



Qualitativ hochwertige Metadaten pflegen und verarbeiten

Handlungsempfehlungen für geodatenhaltende
Stellen und Katalogbetreiber

Arbeitskreis Metadaten

13.09.2018

Version: 1.0

Autor	AK Metadaten
Herausgeber	Koordinierungsstelle GDI-DE
Haftungsausschluss	Diese Handlungsempfehlung wird als ein empfehlendes Papier veröffentlicht. Aussagen daraus können nicht im Rahmen gesetzlicher Verpflichtungen und deren Umsetzung geltend gemacht werden.

Änderungsverzeichnis

Version	Datum	Änderung	Ersteller
0.1 beta	08.09.2011	Testphase innerhalb der GDI-NI (Entwurf)	Kst. GDI-NI
0.1	08.11.2011	Fertigstellung zur Veröffentlichung im LG GDI-DE	AK Metadaten
	14.12.2011	Kennntnisnahme im LG GDI-DE (Beschluss Nr. 49)	Kst. GDI-DE
	20.12.2011	Veröffentlichung im Geodatenportal NI	Kst. GDI-NI
0.2	05.05.2017	Sprachliche Anpassungen (alle Kapitel)	Anja Loddenkemper
0.3	01.12.2018	Formatierung, Layoutbearbeitung	Anja Litka
0.4	02.02.2018	Inhaltliche und sprachlichen Korrekturen (alle Kapitel)	Anja Jacobi, Andreas Berg
0.51	27.04.2018	Inhaltliche Korrekturen, Ergänzungen und Anmerkungen (alle Kapitel)	Peter Kochmann
0.52	27.04.2018	Übernahmen aller inhaltlichen und sprachlichen Korrekturen, Ergänzungen (alle Kapitel)	Anja Loddenkemper
0.6	18.05.2018	Inhaltliche Korrekturen, Ergänzungen und Anmerkungen (alle Kapitel)	Marion Branzk, Anja Jacobi
0.7	23.08.2018	Festlegung Zielgruppe und Zielstellung	AK Metadaten
0.8	24.08.2018	Inhaltliche und sprachlichen Korrekturen (alle Kapitel)	Anja Loddenkemper
0.9	10.09.2018	Review	Andreas Berg
1.0 beta	13.09.2018	Schlussredaktion und Formatierung	Anja Litka
1.0	14.11.2018	Kennntnisnahme im LG GDI-DE	LG GDI-DE

Inhalt

1	Ziel und Zweck des Dokuments.....	4
2	Allgemeine Hinweise zum Umgang mit Metadaten	4
2.1	Metadaten als aussagekräftige Beschreibung der Geodatenressource	4
2.2	Datenbeschreibung - Suchdienst - Suchoberfläche.....	5
3	Erstellen von Datenbeschreibungen	6
3.1	Sprachliche Gestaltung.....	6
3.2	Inhaltliche Gestaltung.....	7
3.3	Überprüfen der angegebenen Kommunikationswege	9
4	Hinweise für Katalogbetreiber	9
4.1	Originäre Kataloge - Kataloge mit direkt erfassten Metadaten.....	10
4.1.1	Technische Änderungen an Metadaten	11
4.1.2	Inhaltliche Änderungen an Metadaten.....	12
4.1.3	Gewährleistung der rückverfolgbaren Qualität von Metadaten.....	12
4.2	Harvesting-Kataloge - Kataloge mit geharvesteten Metadaten.....	13
4.3	Gemischte Kataloge - Kataloge mit Metadaten gemischter Herkunft.....	14
5	Referenzen	15
A	Anhang – Begriffe im Zusammenhang mit Metadaten.....	16

1 Ziel und Zweck des Dokuments

Ein wesentlicher Bestandteil des Erfolges einer Geodateninfrastruktur (GDI) ist das Auffinden der in der GDI vorhandenen Geodatenressourcen. Die Suche basiert dabei auf den jeweils vorliegenden aktuellen Metadaten (Datenbeschreibungen). Sind die Metadaten nicht nur technisch und inhaltlich richtig, sondern zugleich auch verständlich, tragen sie direkt zu der erfolgreichen Suche nach Geodatenressourcen bei. Die Datenbeschreibungen sind damit qualitativ hochwertig.

Dieser Leitfaden dient dazu, qualitativ hochwertige Datenbeschreibungen zu erstellen sowie weiterzuverbreiten und richtet sich in erster Linie an Mitarbeiter/innen, die Kataloge (Suchdienste wie z. B. CSW) betreiben oder die Andere anleiten, Metadaten zu erfassen und inhaltlich zu befüllen. Nur durch das Zusammenspiel dieser beiden Gruppen „Katalogbetreiber“ und „Metadatenerfasser“ können qualitativ hochwertige Metadaten dazu beitragen, Geodaten im Rahmen von Geodateninfrastrukturen auffindbar zu machen und interoperabel bereitzustellen.

2 Allgemeine Hinweise zum Umgang mit Metadaten

Metadaten im Sinne dieses Dokumentes sind Datenbeschreibungen für

Geodatenressourcen. Dies beinhaltet die Beschreibung von Geodaten, Diensten wie WMS, WFS, WCS und ggf. auch kompletten Fachinformationssystemen. Ziel bei der Erstellung von Metadaten ist es, einem zukünftigen Nutzer so viele Informationen wie möglich zur Verfügung zu stellen. Durch die Beschreibung der Geodatenressource soll eine Beurteilung ermöglicht werden, ob die Geodaten zur Erledigung von (fachlichen) Aufgaben genutzt werden können.

