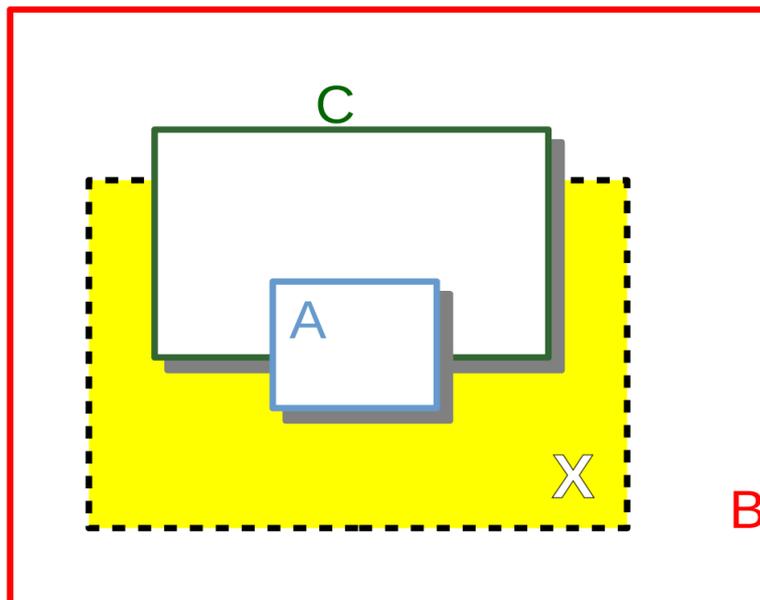


Abb.8: Beispiel der genutzten Lagebeziehungen der angefragten Bounding Box (X) zu möglichen Ergebnis-Bounding Boxen (A,B,C) der Ortssuche, eigene Darstellung

Diese Ergebnisse werden von Lucene erzeugt, über Solr an die „Geodatenuche“ zurückgegeben und werden dann im Geoportal.de angezeigt.

Die Gewichtung der Lagebeziehungen stellt sich bei der Ortssuche wie folgt dar (vgl. Abb.8):

Typ A (contains) > Typ C (intersect) > Typ B (within)

Bei der Suche werden nachfolgende Lagebeziehungen von Bounding Boxen nicht berücksichtigt:

- Die angefragte Bounding Box stimmt mit einer Bounding Box überein (equal to)
- Die angefragte Bounding Box berührt eine Bounding Box (touch)
- Die angefragte Bounding Box liegt außerhalb einer Bounding Box

Innerhalb der Typen A (contains), B (within) und C (intersect) werden die Ergebnisse nicht mehr „geobasiert“ differenziert (z.B. aufgrund der gemeinsamen Fläche).

Das bedeutet auch, dass gleiche Lagebeziehungen (z.B. intersect) bei einer Suchanfrage gleich gewichtet werden, ungeachtet der Verhältnisse innerhalb dieser Lagebeziehung. Eine indexierte Bounding Box (C1), welche die angefragte Bounding Box (X) nur minimal schneidet, wird genau so gewichtet, wie die Bounding Box (C2) mit anteilmäßig größerer Schnittfläche.

Ebenso verhält es sich zwischen den Lagebeziehungen. Gesetzt den Fall, eine Bounding Box (B) ist nur geringfügig größer als die angefragte Bounding Box (X), so wird sie doch stets geringer gewichtet als eine Bounding Box (C), obwohl beispielsweise die Bounding Box (C) die angefragte Bounding Box (X) nur leicht schneidet (vgl. Abb 8) (Rashid-Pour Anhang).

Als Ende des Suchprozesses verrechnet Solr die Ergebnisse der Volltextsuche mit denen der Ortssuche nach einer voreingestellten Gewichtung (Rashid-Pour Anhang).

4. Ist-Zustand der Metadaten im Geodatenkatalog.de

4.1. Die Extraktion der Metadaten

Eine Suche kann nur so gut wie die zugrunde gelegten Metadaten sein. Dies gilt auch für die Metadaten des Geodatenkatalog.de, welche von Bund, Länder, Kommunen und Institutionen bezogen werden (Rashid-Pour Anhang). Dieser Arbeit liegen die Metadaten der Geonetwork-Datenbank (vgl. Abb.3) vom 04.07.2015 zugrunde. Geonetwork gibt automatisch eine Struktur der Datenbank vor. Alle Tabellen in dieser Datenbank werden zunächst von Geonetwork angelegt und dann mit den gesammelten Metadaten befüllt.

Für die Suche im Geodatenkatalog.de verwendet die Applikation „Geodaten suche“ jedoch nur ausgewählte Informationen aus der Geonetwork-Datenbank.

Die Metadaten für den Suchindex im Geodatenkatalog.de werden von der „Geodaten suche“ aus der Tabelle „metadata“ dieser Datenbank extrahiert. Aus dieser Tabelle werden jedoch nur die Einträge des Feldes „data“ verwendet, welche die Informationen über die Metadaten in einem XML-Format enthalten (Rashid-Pour Anhang).

Im Folgenden wird erläutert, mit welchen Hilfsmitteln auf die Datenbank zugegriffen und die Metadaten analysiert wurden.

4.2. Analysewerkzeuge

Die Geonetwork-Datenbank des Geodatenkatalog.de ist eine relationale PostgreSQL-Datenbank. Das hier verwendete PostgreSQL ist ein OpenSource Datenbankmanagementsystem für relationale Datenbanken (OBE, HSU:IX).

Eine relationale Datenbank enthält Tabellen, die miteinander verknüpft sind und so aufeinander verweisen (IT WISSEN).

PostgreSQL bietet eine ganze Reihe von Funktionen, die über das reine Speichern von Daten in einer Datenbank hinausgehen. Hierzu gehören komplexe Datenfilterung, der Import und Export von Daten, Rechenoperationen und die Verwendung von Erweiterungen.

Um diese Funktionen nutzen zu können, stehen verschiedene Administrationswerkzeuge zur Verfügung. Bei der hier durchgeführten Analyse wurde beispielsweise die kostenlose graphische Benutzeroberfläche pgAdmin verwendet (OBE, HSU 2012:2).

Das Programm pgAdmin ermöglicht nicht nur über SQL-Abfragen eine Interaktion mit der PostgreSQL-Datenbank, es erlaubt auch den Zugriff auf die Daten allein durch die Bedienung des Programms mit der Maus (OBE, HSU 2012:43-44).

Ein weiteres Hilfsmittel zur Untersuchung der vorliegenden Metadaten ist die statistische Methode der reinen Zufallsstichprobe. Hierbei werden alle Elemente einer endlichen Grundgesamtheit durchnummeriert. Die Stichprobe wird dann mit einem Lotterieverfahren ermittelt (BAHRENBERG 1975:19).

Die Repräsentativität der Stichprobe hängt vom verwendeten Verfahren und der absoluten Zahl der Fälle ab. Nur zufällige Stichproben garantieren eine Repräsentation der Grundgesamtheit. Beim

Umfang der Stichprobe sollte darauf geachtet werden eine möglichst hohe Zahl an Fällen auszuwählen. Allgemein gelten Stichprobenumfänge von unter 30 als nicht repräsentativ (BAHRENBERG 1975:18).

Als drittes und letztes Hilfsmittel wird bei der Betrachtung der Datenqualität die Applikation der GDI-Testsuite eingesetzt (vgl. Kap. 2.2.). Die Webanwendung der GDI-Testsuite umfasst hierzu die notwendigen Funktionalitäten (GDI-DE TESTSUITE).

4.3. Analysevorgang

Mit den beschriebenen Werkzeugen wurde eine Analyse der Datenqualität der Metadaten des Geodatenkatalog.de durchgeführt.

Mit pgAdmin wurde dafür eine lokale Datenbank mit den Metadaten des Geodatenkatalog.de angelegt. Von den Einträgen der „metadata“-Tabelle in dieser Datenbank wurden dann mithilfe der reinen Zufallsstichprobe 100 Zeilen ausgewählt. Aus diesen 100 Zeilen wurde der Inhalt des „data“-Feldes in ein XML-Dokument extrahiert.

Bei diesem Vorgang wurde ein interner Zufallsalgorithmus von PostgreSQL verwendet, der aus allen „data“-Feldern der „metadata“-Tabelle 100 Einträge zufällig auswählt.

Dieser Vorgang wurde durch ein *Script* initiiert. Mit dessen Hilfe konnten die Daten in einem nächsten Schritt extrahiert werden (Anhang Script). Das Script führte dafür folgende Operationen durch:

- Führe den folgenden Befehl 100 mal aus
- Wähle in PostgreSQL ein zufälliges „data“-Feld aus der Tabelle „metadata“ der betreffenden Datenbank aus
- Speichere den Inhalt des „data“-Feldes in einer Datei mit der Endung .xml und nummeriere sie

Die zu testenden 100 XML-Dokumente wurden jetzt in die Webanwendung der GDI-Testsuite (<http://testsuite.gdi-de.org/gdi/>) geladen und auf ihre Konformität zu den Anforderungen von ISO, GDI-DE und INSPIRE überprüft. Hierzu legt man zunächst einen Testfall im Testmanagement der Testsuite an, bei dem die entsprechende Testklasse und die Konformitätsklassen konfiguriert werden.

Zusätzlich zur Konformitätsanalyse wurden die zufällig ausgewählten Metadaten händisch auf Fehler im Bezug auf ihre Bounding Box untersucht. Hierzu wurden die Angaben im Bereich Ausdehnung in den XML-Dokumenten mit den Angaben in den Bereichen Kurzbeschreibung und Titel der selben Datei verglichen. Stimmt diese nicht überein, handelte es sich um eine fehlerhafte Angabe der Bounding Box.

Falls in den Metadaten das Stichwort „inspireidentifiziert“ auftauchte, so wurde auch dieses erfasst.

Außerdem konnte überprüft werden, ob der Geodatensatz des Geodatenkatalog.de indexiert wurde, indem nach der UUID (vgl. Kap. 2.5) des Datensatzes im Geoportal.de gesucht wurde. Wenn der gesuchte Geodatensatz/ der gesuchte Dienst gefunden wurde, konnte die Bounding Box über den

Punkt „Gebiet“ in der Trefferliste visualisiert werden. Dies half zusätzlich bei der Einschätzung der Richtigkeit der Bounding Box.

Praktisches Vorgehen:

- Öffnen der XML-Datei
- Suche des Schlagworts „inspireidentifiziert“
- Suche der UUID des Geodatensatzes/dienstes über das Schlagwort „fileIdentifier“
- Eingabe der UUID in das Suchfeld des Geoportal.de
 - kein Ergebnis: Geodatensatz/dienst wurde nicht indexiert
→ Überprüfung der Existenz des Anbieters im Geoportal.de
 - mehr als 1 Ergebnis: zusätzliche Eingabe des Titels (im ersten „title“-Abschnitt des XML-Dokuments) in die Suche
 - genau 1 Ergebnis: Vergleich der Bounding Box mithilfe des „Gebiet“-Buttons und des Titels + Kurzbeschreibung
→ Bei Orthofotos den Anbieter bestimmen
- Bewertung der Übereinstimmung („ja“ oder „nein“) der in den Metadaten (im XML-Dokument) angegebenen Bounding Box mit der zu erwartenden Bounding Box

4.4. Ergebnis der Metadatenanalyse

4.4.1. Konformität

Die durchgeführten Tests in der GDI-DE Testsuite zeigen für die gewählte Stichprobe eine ISO-Konformität von 33% und eine GDI-Konformität von 73% (vgl. Abb.9).

