

## Geo-Metadaten in OpenData Portalen

Armin Retterath



### Einleitung

In den letzten Jahren nahm die Verbreitung von OpenData Portalen auf Basis der Software CKAN immer mehr zu. Die von der OKF initiierte Software hat sich dabei zu einem De-facto-Standard entwickelt. Sowohl das OpenData Portal der Europäischen Union, als auch viele OpenData Portale der EU-Mitgliedstaaten und kleinerer Institutionen basieren auf dem in Python entwickelten Softwarepaket CKAN. Als RDBM-Backend hat sich in diesem Umfeld Postgres etabliert. Bei CKAN handelt es sich um ein Metainformationssystem mit einem speziell auf OpenData ausgerichteten Informationsmodell, das über eine "spatial" Extension auch als Geo-Metadatenkatalog genutzt werden kann.

Als das Land Rheinland-Pfalz Ende 2012 beschloss, ebenfalls ein eigenes OpenData Portal einzurichten, wurde zunächst nach möglichen Datenquellen recherchiert. Ziel war es, vorhandene Metadatenbestände - soweit möglich - automatisiert an das neue Portal anzuschließen. Als Datenquellen boten sich hier PortalU.rlp sowie das vollständig auf FOSS basierende GeoPortal.rlp an.

Das OpenData Portal des Landes sollte bis zur CeBIT 2013 frei geschaltet werden. Zunächst wurde versucht, das CSW 2.0.2 Interface von Geonetwork zu nutzen, um die Geo-Metadaten über die CKAN harvest-extension zu "ernten". Während des Versuchs zeigten sich aber starke konzeptionelle wie auch technische Schwächen dieser Vorgehensweise. Aus diesem Grund wurde der Ansatz zugunsten eines vom GeoPortal.rlp angestoßenen "push"-Verfahrens verworfen. Bei diesem Verfahren werden die Geo-Metadaten der im GeoPortal registrierten Webservices in das CKAN spezifische json Format transformiert und dann über die CKAN json API direkt in den OpenData Katalog überführt.

Mit der Umsetzung der Schnittstelle im März 2013 konnte in Deutschland erstmalig die Übertragung der Inhalte eines umfangreichen Geo-Metadatenkatalogs in ein OpenData Portal realisiert werden. Die Daten aus dem GeoPortal.rlp sind derzeit, sowohl im rheinland-pfälzischen **Open Government Data Portal** (<http://www.daten.rlp.de>) als auch unter <http://www.govdata.de> recherchierbar. Ein Anschluss ans Portal der Europäischen Kommission befindet sich in Vorbereitung.

### Modellierung von Geo-Metadaten in CKAN

Im Prinzip handelt es sich bei der Open Source Software CKAN um ein webbasiertes Metainformationssystem, dass darauf ausgerichtet ist, Daten im Internet verfügbar zu machen. Mit CKAN sollen *freie* Datensätze veröffentlicht, geteilt und auffindbar gemacht werden. Besonders wichtig ist dabei die einfache Modellierung von Lizenzen. Der Nutzer eines OpenData Katalogs soll auf den ersten Blick sehen, welche Datensätze über freie

Lizenzen - im Sinne von OpenData<sup>1</sup> - zur Verfügung stehen. Daher wurde beim Entwurf des Datenmodells besonderer Wert darauf gelegt, Lizenzen einheitlich abzubilden und auch die *Offenheit* der Lizenz über ein Attribut zu klassifizieren (*isopen*).

Das Datenmodell<sup>2</sup> selbst ähnelt im Grunde dem eines typischen Geo-Metainformationssystem wie beispielsweise geonetwork-opensource.

Es beinhaltet neben den Datensätzen (*packages*), auch Ressourcen (*resources*), Tags (*tag*), Nutzer (*user*) und Gruppen (*group*) Elemente.

Eine Besonderheit ist, dass es neben den Datensätzen die s.g. Ressourcen gibt. Diese sind unabhängige Verweise auf Zugriffspunkte, die den jeweiligen Datensatz in verschiedenen Formaten zur Verfügung stellen. Ein und derselbe Datensatz könnte damit z.B. sowohl als CSV wie auch als XML bereitstehen. Um diesen Zusammenhang abbilden zu können, werden die Relationen zwischen *packages* und *resources* in der Datenbank verwaltet.

Das Datenmodell ist dabei sehr flexibel und lässt sich über ein Extension-System (js *extras* Objekt) leicht an unterschiedliche Anforderungen anpassen.