2.1 Metadaten als aussagekräftige Beschreibung der Geodatenressource

Les- und interpretierbare Beschreibungen von Geodatenressourcen führen - neben technischen Voraussetzungen, die erfüllt sein müssen - zu interoperabel nutzbaren Geodaten innerhalb der GDI.

Im Wesentlichen wird durch die inhaltlich hochwertigen Datenbeschreibungen der Zeitaufwand für die Recherche nach Geodatenressourcen minimiert und die Einschätzung der Nutzbarkeit für ein Projekt gesteigert.

Je nach Art und Zugänglichkeit hat eine qualitativ hochwertige Beschreibung der vorhandenen Geodatenressourcen u. a. folgende Vorteile:

- Geodatenressourcen werden katalogisiert gelistet und mit Schlüsselwörtern versehen
- Geodatenressourcen können ad hoc gefunden werden
- die Entstehung der Geodatenressource ist sowohl intern als auch extern nachvollziehbar
- Testdaten stehen zur Verfügung
- Geodatenressourcen können in ihrem inhaltlichen Umfeld angeschaut werden
- alle Informationen zu den Daten sind innerhalb der geodatenhaltenden Stelle bekannt
- alle Ansprechpartner sind innerhalb der geodatenhaltenden Stelle bekannt
- eine Recherche nach Geodatenressourcen kann jederzeit und standortunabhängig erfolgen

Aufgrund der plattform- und implementationsunabhängigen Bereitstellung der Metadaten als Extensible Markup Language (XML) ist es möglich, die Datenbeschreibungen für Präsentationen auf Webseiten und/oder für Apps heranzuziehen. Außerdem können die Geodatenressourcen, die durch Metadaten beschrieben sind, direkt in internetbasierte Betrachtungskomponenten oder Portale übernommen werden. Grundsätzlich werden Metadaten zentral von der geodatenhaltenden Stelle gepflegt und sowohl zentral als auch dezentral für unterschiedliche Zwecke ausgelesen und genutzt.

2.2 Datenbeschreibung - Suchdienst - Suchoberfläche

Metadaten werden über sogenannte Suchdienste (CSW) veröffentlicht. Diese Suchdienste können standardisiert an- bzw. abgefragt werden. Üblicherweise wird eine standardisierte Anfrage über eine benutzerfreundlich gestaltete Webapplikation erstellt, die die Anfragen an den Suchdienst (technisch) verarbeitet und ausführt. Hierbei kann mit der Suchoberfläche sowohl auf nur einen Suchdienst z. B. eines einzelnen Anbieters oder aber auf mehrere Suchdienste unterschiedlicher Anbieter zugegriffen werden. Allgemein bekannte Suchoberflächen sind z. B.:

- Geoportal.de
- AdV-MIS
- GOVDATA

Suchdienste stehen, außer in begründeten Ausnahmefällen, kostenfrei und ohne Einschränkungen für den Zugriff öffentlich zur Verfügung. Dies bedeutet, dass die URL der CSW-Schnittstelle des Suchdienstes im Internet eingesehen und genutzt werden kann.

Die Suchoberflächen, mit deren Hilfe Nutzer Geodatenressourcen auffinden, orientieren sich an der jeweiligen Zielgruppe. Für die innerbetriebliche Nutzung kann eine eher komplexe Suchoberfläche angeboten werden. Dagegen kann für die Bürgerinformation eine gesonderte Suchoberfläche mit eventuell reduziertem Inhalt zur Verfügung gestellt werden. Beide Suchoberflächen werden vom gleichen Suchdienst bedient, präsentieren jedoch die Ergebnisse nach der Zielgruppe differenziert.

Zusätzlich zu der Recherche nach selbst geführten Geodaten, muss häufig auch ein Zugriff auf weitere Geodatenressourcen bestehen. Diese werden meist nicht originär in der Institution geführt, jedoch ebenfalls innerbetrieblich verwendet. Beispielsweise können dies Daten aus einem bestimmten Fachgebiet sein. Die Suchoberfläche wird daher nicht nur auf den eigenen Suchdienst zugreifen, sondern zusätzlich den Suchdienst ansprechen, der Informationen zu den Daten aus dem bestimmten Fachgebiet enthält.

Ein Suchportal, das „alle“ Metadaten in Deutschland beinhaltet, präsentiert Geodatenressourcen aus unterschiedlichsten Regionen und Themenbereichen. Dies muss bei der Datenbeschreibung der eigenen Geodatenressourcen beachtet werden. Ein Nutzer, der beispielsweise mit dem Begriff „Straße“ nach Geodaten sucht, findet nicht nur einen Datensatz u. a. mit Gemeindewegen, sondern auch Straßen einer Bundesbehörde, eines Landesamtes oder Gemeinde aus einem anderen Teil Deutschlands.

3 Erstellen von Datenbeschreibungen

Die Empfehlung für das inhaltliche Erstellen von aussagekräftigen Beschreibungen für Geodatenressourcen fokussiert sich auf die sprachliche und inhaltliche Gestaltung der Metadaten. Im Idealfall führt bereits der Prozess der Entstehung von Metadaten zu einer Verbesserung des Kommunikationsflusses und der Arbeitsabläufe in der geodatenhaltenden Stelle. Es gilt dabei stets der Grundsatz, ausschließlich **eigene** Geodatenressourcen zu beschreiben.