Häufige Fehlermeldungen in den Metadaten, welche von der GDI-Testsuite festgestellt wurden, waren:

fehlende Bewertung der Qualität in den Metadaten (Element „dataQualityInfo“)

↳ *ISO + INSPIRE-Anforderung*

fehlende Angabe des Ressourcentyps (z.B. Service) (Element „hierarchyLevel“)

↳ *GDI-DE-Anforderung*

fehlende Daten-Dienste-Kopplung (im Element „srv:connectPoint“)

↳ *GDI-DE-Anforderung*

fehlende Angabe des Abgabeformats und seiner Version (Element „distributionFormat“)

↳ *ISO-Anforderung*

(TESTMANAGEMENT TESTSUITE)

(Anhang Testfehler)

Die INSPIRE-Konformität beläuft sich auf 22% der Stichprobe (vgl. Abb.9). Doch ein Metadatensatz, der als „inspireidentifiziert“ in den Metadaten bezeichnet wurde, ist nicht immer INSPIRE-konform und umgekehrt. In der vorliegenden Stichprobe waren von 23

„inspireidentifizierten“ Metadatenansätzen nur 12 auch INSPIRE-konform.

Auf der anderen Seite waren von 22 INSPIRE-konformen Metadatenansätzen nur 12 zur gleichen Zeit „inspireidentifiziert“ (Anhang Analyse).

Bei der Beurteilung der durchgeführten Konformitätsanalyse müssen jedoch einige Faktoren beachtet werden, welche die hier gemessenen Ergebnisse relativieren könnten.

Erstens tauchen 22 der hier untersuchten Metadatenansätze im Index der Geodatenuche später nicht mehr auf (vgl. Abb.9). Dies kann verschiedene Gründe haben. Die hier verwendeten Metadatenansätze sind dem Geodatenkatalog.de am 04.07.2015 entnommen worden. Die vorliegende Analyse der Metadaten des Geodatenkatalog.de fand jedoch erst am 15.09-16.09.2015 statt. Zwischenzeitlich können also Änderungen des Metadaten-Sammelns seitens des BKGs vorgenommen worden sein. Ein Hinweis darauf ist auch, dass die Anbieter von 17 der 22 betroffenen Metadatenansätzen nicht im Geoportal.de zu finden sind (Anhang Analyse).

Auch von Seiten der Geodatenanbieter könnten die betroffenen Metadatenansätze inzwischen geändert worden sein, bzw. in dieser Form nicht mehr zur Verfügung stehen. Schließlich könnte die Geodatenuche auch Fehler bei Übernahme der Metadaten in den Index gemacht haben.

Zweitens stammen viele XML-Dokumente von einigen wenigen Anbietern. Beispielsweise sind von den 100 untersuchten XML-Dateien 24 Orthofotos von nur einem einzigen Anbieter. Insofern hat dieser Anbieter einen enormen Einfluss auf die prozentuale Konformität der Metadatenansätze in der Stichprobe. Dass dies nicht nur in der Stichprobe der Fall ist, beweisen über 27000 Treffer bei der Suche im Geoportal.de nach Geodatenansätzen von besagtem Anbieter (GEOPORTAL) (Anhang Analyse).

Drittens bestehen Metadatenansätze die Testsuite-Tests oft nur wegen einigen wenigen Mängeln nicht. Die XML-Datei ist in den meisten Fällen nahezu richtig aufgebaut. Immer wieder fehlt jedoch eine „Kleinigkeit“ (z.B. die Bewertung der Qualität des Metadatenansatzes). In die vorliegende Konformitätsanalyse geht jedoch nur das Endergebnis („Bestanden“ oder „Nicht Bestanden“) und nicht die prozentuale Konformität ein (TESTSUITE TESTMANAGEMENT).

Zuletzt ist die geringe Größe der Stichprobe (100) bei einer großen Grundgesamtheit ein Unsicherheitsfaktor. Ein größerer Stichprobenumfang wäre zur Erhöhung der Repräsentativität wünschenswert gewesen. Die absolute Größe der Stichprobe ist hierbei entscheidender als ihr Anteil an der Grundgesamtheit (BAHRENBERG 1975:18).

Aufgrund dieser Punkte ist die Bewertung der Standardkonformität der Metadaten mit der vorliegenden Stichprobe ein problematisches Unterfangen. Trotz der genannten Fehlerquellen, welche das Ergebnis der Untersuchung sicherlich beeinflusst haben, können folgende Aussagen über die Standardkonformität der Metadatenansätze des Geodatenkatalog.de formuliert werden:

1. Nur ein geringer Teil der Metadatenansätze im Geodatenkatalog.de ist standardkonform zu ISO 19115/19119 (33%) oder INSPIRE (22%).
2. Die Konventionen der GDI-DE bezüglich Metadaten werden zum Großteil erfüllt (73%).
3. Metadaten, die als „inspireidentifiziert“ bezeichnet werden, sind nicht immer INSPIRE-konform (nur 54%).

(Anhang Analyse)

4.4.2. Räumliche Bewertung

Der zweite Teil der Qualitätsanalyse der Metadatensätze des Geodatenkatalog.de befasst sich mit ihrer räumlichen Bewertung. Hierzu wurde die Ausdehnung der Geodatenätze und -dienste, welche in den Metadaten im Geodatenkatalog.de angegeben ist, auf ihre Richtigkeit hin überprüft. Anhand der Angaben im Titel und in der Kurzbeschreibung der Metadaten wurde festgestellt, ob die angegebene Bounding Box zweckmäßig gewählt wurde (vgl. Kap.4.3).

Bei den 22 nicht-indexierten Metadatensätzen war dies nicht möglich, sodass der Stichprobenumfang in Bezug auf diese Analyse nur 78 Fälle beträgt.

Von den untersuchten Geodatenätzen und -diensten hatten 73% eine passende Bounding Box. 27% der untersuchten Bounding Boxen waren fehlerhaft, wobei fast alle der fehlerhaften Bounding Boxen als „zu groß/grob“ erfasst wurden (vgl. Abb.9).

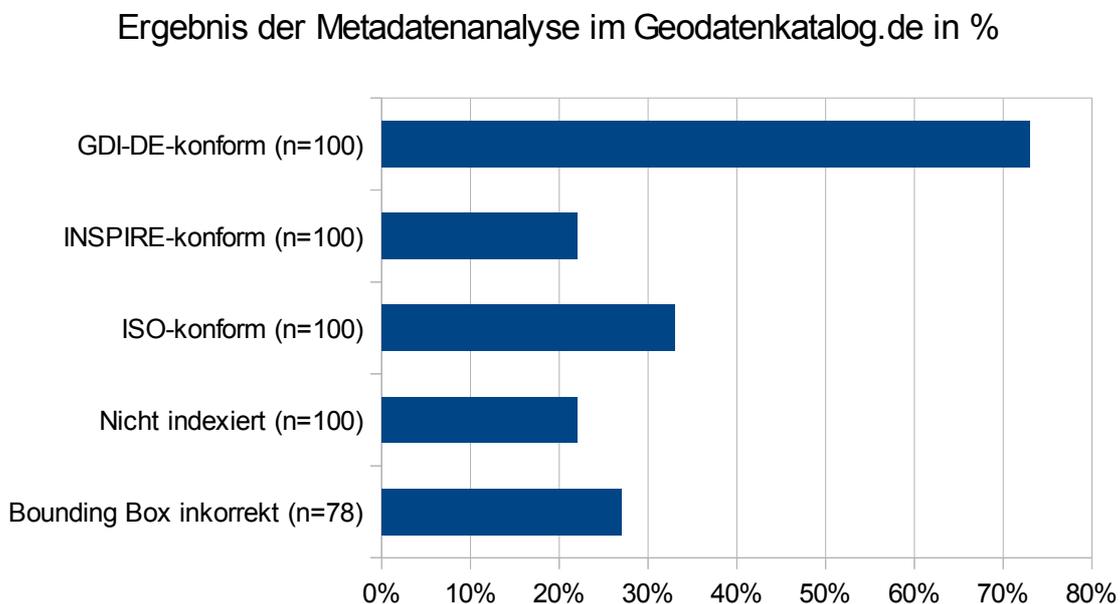


Abb.9: Anteil der untersuchten Parameter an der zufälligen Stichprobe aus der Geonetwork-Datenbank des Geodatenkatalog.de in Prozent (%), eigene Darstellung, Ergebnisse unter (Anhang Analyse)

Es gibt einige Gründe, welche für fehlerhafte Bounding Boxen in Betracht gezogen werden können. Die bei der Metadatenerfassung verwendeten Programme könnten hier ebenso eine Rolle spielen, wie die unsachgemäße Bedienung dieser. Es ist zudem nicht auszuschließen, dass übergroße Angaben der Bounding Box intendiert vorgenommen wurden, damit die Metadaten im Geoportal.de öfter gefunden werden (Rashid-Pour Anhang).

Eine Bewertung dieser Ergebnisse kann nur mit weiteren Einschränkungen geschehen.

Erstens sind die tatsächlichen Geodaten und ihre Lage nicht aus den Metadaten ablesbar (vgl. Kap. 2.5). Ohne die Kenntnis der Geometrien der Geodaten ist eine Einschätzung der Zuverlässigkeit der Bounding Boxen oft schwer - wenn nicht gar unmöglich. Liegen Geodatenätze ohne Verknüpfung

zu einem Darstellungsdienst vor, so kann man die dahinter stehenden Geometrien (im Geoportal.de) nicht visualisieren. Die Anbieter verbinden darüber hinaus oft die Informationen, welche sich aus der Darstellung bzw. der Nutzungen der Geodaten ergeben, mit Nutzungsgebühren. Aus diesen Gründen wird eine Darstellung der hinter den analysierten Metadaten stehenden Geodaten teilweise erschwert (GEOPORTAL.DE).

Zweitens stellt die Geodatenuche im Geoportal.de über den Button „Gebiet“ nur eine kleine Anzeige der Bounding Box bereit (vgl. Abb. 6). Außerdem wird dort die Bounding Box stets im gleichen Maßstab und innerhalb einer Deutschlandkarte angezeigt, was besonders bei kleinen Bounding Boxen (z.B. von Gemeinden oder Kommunen) von großem Nachteil ist, wenn man die Genauigkeit der Bounding Box einschätzen möchte. Es gibt keine Möglichkeit im Geoportal.de die Bounding Box wesentlich genauer ablesen zu können (GEOPORTAL.DE). Dies kann zu einer falschen Einschätzung der Zuverlässigkeit der Bounding Boxen führen.