Auszug aus einem CKAN Datensatz mit angegebener WMS Ressource (rheinland-pfälzisches OGDP<sup>3</sup>):

```
{
  "maintainer": "VerbandsgemeindeGerolstein",
  "point_of_contact": "VerbandsgemeindeGerolstein",
  "id": "1e7454dc-8ca0-444b-be6e-db8c3a41ff7f",
  "metadata_created": "2013-03-01T12: 29: 32.204641",
  ...
  "resources": [
  {
    ...
    "package_id": "1e7454dc-8ca0-444b-be6e-db8c3a41ff7f",
    ...
    "id": "b2d4fa8d-8349-410b-8671-d1e06af2a7db",
    ...
    "description": "WMS Capabilities Link zur Integration in GIS oder Webapplikationen",
    "format": "WMS",
    "created": "2013-05-02T14:56:29.739776",
    "url": "http://www.geoportal.rlp.de/mapbender/php/wms.php?layer_id=35236&REQUEST=GetCapabilities&VERSION=1.1.1&SERVICE=WMS",
    ...
    "resource_type": "visualization"
  }
  ],
  "tags": [
    "Bauleitplan",
    ...
  ]
}
```

1 <http://opendefinition.org/okd/deutsch/>

2 Die den Untersuchungen zugrundeliegende Software basiert auf der Version 1.8.1b von CKAN. Änderungen des Datenmodells der neuen Versionen 2.0 bzw. 2.1 sind noch nicht berücksichtigt.

3 **OpenGovernmentDataPortal** – <http://www.daten.rlp.de> - <http://www.daten.rlp.de/api/rest/dataset/00158f6b-9ab8-44c4-58f3-c64d98a5c8e3>

## Geo-Metadaten in OpenData Portalen

```
},
"point_of_contact_address": {
  "email": "rathaus@gerolstein.de",
  ...
},
"groups": [
  "gdi-rp",
  ...
],
"name": "00158f6b-9ab8-44c4-58f3-c64d98a5c8e3",
"isopen": false,
"url": null,
"ckan_url": "http://daten.rlp.de/dataset/00158f6b-9ab8-44c4-58f3-c64d98a5c8e3",
"notes": "BebauungsplänerStadtGerolstein\r\nBebauungsplan\"KombinierteSatzungGerolstein-Michelbach\"derOrtsgemeindeGerolstein",
"title": "Gerolstein-Ebene: KombinierteSatzungGerolstein-Michelbach",
"maintainer_address": {
  ...
},
"terms_of_use": {
  "license_id": "dl-de-by-nc-1.0",
  ...
},
"extras": {
  "temporal_coverage_to": "2012-09-0600: 00: 00",
  ...
},
"revision_id": "ca4f8aa5-c36e-4f4f-a5a7-09fb6db1c052"
}
```

Es gibt für CKAN eine Erweiterung<sup>4</sup>, die die zusätzliche Verwaltung von Geometrien erlaubt. Dabei wird das Postgres Datenbankschema um die PostGIS Funktionen erweitert und CKAN erhält Plugins, die eine räumliche Suche, das Anzeigen von WMS im Frontend (einfacher OpenLayers-Client) sowie eine CSW Schnittstelle implementieren.

### Weitere Besonderheiten von CKAN:

- Es gibt sowohl ein **name** als auch **id** Attribut, wobei **id** von der jeweiligen CKAN Instanz beim INSERT vergeben wird.
- CKAN verwaltet die Daten grundsätzlich mit einer **Vollhistorie** – jede Änderung wird aufgezeichnet.
- Für die Suche verwendet CKAN den **Apache Solr**<sup>5</sup> Index. Es gibt eine mit Konfigurationsdatei, die das Datenmodell auf den Index mapped.
- Das **group** Element werden sowohl für die Klassifizierung von Datensätzen, als auch als Autorisierungselement (Gruppe/Organisation) verwendet.

---

4 ckanext-spatial: <https://github.com/okfn/ckanext-spatial>

5 <http://lucene.apache.org/solr/>

- CKAN verfügt über drei verschiedene json **API's**<sup>6</sup>, die genutzt werden können, um den Katalog in beliebige CMS zu integrieren oder automatisch befüllen zu lassen.
- Über die Erweiterung ckanext-harvest<sup>7</sup> lassen sich grundsätzlich auch Daten von CSW 2.0.2 API ISO 1.0 Interfaces (ISO19139 Metadaten) automatisiert in eine CKAN Instanz integrieren.

