

Gegenüberstellung der OGD-Metadaten Ausarbeitungen von Deutschland und Österreich

basierend auf den aktuellen Versionen der Metadaten-Ausarbeitungen(26.02.2013)

AutorInnen

Johann Höchtl
Beratendes Mitglied der Cooperation OGD Österreich
Donau-Universität Krems
Dr. Karl-Dorrek-Straße 30
3500 Krems

Christian Habernig
Arbeitsgruppenleiter AG Metadatenstruktur der Cooperation OGD Österreich
Magistrat der Stadt Wien, Magistratsabteilung 14
Rathausstraße 1
1082 Wien

mit Kommentaren von

Ina Schieferdecker
Kordinatorin Open Data @ Fraunhofer FOKUS
Kaiserin-Augusta-Allee 31
10589 Berlin

Florian Marienfeld,
Open Data Strategie Fraunhofer FOKUS
Kaiserin-Augusta-Allee 31
10589 Berlin

Brigitte Lutz
Cooperation OGD Österreich
Magistratsdirektion der Stadt Wien
Rathausstraße 8
1010 Wien

Zielsetzung. Dieses Dokument bietet eine Gegenüberstellung der Metadaten-Ausarbeitungen von Österreich und Deutschland. Damit wird das Ziel verfolgt, die Deckungsgleichheit der jeweiligen Ausarbeitungen zu identifizieren und strukturelle und inhaltliche Unterschiede aufzuzeigen. Längerfristig wird eine weitgehende Harmonisierung der Metadatenstandards angestrebt, um eine Zusammenführung auf einem übergeordneten Datenportal, das im Raum D-A-CH-LI oder auf EU-Ebene angesiedelt sein könnte, zu vereinfachen.

Basisdokumente

Österreich: Arbeitsgruppe Metadatenstruktur der Cooperation OGD Österreich: OGD Metadaten – 2.1 (Entwurf), verfügbar unter http://reference.e-government.gv.at/uploads/media/OGD-Metadaten_2_1_2012_10.pdf (abgerufen am 26.02.2013)

Deutschland: ogd-metadata: Metadata structure and documentation to be used by the German Open Government Platform (github commit 0b78228c08aa380764287ad308d5e22eeb98f4e9), verfügbar unter <https://github.com/fraunhoferfokus/ogd-metadata> (abgerufen am 26.02.2013, weiterführende Informationen unter <http://open-data.fokus.fraunhofer.de/?p=643>)

Organisatorische Rahmenbedingungen

Österreich. Die Metadatenspezifikation Österreichs wird von der Arbeitsgruppe AG Metadatenstruktur der Cooperation OGD Österreich (<http://data.gv.at/hintergrund-infos/cooperation-ogd-oesterreich/>) mit dem Ziel erarbeitet, eine Beschreibung für offene Daten und Services der öffentlichen Verwaltung Österreichs zu schaffen. Die Mitglieder der Arbeitsgruppe setzen sich aus Mitgliedern OGD-anbietender Verwaltungen, GIS und INSPIRE-Beauftragten, Semantic-Web ExpertInnen und VertreterInnen der OGD-Community zusammen. Letztere sollen bekannte good practices bestehender OGD-VorreiterInnen wie data.gov oder data.gov.uk, die europäische Perspektive (ADMS, DCAT, W3C Arbeitsgruppe) und die Eigenheiten der de-facto Standardplattform CKAN mitberücksichtigen. Bei Detailfragen werden anlassbezogen ExpertInnen zugezogen. Ausarbeitungen werden als „Sprints“ durchgeführt, d.h. wird der Bedarf nach einer Ausarbeitung, Korrektur oder Präzisierung erkannt, arbeiten die Mitglieder der Arbeitsgruppe in sehr kurzen Zeitintervallen, bis das gesteckte Ziel erreicht ist.

Die Spezifikationsarbeit erfolgt „am Tisch“, in wenigen Ausnahmen werden Telefonkonferenzen einberufen. Die Ausarbeitung liegt bis zum Zeitpunkt der arbeitsgruppeninternen Freigabe an die Cooperation OGD Österreich nur den Mitgliedern vor. Die Ausarbeitung erfolgt in Textverarbeitung und wird als PDF veröffentlicht.