Drittens fehlen im Titel und/oder der Kurzbeschreibung des Geodatensatzes in den Metadaten in einigen Fällen wichtige Angaben. Der Titel sagt nicht unbedingt immer etwas über das betroffene Gebiet der Geodaten aus (z.B. Dop20rgb_32645_8263_1_sh, DOP 20). Und auch die Kurzbeschreibung enthält zum Teil keine räumliche relevanten Informationen, die bei der Lokalisierung des Datensatzes helfen.

Insofern sind auch die Ergebnisse aus der räumlichen Analyse der Metadatenätze des Geodatenkatalog.de mit Vorsicht zu behandeln. Sie sind keinesfalls als gesichert zu betrachten. Trotzdem lassen sich generell folgende Tendenzen aus der räumlichen Analyse ziehen:

1. Die inkorrekte räumliche Verortung von Geodatensätzen und -diensten stellt ein wesentliches Problem in den Metadaten des Geodatendatenkatalog.de dar.
2. Der vorherrschende Fehler in Verbindung mit der räumlichen Ausdehnung von Geodatensätzen und -diensten im Geodatenkatalog.de ist eine zu große Angabe des Begrenzungsrechtecks (Bounding Box).
3. Fehler bei der Angabe des Begrenzungsrechtecks (Bounding Box) lassen sich anhand der Metadaten schwer nachvollziehen.

(Anhang Analyse)

5. Konzept einer Suche in Metadaten

5.1. Vorgehen und Bewertungskriterien

Im Folgenden werden Verbesserungsvorschläge gemacht, die eine effiziente Suche in Metadaten im Geodatenkatalog.de ermöglichen. Als Ausgangspunkt wird der Zustand der Geodatenuche im Geodatenkatalog.de/Geoportal.de (Stand: 01.07.2015) angenommen.

Folgende Anforderungen wurden von Seiten des BKG an die Metadatenuche im Geodatenkatalog.de formuliert:

1. Grundlage für eine effiziente Suche im Geodatenkatalog.de ist eine gute Datenqualität der Metadaten insbesondere in Bezug auf:
 - a) Die Bounding Box-Angaben
 - b) eindeutige Angaben in Titel, Kurzbeschreibung und verwendeten Schlüsselwörtern (Keywords)
2. Bei einer räumlichen Eingrenzung der Suche sollen Ergebnisse mit einer möglichst großen Abdeckung innerhalb der verwendeten räumlichen Ausdehnung zuerst angezeigt werden.
3. Maßnahmen, welche es für den Nutzer ermöglichen textliche Suchprozesse schneller und komfortabler durchzuführen wie z.B. die Verwendung von Thesauri, Autokorrektur, Highlighting, zusätzlicher Datenbestände (PLZ, Ortsvorwahlen, geographische Namensbestände etc.), Rankingverfahren, sind zu begrüßen.
4. Bei einer guten Datenqualität (s. Punkt 1) soll sichergestellt werden, dass die gesuchten Geodaten bzw. Dienste möglichst als erstes (mindestens jedoch auf der ersten Seite) angezeigt werden.
5. Die räumliche Eingrenzung der Suche soll möglichst intuitiv und komfortabel für den Nutzer gestaltet werden.

(Heimbürger Anhang)

Ein Handlungsbedarf ergibt sich aus der Verbindung dieser Anforderungen an eine Suche in Metadaten im Geodatenkatalog.de („Soll-Zustand“) mit dem Ist-Zustands der Suche (vgl. Kap. 3.4.). Die Ergebnisse aus der Metadatenanalyse (vgl. Kap.4.4) zeigen eine darüber hinaus eine Diskrepanz im Bereich der Datenqualität zwischen Soll- und Ist-Zustand auf.

Mit Bezug zu den Anforderungen (1-5) wird ein Maßnahmenkatalog entwickelt, mit welchem eine effiziente Suche in Metadaten umgesetzt werden kann. Der „Status quo“ der Geodatenuche und der gegenwärtigen Architektur des Geodatenkatalog.de soll hierbei als „Grundgerüst“ für die vorgeschlagenen Maßnahmen dienen. Alle Vorschläge stellen dahingehend eine Ergänzung des aktuellen Standes dar, mit dem Ziel, eine effektive Suchfunktion am Beispiel des Geodatenkatalog.de zu konzipieren.

5.2. Optimierung der Datenqualität

Zur Optimierung der Datenqualität des Geodatenkatalog.de, im Sinne einer oder mehrerer der genannten Anforderungen, sind folgende Maßnahmen zu empfehlen:

Maßnahme	Begründung*	Umsetzung
Metadaten-Editor	[1a] [1b] Entscheidend für eine gute Suchfunktion in Metadaten ist die Qualität der Metadaten (Rashid-Pour Anhang, Heimbürger Anhang). Die Ergebnisse der Analyse des Ist-Zustands der Metadaten offenbaren dabei substantielle Mängel (vgl. Kap.4.4.). Um die Anforderungen an die Metadatensuche im Geodatenkatalog.de umzusetzen, müssen die Metadaten in korrekter Form bereitgestellt werden.	Den geodatenhaltenden Stellen sollte ein (kostenloses) Werkzeug zur benutzerfreundlichen, standardkonformen und einheitlichen Erfassung ihrer Metadaten für Geodatenätze und -dienste zur Verfügung stehen.
Metadaten-Editor - Vorgaben zu Titel- Kurzbeschreibungs- und Schlüsselwortelementen	[1b] zahlreiche Titelbezeichnungen, Kurzbeschreibungen und Schlüsselwörter in den Metadaten des Geodatenkatalog.de sind unpräzise. Sie enthalten oftmals keine eindeutigen bzw. zweckmäßigen Informationen zum Inhalt des Geodatensatzes (vgl. Kap. 4.4.).	In den Metadaten-Editor werden Vorgaben für diese Elemente integriert und Beispiele dazu angegeben. Hinweis im Metadaten-Editor, dass diese Elemente <u>keine</u> : Abkürzungen, Fachausdrücke, Unterstriche, Maßstabsangaben, Angaben über INSPIRE-Konformität, Fremdsprachen und Fremdwörter enthalten sollten. Die betreffenden Metadatenelemente sollen in Feldern erfasst und dort mithilfe einer Dropdown-Liste zur Erhöhung der Benutzerfreundlichkeit ausgewählt werden können.

* Kennzahl der von der Maßnahme betroffenen Anforderung(en) des BKG in [eckigen Klammern]

Maßnahme	Begründung*	Umsetzung
Metadaten-Editor - Aufnahmehilfe der Bounding Box	<p>[1a] Inkorrekte Bounding Boxen stellen ein wesentliches Problem der Metadaten im Geodatenkatalog dar. Der Zugriff auf die hinter den Metadaten liegenden Geometrien zu Überprüfungszwecken ist jedoch oft nicht realisierbar oder nicht hilfreich (vgl. Kap. 4.4.). Deshalb sollte schon bei der Erfassung der Bounding Box auf ihre Richtigkeit geachtet werden.</p>	<p>Im Metadaten-Editor soll eine Ortssuche, ähnlich zu der Ortssuche im Geoportal.de, bei der Aufnahme der Bounding Boxen eingesetzt werden.</p> <p>Folgende Funktionen sollen umgesetzt werden:</p> <ul style="list-style-type: none"> a) automatische Erzeugung der Bounding Box bei der Eingabe eines Ortes oder Gebietes auf einer vorgegebenen Hintergrundkarte b) Visualisierung des betreffenden Kartenausschnittes bzw. der Bounding Box bei der Eingabe eines Ortes oder Gebiets c) Möglichkeit zur Verschiebung, Vergrößerung und Verkleinerung des Kartenausschnitts der Hintergrundkarte d) dynamische Änderung der Bounding Box bei einer Verschiebung, Vergrößerung oder Verkleinerung des Kartenausschnittes e) händisches Aufziehen und Bearbeiten der Bounding Box innerhalb der Hintergrundkarte mithilfe einer Editierfunktion
Metadaten-Editor – Validierung der Eingaben	<p>[1a] [1b] Viele Metadatensätze im Geodatenkatalog.de sind fehlerhaft; sowohl was die Konformität zu Standards, als auch die Größe/Lage der Bounding Box angeht (vgl. Kap.4.4.). Eine Überprüfung der Angaben im Metadaten-Editor hilft bei der Bereitstellung von qualitativ hochwertigen Metadatensätzen.</p>	<p>Die Validierung soll mithilfe folgender Funktionen im Metadaten-Editor realisiert werden:</p> <ul style="list-style-type: none"> a) Einbettung der der GDI-DE-Testsuite in den Metadaten-Editor b) Begrenzung der Zeichenzahl im Titel und den Schlüsselwörtern c) Begrenzung der zulässigen Bounding Box-Werte auf sinnvolle Werte d) Überprüfung, ob die standardrelevanten Felder befüllt sind. e) Validierung der jeweils erlaubten Feldinhalte

Maßnahme	Begründung*	Umsetzung
Meldefunktion im Geoportal.de	<p>[1a] [1b] Im Geoportal.de werden bislang nur die Ergebnisse aus der Suchanfrage angezeigt (vgl. Kap.3.2.). Eine Meldefunktion für fehlerhafte Angaben der Metadaten würde zu einer Qualitätsverbesserung der betreffenden Metadaten führen. Der Nutzer des Geoportal.de kann dadurch dem Anbieter der Metadaten eine Rückmeldung geben, worauf hin dieser die Möglichkeit bekommt, seine Metadaten zu verbessern.</p>	<p>Diese Meldefunktion soll hierzu in der Applikation Geodatenuche verwirklicht werden (vgl. Kap. 3.3).</p> <p>Unter Rückgriff auf die Kontaktinformationen des Betreibers des betroffenen Geodatenatzes bzw. Dienstes in den Metadaten (vgl. Kap 2.5.), soll eine E-Mail an den Betreiber versendet werden. Der Absender soll hierbei die Möglichkeit haben, anonym zu bleiben. Außerdem sollen gleiche Meldungen zu einer einzigen Meldung an den Betreiber/Anbieter der Geodaten bzw. des Dienstes zusammengefasst werden, um redundante Meldungen zu vermeiden.</p> <p>Der Meldevorgang stellt sich dabei wie folgt dar: Der Nutzer kann auf einen Button „Melden“ klicken, worauf hin sich ein Dialogfenster öffnet. Hierhin muss er eine Meldekategorie auswählen und kann in einem Textfenster Informationen hinzufügen. Nun kann er die Meldung mit einem „Senden“-Button verschicken.</p> <p>Mögliche Meldekategorien:</p> <ul style="list-style-type: none"> - Fehlerhafte Bounding Box - Unpräzise Angaben in den Metadaten - Fehlende Angaben in den Metadaten - Fehler bei der Darstellung der Geodaten (nur bei Darstellungsdiensten auswählbar) - Fehler bei weiteren Funktionen des Geoportal.de (Bestellen, Download, Drucken, XML, KML oder WMS) - Sonstiges

Maßnahme	Begründung*	Umsetzung
Workshops zur Metadaterfassung	[1a] [1b] Der Umgang mit der standardkonformen Metadaterfassung und -pflege ist in Deutschland mit erheblichen Problemen behaftet. Dies zeigt sowohl das jährliche INSPIRE-Monitoring, als auch die hier vorgenommene Qualitätsanalyse der Metadaten (vgl. Kap.2.3. und Kap. 4.4.). Zur Lösung dieser Defizite schlägt die European Environment Agency unter anderem die Durchführung von Workshops vor (EUROPEAN ENVIRONMENT AGENCY:5-6). Diese helfen den betroffenen Behörden die Bereitstellung von standardisierten, brauchbaren und aktuellen Metadaten sicherstellen.	Innerhalb der Architektur der GDI-DE (vgl. Kap.2.2.) erteilt das Lenkungsgremium GDI-DE der Koordinierungsstelle GDI-DE den Auftrag solche Workshops durchzuführen. Die Koordinierungsstelle wiederum organisiert in Zusammenarbeit mit den Kontaktstellen von Bund und Ländern, sowie den Partnern aus der Wirtschaft und Forschung Workshops zu diesem Thema. Ziel des Workshops ist: - das Erlernen des Erfassungsvorgangs von Metadaten mit einem geeigneten Metadateneditor - das Verständnis für die Relevanz von standardisierten Metadaten - die Optimierung der Metadaterfassung und -pflege Der Workshop richtet sich an alle Geodatenanbieter - insbesondere an von der INSPIRE-Richtlinie betroffenen Behörden.