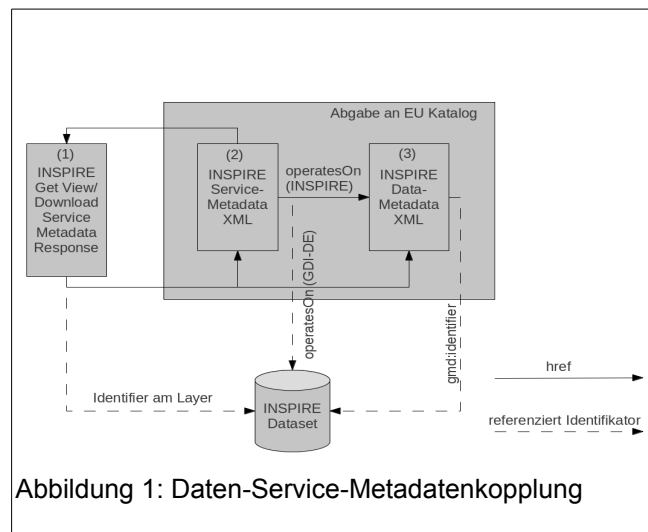
Zusammenfassend kann festgestellt werden, dass die Software CKAN *prinzipiell* dazu geeignet ist, um Geo-Metadaten zu verwalten. Ihre Stärken hat sie bei den einfachen API's, der Erweiterungsmöglichkeit des Datenmodells sowie der stark wachsenden Community im OpenData Umfeld. Auf die konzeptionellen und technischen Probleme, die sich bei genauerer Untersuchung gezeigt haben, wird später näher eingegangen.

## Modellierung von standardisierten Geo-Metadaten

Im Gegensatz zum (Meta-)Datenmodell von CKAN geht es bei standardisierten Metadaten aus dem Geo-Umfeld um die Beschreibung von Geodaten und Geodiensten.

Hier hat sich mit der Einführung der OGC Catalogue Service IS 2.0.2<sup>8</sup> im Sommer 2007 und der Übernahme der Vorgabe des zugehörigen ISO Applikationsprofils<sup>9</sup> in das Metadaten Guidance Paper<sup>10</sup> für die Umsetzung der INSPIRE-Richtlinie einiges getan. Gerade die Vorgaben der EU haben dazu geführt, dass die Modellierung sukzessive vereinheitlicht wird. Durch die Anpassung der Softwaresysteme entsteht hier gerade ein De-facto-Standard, der sich momentan auch weltweit verbreitet.

Das Modell selbst ist in sich etwas komplex<sup>11</sup>. Es basiert grundsätzlich auf den Normen ISO19115, ISO19119 sowie ISO19139. Ein Datensatz und ein zugehöriger Dienst werden dabei mit zwei unterschiedlichen xml-codierten Metadatensätzen<sup>12</sup> beschrieben. Die Relation, die beim CKAN Modell im json Objekt implizit vorhanden ist, ist im Falle



6 <http://docs.ckan.org/en/ckan-1.8/api-v2.html#model-api>, <http://docs.ckan.org/en/ckan-1.8/apiv3.html>

7 <https://github.com/okfn/ckanext-harvest>

8 [http://portal.opengeospatial.org/files/?artifact\\_id=20555](http://portal.opengeospatial.org/files/?artifact_id=20555)

9 [http://portal.opengeospatial.org/files/?artifact\\_id=21460](http://portal.opengeospatial.org/files/?artifact_id=21460)

10 [http://inspire.jrc.ec.europa.eu/documents/Metadata/INSPIRE\\_MD\\_IR\\_and\\_ISO\\_v1\\_2\\_20100616.pdf](http://inspire.jrc.ec.europa.eu/documents/Metadata/INSPIRE_MD_IR_and_ISO_v1_2_20100616.pdf)

11 [http://www.geoportal.rlp.de/portal/fileadmin/user\\_upload/Download/Dokumente/service\\_metadata\\_where2b.pdf](http://www.geoportal.rlp.de/portal/fileadmin/user_upload/Download/Dokumente/service_metadata_where2b.pdf)

12 Beispiel für Naturschutzgebiete in RP:

**Service:** [http://www.geoportal.rlp.de/mapbender/php/mod\\_layerISOMetadata.php?SERVICE=WMS&outputFormat=iso19139&Id=30825](http://www.geoportal.rlp.de/mapbender/php/mod_layerISOMetadata.php?SERVICE=WMS&outputFormat=iso19139&Id=30825)

**Data:** [http://www.geoportal.rlp.de/mapbender/php/mod\\_dataISOMetadata.php?outputFormat=iso19139&id=e9d22d13-e045-f0e0-25cc-1f146d681216](http://www.geoportal.rlp.de/mapbender/php/mod_dataISOMetadata.php?outputFormat=iso19139&id=e9d22d13-e045-f0e0-25cc-1f146d681216)

## Geo-Metadaten in OpenData Portalen

der Geo-Metadaten nur über gleiche Identifikatoren bzw. Links in den XML Dateien realisiert. Abbildung 1 verdeutlicht die Beziehungen der in Geodateninfrastrukturen verwendeten XML Dokumente untereinander.