Die Dokumente werden über die Kooperation-Bund-Länder-Städte-Gemeinden (BLSG) als E-Government - Konvention bzw. White Paper öffentlich am österreichischen E-Government Reference Server publiziert (<http://reference.e-government.gv.at/Veroeffentlichte-Informationen.2774.0.html>).

Deutschland. Die deutsche Metadaten-Struktur ist als Teilergebnis des Prototyps GovData.de zur Studie Open Government Data Deutschland entstanden und wird entlang des Pilotbetriebs weiterentwickelt. Für die **initiale** Fassung wurde das Format JSON-Schema und der Umfang der Felder aus dem Best-Practice-Abschnitt der Studie abgeleitet. Diese wurden außerdem im Rahmen der Bund-Länderarbeitsgruppe Open Government abgestimmt, dort speziell in der Unterarbeitsgruppe Technik und in einem Workshop mit maßgeblichen Datenbereitstellern, vor allem des Geodatenportal Deutschland, PortalU für Umweltdaten und dem Statistischen Bundesamt (<http://open-data.fokus.fraunhofer.de/?p=877>).

Die Erarbeitung der technischen Spezifikation ist über den gesamten Entstehungsverlauf öffentlich unter <https://github.com/fraunhoferfokus/ogd-metadata> einsichtig. Die Plattform unterstützt prinzipiell gemeinsames Arbeiten (fork & pull requests), neuer Ergebnisse und Zwischenstände stehen unmittelbar der Community zur Diskussion bereit.

Struktureller Aufbau

Österreich. Die OGD-Metadatenstruktur besteht aus dem **Metadatenkern** mit 11 Pflichtfeldern und 20 zusätzlich empfohlenen **optionale Metadatenfeldern**. Nur wenn ein Datensatz mit den 11 Pflichtfeldern beschrieben ist, wird er zum OGD-Datensatz.

Die Felder sind spezifiziert durch

- numerische ID
- deutschsprachigen, menschenlesbaren Bezeichner
- ODG-Kurznamen (identifizier)
- Mapping zum entsprechenden CKAN-Feld
- Häufigkeit des Auftretens (Kardinalität, 1 oder n)
- Definition in deutscher Sprache
- Erläuterung in deutscher Sprache. Die Erläuterungen sind so formuliert, dass sie als Hilfestellung in Eingabemasken von OGD-Verwaltungswerkzeugen verwendet werden könnten.
- Beispiel: Eine beispielhafte Feldinstanz
- ON A 2270:2010: Verweis auf die korrespondierende Metadatenbeschreibung in der ÖNORM A 2270 (DIN EN ISO 19115) als Durchführungsbestimmung der ISO 19115 („Geographic Information - Metadata“, ISO-Norm)
- ON/EN/ISO 19115:2003: Verweis auf das Metadatenfeld der ISO 19115
- RDF-Property: RDF-Vokabular-Verweis
- Definition in englischer Sprache

Die Kardinalitätsangaben ergeben eine zweistufige Gliederungsebene durch Felder mit der Kardinalität N, die eine implizite zweite Gliederungsebene aufspannt.

Tabelle 1: Gliederungsebene der 11 OGD Österreich Pflichtfelder

Ebene 1, Kardinalität 1	Ebene 2, Kardinalität N	Beschreibung
Eindeutiger Identifikator		Eindeutiger Identifikator für den Metadatensatz. Es sollte eine UUID (nach RFC:4122) gewählt werden
Datum des Metadatensatzes		Datum, an dem der Metadatensatz erzeugt bzw. aktualisiert wurde als YYYY-MM-DD
Titel des Metadatensatzes		Titel des beschriebenen Metadatensatzes.
Beschreibung		Inhaltliche Beschreibung des Datensatzes bzw. Dienstes
datenverantwortliche Stelle		Bezeichnung bzw. Name der für den Datensatzes oder Dienst zuständigen Organisation bzw. Person
Lizenz		Rechtliche Nutzungsinformationen für die Verwendung des Datensatzes oder Dienstes.
zeitliche Ausdehnung (Anfang)		Element zur Erfassung des Beginns der Gültigkeit eines Datensatzes oder Dienstes als YYYY-MM-DDThh:mm:ss