* Kennzahl der von der Maßnahme betroffenen Anforderung(en) des BKG in [eckigen Klammern]

5.3. Optimierung des Suchvorgangs

Zur Verbesserung des Suchvorgangs des Geodatenkatalog.de, im Sinne einer oder mehrerer der genannten Anforderungen, werden folgende Maßnahmen vorgeschlagen:

Maßnahme	Begründung*	Umsetzung
Solr-Upgrade	[2] [3] [4] [5] Die aktuelle Solr-Version der Geodatenuche (vgl. Kap. 3.1.), kann die genannten Anforderungen nicht optimal erfüllen. Ein Upgrade beinhaltet neben neuen Funktionen und Bugfixes auch eine verbesserte Performance und höhere Sicherheit (Rashid-Pour Anhang) (Solr Features).	Upgrade der aktuellen Solr-Version (3.3.0) auf 5.3.1 und Anpassung der Applikation der Geodatenuche an die neue Solr-Version.

* Kennzahl der von der Maßnahme betroffenen Anforderung(en) des BKG in [eckigen Klammern]

Maßnahme	Begründung*	Umsetzung
Solr-Features - Highlighting in der Kurzbeschreibung	<p>[3] Bei der Ergebnisausgabe im Geoportal.de findet momentan kein „Highlighting“ statt (GEOPORTAL).</p> <p>Beim „Highlighting“ werden übereinstimmende Textteile der Suchanfrage bei der Suchantwort hervorgehoben ausgegeben (GUIDE SOLR:286).</p> <p>Dies hilft bei der Erfüllung der Anforderung [3] des BKG.</p> <p>Das Hervorheben der Suchergebnisse soll nur auf die Kurzbeschreibung angewendet werden, da sonst die Gefahr einer Unübersichtlichkeit bei der Darstellung der Suchergebnisse besteht.</p>	<p>Die Umsetzung geschieht durch die Implementierung der „Highlighting“-Funktionalität in die Geodatenuche:</p> <p>Hierzu wird beim Anlegen des Lucene-Indexes eine einzige Spalte nur für den Inhalt der Kurzbeschreibung angelegt, auf welche die „Highlighting“-Funktion angewendet werden kann. Bisher lagen die Textinhalte von Kurzbeschreibung, Alternativtitel und Schlüsselwörtern gemeinsam in einer Indexspalte vor (vgl. Kap. 3.4.).</p> <p>Die Konfiguration des „Highlighting“ wird in Solr durch die Festlegung von Highlighting-Parametern realisiert (GUIDE SOLR:287-289).</p>
Solr-Features - Synonyme einbinden	<p>[3] Die Geodatenuche verwendet gegenwärtig keine Synonyme bei der Indexierung der Metadaten (GEOPORTAL).</p> <p>Synonyme werden beim Suchprozess zusätzlich eingebunden, falls ein Wort(teil) mit ähnlicher oder gleicher Bedeutung gesucht wird. (z.B. Population bei Bevölkerung) (GUIDE SOLR: 122).</p> <p>Somit wird die Quantität der Suchergebnisse erhöht und die Nutzerfreundlichkeit verbessert.</p>	<p>Die Umsetzung geschieht durch die Implementierung von Synonymlisten in Solr. Zunächst muss die Synonymfunktion in der Konfigurationsdatei (schema.xml) in Solr definiert und konfiguriert werden, bevor sie bei der Suchanfrage eingesetzt werden kann.</p> <p>In diesem Fall bedeutet dies das Einpflegen oder Einbinden von Synonymlisten in die Synonymfunktion (z.B. Solr.SynonymFilterFactory) (GUIDE SOLR 122-123).</p>
Solr-Features - Verbesserung des „autosuggest“	<p>[3] [5] Eine gute Ergänzungsfunktion im Geoportal.de führt den Nutzer ziel-sicher zu den gewünschten Informationen. Dabei werden bisher keine vorher eingegebenen Begriffe im Suchfeld berücksichtigt, außerdem fließt die räumliche Einschränkung nicht in die Ergänzungsvorschläge ein (Kap. 3.5.). Dies soll hiermit angepasst werden.</p>	<p>Durch eine Anpassung des Suchstrings der Geodatenuche können vorher eingegebene, und durch Leerzeichen getrennte Wörter berücksichtigt werden und so als Einschränkung für die weiteren Vorschläge der Ergänzungsfunktion dienen.</p> <p>Dieser eingebaute Suchfilter kann ebenso auch in Verbindung mit der Bounding Box angewendet werden (Rashid-Pour Anhang).</p>

Maßnahme	Begründung*	Umsetzung
Solr-Features - Meinten Sie?...	<p>[3] [5] Im Augenblick fehlt der Geodaten-suche des Geodatenkatalog.de eine Rechtschreibprüfung bzw. Vorschlagsfunktion (z.B. Meinten Sie..?) für alternative Suchanfragen im Geoportal.de (GEOPORTAL).</p> <p>Mit der Vorschlagsfunktion werden dem Nutzer bei der Ergebnisausgabe der Suche Vorschläge zu ähnlichen Wörtern gemacht (SOLR GUIDE:295).</p> <p>Um die Nutzerfreundlichkeit zu verbessern, ist eine solche Funktion eine sinnvolle Ergänzung der bisherigen Suche.</p>	<p>Um eine Rechtschreibprüfung (Spell Checking) in die Geodaten-suche zu integrieren, können unter anderem Felder des Lucene-Indexes oder auch externe Textdateien verwendet werden. Hierzu muss die Datenquelle für die Vorschläge in der schema.xml festgelegt werden.</p> <p>Die Rechtschreibprüfungsfunktion muss dann in der XML-Konfigurationsdatei von Solr (solrconfig.xml) definiert und angepasst werden, je nachdem woher die Daten für die Rechtschreibprüfungsfunktion stammen und welche Parameter spezifiziert werden (SOLR GUIDE:295-303).</p>
Solr-Features - Overlap Sorting	<p>[2] [4] Zur Zeit arbeitet die Geodaten-suche bei einer räumlichen Suchanfrage nur mit den Lagebeziehungen contains (enthält), intersect (schneidet) und within (liegt innerhalb). Ohne eine weitere räumliche Differenzierung der Suchergebnisse treten Ungenauigkeiten auf. Damit wird die Entdeckung der gesuchten Geodaten-sätze und -dienste erschwert (vgl. Kap. 3.5.).</p> <p>Um räumliche Aspekte nachvollziehbar und zweckdienlich in die Geodaten-suche einfließen zu lassen, ist das Overlap Sorting eine geeignete Möglichkeit. Hierbei werden die Bounding Boxen jener Daten im Index höher gewichtet, welche prozentual eine höhere Übereinstimmung mit der angefragten Bounding Box besitzen (SOLR GUIDE: 347).</p> <p>Unter der Prämisse einer guten Metadatenqualität lassen sich so die gesuchten Informationen passgenau ermitteln.</p>	<p>Im Lucene-Index wird zunächst ein Feld angelegt, welches die Bounding Box-Angaben aus den Metadaten enthält. Die Werte in diesem Feld (BboxField) müssen beim Befüllen eine bestimmte Syntax (WKT/CQL Envelope) vorweisen.</p> <p>Auch hier müssen Anpassungen in Solr (schema.xml) durchgeführt werden.</p> <p>Bei einer Suchanfrage berechnet Solr dann nur die prozentuale Überschneidung (intersect) der angefragten Bounding Box mit allen indexierten Bounding Boxen. Die Sortierung wird hierbei durch den Einsatz des Parameters „score=overlapRatio“ umgesetzt (SOLR GUIDE: 347).</p> <p>Das Overlap Sorting stellt damit eine einfache Lösung für ein gewichtiges Problem der bisherigen Suche dar. Die Umsetzung ist zudem relativ unkompliziert im Vergleich zu der bisherigen räumlichen Suche (vgl. Kap. 3.5.). Voraussetzung für diese Funktion ist eine Solr-Version von mindestens 4.1.0. (SOLR CHANGELOG).</p>

Maßnahme	Begründung*	Umsetzung
Ergänzung - mehrere Bounding Boxen indexieren	<p>[2] [4] In den Lucene-Index wird bisher nur die erste Bounding Box aus den Metadaten übernommen (Kap. 3.5.). Es können jedoch mehrere Bounding Boxen in den Metadaten enthalten sein (Kap. 2.5.). Dies macht dann Sinn, wenn die Geodaten in räumlich voneinander getrennten Klumpen bzw. Clustern vorliegen (Rashid-Pour Anhang). Der Nutzer kann diese Geodaten momentan nur über eine einzige Bounding Box suchen. Aus diesem Grund sollten alle in den Metadaten angegebenen Bounding Boxen in Zukunft in den Index aufgenommen werden.</p>	<p>Die Geodatenuche muss so konfiguriert werden, dass sie alle Bounding Boxen zum Indexieren an Solr weitergibt. In Solr selbst muss beim betreffenden Bounding Box Feld (BboxField) darauf geachtet werden, dass bei den Einstellungen zum Feldtyp (Field Default Properties) der Parameter „multiValued“ auf „true“ gesetzt ist. Dadurch kann dieses Feld mehrere (Bounding Box-) Werte enthalten (Solr guide:46).</p>
Ergänzung - Kataloge in den Index einbauen + Filteroptionen einbauen	<p>[3] [4] Aufgrund der überschaubaren Zahl an Bezugsquellen für die Metadaten des Geodatenkatalog.de (INFOS GEODATENKATALOG) ist eine Filterfunktion im Geoportal.de von großem Nutzen. Auch für die Suche nach Daten von einem bestimmten Anbieter müssen die Herkunftskataloge und Organisationen in den Index aufgenommen werden - bisher ist dem nicht so (Kap. 3.4.).</p> <p>Mit diesen Maßnahmen können die Nutzer des Geoportal.de die Suchergebnisse zusätzlich eingrenzen und so gesuchte Informationen schneller finden, vorausgesetzt sie wissen um die Herkunft der gesuchten Daten.</p>	<p>Die hierzu benötigten Anpassungen sind abermals mit Änderungen der Geodatenuche und des Solr-Suchservers verbunden.</p> <p>Die Geodatenuche muss dazu zunächst die Metadaten aus der Tabelle „metadata“ mit den Daten aus der Tabelle „sources“ aus der Geonetwork-Datenbank koppeln. Dies geschieht über das source-Feld in der metadata-Tabelle und das uuid-Feld in der sources-Tabelle, die beide mit einer identischen UUID verknüpft sind. Nun kann man einem Metadatenatz jeweils einen „Herkunftskatalog“ zuordnen. Solr muss nun so konfiguriert werden, dass es nun auch das „neue“ Katalogfeld aus der bereinigten Datenbank indexiert.</p> <p>In der Solr-Konfiguration (schema.xml) wird ein neues Feld angelegt. Außerdem müssen die Angaben zur Herkunft des Datensatzes in einen der Datentöpfe für die Volltextsuche übernommen werden, damit sie später gefunden werden können (vgl. Kap.3.3.).</p>