Die konkrete Relation über den eindeutigen Datensatzidentifikator kann den folgenden Auszügen aus zwei gekoppelten INSPIRE Metadatensätzen entnommen werden<sup>13</sup>:

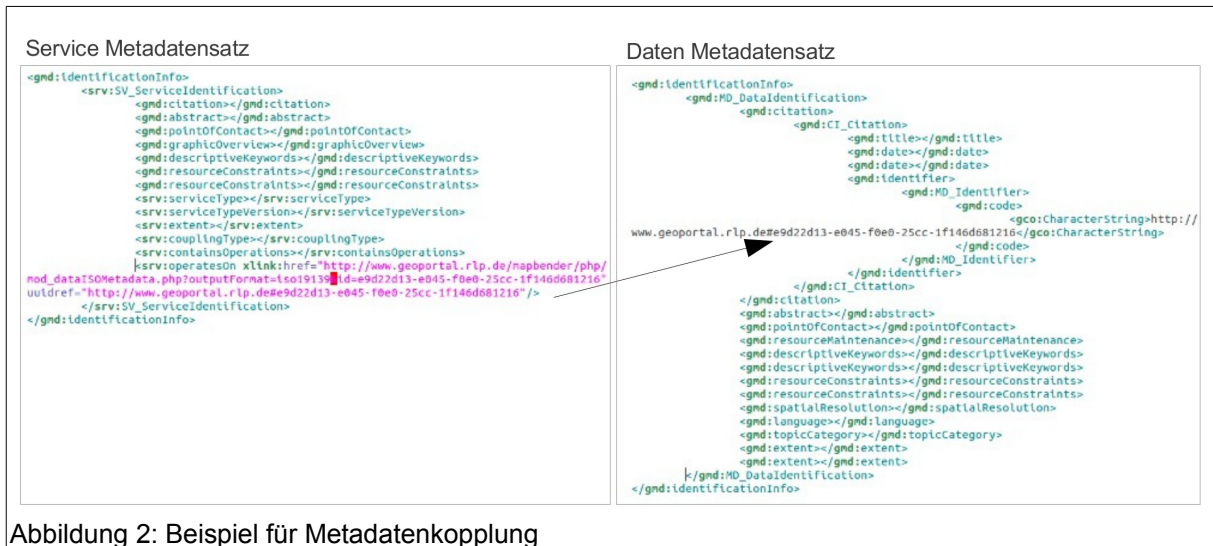


Abbildung 2: Beispiel für Metadatenkopplung

An dieser Stelle ist anzumerken, dass bei der Modellierung von Metadaten im Geo-Bereich beide Instanzen - sowohl der Datensatz als auch der Dienst – über eigene Angaben zu Nutzungsbedingungen und Zugriffsbeschränkungen verfügen. Außerdem sind die Lizenzen nicht harmonisiert. Im Umfeld von INSPIRE spricht man hier von **Constraints related to access and use**. Für das Mapping von INSPIRE konformen Geo-Metadaten in das CKAN Datenmodell sind insbesondere die s.g. **Conditions for applying for access and use** relevant. Diese werden aber grundsätzlich in Form von Freitext angegeben. Der XPath im INSPIRE Metadatensatz ist *identificationInfo[1]\*/resourceConstraints/\*/useLimitation* und kann mehrfach vorkommen.

Zusammenfassend gilt hier:

- Geo-Metadaten liegen grundsätzlich in verschiedenen XML Dateien vor.
- Die Beziehung untereinander ist nur über einen Identifikator, bzw. bei INSPIRE eine zusätzliche URL vom Service-Metadatensatz auf den Daten-Metadatensatz realisiert.
- Die Nutzungsbedingungen liegen i.A. nur als Freitext vor.
- Die Nutzungsbedingungen für Daten und Dienste sind i.d.R. unterschiedlich.

<sup>13</sup> Hinweis: Das Beispiel beschreibt nicht die vom deutschen AK Metadaten empfohlene Verknüpfung. Dort wird empfohlen den Datensatzidentifikator im Attribut xlink:href abzulegen.

# Mapping der Datenmodelle und Synchronisierung

## Erste Versuche

Um nun ausgewählte Geo-Metadaten der GDI-RP in das OpenData Portal des Landes zu überführen, wurde zunächst ein CKAN in der Version 1.8.1b, wie er auch für das Produktivsystem verwendet werden sollte, lokal installiert. Man ging davon aus, dass sich über die für das <http://data.gov.uk> Projekt entwickelte CSW-Schnittstelle (Teil von *ckanext-harvest*), ISO19139 Metadatenätze automatisiert in den CKAN überführen lassen würden.