Datensatz oder Dienst-Link	URL für den Zugriff auf den Datensatz oder Dienst
Datensatz oder Dienstformat	Angabe zum Format des Datensatzes oder Dienstes. Diese ist als Dateiformat, Download- oder Service-Link anzugeben. Liste
Schlagworte	Beschlagwortung des Datensatzes bzw. Dienstes.
Kategorie	Kategorisierung des Datensatzes bzw. Dienstes. Vorgegebenes Kategorienschema für Österreich. Liste

Kategorien:

1. Arbeit
2. Bevölkerung
3. Bildung und Forschung
4. Finanzen und Rechnungswesen
5. Geographie und Planung
6. Gesellschaft und Soziales
7. Gesundheit
8. Kunst und Kultur
9. Land- und Forstwirtschaft
10. Sport und Freizeit
11. Umwelt
12. Verkehr und Technik
13. Verwaltung und Politik
14. Wirtschaft und Tourismus

Formate:

txt, csv, html, xml, rdf, odt, ods (...), rss, atom, json, png, jpg, jpeg, jp2, svg, gml, gpx, kml, shp, shx, dbf, prj, rss, wms, wfs, wmts, wcas csw

Deutschland. OGD-Metadaten gliedern sich in 15 Pflichtfelder und 35 optionale Felder.

Die Felder werden spezifiziert durch

- Metadaten-Eintrag: Bezeichnung in deutscher Sprache
- CKAN-Bezeichner
 - Bezeichnung der 2. Gliederungsebene
 - Bezeichnung der 3. Gliederungsebene
- Pflichtfeld: Ja oder Nein
- Datentyp: Grundlegende JSON-Datentypen (vgl. <http://www.json.org>, RFC 4627)
- Format / Pattern: Wahlweise regulärer Ausdruck zur Beschreibung der gültigen Werte, Aufzählungsfeld (Enumeration) oder Spezifikations-Identifizier (z.B. URI, date-time)
- Beschreibung: menschen-lesbare Beschreibung zur Verwendung des Feldes, enthält Verweise auf Standards und Normen, die eine weitere Einschränkung zum Gültigkeitsbereich der Felder liefern (z.B. MIME für Formate, ISO 639-1 zur Sprachspezifikation, ...)

Tabelle 2: Gliederungsebene OGD Deutschland

Ebene 1	Ebene 2	Ebene 3	Beschreibung
Name			Ein für Menschen lesbarer Bezeichner des Datensatzes, des Dokuments oder der Apps, der für eine Identifizierung genutzt werden kann
Titel			Der Titel beschreibt den Datensatz, das Dokument oder die App prägnant
Veröffentlichende Stelle			Die Behörde, von der die Daten stammen. Bei Apps der App-Hersteller.
Beschreibung			Beschreibung und weitere Informationen zum Datensatz, zum Dokument oder zur App, wird auf der Detailseite dargestellt und kann mehrere Absätze umfassen.
Kategorien			Kategorien für die Datensätze und Dokumente. Die Kategorien sind statisch und werden vom Portalbetreiber langfristig gepflegt. Liste
Typ			Zeigt an, ob ein Datensatz, ein Dokument, eine App beschrieben wird.
Lizenz-ID			Lizenz aus einer festen Liste
Resources			
	url		Verweis auf die eigentliche Datendatei, das Dokument oder die Anwendung
	format		Typ der Ressource (MIME Types)
	description		Erläuterung, welche Rolle die Ressource für den Datensatz, das Dokument bzw. die App spielt (z. B. ob es sich dabei um eine bestimmte Zeitscheibe oder Schlüsseliste handelt. Der Beschreibungstext, wird mit dem Link zur Ressource unterlegt.)
Extras			
	contacts		Während der Containertyp verpflichtend ist, ist es keines der Kind-Elemente? D: prüfen

	role	Rolle, die dieser Kontakt für das Metadatum innehat. Enum
	name	Name der Person oder Funktion
	url	Webseite des Kontakts
	email	Email-Adresse des Kontakts, wird auch für Benachrichtigungen genutzt
	address	Postalische Adresse des Kontakts
dates		
	role	Rolle, die dieses Datum für den Datensatz spielt
	date	Date-time, das Kalender-Datum
terms_of_use		
	other	Beschreibung, falls keiner der Lizenzwerte zutrifft
	license_url	URL, auf der die Lizenz des Datensatzes, Dokuments oder der App beschrieben und erklärt ist.
	is_free_to_use	Die Information, ob der Datensatz/das Dokument/die App kostenfrei und zweckoffen ist (muss gesetzt werden, wenn Freitext bei der Lizenzangabe genutzt wird)