Maßnahme	Begründung*	Umsetzung
Komponenten - Bewertung	<p>[1] [3] Die Facette „Bewertung“ des Geoportal.de kann bislang nicht Erfolg versprechend genutzt werden, denn es sind bisher nur 13 Geodatenätze oder -dienste des gesamten Geodatenkatalog bewertet worden (GEOPORTAL). Deshalb sollte die Bewertungsfunktion so verändert werden, dass mehr Nutzer des Geoportal.de die Geodaten bewerten können, ohne jedoch manipulierte Bewertungen zu einfach zu machen.</p> <p>Eine gute Bewertungsfunktion hilft damit sowohl dem Nutzer qualitativ hochwertige Daten zu finden und mangelhafte Daten auszusortieren, als auch dem Datenanbieter. Er kann damit die Qualität der Daten durch die Nutzer überprüfen lassen und wird durch eine hohe Bewertung für fehlerfreie Geodaten belohnt.</p>	<p>Zunächst einmal sollte die Bewertung auch ohne Anmeldung funktionieren. Hiermit erleichtert man dem Bürger die Mitwirkung an der GDI-DE im Sinne des Open Government (LUCKE:1).</p> <p>Sobald ein Nutzer des Geoportal.de eine Bewertung eines Datensatzes abgeben möchte, öffnet sich ein Dialogfenster. Hierin wird er aufgefordert in ein Pflichtfeld seine Mailadresse für eine Verifikationsmail anzugeben. Außerdem kann er in einem Textfenster optional den Datensatz bzw. Dienst kommentieren und seinen Namen angeben. Man bekommt dann eine Verifikationsmail zugeschickt, bei der man auf einen inbegriffenen Link klicken muss um die Bewertung zu bestätigen.</p> <p>So wird es möglich, ohne Anmeldung bewerten und kommentieren zu können, um die Zahl der bewerteten Geodatenätze und -dienste im Geodatenkatalog.de zu erhöhen und gleichzeitig den Missbrauch der Bewertungsfunktion zu erschweren.</p>
Komponenten - Log-Auswertung	<p>[1] [4] Was Solr im direkten Vergleich zum Suchserver „Elasticsearch“ lange Zeit fehlte, war die Möglichkeit, über Extensions <u>Log</u>-Dateien zu indexieren und daraufhin auszuwerten bzw. zu visualisieren (KLOSE, WRINGLEY 2014: 284).</p> <p>Mit einer Implementierung einer solchen Extension in den Geodatenkatalog.de kann man Fehler, Performance und Details zu Suchanfragen überwachen (KLOSE, WRINGLEY 2014:284).</p> <p>Damit erhält man Anhaltspunkte zur Verbesserung der Suche im Geodatenkatalog.de und ein wichtiges Werkzeug der Qualitätssicherung .</p>	<p>Es gibt inzwischen zahlreiche Möglichkeiten Logs in Solr zu indexieren und zu visualisieren (SEMATEXT). Eine Möglichkeit ist der Einsatz von <u>SiLK</u>. Mit der Installation des SiLK-Paketes werden mehrere vorkonfigurierte Komponenten mitgeliefert. Folgende Funktionen sind damit verbunden:</p> <ul style="list-style-type: none"> - Logstash sammelt, modifiziert und indexiert die Log-Dateien von Solr - Banana ist eine Java-Webapplikation, welche die Log-Dateien auswertet und visualisiert. (SiLK) - Voraussetzung hierfür ist mindestens eine Solr-Version von 4.4.0 (SiLK).

Maßnahme	Begründung*	Umsetzung
Optimierung - Solr	<p>[3] [5] Solr besitzt von sich aus etliche Einstellungen, welche die Indexierung und Suche beeinflussen.</p> <p>Hier ist ein Optimierungspotenzial vorhanden, welches mit den Standardeinstellungen nicht ausgeschöpft wird. Eine Anpassung der entsprechenden Parameter kann Indexierung und den Suchvorgang schneller und zielgerichteter machen (KLOSE, WRINGLEY:135-137) (Rashid-Pour Anhang).</p> <p>Eine erweiterte Log-Auswertung kann zu einer optimalen Anpassung der betreffenden Parameter beitragen und die Effekte bei einer Änderung der Einstellungen nachvollziehbar feststellen.</p>	<p>Die Anpassung der diesbezüglichen Einstellungen findet in Solr in der allgemeinen Konfigurationsdatei (solrconfig.xml) statt. Hier können zum Beispiel in den Elementen admin, indexConfig, Query, requestParsers und httpCaching Änderungen durchgeführt werden (KLOSE, WRINGLEY 2014:135-137).</p> <p>Vorgeschlagene Anpassungen:</p> <ul style="list-style-type: none"> - maximale Verfügbare Speichergröße beim Indexierungsprozess erhöhen - zielgerichtetes <u>Caching</u> durch „enableLazyFieldLoading“, Anpassungen des „filterCache“- und „queryResultMaxDocsCached“-Parameters
Optimierung - Festplatte	<p>[3] [5] Eine naheliegende Möglichkeit die Geschwindigkeit des Such- und Indexierungsprozesses des Geodatenkatalog.de zu verbessern, stellt der Einsatz von <u>SSD</u>-Festplatten für die betreffenden Server dar (Rashid-Pour Anhang).</p> <p>Zur Zeit sind ausschließlich <u>HDD</u>-Festplatten im Einsatz (Heimbürger Anhang), welche eine weitaus geringere Lese- und Schreibgeschwindigkeit und höhere Zugriffszeiten besitzen.</p> <p>SSDs besitzen zudem weitere Vorteile: höhere Zuverlässigkeit, lautloser Betrieb, geringerer Energieverbrauch und größere Sicherheit. Sie kosten jedoch mehr, haben eine limitierte Schreibleistung und verfügen über eine geringere maximale Speicherkapazität (SSD-TEST).</p>	<p>Die Umsetzung findet durch den sukzessiven Einbau geeigneter SSD-Festplatten in die Serversysteme der Komponenten des Geodatenkatalog.de und des Geoportal.de statt.</p>

* Kennzahl der von der Maßnahme betroffenen Anforderung(en) des BKG in [eckigen Klammern]

6. Fazit

Für den Erfolg einer Suche in Metadaten ist sowohl die Datenqualität, als auch die eigentliche Suchfunktion von maßgeblicher Bedeutung

Im konkreten Fall der Suche des Geodatenkatalog.de wurden hierbei Defizite in beiden Bereichen identifiziert (vgl. Kapitel 3 und 4), wobei die Metadatenqualität im Wesentlichen von den geodatenhaltenden Behörden in ganz Deutschland abhängt.

Natürlich funktioniert eine effiziente Suche im Geodatenkatalog.de auch ohne Kenntnis von INSPIRE-Richtlinien und -Bestimmungen. Dennoch ist diese Suche eingebettet in das Thema INSPIRE und dadurch erst gewinnt diese Suche an besonderer Bedeutung. Außerdem entstand die Suche in den Metadaten des Geodatenkatalog.de erst im Rahmen der INSPIRE-Richtlinie. (VERWALTUNGSVEREINBARUNG GDI:4, Anlage 2:1).

Im Zusammenhang mit der Umsetzung von INSPIRE in Deutschland muss und wird in Zukunft noch viel getan werden. Vor allem die technische Durchführung der Richtlinie ist und bleibt für Behörden und insbesondere Kommunen eine große Herausforderung (KOORDINIERUNGSSTELLE GDI-DE²:17). Mit Blick auf den INSPIRE-Zeitplan (INSPIRE-ROADMAP) spielen zudem Kosten, Zugangsprobleme und Koordinierungsschwierigkeiten eine wichtige Rolle (KOORDINIERUNGSSTELLE GDI-DE²:23, 16, 26).

Ob das Ziel, eine europaweite, interoperable Verwendung von Geodaten zu ermöglichen, die beschlossenen Maßnahmen, trotz der genannten Probleme rechtfertigt, bleibt abzuwarten. Denn erst wenn alle geforderten Maßnahmen umgesetzt sind, können quantifizierbare und fundierte Aussagen hierzu getroffen werden (vgl. KOORDINIERUNGSSTELLE GDI-DE²:25). Einen Beitrag dazu leistet für Deutschland die GDI-DE. Gestützt durch das Geodatenzugangsgesetz und die Gesetze der Bundesländer zur Umsetzung der INSPIRE-Richtlinie arbeiten hier Bund, Länder und Kommunen an der Realisierung dieses Ziels (AK METADATEN:12).

Innerhalb dieses Prozesses kommt dem Geodatenkatalog.de eine besondere Bedeutung zu, denn er ermöglicht es, über die dort bereitgestellten Metadaten, Geodaten und Dienste zentral zu recherchieren (INFOS GEODATENKATALOG.DE). Der Zugang zu den Informationen wird durch das Geoportal.de umgesetzt. Hier sollen über das Internet idealerweise alle behördlichen Geodatenätze- und Dienste Deutschlands für jedermann zugänglich gemacht werden.

Der Weg zu den Geodaten führt jeden Nutzer des Geoportal.de über die Applikation der Geodatenuche, welche in den Metadatenätzen des Geodatenkatalog.de recherchiert. Um die Anforderungen an eine effiziente Suche zu erfüllen (vgl. Heimbürger Anhang), sollte diese bestmöglich umgesetzt werden.

Die vorliegende Arbeit identifiziert in dieser Beziehung vielfältige Entwicklungsmöglichkeiten und schlägt dementsprechende praktische Maßnahmen vor.