Die GDI-RP betreibt für die Abgabe der Metadaten an den Geodatenkatalog.DE und damit auch an INSPIRE eine *geonetwork-opensource* Instanz<sup>14</sup>, die über eine CSW-Schnittstelle verfügt und über die schon seit mehreren Jahren INSPIRE-konforme Metadaten abgegeben werden.

Beim Versuch auf diese produktive Schnittstelle zuzugreifen, zeigte sich, dass das CSW-Harvestingverfahren von CKAN prinzipiell funktionsfähig ist. Leider waren die im CKAN Quelltext verankerten Funktionen zur Überführung der Datenmodelle zu sehr an das britische Metadatenprofil angepasst, was zu Fehlern beim Mapping führte. Der Harvestingprozess dauerte für 3500 Datensätze mehrere Stunden. Die Daten wurden aufgrund des beschriebenen fehlerhaften *fetching-process* (Validierung & Modelltransformation) nicht in der CKAN Oberfläche angezeigt.

Um die Tests zu beschleunigen, wurde eine lokale *geonetwork-opensource* Instanz installiert und verschiedene Beispieldateien eingespielt. Dieser Katalog konnte in einem vernünftigen Zeitrahmen vom CKAN Harvester ausgelesen werden, jedoch kam es beim Mapping der Datenmodelle immer noch zu Fehlern. Die Idee war ursprünglich, die ISO19139 Daten sukzessive so anzupassen, dass der CKAN Harvester diese anstandslos übernehmen würde. Da aber im Github zu erkennen war, dass das CSW Harvesting wohl noch *work in progress* ist und man außerdem das Problem mit den Freitext-Lizenzfeldern mit einem von CKAN ausgelösten Harvesting-Prozess nicht in den Griff bekommt, musste nach einer anderen Lösung gesucht werden.

## Umsetzungskonzept für Rheinland-Pfalz

### Grundlagen

In Rheinland-Pfalz basiert die Architektur der GDI, im Gegensatz zu den meisten anderen deutschen Bundesländern, nicht auf der Verwendung eines herkömmlichen Geo-Metadatenkatalogs. Das Konzept für die GDI-RP wurde im Jahr 2005 zu einem Zeitpunkt erstellt, zu dem es noch keine funktionsfähigen Standards für den Austausch von Geo-Metadaten zu Daten und Diensten gab. Es basiert auf dem Prinzip der Extraktion und des Caching von Dienst- und Daten-Metadaten aus OGC Capabilities Dokumenten und erlaubt Anbietern von OWS, diese Metadaten in einem zentralen *Metadatenproxy* zu editieren sowie mit weiteren Informationen anzureichern. Die für die INSPIRE Umsetzung und den Aufbau der GDI-DE benötigten ISO19139 Metadaten werden dabei grundsätzlich *on the fly* bzw. *on demand* aus dem internen Datenmodell erzeugt.

---

14 <http://www.gdi-rp-dienste.rlp.de/geonetwork/>

## Harmonisierung der Lizenzen

Das zentrale Informationsmodell wurde, seit der Inbetriebnahme 2007, immer wieder an die Anforderungen der Praxis angepasst. Eine dieser Anforderungen war es, die veröffentlichten Dienste im GeoPortal.rlp mit vordefinierten Lizenzbedingungen derart anzureichern, dass ein Nutzer nach *freien* Datenquellen recherchieren kann<sup>15</sup>. Dieses Verfahren selbst wurde jedoch bis Anfang 2013 nur sporadisch genutzt; die meisten Anbieter blieben bei der Bereitstellung eigener Nutzungsbedingungen über die schon zuvor erläuterten Freitext-Felder.

Im Zusammenhang mit dem Austausch von Metadaten zwischen Geodateninfrastrukturen und OpenData Portalen, war das o.g. Konzept jedoch zielführend. Hier besteht der Bedarf, eindeutig lizenzierte Daten automatisiert an andere Plattformen zu übertragen. Die Verantwortung für die Angabe von Nutzungsbedingungen liegt beim Datenanbieter selbst. Das Datenmodell der GDI-RP wurde im Februar 2013 dahingehend erweitert, dass – ähnlich wie beim CKAN – die im Portal hinterlegten Lizenzen zusätzlich ein *isopen* Attribut erhielten. Nach der Umsetzung der neuen Möglichkeit im Backend wurden die Anbieter von Diensten über die Änderungen informiert und darauf hingewiesen, dass im Falle der Zuordnung einer von vier abgestimmten Lizenzen, ihre Daten dann auch im OpenData Portal des Landes auffindbar seien. Diese Möglichkeit wurde dankend angenommen und viele Anbieter stellten daraufhin ihre Daten unter eine weniger restriktive Lizenz.