Listwerte für Kategorien:

1. wirtschaft_arbeit
2. transport_verkehr
3. umwelt_klima
4. geo
5. gesundheit
6. verbraucher
7. infrastruktur_bauen_wohnen
8. bildung_wissenschaft
9. verwaltung
10. gesetze_justiz
11. bevoelkerung
12. politik_wahlen
13. soziales
14. kultur_freizeit_sport_tourismus

Listwerte für Lizenz:

1. apache
2. bsd-license
3. cc-by
4. cc-by-sa
5. cc-nc
6. cc-zero
7. gfdl
8. gpl-3.0
9. mit-license
10. mozilla
11. mozilla1.1
12. odc-odbl
13. other-closed
14. other-open
15. app_openSource
16. app_freeware

Listwerte für Typ:

1. datensatz
2. dokument
3. app

Enum-Werte für Contact-Role:

1. veroeffentlichende_stelle
2. ansprechpartner
3. vertrieb
4. autor

Listwerte für dates-role:

1. erstellt
2. veroeffentlicht
3. aktualisiert

Inhaltliche Gegenüberstellung der OGD-Metadaten schemata

(beschränkt auf Pflichtfelder, unter Angabe von optionalen Feldern zur Aufnahme von Daten, die nicht 1:1 gemappt werden können bzw. falls das entsprechende Pflichtfeld fehlt)

Diese Gegenüberstellung auf Feldebene dient der Veranschaulichung des Handlungsbedarfs um ein eindeutiges Mapping von offenen Verwaltungsdaten aus Österreich und Deutschland zu ermöglichen. Im Vordergrund der Überlegung stehen ein mögliches Datenportal im Raum D-A-CH-LI und die Möglichkeit, einer länderübergreifenden Datenzusammenführung.

Tabelle 3: Gegenüberstellung der OGD-Metadaten schemata

Österreich	Deutschland	Bemerkung
Eindeutiger Identifikator	Optional: metadata_original_portal und Optional: metadata_original_id	Ein eindeutiger Identifikator ist in den Pflichtfeldern des Metadaten schemas Deutschland nicht vorgesehen, allerdings wird Rücksicht auf CKAN-Kompatibilität gelegt. CKAN vergibt intern eine ID, sodass diese herangezogen werden könnte oder aus den optionalen Feldern D:metadata_original_portal und D:metadata_original_id bezogen werden kann. Bei der Übernahme von Daten aus Deutschland kann eine ID „on the fly“ generiert werden.
Datum des Metadaten satzes	Extras:dates:role=erstellt; Extras:dates:date	1:1 Mapping möglich, D spezifiziert kein Format für DateTime, daher evtl. Konversion notwendig. Achtung UTC vs. lokaler Zeit?
Titel des Metadaten satzes	Name	1:1
	Titel	Deutschland vergibt sowohl Name als auch Titel, wobei „Titel beschreibt den Datensatz, das Dokument oder die App prägnant“. Für D:Titel gibt es kein Äquivalent in der österreichischen Metadatenbeschreibung (sowohl in den Pflicht- als auch optionalen Feldern).
Beschreibung	Beschreibung	1:1
Datenverantwortliche Stelle	Datenverantwortliche Stelle	1:1