Von erhöhter Wichtigkeit sind hierfür meines Erachtens nachfolgende Maßnahmen (siehe Kap.5.3.):

- Solr-Upgrade auf die aktuelle Version 5.3.1. um den vollen Funktionsumfang für eine Suchanwendung nutzen zu können und die Geschwindigkeit und Zuverlässigkeit der Suche zu verbessern

- Einbau des Overlap Sortings in die räumliche Komponente der Metadatenuche zur effektiven und effizienten Weiterentwicklung dieser.
- Auswertung der Logging-Ergebnisse von Solr um auch in Zukunft Fehler, Schwachstellen und Anhaltspunkte zur Verbesserung der Suche des Geodatenkatalog.de zu erhalten

Für die Qualitätssteigerung der Metadaten werden insbesondere folgende Maßnahmen empfohlen (siehe Kap 5.2.):

- Bereitstellung eines adäquaten Metadaten-Editors für alle geodatenhaltenden Behörden zur Verbesserung der Metadatenqualität
- Durchführung von Workshops für geodatenhaltende Behörden zum Thema Metadatenerfassung, um eine Qualitätssteigerung der Metadaten und eine Optimierung des Erfassungsvorgangs zu ermöglichen
- Integration einer Meldefunktion im Geoportal.de zur Realisierung einer Rückmelde-möglichkeit für den Nutzer des Geoportal.de

Das Ziel eine effiziente Suche im Geodatenkatalog.de zur Verfügung zu stellen, wird durch die Arbeit unterstützt und konkretisiert. Es bleibt zu hoffen, dass mit der Umsetzung, der in dieser Arbeit vorgestellten Maßnahmen und der hier gewonnenen Erkenntnisse, ein Beitrag zum Aufbau der Geodateninfrastruktur in Deutschland geleistet werden kann.

7. Quellen:

7.1. Gesetze und Richtlinien:

GEODATENZUGANGSGESETZ: Bundestag (Hrsg.) (2009): Gesetz über den Zugang zu digitalen Geodaten (Geodatenzugangsgesetz – GeoZG). BGBl. I S.278. Stand: 07.11.2012.

Abrufbar unter: <http://www.gesetze-im-internet.de/bundesrecht/geozg/gesamt.pdf> (letzter Aufruf: 10.10.2015)

INSPIRE: Europäisches Parlament und der Rat der europäischen Union (Hrsg.) (2007): RICHTLINIE 2007/2/EG DES EUROPÄISCHEN PARLAMENTS UND DES RATES vom 14. März 2007 zur Schaffung einer Geodateninfrastruktur in der Europäischen Gemeinschaft (INSPIRE). Straßburg.

Abrufbar unter: <http://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:32007L0002&from=EN> (letzter Aufruf: 10.10.2015)

INSPIRE INTEROPERABILITÄT: Die Kommission der Europäischen Gemeinschaften (Hrsg.) (2010): VERORDNUNG (EG) Nr. 1089/2010 DER KOMMISSION vom 23. November 2010 zur Durchführung der Richtlinie 2007/2/EG des Europäischen Parlaments und des Rates hinsichtlich der Interoperabilität von Geodatenätzen und -diensten. Brüssel.

Abrufbar unter: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:323:0011:0102:DE:PDF> (letzter Aufruf: 10.10.2015)

INSPIRE METADATEN: Die Kommission der Europäischen Gemeinschaften (Hrsg.) (2008): VERORDNUNG (EG) Nr. 1205/2008 DER KOMMISSION vom 3. Dezember 2008 zur Durchführung der Richtlinie 2007/2/EG des Europäischen Parlaments und des Rates hinsichtlich Metadaten. Brüssel.

Abrufbar unter: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:326:0012:0030:DE:PDF> (letzter Aufruf: 10.10.2015)

INSPIRE MONITORING: Die Kommission der Europäischen Gemeinschaften (Hrsg.) (2009): 2009/442/EG: Entscheidung der Kommission vom 5. Juni 2009 zur Durchführung der Richtlinie 2007/2/EG des Europäischen Parlaments und des Rates hinsichtlich Überwachung und Berichterstattung. Brüssel.

Abrufbar unter: <http://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:32009D0442&from=DE> (letzter Aufruf: 10.10.2015)

INSPIRE ROADMAP: Europäisches Parlament und der Rat der europäischen Union (Hrsg.) (2007): RICHTLINIE 2007/2/EG DES EUROPÄISCHEN PARLAMENTS UND DES RATES vom 14. März 2007 zur Schaffung einer Geodateninfrastruktur in der Europäischen Gemeinschaft (INSPIRE). Straßburg. Stand: Verordnung (EG) Nr. 1312/2014. Abrufbar unter: <http://inspire.ec.europa.eu/index.cfm/pageid/44> (letzter Aufruf: 10.10.2015)

VERWALTUNGSVEREINBARUNG (VVE) GDI: Vereinbarung zwischen dem Bund und den Ländern zum gemeinsamen Aufbau und Betrieb der Geodateninfrastruktur Deutschland (Hrsg.) (2013). Abrufbar unter: http://www.geoportal.de/SharedDocs/Downloads/DE/GDI-DE/Verwaltungsvereinbarung_2013.pdf?__blob=publicationFile (letzter Aufruf: 10.10.2015)

7.2. Literatur:

AK METADATEN (Hrsg.) (2015): Architektur der Geodateninfrastruktur Deutschland – Konventionen zu Metadaten. Abrufbar unter: http://www.geoportal.de/SharedDocs/Downloads/DE/GDI-DE/Architektur_GDI_DE-Konventionen_%20Metadaten.pdf?__blob=publicationFile (letzter Aufruf: 10.10.2015)

BAHRENBERG, G., GIESE, E. (1975): Statistische Methoden und ihre Anwendung in der Geographie. (B.G. Teubner) Stuttgart.

DEUTSCHER LANDKREISTAG BERLIN (Hrsg.) (2009): Geodaten sinnvoll nutzen. In: Band 81 der Veröffentlichungen des Vereins für Geschichte der Deutschen Landkreise e.V. Berlin. Abrufbar unter: http://www.kreise.de/_gdi/dok/bd-81.pdf (letzter Aufruf: 10.10.2015)

EUROPEAN ENVIRONMENT AGENCY (EEA) (2014): Mid-term evaluation report on INSPIRE implementation. Luxemburg. S.5-6
Abrufbar unter: http://www.eea.europa.eu/publications/midterm-evaluation-report-on-inspire-implementation/at_download/file (letzter Aufruf: 10.10.2015)

HÜBEN, M. (2011): Harvesting und Recherche im Geodatenkatalog-DE. In: STOBL, J., BLASCHKE, T., GRIESEBNER, G.: Angewandte Geoinformatik 2011. Berlin/Offenbach. S.227-232.

KLOSE, M., WRIGLEY, D. (2014): Einführung in Apache Solr. Praxiseinstieg in die innovative Suchtechnologie. (O'Reilly Verlag GmbH & Co. KG) Köln.

- KOORDINIERUNGSSTELLE GDI-DE¹ (Hrsg.) (2011) auf Basis des Leitfadens „Qualitativ hochwertige Metadaten pflegen und verarbeiten“ der Koordinierungsstelle GDI-NI beim Landesamt für Geoinformation und Landentwicklung (LGLN): Qualitativ hochwertige Metadaten pflegen und verarbeiten (Entwurf). Hinweise für Metadatenerfasser und Katalogbetreiber. Abrufbar unter: http://www.geoportal.de/SharedDocs/Downloads/DE/GDI-DE/Qualitativ%20hochwertige%20Metadaten%20pflegen%20und%20verarbeiten%20-%20Entwurfsdokument.pdf?__blob=publicationFile (letzter Aufruf: 10.10.2015)
- KOORDINIERUNGSSTELLE GDI-DE² (Hrsg.) im Auftrag des LG GDI-DE (2013): Bericht Mitgliedstaat: Deutschland, 2013. Abrufbar unter: http://cdr.eionet.europa.eu/de/eu/inspire/reporting/envuztk9g/INSPIRE_Bericht_Mitgliedstaat_DE_2013.pdf (letzter Aufruf: 10.10.2015)
- LAKES, T. (2011): Geodaten in Deutschland: eine erste Bestandsaufnahme von Geodaten und ihren Zugangsmöglichkeiten. Berlin.
Abrufbar unter: http://www.ratswd.de/Geodaten/downloads/Kurzbericht_Lakes.pdf (letzter Aufruf: 10.10.2015)
- LENKUNGSGREMIUM GDI-DE (Hrsg.) (2007): Memorandum of Understanding zwischen AK Metadaten und Lenkungsgremium GDI-DE. Abrufbar unter: http://www.geoportal.de/SharedDocs/Downloads/DE/GDI-DE/Memorandum%20of%20Understanding.pdf;jsessionid=48C49D10320BBE02639AA16744D6D1A8?__blob=publicationFile (letzter Aufruf: 10.10.2015)
- OBE, R., HSU, L. (2012): PostgreSQL: Up and Running. A Practical Guide to the Advanced open Source Database. (O'Reilly Media Inc.) Sebastopol.
- VON LUCKE, B. (2012): Open Government, das offene Regierungs- und Verwaltungshandeln und die Open Government Partnership. In: BBE-Newsletter 2013/23. Friedlichshafen.
Abrufbar unter: http://www.b-b-e.de/fileadmin/inhalte/aktuelles/2013/11/NL23_Gastbeitrag_Lucke.pdf (letzter Aufruf: 10.10.2015)
- WALTHER, J. (2011): Geodatenkatalog-DE – Architektur und Inhalte. In: STOBL, J., BLASCHKE, T., GRIESEBNER, G.: Angewandte Geoinformatik 2011. Berlin/Offenbach. S.271-274.

7.3. Internetquellen:

BVA: Der GSB im Überblick, Abrufbar unter:

http://www.bva.bund.de/DE/Organisation/Abteilungen/Abteilung_BIT/Leistungen/IT_Produnkte/GSB/Produktinfos/node.html;jsessionid=C77521BFBC26052112FAFCBDD91C181A.1_cid392 (letzter Aufruf: 10.10.2015)

CSW-FEATURES: Catalog Services for the Web (CSW) features. Abrufbar unter:

<http://docs.geoserver.org/latest/en/user/extensions/csw/features.html> (letzter Aufruf: 10.10.2015)

DATENGRUNDLAGE GEOPORTAL: Welche Datengrundlage nutzt die Ortssuche?. Abrufbar unter:

<http://www.geoportal.de/SharedDocs/FAQs/DE/Welche-Datengrundlage-nutzt-die-Ortssuche.html> (letzter Aufruf: 10.10.2015)

INFOS GEODATENKATALOG.DE: Geodatenkatalog.de. Abrufbar unter: <http://www.geoportal.de/DE/GDI-DE/Komponenten/Geodatenkatalog-DE/geodatenkatalog-de.html?lang=de> (letzter Aufruf: 10.10.2015)

ISO: About ISO. Abrufbar unter: <http://www.iso.org/iso/home/about.htm> (letzter Aufruf: 10.10.2015)

IT WISSEN: Relationale Datenbank. Abrufbar unter: <http://www.itwissen.info/definition/lexikon/Relationale-Datenbank-database-relational.html> (letzter Aufruf 27.07.2015)

GEOPORTALBUND: Das Geoportal des Bundes. Abrufbar unter:

https://www.bmi.bund.de/cae/servlet/contentblob/887742/publicationFile/56098/geoportal_bund.pdf (letzter Aufruf: 10.10.2015)

GEOPORTAL.DE: Bundesamt für Kartographie und Geodäsie: Suche. Abrufbar unter: <http://www.geoportal.de/DE/Geoportal/Suche/suche.html?lang=de> (letzter Aufruf: 10.10.2015)

GDI-DE: INSPIRE Durchführungsbestimmung Überwachung (Monitoring) & Berichterstattung (Reporting). Abrufbar unter: <http://www.geoportal.de/DE/GDI-DE/INSPIRE/Direktive/Monitoring-und-Reporting/monitoring-und-reporting.html?lang=de> (letzter Aufruf: 10.10.2015)

GDI-NI: Metadaten – integraler Bestandteil einer Geodateninfrastruktur. Abrufbar unter:

http://www.geodaten.niedersachsen.de/portal/live.php?navigation_id=8651&article_id=25492&_psmand=28 (letzter Aufruf: 10.10.2015).