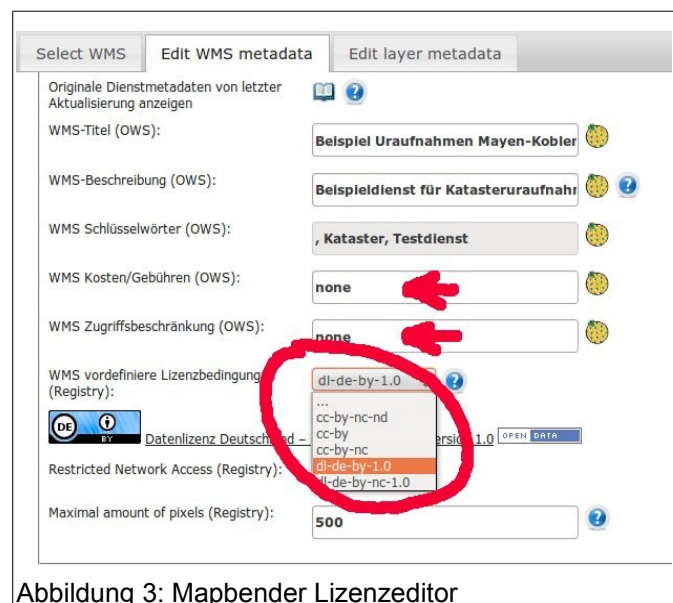


Abbildung 3: Mapbender Lizenzeditor

## Übertragung der Metadaten

Für die Übertragung der Metadaten aus dem GeoPortal.rlp in das rheinland-pfälzische OpenData Portal wurde der Weg über die CKAN API gewählt. Das Verfahren basiert darauf, dass die Geo-Metadaten des Informationsmodells der GDI-RP zunächst in das CKAN Datenmodell (json - UTF8) transformiert<sup>16</sup> und dann über insert, update und delete Funktionen der CKAN API verwaltet werden. Die CKAN verfügt über drei verschiedene API's<sup>17</sup>, die alle auf den einfachen http Methoden GET, POST, PUSH basieren und deren Parameter und Rückgabewerte in den meisten Fällen json-Objekte sind.

- API 1 – Ansprechen von Objekten über *id*, *RESTful*
- API 2 – Ansprechen von Objekten über *name*, *RESTful*
- API 3 (action API), auch Rückgabe von *results*, *errors* und *help* Informationen

<sup>15</sup> OpenData war 2008 noch nicht im Fokus der Politik.

<sup>16</sup> [http://www.geoportal.rlp.de/mapbender/php/mod\\_exportMapbenderLayer2CkanObjects.php](http://www.geoportal.rlp.de/mapbender/php/mod_exportMapbenderLayer2CkanObjects.php)

<sup>17</sup> <http://docs.ckan.org/en/ckan-1.8/api.html>

Für diese API gibt es Clients in verschiedenen Programmiersprachen. Da die dem GeoPortal.rlp zugrundeliegende Software *mapbender* auf php basiert, wurde ein unter MIT Lizenz stehende Client<sup>18</sup> adaptiert und als Klasse in den *mapbender* Funktionsumfang integriert.

Grundsätzlich war der Aufwand bei der Implementierung der API überschaubar. Probleme können jedoch auftauchen, wenn im OpenData Portal Anpassungen am CKAN Quelltext vorgenommen werden. Leider ist das in Rheinland-Pfalz passiert und hat dazu geführt, dass zur Synchronisierung der Datensätze alle 3 API's verwendet werden müssen.

Der Synchronisierungsprozess aktualisiert momentan alle Datensätze des GeoPortal.rlp, gleichgültig ob diese sich geändert haben oder nicht. In Zukunft ist es vorgesehen, dass hier vor einem *update* ein Vergleich von timestamps erfolgt.

Die Erzeugung eines CKAN Datensatzes aus dem GeoPortal.rlp Informationsmodell kann grundsätzlich über zwei verschiedene Ansätze erreicht werden:

1. Erzeugung eines CKAN *packages* aus einem Geo-Daten-Metadatenatz. Anschliessend müssen alle verknüpften Geo-Service-Metadatenätze in jeweils ein CKAN *resource* Objekt überführt und an das *package* angehängen werden.
2. Erstellung eines CKAN *packages* für ein WMS Layer Objekt und Anreicherung von Ressourcen für verschiedene Zugriffsoptionen auf dieses Layer Objekt.