Der Hersteller von Geodatenressourcen ist der fachlich kompetente Ansprechpartner und ermöglicht durch verständliche Metadaten, den Aufwand für Nachfragen und Auskünfte auf ein Minimum zu reduzieren. Jede persönliche Nachfrage eines Suchenden sollte als Hinweis aufgefasst werden, dass die vorhandenen Datenbeschreibungen ggf. weiter zu verbessern sind. Aus diesem Grund muss jede Beschreibung einer Geodatenressource einem kontinuierlichen Review-Prozess unterliegen.

3.1 Sprachliche Gestaltung

Im Folgenden sind einige Hinweise zur sprachlichen Gestaltung von Metadaten aufgeführt, die einen unmittelbaren Einfluss auf eindeutige Suchergebnisse haben können.

Fließtext schreiben

In der Beschreibung darf ausschließlich reiner Text ohne Formatierung verwendet werden. Mit Hilfe von HTML gestaltete Überschriften oder Tabellen werden meist nur im eigenen System korrekt wiedergegeben. Von daher dürfen Elemente von Auszeichnungssprachen nicht in der Beschreibung verwendet werden.

Orthografisch korrekt schreiben

Die korrekte Beschreibung der Geodatenressourcen ist ein essenzieller Bestandteil, um die angebotenen Geodaten such- und auffindbar zu machen. Das Verwenden einer durchgängig richtigen Rechtschreibung, Grammatik und Interpunktion ist unerlässlich.

Abkürzungen vermeiden

Grundsätzlich sollten ausgeschriebene Begriffe verwendet werden. Eine Geodatenressource mit dem Titel „UKN_WA_NDS“ kann für den Fachexperten durchaus les- und interpretierbar sein. Nutzer außerhalb dieser Fachgruppe bevorzugen allgemein verständliche Titel wie „Übersichtskarte Waldflächen in Niedersachsen“. Die Abkürzungen können in Klammer gesetzt bzw. in den Schlüsselwörtern eingetragen werden, wie z. B.: „Geoinformationssystem Boden (BoGIS)“ oder „Geoinformationssystem Böden und Altlasten (BoGIS)“. Besonders häufig werden Abkürzungen für hausinterne Organisationseinheiten verwendet. Hierbei empfiehlt es sich immer, die Organisationseinheiten auszuschreiben und die Kurzbezeichnung in Klammern hinter dem vollständigen Begriff zu nennen.

Gängige Begriffe verwenden

Geodatenressourcen beschreiben zum Teil hoch komplexe Sachverhalte, die nur mit entsprechender Expertise fachlich korrekt und vollständig beschrieben werden können. So können die Ergebnisse einer Probebohrung mit „sondiert“ beschrieben werden. Andere Institutionen könnten den gesamten Vorgang jedoch mit „Beprobung“, „Analyse“ oder „Testergebnis“ beschreiben.

Das Hauptaugenmerk bei der Beschreibung liegt beim Verständnis für alle Laien und Fachnutzer. Deshalb sind sowohl Fachtermini als auch allgemein, gängige Begriffe für Jedermann zu verwenden.

Einer Systematik folgen

Metadaten sollten sich immer an der Handhabung der Daten innerhalb der eigenen Organisation orientieren. Beschreibungen verschiedener Geodatenressourcen, die in einem thematischen oder technischen Zusammenhang stehen, sollten diesen Bezug bereits im ähnlich aufgebauten Titel widerspiegeln - beispielsweise: „Übersichtskarte Wald im Landkreis“, „Übersichtskarte Boden im Landkreis“ bzw. „Übersichtskarte Gewässer im Landkreis“.

Datenbeschreibung „spezifisch“ machen

Bei der Beschreibung von einer Geodatenressource sollte immer auf die Unterscheidbarkeit zu gleichartigen Geodatenressourcen anderer Behörden geachtet werden. Bei Verwaltungen mit einem begrenzten räumlichen Wirkungskreis ist es wichtig, im Titel der Geodatenressource auf diese räumliche Begrenzung hinzuweisen. Möglicherweise ist ein zusätzlicher deutlicher Hinweis im ersten Satz der Beschreibung der Geodatenressource erforderlich und nützlich. Vorschäbilder tragen dazu bei, Metadaten verständlicher zu gestalten. INSPIRE empfiehlt, alle wichtigen Informationen bis hin zur rechtlichen Grundlage für die Daten bereits in den ersten 256 Zeichen der Datenbeschreibung darzustellen.



The screenshot shows a metadata entry for a geospatial dataset. On the left is a small thumbnail image of a magnetic field map. To the right of the image is the title 'Hubschrauber-Elektromagnetik (HEM) Gebiet 130 Borkum'. Below the title is the date '2009-12-16' and the source 'Bundesanstalt für Geowissenschaften und Rohstoffe (BGR)'. There are five stars and '(0) kommentieren' next to the source. The main description text reads: 'Die BGR führte im Projekt „Deutschlandweite Aerogeophysik-Befliegung zur Kartierung des nahen Untergrundes und seiner Oberfläche“ (D-AERO) flächenhafte Befliegungen an der deutschen Nordseeküste durch. Das Messgebiet der Ostfriesischen Insel Borkum (2008) ist Südwesten begrenzt durch die [...]'. At the bottom of the entry, there is a row of icons for 'Gebiet', 'Bestellen', 'Download', 'Drucken', 'XML', and 'KML', followed by a 'Mehr' button with a right-pointing arrow.

