

Cooperation OGD Österreich: Arbeitsgruppe Metadaten		White Paper	
		OGD Metadaten – 1.1	
		Ergebnis der AG	
Kurzbeschreibung	<p>Primäre Aufgabe der Arbeitsgruppe Metadaten (AG) war die Entwicklung einer Metadatenstruktur, die als Empfehlung für OGD in Österreich verwendet werden soll. Es wurde auch die INSPIRE-Relevanz berücksichtigt.</p> <p>Als Ergebnis liegen ein sogenannter Metadatenkern, zusätzlich optionale Attribute und das Vokabular zum Metadatenkatalog Österreich vor.</p> <p>Die wichtigsten offenen Entscheidungen und noch zu beauftragenden Folgeaktivitäten sind im Kapitel „Weitere Vorgangsweise“ angeführt.</p>		
Autor(en):	Christian Habernig (Stadt Wien), Martin Kaltenböck (Semantic Web Company), Brigitte Lutz (Stadt Wien)	Projektteam / Arbeitsgruppe Arbeitsgruppe Metadatenstruktur der Cooperation OGD Österreich	
Beiträge von:	Thomas Burg (STATISTIK AUSTRIA), Stefan Dürauer (Stadt Wien), Gregor Eibl (Bundeskanzleramt), Johannes Forsthuber GmbH, Roland Grillmayer (FH Wiener Neustadt), Gerhard Hartmann (Stadt Wien), Johann Höchtl (Donau-Universität Krems), Wolfgang Jörg (Stadt Wien), Dominik Klauser (Bundeskanzleramt), Gustav Lebhart (Stadt Wien), Rudolf Legat (Umweltbundesamt), Manfred Mittlböck (Research Studios Austria Forschungsgesellschaft mbH), Erik Obersteiner (Umweltbundesamt), Stefan Pawel (Stadt Linz), Carl-Markus Piswanger (Bundesrechenzentrum GmbH), Rainer Prager (Land Niederösterreich), Michael Rederer (Stadt Wien), Thomas Thurner (Semantic Web Company), Wolfgang Tinkl (Land-, forst- und wasserwirtschaftliches Rechenzentrum GmbH)		

Version 1.0: **24.10.2011**

Version 1.1: **12.03.2012**

Inhaltsverzeichnis

(1) ALLGEMEINES	4
(2) MANAGEMENT SUMMARY	4
(3) WEITERE VORGANGSWEISE.....	5
(4) METADATENKERN.....	6
(5) ZUSÄTZLICHE ALTERNATIVE ATTRIBUTE.....	8
(6) GRUNDLAGEN ZUR METADATENSTRUKTUR	11
(7) VOKABULAR ZUR METADATENSTRUKTUR.....	12
A. RDF PROPERTY	12
B. AKTUALISIERUNGSZYKLUS	14
C. OGD-FORMATE.....	15
D. KATEGORIEN	16

Cooperation OGD Österreich: Abschlussbericht der Arbeitsgruppe Metadaten

(1) Allgemeines

Die Cooperation OGD Österreich hat beim Treffen am 13. Juli 2011 in Wien beschlossen, eine Arbeitsgruppe Metadaten (AG Metadaten) ins Leben zu rufen, welche eine Empfehlung für die Metadatenstruktur für Open Government Data (OGD) in Österreich erarbeiten soll.

Nach 5 Treffen der Arbeitsgruppe unter der Leitung der Stadt Wien zwischen 1.8.2011 und 28.9.2011 wurde der Abschlussbericht als White Paper „OGD Metadaten - 1.0“ veröffentlicht.