GDI-TESTSUITE: Qualitätssicherung für Geodaten und Geodienste: GDI-DE Testsuite. Abrufbar unter: <https://wiki.gdi-de.org/display/test/GDI-DE+Testsuite#> (letzter Aufruf: 10.10.2015)

GUIDE SOLR: Apache Solr Reference Guide. Covering Apache Solr 5.3, Abrufbar unter: <http://mirror.arcor-online.net/www.apache.org/Lucene/Solr/ref-guide/apache-Solr-ref-guide-5.3.pdf> (letzter Aufruf: 10.10.2015)

LUCENE: Lucene vs Solr. Abrufbar unter: <http://www.Lucenetutorial.com/Lucene-vs-Solr.html> (letzter Aufruf: 10.10.2015)

LVERMGEO: Landesamt für Vermessung und Geoinformation Sachsen-Anhalt: Monitoring. Abrufbar unter: <https://www.lvermgeo.sachsen-anhalt.de/de/gdi-lsa/Informationen/Monitoring/Monitoring.htm> (letzter Aufruf: 10.10.2015)

MAPBENDER3: Was ist Mapbender3? Abrufbar unter: <http://www.mapbender3.org/?q=de/was-ist-mapbender3> (letzter Aufruf: 10.10.2015)

SEMATEXT: Sematext: Presentation: Solr for Indexing and Searching Logs. Abrufbar unter: <http://blog.sematext.com/2013/11/11/presentation-Solr-for-indexing-and-searching-logs/> (letzter Aufruf: 10.10.2015)

SiLK: SiLK Installation. Abrufbar unter: <https://docs.lucidworks.com/display/SiLK/SiLK+Installation> (letzter Aufruf: 27.07.2015)

SOLR FEATURES: Features. Abrufbar unter: <http://lucene.apache.org/Solr/features.html> (letzter Aufruf: 10.10.2015)

SOLR WEBSITES: Public Websites using Solr. Abrufbar unter: <https://wiki.apache.org/Solr/PublicServers> (letzter Aufruf: 10.10.2015)

SQL: What is SQL? Abrufbar unter: <http://www.sqlcourse.com/intro.html> (letzter Aufruf: 10.10.2015)

SSD-TEST: Warum SSD? Abrufbar unter: <http://www.ssd-test.de/warum-ssd-die-vorteile-und-nachteile.html> (letzter Aufruf: 10.10.2015)

QPS: Quality Positioning Service: World Geodetic System 1984 (WGS84), Abrufbar unter: <https://confluence.qps.nl/pages/viewpage.action?pageId=29855173> (letzter Aufruf: 10.10.2015)

8. Glossar

- Caching: Caching bezeichnet einen sehr schnellen Speichermechanismus, der entweder im Hauptarbeitsspeicher oder auf einem unabhängigen, hochperformantem Speichermedium stattfindet.
(<http://www.webopedia.com/TERM/C/cache.html>)
- CSS: CSS steht als Abkürzung für „Cascading Style Sheets“, ist eine Ergänzung zu HTML und kann auf Webseiten zur Gestaltung eingesetzt werden.
(<http://www.webopedia.com/TERM/C/CSS.html>)
- CSW: Mit dem „Catalog Service for the Web“ (CSW) können Metadaten über ein standardisiertes Protokoll gefunden, ausgetauscht und fortgeführt werden.
(<http://www.geoportal.de/DE/GDI-DE/Glossar/themenkatalog.html?lang=de&lv2=64870&lv3=65238>)
- Geonetwork: Geonetwork ist eine standardkonforme OpenSource-Katalogsoftware zum suchen, sammeln und bearbeiten von Metadaten.
(<http://geonetwork-opensource.org/>)
- GetCapabilities: GetCapabilities bezeichnet eine Anfrage an einen Dienst, worauf dieser Informationen zu dem Dienst im XML-Format zurückgibt.
(http://www.geoportal.rlp.de/portal/servicebereich/glossar.html?tx_lexicon%5Bletter%5D=71)
- Harvest: Bezeichnet einen Vorgang bei dem Verteilt liegende, verfügbare und relevante Daten eingesammelt werden. Metadaten sind meistens die einzigen Daten, die in einer GDI geharvestet werden.
<http://www.geoportal.de/DE/GDI-DE/Glossar/themenkatalog.html/lang=de&lv2=65130&lv3=65354>)?lang=de
- HDD: HDD ist die Abkürzung für „Hard disk drive“ und bezeichnet den Mechanismus mit welchem Lese- und Schreibvorgänge auf (magnetische) Festplatten durchgeführt werden.
(http://www.webopedia.com/TERM/H/hard_disk_drive.html)
- HTML: HTML („HyperText Markup Language“) bezeichnet eine Auszeichnungssprache um Webseitendokumente zu strukturieren und zu gestalten.
(<http://www.webopedia.com/TERM/H/HTML.html>)

- JavaScript: JavaScript ist eine Skriptsprache zum programmieren interaktiver Webseiten.
(<http://www.webopedia.com/TERM/J/JavaScript.html>)
- Json: Json bedeutet „JavaScriptObjektNotation“ und bezeichnet einen leichtgewichtigen Datenaustauschformat, der für Menschen wie auch Maschinen einfach zu lesen bzw. zu analysieren ist.
(<http://www.webopedia.com/TERM/J/JSON.html>)
- Layer: Layer ist ein Teil eines geographischen Datensatzes, der auf einer Datenebene dargestellt wird
(<http://www.spektrum.de/lexikon/geographie/layer/4662>)
- Log-Datei: Eine Datei, welche stattfindende Vorgänge und Aktionen auflistet bzw. protokolliert.
(http://www.webopedia.com/TERM/L/log_file.html)
- OpenSource: Ein Programm was unter der Open-Source-Lizenz veröffentlicht wurde, ist frei verfügbar, nutzbar und der Quellcode ist von jedem Nutzer angepasst werden. Jede Software, die mit dieser Lizenz arbeitet muss ebenfalls so behandelt werden.
(http://www.webopedia.com/TERM/O/open_source.html)
- Php: Serverseitige Programmiersprache für dynamische Webseiten
(<http://www.webopedia.com/TERM/P/PHP.html>)
- Script: Eine Liste aus Befehlen, welche ohne Nutzerinteraktion ausgeführt werden. Ein Skript ist meist mit einer einfachen Programmiersprache geschrieben.
(<http://www.webopedia.com/TERM/S/script.html>)
- SiLK: SiLK steht für „Solr integrated with Logstash and Kibana“ und ist ein maßgeschneidertes Paket aus den Komponenten Solr, Banana und dem Solr Writer for LogStash, welches unter anderem eine Solr Log-Auswertung erlaubt.
(<https://docs.lucidworks.com/display/SiLK/SiLK>)
- SQL: SQL ist die Abkürzung für „structured query language“. Es ist eine standardisierte Abfragesprache um Informationen aus einer Datenbank zu erhalten oder zu verändern.
(<http://www.webopedia.com/TERM/S/SQL.html>)
- SSD: SSD ist die Abkürzung für „solid state disk“. Eine SSD ist eine hochperformantes Speichermedium, welche keine bewegten Teile enthält.
(http://www.webopedia.com/TERM/S/solid_state_disk.html)

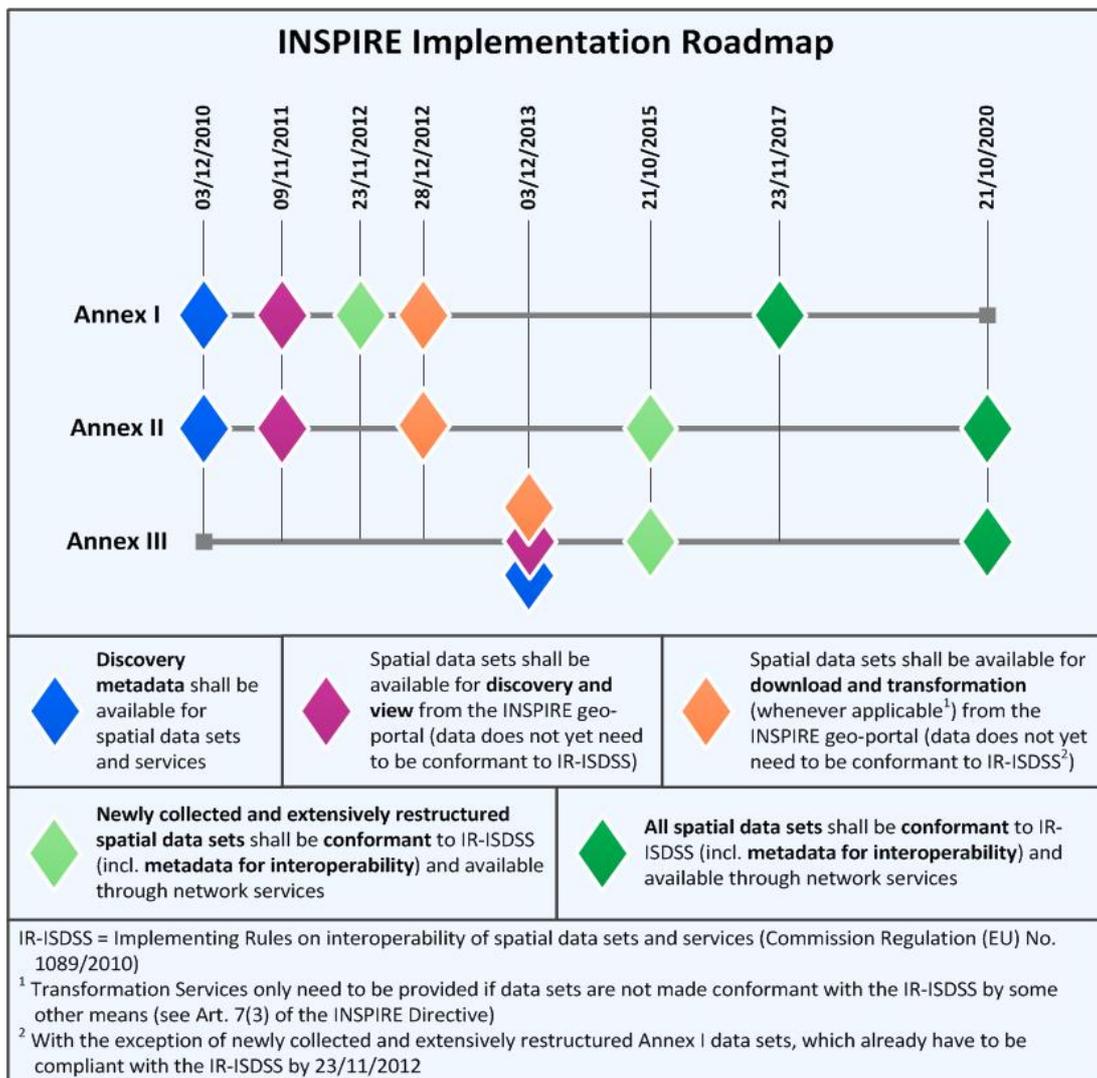
UTF-8: UTF-8 („Universal Transformation Format“) ist eine Methode um Unicode-Schriftzeichen in einen 8-bit-Format zu kodieren.
<http://www.webopedia.com/TERM/U/UTF.html>