Abb. 1 - Beispiel „Datenbeschreibung“ mit Vorschäbilde

3.2 Inhaltliche Gestaltung

Bei der inhaltlichen Gestaltung von Metadaten sollte überlegt werden, was für den Suchenden eine hilfreiche Information ist. Beispielsweise kann die Möglichkeit genutzt werden, auf Testdatensätze zu verlinken oder einen Link auf ein Fachinformationssystem zu setzen, in dem die Geodaten eingebunden sind. So bleibt es dem Suchenden überlassen, ob er diese zusätzliche Information nutzt oder den persönlichen Kontakt per E-Mail oder Telefon vorzieht.

3.3 Überprüfen der angegebenen Kommunikationswege

Neben der umfassenden sprachlich-inhaltlichen Darstellung der Geodatenressource sollte das Augenmerk ebenfalls auf die Kontaktdaten gerichtet werden.

Um einen reibungslosen innerbetrieblichen Ablauf bei einer Anfrage zu gewährleisten, sollen folgende Kontaktdaten in den Metadaten verzeichnet sein:

- Kontakt für fachliche Fragen zur Geodatenressource
- Kontakt für den Vertrieb der Geodatenressource
- Kontakt für den Inhalt des Metadatensatzes

Bei den angegebenen Kontaktpersonen bzw. -ressorts soll Folgendes sichergestellt werden:

- Der Kontakt für die Geodatenressource kann verständlich über diese Geodaten Auskunft geben, ggf. ist eine Weiterleitung an den Vertrieb zur Bereitstellung der Daten möglich.
- Der Vertrieb kann Informationen zu den Kosten, dem Zugang und zur Nutzung der Geodatenressource geben.
- Der Kontakt für den Metadatensatz kann inhaltliche bzw. Schreibfehler im Metadatensatz berichtigen.
- Es ist empfehlenswert, einen Qualitätsstandard zu definieren, der die Antwortzeiten und ggf. -inhalte auf die Anfragen der Kunden festlegt. Durch ein festgeschriebenes Qualitätsmanagement vereinfacht und verbessern sich das Zusammenspiel aller Beteiligten in der Organisation, die Qualität der Produkte und die Außendarstellung der geodatenhaltenden Stelle.

4 Hinweise für Katalogbetreiber

Ein Katalogbetreiber sollte immer einen reibungslosen Betrieb des CSW sowie einen gleichbleibenden inhaltlichen und technischen Qualitätsstandard aller Metadaten im Katalog gewährleisten. Hinweise für einen Katalogbetreiber sind in den nachfolgenden Kapiteln aufgeführt.

Beschreibungen fremder Geodatenressourcen können mittels „Harvesting“ in regelmäßigen Abständen in den eigenen Katalog importiert werden. Die Praxis zeigt, dass die Grenzen zwischen Katalogen mit originären Metadaten (originäre Kataloge) und Katalogen mit geharvesteten Metadaten (gemischte Kataloge) fließend sind. Soll ein Katalog betrieben werden, der ausschließlich mit geharvesteten Metadaten verschiedener Institutionen (Harvesting-Katalog) befüllt ist, ergeben sich Anforderungen, die ebenfalls in diesem Kapitel kurz vorgestellt werden.

Grundsätzlich gilt, dass ein originärer Katalog immer über eine CSW-Schnittstelle verfügen muss, über die Daten abgegeben werden können. Harvesting-Kataloge werden zumeist für bestimmte Aufgaben zur Bündelung von Informationen betrieben und müssen CSW-Schnittstellen vor allem lesen können. Eine Bereitstellung von Metadaten per CSW ist aber für Harvesting-Kataloge nicht zwingend erforderlich.

Deutschland- und europaweit werden die technischen Metadatenprofile weiterentwickelt. Daher sollten Katalogbetreiber in Kontakt mit der jeweils zuständigen Stelle für GDI stehen, um sich über Neuerungen informieren zu können.

Für die Bereitstellung von Metadaten auf der europäischen Ebene ist es unerlässlich, die CSW-Schnittstelle der zuständigen Stelle für GDI bekannt zu geben. Nur somit ist sichergestellt, dass die Vernetzung der Katalogdienste der Länder über den zentralen Katalogdienst auf Bundesebene erfolgt.