Lizenz	license_id	Deutschland erlaubt beliebige Lizenzen, 16 Lizenztypen sind fix vorgegeben. Eine Zuordnung von D → A ist nur dann möglich, wenn Extras:terms_of_use:license_id='cc-by', Mapping A→D möglich.
zeitliche Ausdehnung (Anfang)	Optional: temporal_coverage_from	Das Äquivalent für A: zeitliche Ausdehnung (Anfang) ist in Deutschland nur im optionalen Feld D:temporal_coverage_from gegeben. Mapping A→D möglich (Feld wird in der deutschen Metadatenbeschreibung in das optionale Feld D:temporal_coverage_from übernommen), Mapping D→A nur dann möglich, wenn im deutschen Schema das optionale Feld D:temporal_coverage_from befüllt ist.
Datensatz oder Dienst-Link	Ressources:url	1:1
Datensatz oder Dienstformat	Ressources:format	1:1 über Mapping möglich. Deutschland verwendet als Formatspezifikation das generische MIME-Modell, ohne allerdings Einschränkungen in den Formaten zu geben. Ein 1:1 Mapping von MIME zur Werteliste in Österreich ist möglich.
Schlagworte	Optional: Schlagwörter	Kein Äquivalent in den Pflichtfeldern Deutschlands, allerdings optionales Feld „Schlagwörter“. Da Schlagworte in Österreich mit der Kardinalität ‚N‘ ausgezeichnet ist, kann das Mapping D → A dieses Feldes als optional behandelt werden oder ggf. aus dem optionalen Feld D: „Schlagwörter“ erfolgen, A→D ist möglich und kann in das optionale Feld „Schlagwörter“ übernommen werden.
Kategorie	Kategorien	Das Metadatenchema Deutschland beschreibt 14 förderale Kategorien (für weitere existieren Mappings, z.B. Bremen → D), das österreichische Metadatenchema beschreibt ebenfalls 14 Kategorien

Mögliches Mapping der OGD-Kategorien

Deutschland zu Österreich

D \ A	Bevölkerung	Bildung und Forschung	Finanzen und Rechnungswesen	Geographie und Planung	Gesellschaft und Soziales	Gesundheit	Kunst und Kultur	Land- und Forstwirtschaft	Sport und Freizeit	Umwelt	Verkehr und Technik	Verwaltung und Politik	Wirtschaft und Tourismus	Arbeit
bevoelkerung														
bildung_wissenschaft														
infrastruktur_bauen_wohnen														
geo														
verbraucher														
gesundheit														
soziales														
verwaltung														
gesetze_justiz														
umwelt_klima														
transport_verkehr														
politik_wahlen														
kultur_freizeit_sport_tourismus														
wirtschaft_arbeit														

Österreich zu Deutschland

A \ D	bevoelkerung	bildung_wissenschaft	infrastruktur_bauen_wohnen	geo	verbraucher	gesundheit	soziales	verwaltung	gesetze_justiz	umwelt_klima	transport_verkehr	politik_wahlen	kultur_freizeit_sport_tourismus	wirtschaft_arbeit
Bevölkerung														
Bildung und Forschung														
Finanzen und Rechnungswesen														
Geographie und Planung														
Gesellschaft und Soziales														
Gesundheit														
Kunst und Kultur														
Land- und Forstwirtschaft														
Sport und Freizeit														
Umwelt														
Verkehr und Technik														
Verwaltung und Politik														
Wirtschaft und Tourismus														
Arbeit														

Erläuterung: Mit grüner Farbe markierte Mappings können vermutlich 1:1 gemappt werden (inhaltlich-semantic Abgleich durch gemeinsame Arbeitsgruppe notwendig), als gelb markierte mappings müssen als 1:n ausgeführt werden (möglich, da Kategorie sowohl in Deutschland als auch Österreich die Kardinalität N hat).

Fazit

- Das Mapping von OGD-Metadaten im Kernbereich (Pflichtelemente) ist möglich.
- Die Metadatenbeschreibung von Deutschland und Österreich unterscheidet sich kritisch im Pflichtfeld D:Titel
- Es gibt syntaktische Unklarheiten, die eine inhaltliche Zuordnung unter Umständen erschweren, so ist beispielsweise nicht spezifiziert wie D:dates:date gespeichert wird.

Mögliche Folgeaktivitäten

- Abstimmung der semantischen Bedeutung und des Mappings der Kategorien
- Deutschland
 - Detailliertere technische Festlegung der Syntax von Serialisierungsformaten, z.B. Darstellung Datum / Uhrzeit
- Österreich
 - URI für Schlüssel zu Fremdsystem aufnehmen. Bei österreichischen Daten sollte hier der Schlüssel auf die originale Metadatenquelle abgespeichert werden, bei der Übernahme aus Fremdsystemen der Fremdschlüssel (*canonical URI*). Die derzeitige Metadatenbeschreibung Deutschlands ist dazu in der Lage.