Unter Einbeziehung der INSPIRE-Stakeholder haben am 20.01.2012 und am 01.03.2012 weitere Sitzungen stattgefunden. Aus den Ergebnissen und Abstimmungen mit allen Beteiligten wurde das vorliegende Dokument erarbeitet

Folgende Organisationen wurden in die Arbeitsgruppe Metadaten eingeladen und haben an den Treffen der AG teilgenommen und an dieser Empfehlung mitgearbeitet:

- Bundesamt für Eich- und Vermessungswesen
- Bundeskanzleramt Österreich, BKA I/11
- Bundesrechenzentrum GmbH
- DI Forsthuber GmbH
- FH Wiener Neustadt
- Land Niederösterreich
- Land-, forst- und wasserwirtschaftliches Rechenzentrum GmbH
- Research Studios Austria Forschungsgesellschaft mbH
- Stadt Wien: MD-OS/IKT, MA 14, MA 23, MA 41, MA 53
- Stadt Salzburg
- Stadt Linz, Open Commons Region Linz
- STATISTIK AUSTRIA Bundesanstalt Statistik Österreich
- Umweltbundesamt
- Donau Universität Krems
- Open Knowledge Foundation / Semantic Web Company

(2) Management Summary

Primäre Aufgabe der Arbeitsgruppe Metadaten (AG) war die Entwicklung einer **Metadatenstruktur**, die als Empfehlung für OGD in Österreich verwendet werden soll.

Als Ergebnis liegen ein sogenannter **Metadatenkern**, empfohlene optionale **Metadatenfelder**, ein **Kategorienkatalog** und eine **Übersichtstabelle** (siehe Anhang) vor.

Die wichtigsten offenen Entscheidungen und noch zu beauftragenden Folgeaktivitäten sind im Kapitel „weitere Vorgangsweise“ angeführt.

(3) Weitere Vorgangsweise

Die folgenden Aktivitäten sind aus Sicht der AG Metadaten durch die Cooperation OGD Österreich zu entscheiden bzw. zu beauftragen:

- Beauftragung eines **Erfassungslitfadens**, der begleitend zur Entwicklung des Erfassungsformulars des Bundesportals erstellt werden sollte.
- Beauftragung der Entwicklung eines **URL und/oder URI Schema** für OGD in Österreich
 Ein URL Schema bildet hierbei die empfohlene Struktur der referenzierbaren URL eines Datensatzes auf einem Datenportal ab, welche hinsichtlich einer durchgängigen Referenzierbarkeit empfohlen wird, wie beispielsweise:
 data.name.gv.at/katalog/kategorie/name-des-datensatzes
 Ein URI Schema dient zur eindeutigen Referenzierbarkeit für ‚Dinge‘ wie beispielsweise Datenhalter (Ministerien, Städte et al) und weiterführend für ‚Dinge‘ welche in Datensätzen immer wieder als Daten verwendet werden und über eine eindeutige URI klar und eindeutig referenziert (verlinkt) werden können wie beispielsweise: Schulen oder Kindergärten. Siehe dazu auch die Empfehlung der britischen Regierung auf data.gov.uk unter: <http://data.gov.uk/resources/uris>
- Beauftragung der Entwicklung (Evaluierung und Empfehlung) eines **Open Government Data / eGovernment Thesaurus** unter Berücksichtigung bestehender Vokabulare und Thesauri (z.B. GEMET, INSPIRE, BLSG) für die Unterstützung der Harmonisierung der Metadaten auf diversen OGD Portalen in Österreich sowie zur Sicherstellung der Interoperabilität mit internationalen OGD Initiativen und Portalen (publicdata.eu, Pan-Europäisches Datenportal 2013 et al). Es wird empfohlen sich an bestehende Schlagwortkataloge zu orientieren, z.B. Thesauri aus eGov, Finanz, Recht etc. oder auch INSPIRE und GEMET Schlüsselwörtern. Bei einem Teil-Harvesting von INSPIRE ist es wünschenswert, deren Schlagworte für die bessere Auffindbarkeit nutzen zu können.
- Beauftragung der Harmonisierung der **finalen (gesamten) Empfehlung zu Metadaten, OGD Thesaurus und des URI Schemas** auf nationaler und internationaler Ebene (wie beispielsweise von Europeana via Österreichischer Nationalbibliothek oder dem Abgleich mit der W3C eGovernment Working Group sowie anderen Stellen, welche im Detail seitens der Arbeitsgruppe evaluiert und identifiziert werden) – dies ist auch von besonderer Wichtigkeit, um die österreichischen Empfehlungen für die Wiederverwendung in anderen Ländern bereit zu stellen.
- **Beschreibung des OGD Österreich Metadaten-Schemas mit ADMS**, um international interoperabel zu sein und damit wiederverwendet zu werden (siehe: <http://joinup.ec.europa.eu/asset/adms/release/08>)