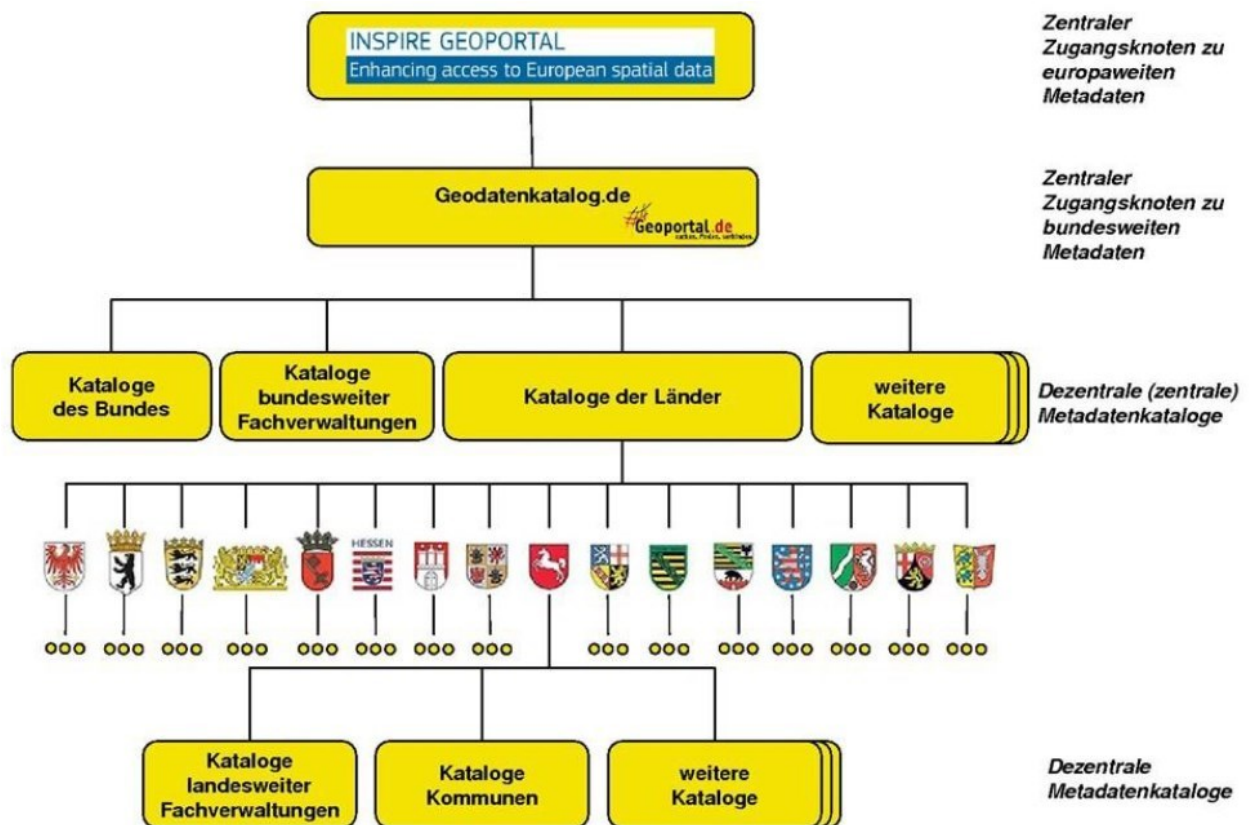


Abb. 3 - Topologie der Metadatenkataloge gemäß Architekturkonzept der GDI-DE

4.1 Originäre Kataloge - Kataloge mit direkt erfassten Metadaten

Falls in einem Katalog originäre Metadaten geführt werden, so ist dies die Originalapplikation. Es ist möglich, eine andere Software als Erfassungsoberfläche zu verwenden als jene, mit der die Metadaten als CSW abgeben werden.

Folgende Erfassungsmöglichkeiten im originären Katalog sind vorstellbar:

- Vollständige Ableitung der Metadaten aus Datenbanken, Geoinformationssystemen (GIS / WebGIS) oder bestehenden Archiven
- Teilweise Ableitung der Metadaten aus o.g. Systemen mit händischer Ergänzung der Metadaten
- Vollständige Erfassung der Metadaten per Hand
- Erfassung der Metadaten per Vorlage bzw. Script, Vorteil: einheitliche Datenbeschreibungen

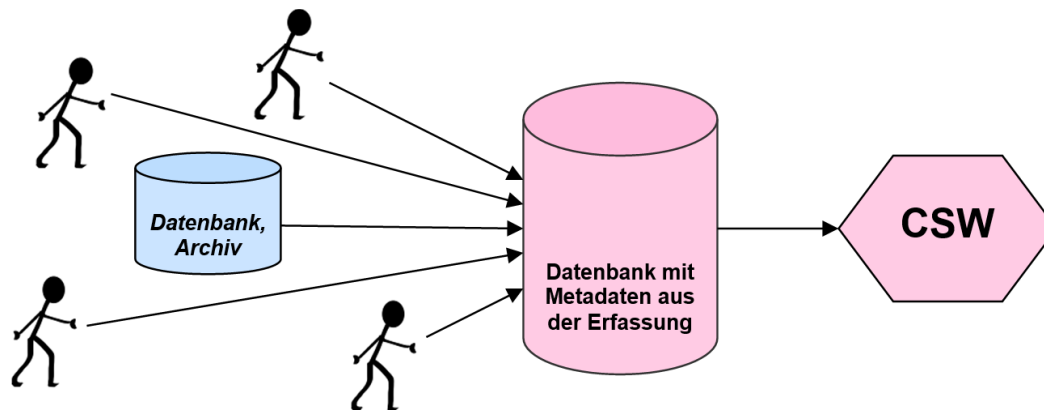


Abb. 4 - Katalog mit direkt erfassten oder erstellten Metadaten (originärer Katalog)

Der Katalogbetreiber gewährleistet gegenüber den Metadaten erfassenden Institutionen, dass die Metadaten den geltenden Vorgaben oder Profilen entsprechen. Bei Änderungen ist es empfehlenswert, die Ansprechpartner der Metadaten per E-Mail darüber zu informieren. Dazu kann beispielsweise zählen, nur noch bestimmte Schlüsselwörter zu verwenden, um Datensätze automatisiert für bestimmte Berichtspflichten zu melden. Ein Katalogbetreiber sollte grundsätzlich in regelmäßigem Kontakt mit den Erfassern und der zuständigen Stelle für GDI stehen. Probleme der Erfassung und des Harvesting mit den Stellen können somit direkt geklärt werden.