XML: XML ist die Abkürzung für „Extensible Markup Language“. Es bezeichnet eine Spezifikation zum Aufbau von (Web-)Dokumenten und wird besonders für den Austausch von Daten über das Internet eingesetzt.
<http://www.webopedia.com/TERM/X/XML.html>

9. Anhang

(Anhang Zeitplan)

INSPIRE Roadmap:



(Anhang Script)

Script zur Extraktion der Metadaten XMLs aus der Geonetwork-Datenbank:

```
#!/bin/sh
for i in `seq 1 100`
do echo 'select data from metadata order by random() limit 1;' | psql -h localhost -U postgres -d
    postgres --quiet --no-align --no-readline --tuples-only -P footer=off > $i.xml
echo "Generiere $i";
done
```

(Anhang Testfehler)

Häufigste Fehler Testsuite:

INSPIRE-Test:

md_281: A conformity statement with a result of conformance must be given. (at /gmd:dataQualityInfo/*/gmd:report/*/gmd:result/*/gmd:pass) (this also applies for the ISO-Requirements)

md_282: A citation of the product specification or user requirement against which the data is being evaluated should be given. (at /gmd:dataQualityInfo/*/gmd:report/*/gmd:result/*/gmd:specification) (this also applies for the ISO-Requirements)

GDI-Test:

hierarchy level: The metadata document must give a hierarchy level which uniquely defines the type of the resource. The type must be one of dataset, series, service or application.

metadata identifier: The service-metadata document must reference a metadata document about the data it operates on. Furthermore, a metadata identifier must be given and must be identical to the identifier of the referenced data-metadata document.

ISO-Test:

Format Version: The distributor format must have a version or must be marked as missing, inapplicable, template, unknown or withheld using the gco:nilReason attribute. If a distribution is given, the distributor format is optional. (at gmd:distributionInfo/*/gmd:distributor/*/gmd:distributorFormat/*/gmd:version/gco:characterString)

Format Name: The distributor format must have a name or must be marked as missing, inapplicable, template, unknown or withheld using the gco:nilReason attribute. If a distribution is given, the distributor format is optional. (at gmd:distributionInfo/*/gmd:distributor/*/gmd:distributorFormat/*/gmd:name/gco:characterString)

(Anhang Analyse)

Ergebnis der Metadatenanalyse von 100 zufällig ausgewählten XML-Dateien der Geonetwork-Datenbank:

<u>INSPIRE-konform</u>	<u>inspireidentifiziert</u>	<u>inspireidentifiziert und INSPIRE-konform</u>
22	23	12
<u>ISO-konform</u>	<u>GDI-DE-konform</u>	
33	73	
<u>Nicht-Indexiert</u>	<u>Nicht-Indexiert weil</u>	<u>Anbieter nicht im Geoportal.de</u>
22	17	
<u>Fehler Bbox</u>		
21		
<u>Orthophoto Landesamt für Vermessung und Geoinformation Schleswig-Holstein</u>		
24		

(Heimbürger Anhang)

Informationen aus dem Email-Verkehr mit Dr. Olaf Heimbürger vom BKG:

Anforderungen an die Suche:

1. Grundlage für eine effiziente Suche im Geodatenkatalog.de ist eine gute Datenqualität der Metadaten insbesondere in Bezug auf:
 - a) Die Bounding Box-Angaben
 - b) eindeutige Angaben in Titel, Kurzbeschreibung und verwendeten Schlüsselwörtern (Keywords)
2. Bei einer räumlichen Eingrenzung der Suche sollen Ergebnisse mit einer möglichst großen Abdeckung innerhalb der verwendeten räumlichen Ausdehnung zuerst angezeigt werden.
3. Maßnahmen, welche es für den Nutzer ermöglichen textliche Suchprozesse schneller und komfortabler durchzuführen wie z.B. die Verwendung von Thesauri, Autokorrektur, Highlighting, zusätzlicher Datenbestände (PLZ, Ortsvorwahlen, geographische Namensbestände etc.), Rankingverfahren, sind zu begrüßen.
4. Bei einer guten Datenqualität (s. Punkt 1) soll sichergestellt werden, dass die gesuchten Geodaten bzw. Dienste möglichst als erstes (mindestens jedoch auf der ersten Seite) angezeigt werden.
5. Die räumliche Eingrenzung der Suche soll möglichst intuitiv und komfortabel für den Nutzer gestaltet werden.

Technische Details zu den Suchservern:

Ram: 8-12 GB

Mindestens 4 CPU (1 Core) 64bit

HD: 300GB

OS: Linux 64bit

Apache2.2/Tomcat7/Java1.7/PostgreSQL 9.x

Solr: 3.3.0.

- CSW als Suche lief nur im Geoportal.Bund

(Rashid-Pour Anhang)

Notizen aus zahlreichen persönlichen Gesprächen:

Aktuelle räumliche Suchfunktion:

- Bbox komplett drin > schneiden Bbox > Teil der Bbox / Außerhalb → weg
- Geosuche findet anhand der ersten eingetragenen BBOX statt. Metadaten können mehrere BBOXen beinhalten. Die Geodatenuche nimmt nur die erste.
- Die Geodatenuche nimmt zwar alle Bboxen aus der Geonetwork Datenbank, gibt jedoch nur die erste an Solr zum indexieren weiter
- zu große Bounding Boxen können zur Beeinflussung der Suche vorgekommen worden sein (öfter finden, da die Bboxen dann öfter in irgendeiner Lagebeziehung zu Such-Bounding Box stehen)
- within<intersect<contains → kein weiterer Rankingmechanismus (weder innerhalb der Lagebeziehung noch dazwischen, d.h. contains zum Beispiel immer vor intersect)
- Gewichtungsfaktor ist in der Administration der Geodatenuche einstellbar.

Aktuelle Volltextsuche:

- trunziert und intern von Solr gewichtet (text > text* > *text > text)
- Es gibt drei Datentöpfe (k1, k2 und k3) die bei der Suche unterschiedlich gewichtet werden.
k1 = title k2 = alternatetitle, abstract, keywords k3 = topicCategory, provider. Gewichtungsfaktor ist für jeden Datentopf in der Administration der Geodatenuche einstellbar.

Generelles:

- lucene= Volltextsuche in Java → daraus → Solr-Suchserver (REST-Interface) → nimmt Anfrage (POST+GET) auf und benutzt lucene
- CSW ist keine Suche sondern nur Ausgabemöglichkeit zur Übertragung von Metadaten (csw-Protokoll)
- CSW nicht performant und modern (kein autokomplete, Highlighting...), Filter nur auf definierte Felder möglich → Ranking, Filterfunktionen und gute Administration war wichtig
- deshalb php-Suchapplikation entwickelt worden

- Geoportal besteht aus GSB (Gouvernement Site Builder (für Behörden) = alles im Geoportal.de außer Karten und Suche) Mapbender3 (Kartendienste, Darstellung) und Geodatenuche (Ergebnisausgabe der Suche)
- Solr-Suchserver bekommt Daten → Index (ist eine große Tabelle in der die Daten eingefügt werden)
- Geodatenuche baut Suchstring auf → schickt Anfrage an Solr
- Suchstring jedes mal (bei jeder Suchanfrage) neu aufgebaut
- Json-Antwort als Ergebnisdokument von Solr geliefert → mithilfe eines templates → Styling der Suchergebnisse im Geoportal.de
- GDS ist Bindeglied von Geoportal und Geodatenkatalog
- Solr 5 hat viele topaktuelle Funktionen
- UUID taucht auch im Index auf
- Interne Verrechnung der Orts- und Volltextsuche bei der bisherigen Suche

Bereinigung der Geonetwork-Datenbank:

- Daten-dienste-kopplung: Layer-Ids in Metadaten und Layer-Ids + Layername in Dienste-GetCapabilities (über URL abfragen) → Layer-Name + URL in Index
- Zeichenkodierung UTF-8 (Transformation wenn nicht), HTML, Java-Script, CSS usw. aus Metadaten entfernen, Bbox korrigieren, wenn nicht Europa → wird entfernt.
- Die Bboxen müssen in 4326 (WGS-84) vorliegen. Wenn es etwas anderes sein sollte → ignoriert (nicht in Datenbank übernommen)
- Wenn Bboxen nicht im richtigen Wertebereich → nicht übernommen

-Verbesserungsvorschläge:

- Solr updaten (mehr Sicherheit, mehr Funktionen, mehr Bugfixes)
- Flächenberechnung der Bbox und Übereinstimmung mit angefragter Bbox bei Suchergebnisausgabe
- Kataloge (werden nicht gemerkt) → keine Sucheinschränkung möglich (nach Herkunft/Organisation)
- Autosuggest integrieren (Berücksichtigung der vorherigen Sucheinschränkung durch Anpassung des Suchstrings (Filters einbauen)
- „autosuggest“-Funktion soll verbessert werden (bisher wird bei Leerzeichen die vorherige Eingabe „vergessen“ → sollte als Filter genutzt werden)
- ISO/GDI-DE/INSPIRE-Konformität als Filteroption in der Sucheinschränkung
- Metadatensätze je nach Zeitstempel überprüfen und aktualisieren (nicht alle täglich)
- Solr optimieren/konfigurieren → Parameter an die Suche anpassen damit diese besser funktioniert
- viele Einstellungen haben großes Potenzial die Suche zu optimieren; muss angepasst werden
- eventuell sollten alle Bboxen in den Metadaten berücksichtigt werden
- SSDs einsetzen

Datenqualität:

- die Suche kann nur so gut sein wie die Daten
- inspireidentifiziert ist nur ein Keyword; heißt nicht inspirekonform

Extraktion:

- nur aus dem „data“-Feld in der „metadata“-Tabelle

Erklärung

Ich versichere, dass ich die Arbeit selbstständig verfasst habe, dass ich keine anderen Quellen und Hilfsmittel als die angegebenen benutzt und die Stellen der Arbeit, die anderen Werken dem Wortlaut oder dem Sinn nach entnommen sind, in jedem Fall als Entlehnung kenntlich gemacht habe. Das Gleiche gilt auch für beigegebene Zeichnungen, Kartenskizzen und Abbildungen.

Bonn, den (Datum)

(Unterschrift)