(4) Metadatenkern

12 empfohlene Pflichtfelder bilden den „Metadatenkern“, wobei der „Eindeutige Identifikator“ und das „Datum des Metadatensatzes“ automatisch vom System generiert werden sollten.

ID	Bezeichner <i>OGD-Schema Kurznahme</i>	Anz	Beschreibung Deutsch	Beschreibung englisch	Kommentar
1	Eindeutiger Identifikator <i>identifier</i>	1	Eindeutiger Identifikator für den Metadatensatz	Unique identifier to the meta dataset	Soll automatisch vom System generiert werden. Falls schon vorhanden, muss dieser übernommen werden (Harvesten bestehender Metadatensätze).
5	Datum des Metadaten-satzes <i>metadata_date</i>	1	Datum, an dem der Metadaten-satz erzeugt bzw. aktualisiert wurde.	Date on which the metadata record created or updated.	Soll automatisch vom System generiert werden. JJJJ-MM-TT
8	Titel <i>title</i>	1	Titel des Metadaten-satzes	A name given to the meta dataset .	
9	Beschreibung <i>description</i>	1	Kurze Beschreibung für BenutzerInnen, keine Wiederholung anderer Datenfelder, Kontext des Metadaten-satzes zusammenfassend beschrieben	Data abstract giving a short, human-readable description of the resource including domain contextualisation.	Inhaltliche Beschreibung des Metadaten-satzes
10	Kategorie <i>categorization</i>	N	Merkmal zur Einordnung der Ressource in ein vorgegebenes Kategorienschema	Categories to categorise a resource	Ziel: harmonisierte Ontologie. Erster Schritt: an Lebenslagen-Kategorien anlehnen und veröffentlichen, sofern es Abweichung gibt
11	Schlagworte <i>keyword</i>	N	Schlagworte, die den Inhalt der Ressource beschreiben	Free keywords that describe the resource until harmonized using an Austrian OGD / eGovernment thesaurus	Schlagworte müssen nicht zwingend im Datensatz vorkommen, können aber durchaus vorkommen. Freitext, solange kein OGD Thesaurus verwendbar ist
14	Ressourcen-Link <i>resource_url</i>	N	URL für den Zugriff auf die Ressource	Permanent and unique URL to resource	Entspricht auch in manchen Fällen dem optionalen Feld URI zu den Metadaten

ID	Bezeichner <i>OGD-Schema Kurzname</i>	Anz	Beschreibung Deutsch	Beschreibung englisch	Kommentar
15	Ressourcen-Format <i>resource_format</i>	N	Dateiformat, Schnittstelle; Schreibweise wird aus den Cooperation OGD Rahmenbedingungen übernommen	File, transport or encapsulation format as specified by OGD Austria Framework	Ohne Punkt / without path delimiters, in Kleinbuchstaben / in lower cases Auflistung siehe Anhang / listing
16	Ressourcen-Titel <i>resource_description</i>	N	Nähere Beschreibung der einzelnen Ressource	Detailed resource abstract	Beispiel: „Hauptwohnsitzbevölkerung 2010“, „Hauptwohnsitzbevölkerung 2011“,...
19	Datenverantwortliche Stelle <i>author</i>	1	Name der für die Ressource zuständige Person oder Organisation	Data accountable	Kann in kleinen Organisationen gleichzeitig die (siehe optional) veröffentlichende Stelle sein.
21	Lizenz <i>licence</i>	1	Lizenz, unter welcher der Datensatz verwendet, verändert und weitergegeben werden kann.	Link to licence; in Austria CC BY 3.0 AT is recommended for open data	In Österreich CC-BY 3.0 AT bei kostenfreien Datensätzen im Sinne von Open Data. Link auf die Lizenz.
24	Zeitliche Ausdehnung (Anfang) <i>begin_datetime</i>	1	Beginn der Gültigkeit einer Ressource	Start of validity of a resource	DATUM in Form JJJJ-MM-TT (hh:mm:ss) Zumindest JJJJ