4.1.1 Technische Änderungen an Metadaten

Über die Katalogsoftware ist festgelegt, inwieweit technisch bedingte Änderungen am Metadatensatz dazu führen, dass dieser ein aktualisiertes Datum (dateStamp) erhält. Hierbei ist zu beachten, dass ein Metadatensatz nur dann von fremden Katalogen geharvestet wird, wenn das Datum aktualisiert wurde.

Im Folgenden ist ein Beispiel für eine technisch bedingte inhaltliche Änderung an den Metadatensätzen eines Kataloges aufgeführt:

Versehentlich wurde eine ISO codeList für einen angegebenen Wert beim Anlegen eines neuen Metadatensatzes durch die verwendete Software nicht gesetzt, sondern die entsprechende Angabe blieb leer:

...

```
<gmd:characterSet>
  <gmd:MD_CharacterSetCode codeList="" codeListValue="utf8"/>
</gmd:characterSet>
```

...

Über die gesamte Datenbank musste dieser Parameter nun mit dem zutreffenden Wert für die codeList versehen werden. Es wurde daraufhin ein SQL ausgeführt, welches in allen Metadatensätzen den Fehler berichtigte.

...

```
<gmd:characterSet>
  <gmd:MD_CharacterSetCode codeList="MD_CharacterSetCode"
    codeListValue="utf8"/>
</gmd:characterSet>
```

...

Der Fehler im Erfassungsprogramm wurde zwischenzeitlich behoben, so dass neu erstellte Datensätze den Fehler nicht mehr aufwiesen und die codeList wie vorgesehen befüllt wurde.

4.1.2 Inhaltliche Änderungen an Metadaten

Grundsätzlich dürfen durch den Katalogbetreiber keine inhaltlichen Änderungen an einem Metadatensatz erfolgen. Dieser stellt zwar das Erfassungssystem zur Verfügung, die inhaltlichen Rechte am Metadatensatz hält jedoch der Metadatenkontakt.

Insbesondere bei der ersten Erfassung von Metadaten durch einen Nutzer kann die direkte Unterstützung durch den Katalogbetreiber erforderlich oder sogar zwingend notwendig sein. Ein geeignetes Mittel, inhaltlich notwendige Änderungen als Katalogbetreiber durchzusetzen, ist es, den Metadatensatz vorübergehend aus dem Katalogangebot zu entfernen und dem Erfasser den Änderungswunsch aufzuzeigen.

4.1.3 Gewährleistung der rückverfolgbaren Qualität von Metadaten

Der Katalogbetreiber sollte Regeln erarbeiten, mit denen die Qualität und die Aktualität der Metadaten im eigenen Katalog gewährleistet werden. Dazu zählen:

- inhaltliche Prüfung der hinterlegten Metadaten
- Nutzung einer Erfassungsmaske, inklusive vorhandener Listenauswahl und Plausibilitätsprüfungen
- Festlegungen, dass der Erfasser die inhaltlich korrekte Beschreibung garantiert
- Festlegungen, dass der Katalogadministrator gewährleistet, dass die inhaltlich korrekten Metadatenelemente vorzufinden sind.

Der Katalogbetreiber sollte an einer geeigneten Stelle folgende Informationen bekanntgeben:

- Nutzungsbedingungen der CSW-Schnittstelle
- Intervalle für Wartungsarbeiten
- Zeiten des Harvestings

Im Rahmen des Harvesting durch Dritte muss häufig ein Unterscheidungsmerkmal zwischen den Metadaten des eigenen Kataloges und den Metadaten aus anderen Katalogen benannt werden. Hierfür gibt es im Metadatensatz den so genannten „Metadatenkontakt“. In diesem Kontakt kann eine URL auf das Erfassungssystem dieses Metadatensatzes hinterlegt werden. An dieser Stelle sollte in jedem Metadatensatz die URL „von diesem Metadatensatz selbst im Katalog der Herkunft“ angegeben werden. Es handelt sich hierbei um das Metadatenelement `MD_Metadata.contact.CI_ResponsibleParty.contactInfo.CI_Contact.onlineResource.CI_onlineResource.linkage`. Dieser Link könnte beispielsweise durch einen automatisierten Abgleich auf die Datenbank standardmäßig gesetzt werden.

Die CSW-Schnittstelle sollte sowohl in einem Capabilities-Dokument als auch mit einem Dienst-Metadatensatz beschrieben sein und zu folgenden Sachverhalten Auskunft geben:

- Name der führenden Institution
- Namen der erfassenden Institution(en)
- Nutzbare Qualitätskontrollen während bzw. beim Abschluss der Erfassung

4.2 Harvesting-Kataloge - Kataloge mit geharvesteten Metadaten

Ein Katalog mit geharvesteten Metadaten enthält ausschließlich Metadaten fremder CSW. Das Einrichten eines Harvesting-Kataloges ist immer dann sinnvoll, wenn Datenbeschreibungen aus sehr vielen unterschiedlichen Quellen zusammengetragen werden müssen, um z. B. eine Suchoberfläche für eine bestimmte Nutzergruppe bereitzustellen. Der Betreiber dieses Katalogs trägt die Verantwortung für die Bereitstellung der geharvesteten Metadaten in der Originalfassung.