Grundsätzlich ist der erste Weg zu bevorzugen. Es gibt aber insbesondere drei Gründe die das verhindern:

- Die Daten-Service-Metadatenkopplung ist nicht für alle im GeoPortal.rlp registrierten Dienste vorhanden. Nur für die INSPIRE Daten ist sie verpflichtend (75 von 4200 Datensätze!). Für die anderen Ressourcen müsste sie erst durch die Datenanbieter generiert werden.
- Das Geo-Metadatenmodell sieht unterschiedliche Nutzungsbedingungen auf Datensatz- und Diensteebene vor. Diese lassen sich nicht in das CKAN Datenmodell abbilden.
- Eine saubere Bereitstellung von Nutzungsbedingungen auf Ebene des Datensatzes ist derzeit nicht möglich. Hier fehlt noch eine Abstimmung, wie man klassifizierte Lizenzbedingungen in ISO19139 Dateien abbildet (das sollte dann aber weltweit einheitlich erfolgen).

---

18 [https://github.com/jeffreybarke/Ckan\\_client-PHP](https://github.com/jeffreybarke/Ckan_client-PHP)



## Ergebnisse

Mit der beschriebenen Vorgehensweise konnte in Deutschland *erstmalig* eine automatisierte Übertragung von Geo-Metadaten aus einem Länderkatalog in ein OpenData Portal realisiert werden. Damit war keine redundante Neuerfassung von Metadaten über HTML-Frontends oder mit Hilfe von Excel-Tabellen mehr notwendig.

Der Weg, der dabei beschrrieben wurde, erlaubt es zusätzlich auch weitere Nutzungsmöglichkeiten (*ressourcen*) im OpenData Portal zur generieren, z.B.:

- Direkter Aufruf eines Kartenviewers mit Anzeige der jeweiligen Datenquelle
- Verlinkung der speziellen Geo-Metadatenanzeige des GeoPortal.rlp:



Außerdem wurde ein großer Schritt in Richtung mehr Transparenz bezüglich der Nutzungsbedingungen von Geodatendiensten geschaffen. Es wird demnächst möglich sein auch die Frage „zeige mir nur alle frei zugänglichen Datenquellen an“ zu beantworten.

Die aktive Synchronisierung des CKAN über die API ermöglicht hochaktuelle Metadaten – im Falle von Änderungen könnten update-Prozesse getriggert werden.

Aktuell stammen 788 von insgesamt 1503 Datensätzen im OpenData Portal Rheinland-Pfalz aus dem GeoPortal.rlp!

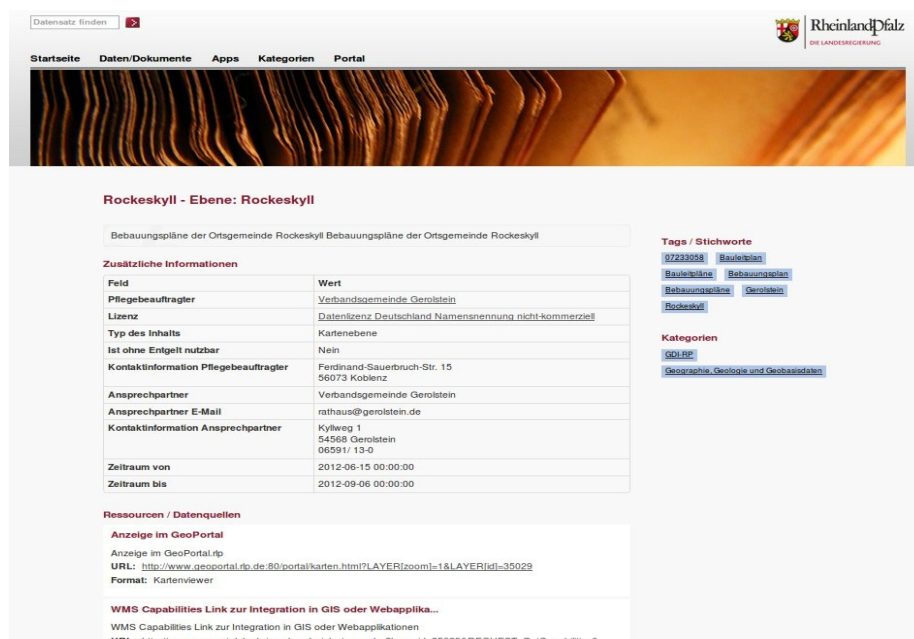


Abbildung 4: Metadaten zu einem WMS Layer Bebauungsplan im OGD RP

## Lessons learned / Pitfalls

### **Meta-Datenmodelle**

Eine saubere Überführung des standardisierten Geo-Metadatenmodells, wie es z.B. von INSPIRE gefordert wird, ist in das CKAN Modell aus folgenden Gründen derzeit **nicht** ohne weiteres möglich:

- Es gibt unterschiedliche Geo-Metadatenprofile, die einer Entwicklung eines einheitlichen Harvesting-Verfahrens im Wege stehen.
- Das Vorhandensein unterschiedlicher, nicht maschinell auswertbarer Nutzungsbedingungen auf Daten- und Diensteebene der Geo-Metadaten.
- Fehlerhafte und/oder nicht vorhandene Relationen zwischen Daten- und Dienst-Metadaten auf Seite der Geodatenkataloge.
- Fehlendes einheitliches Mapping zwischen ISO Kategorien und CKAN *group* Elementen.