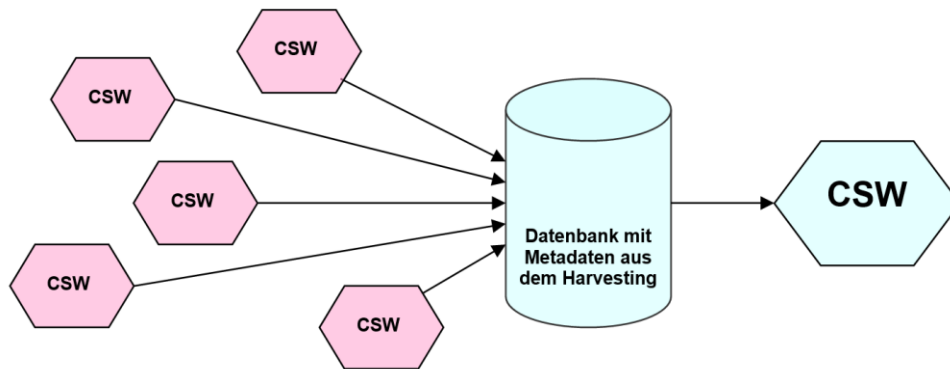


Abb. 5 - Katalog mit geharvesteten Metadaten (Harvesting-Katalog)

Die CSW-Schnittstelle sollte sowohl in einem Capabilities-Dokument als auch in einem Dienst-Metadatensatz beschrieben sein und zu folgenden Sachverhalten Auskunft geben:

- Name und URL für diesen Harvesting-Katalog
- Name der führenden Institution
- Namen und ggf. Links der geharvesteten Kataloge
- Zielgruppe(n) dieses Kataloges
- Turnus des Harvesting der angeschlossenen Kataloge

Beim Harvesting sollte folgende Reihenfolge eingehalten werden:

1. die Kataloge mit originären Metadaten
2. gemischte Kataloge (mit originären und geharvesteten Daten)
3. Kataloge die ausschließlich geharvestete Daten enthalten und nur, wenn diese Metadaten wirklich benötigt werden.

Durch die Beachtung dieser Hinweise werden Kataloge auf nationaler (z.B. bei der GDI-DE) und europäischer Ebene (z. B. INSPIRE) bestmöglich unterstützt.

Katalogbetreiber der geharvesteten CSW sollten über das beabsichtigte Harvesting der Daten informiert werden. Der Turnus (festgelegte/r Termin/e und Uhrzeit) des Harvestings wird zwischen den einzelnen Betreibern abgesprochen. Stichprobenhafte Kontrollen, sowohl nach inhaltlichen Gesichtspunkten der Metadaten als auch nach der Anzahl der geharvesteten Daten, sollten durch die Betreiber der Kataloge in regelmäßigen Abständen stattfinden.

Es ist zu empfehlen, nur geharvestete Metadaten, die den Qualitätsanforderungen genügen, mittels Filter in den eigenen Katalog aufzunehmen. Denkbar sind hier entweder Profil- bzw. inhaltliche

Filter oder das Anlegen einer Blacklist mit UUIDs der Metadatenatz-Identifikatoren von unerwünschten Metadaten. Sofern Metadatenätze eine mangelhafte Qualität aufweisen, sollte dies dem Katalogbetreiber des originären Kataloges mitgeteilt werden. Die UUIDs einzelner, nicht konformer Metadatenätze können über Logfiles (s. Abb. 6) für die Betreiber der geharvesteten Kataloge bereitgestellt werden.

```

1 2018-04-27 10:15:00,004 INFO [Niedersachsen] - Starting harvesting of Niedersachsen
2 2018-04-27 10:15:00,011 INFO [Niedersachsen] - Started harvesting from node : Niedersachsen (CswHarvester)
3 2018-04-27 10:15:00,011 INFO [Niedersachsen] - Retrieving capabilities file for : Niedersachsen
4 2018-04-27 10:15:00,211 INFO [Niedersachsen] - Re-trying the search with another HTTP method.
5 2018-04-27 10:24:54,259 INFO [Niedersachsen] - Records added to result list : 12824
6 2018-04-27 10:24:54,275 INFO [Niedersachsen] - Start of alignment for : Niedersachsen
7 2018-04-27 10:29:49,266 INFO [Niedersachsen] - Ignoring invalid metadata with uuid 0b3a024e-225b-47b7-a75a-3d26dcd0de1d
8 2018-04-27 10:30:22,077 INFO [Niedersachsen] - Ignoring invalid metadata with uuid 277a1af7-4d9f-49ba-8229-4b84ea537be0
9 2018-04-27 10:30:41,708 INFO [Niedersachsen] - Ignoring invalid metadata with uuid 3e9d5dea-0517-4caf-853a-7a5879f5e74d
10 2018-04-27 10:44:44,685 INFO [Niedersachsen] - Ignoring invalid metadata with uuid f859cf67-eaab-4dc9-8c2f-cd250ee09418
11 2018-04-27 10:44:58,416 INFO [Niedersachsen] - Ignoring invalid metadata with uuid d2696305-7dfc-42ee-aace-864c43d4ec98
12 2018-04-27 10:50:48,360 INFO [Niedersachsen] - Ignoring invalid metadata with uuid e4dd636f-0b30-4753-8913-f9624bebb447
13 ...
23 2018-04-27 11:06:44,719 INFO [Niedersachsen] - Ignoring invalid metadata with uuid 3d8e464c-bf37-442f-a126-3160a10fad41
24 2018-04-27 11:06:59,819 INFO [Niedersachsen] - End of alignment for : Niedersachsen
25 2018-04-27 11:06:59,820 INFO [Niedersachsen] - Ended harvesting from node : Niedersachsen (CswHarvester)
26 2018-04-27 11:07:00,270 INFO [Niedersachsen] - Harvester mail notification - start
27 2018-04-27 11:07:00,270 INFO [Niedersachsen] - Harvester mail notification - Retrieving mail addresses
28 2018-04-27 11:07:00,270 INFO [Niedersachsen] - Harvester mail notification - Retrieving subject
29 2018-04-27 11:07:00,270 INFO [Niedersachsen] - Harvester mail notification - Retrieving message
30 2018-04-27 11:07:00,270 INFO [Niedersachsen] - Harvester mail notification - retrieve harvest result
31 2018-04-27 11:07:00,270 INFO [Niedersachsen] - MHarvester mail notification - processing harvest result
32 2018-04-27 11:07:00,271 INFO [Niedersachsen] - Mail - mail prepared
33 2018-04-27 11:07:00,271 INFO [Niedersachsen] - Mail - to: [xxxx@bkg.bund.de, xxxx@bkg.bund.de]
34 2018-04-27 11:07:00,271 INFO [Niedersachsen] - Mail - subject: [csw] Niedersachsen finished harvesting
35 2018-04-27 11:07:00,271 INFO [Niedersachsen] - Mail - message: <p><b>Imported:</b>12824</p>
36 <p><b>Added:</b>13</p>
37 <p><b>Updated:</b>12780</p>
38 <p><b>Removed:</b>2</p>
39 <p><b>doesNotValidate:</b>17</p>
40
41
42 2018-04-27 11:07:00,271 INFO [Niedersachsen] - Harvester mail notification - end
43 2018-04-27 11:07:00,272 INFO [Niedersachsen] - Ended harvesting of Niedersachsen

```

Abb. 6 - Logfile mit dem Nachweis fehlerhafter Metadatenätze durch Angabe der UUID

4.3 Gemischte Kataloge - Kataloge mit Metadaten gemischter Herkunft

Gemischte Kataloge entstehen durch das Harvesting von bestehenden Katalogen, durch die direkte Datenerfassung, als auch durch Importprozesse von automatisiert erzeugten Metadaten oder Metadaten aus Datenbanken oder Dateisystemen.

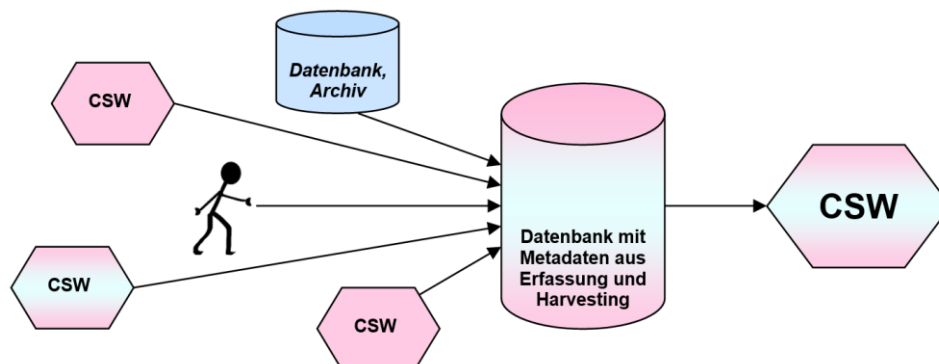


Abb. 7 - Katalog mit sowohl direkt erfassten als auch geharvesteten Metadaten (gemischter Katalog)

Kataloge mit Metadaten gemischter Herkunft dienen häufig dazu, bestimmte Systeme mit Datenbeschreibungen zu bestücken, beispielsweise ein übergreifendes Shop-System.

Gemischte Kataloge sollten eine öffentlich zur Verfügung stehende CSW-Schnittstelle besitzen, die direkt erfasste Metadaten und geharvestete Metadaten aus Fremdkatalogen separat anbietet. Diese Handhabung gestaltet sich nach dem derzeitigen technischen Entwicklungsstand in der Umsetzung schwierig. Es wird empfohlen, für die originär erfassten Metadaten eine separate CSW-Schnittstelle anzubieten.

5 Referenzen

Dieser Leitfaden wird auf Basis des Leitfadens „Qualitativ hochwertige Metadaten pflegen und verarbeiten“ (Stand 2011, 2014) der Koordinierungsstelle GDI-NI beim Landesamt für Geoinformation und Landesvermessung Niedersachsen (LGLN) veröffentlicht.

Architekturkonzept der Geodateninfrastruktur Deutschland, Konventionen zu Metadaten, Version 1.2.0 vom 18.04.2017.

A Anhang – Begriffe im Zusammenhang mit Metadaten

Geodatenressourcen

Geodatenressourcen sind alle Daten, die innerhalb einer Geodateninfrastruktur als Daten mit räumlichem Bezug (Geodaten) bezeichnet werden. Üblicherweise wird zwischen Geodaten (z. B. Shape- oder TIFF-Dateien), Diensten (z. B. WMS, WFS, CSW) und Fachinformationssystemen (z. B. Viewer für ein spezielles Fachthema) unterschieden. Es können aber auch Tabellen oder Listen mit geographischem Bezug Merkmale aufweisen, die diese für die Nutzung in einer Geodateninfrastruktur qualifizieren. Um alle mögliche Datenquellen unter einem Begriff subsumieren zu können, wird in diesem Dokument die Bezeichnung „Geodatenressourcen“ verwendet.

Weitere Begriffe im Glossar im Geoportal.de unter <http://www.geoportal.de/DE/GDI-DE/Glossar/A/Glossar.html>