### **Technik**

- Das automatisierte Update schon kleiner Datenmengen über die CKAN API (ein paar hundert Datensätze) kann durch inkrementelle Fortführung des *apache-solr* zu Speicherüberläufen führen.<sup>19</sup>
- Es ist wichtig, bei der Synchronisierung eindeutige Datensatzidentifikatoren auf der Seite des führenden Systems zu haben, die dann als CKAN *name* genutzt werden.
- CKAN Objekte können nicht gelöscht werden, sondern bekommen den *status*: „*deleted*“. Danach ist es nicht mehr änderbar! Das bedeutet, dass man um ein CKAN Objekt aus der Suche zu entfernen und später wieder mit gleichem *name* anzulegen, dessen *name* ändern muss (bspw.: *new\_name: old\_name\_timestamp*).

### **Sonstiges**

Die OpenData Informationen gelangen auf vielen Wegen in die CKAN Kataloge. Teils werden sie von Hand eingestellt, teils über automatisierte Schnittstellen generiert. Es kann z.B. vorkommen, dass eine Stelle Metadaten für ein Online-Auskunftssystem, das auf einer Datenbank arbeitet über ein HTML-Frontend erfasst und auf der anderen Seite Metadaten für einen WMS Dienst, der auf dieselbe Datenbank zugreift, über eine automatisierte Schnittstelle in das OpenData Portal übertragen werden. Aus Sicht des OpenData Portals gibt es jetzt zwei *package* Objekte mit unterschiedlichen *id* und *name* Werten. Das ist natürlich unsinnig. Eine Lösung könnte sein, dass Geo-Metadaten ein „*genuine\_name*“ Feld bekommen, in das dann eine UUID eingetragen werden kann. Diese kann von einer ganz anderen Stelle vergeben worden sein. Das Synchronisierungsverfahren könnte aber dann bei Bedarf nach einem Datensatz mit diesem *name* suchen und die verknüpften Dienste dort als *resource* Objekte anhängen.

---

<sup>19</sup> Hängt ggf. auch mit der Vollhistorie des CKAN zusammen.

## Fazit

Man muss feststellen, dass es beim Versuch des automatischen Abgleichs von Geo-Metadaten mit OpenData Katalogen auf CKAN Basis einige semantische, inhaltliche und technische Probleme gibt, die bisher noch nicht gelöst sind. Manche der Probleme lassen sich durch pragmatische Ansätze umgehen, andere müssen wahrscheinlich auf der Ebene der Standardisierung angegangen werden.

Um den aktuellen politischen OpenData Prozess voranzubringen und damit mehr und mehr Daten *frei* zugänglich zu machen, ist es notwendig, die Datenmodelle bestehender Metainformationssysteme und OpenData Kataloge abzugleichen und deren Inhalte automatisiert zu synchronisieren. Das gilt natürlich nicht nur für Geo-Metadaten sondern auch für allgemeine Metadatenkataloge, wie z.B. das Umweltinformationssystem PortalU<sup>20</sup>.

Bei den Anbietern von Daten muss weiterhin das Bewusstsein geschaffen werden, dass man einen wirklichen volkswirtschaftlichen Mehrwert nur dann erhält, wenn die Hürden für die Nutzung der Daten so niedrig wie möglich sind.

Die Integration bestehender Geodateninfrastrukturen in den OpenData Prozess bietet hier Chancen, die genutzt werden sollten.

Kontakt zum Autor:

Armin Retterath  
Zentrale Stelle Geodateninfrastruktur Rheinland-Pfalz  
LVerGeo RP  
Ferdinand-Sauerbruch-Straße 15  
56073 Koblenz  
0261/492-466  
armin.retterath@lvermgeo.rlp.de

Weitere Informationen im Netz

- [1] <http://www.geoportal.rlp.de>
- [2] <http://ckan.org/>
- [3] <http://www.daten.rlp.de>
- [3] <http://www.govdata.de>
- [4] <http://publicdata.eu>
- [5] <http://inspire.jrc.ec.europa.eu>
- [6] <http://lucene.apache.org/solr/>
- [7] <http://www.opengeospatial.org/>
- [8] <http://okfn.org/>
- [9] [http://www.mapbender.org/Handbuch\\_Geoportal](http://www.mapbender.org/Handbuch_Geoportal)

---

20 <http://www.portalu.de>