ID ... laufende Nummerierung im OGD-Metadatenkatalog Österreich

Anz ... 1= Single-Value-Feld N=Multi-Value-Feld

(5) Zusätzliche alternative Attribute

Zusätzlich sind 18 weitere Attribute empfohlen, jedoch kein Bestandteil des Metadatenkerns

ID	Bezeichner <i>OGD-Schema Kurzname</i>	Anz	Beschreibung Deutsch	Beschreibung englisch	Kommentar
2	Bezeichnung der Metadatenstruktur <i>schema_name</i>	1	OGD Austria Metadata 1.1	OGD Austria Metadata 1.1	Konsistente Metadatenführung erleichtert. Dieses Metadatenelement sollte automatisch vom System befüllt werden
3	Sprache des Metadatensatzes <i>schema_language</i>	1	Sprache des Metadatensatzes: ger	Language of meta data: ger	ISO 639-2 dreistelliger ISO Sprachcode für den Metadatensatz: ger
4	Character Set Code des Metadatensatzes <i>schema_characterset</i>	1	Zeichensatz des Metadatensatzes: utf8	Character Set Code : utf8	Konsistente Metadatenführung erleichtert. Dieses Metadatenelement sollte automatisch vom System befüllt werden
6	Weiterführende Metadaten <i>metadata_linkage</i>	1	URL zur weiteren ausführlicheren Beschreibung der Metadaten.	URL for more detailed description of the metadata (e.g. INSPIRE meta data catalogue)	Soll angegeben werden wenn in einem anderen Metadatenkatalog zusätzliche Informationen verfügbar sind (z.B. INSPIRE Metadatenkatalog)
7	URI zu den Metadaten <i>metadata_uri</i>	1	Permanente, eindeutige URL auf die Metadaten eines Datensatzes (das Datenblatt selbst)	Permanent and unique URL to meta data, possibly a referrer to itself	Entspricht in manchen Fällen dem Feld Ressourcen-URL (ein Zeiger auf sich selbst)
12	Attributbeschreibung <i>attribute_description</i>	1	menschenlesbare Beschreibung der Bedeutung der Datenfelder in einem Datensatz	human-readable description of the meaning of the data fields in a record	zB die Spaltennamen und deren Bedeutung in einer CSV Datei
13	Kontaktseite der datenverantwortlichen Stelle <i>maintainer_link</i>	1	Link auf Ansprechpartner/-stelle, Link auf den Erzeuger	Link to contact of the resource accountable	

ID	Bezeichner <i>OGD-Schema Kurzname</i>	Anz	Beschreibung Deutsch	Beschreibung englisch	Kommentar
17	Veröffentlichungsdatum <i>date_released</i>	1	Datum der Veröffentlichung der Ressource.	Date on which the resource was created	DATUM in Form JJJJ-MM-TT (hh:mm:ss)
18	Änderungsdatum <i>date_updated</i>	1	Datum der letzten Aktualisierung der Ressource.	Date of last update of the resource	Soll automatisch vom System generiert werden DATUM in Form JJJJ-MM-TT (hh:mm:ss)
20	Veröffentlichende Stelle <i>publisher</i>	1	Name der für die Veröffentlichung der Ressource zuständige Person oder Organisation	publisher	Unterschied zu datenverantwortlicher Stelle (siehe Pflichtfeld)
22	Geographische Bezugsebene <i>geographic_toponym</i>	1	Menschenlesbare Beschreibung der räumlichen Lage einer Ressource	Human readable, habituated naming of spatial coverage	z.B. „Österreich“, „Wien“
23	Geographische Ausdehnung <i>geographic_bbox</i>	1	Dokumentation der geographischen Ausdehnung eines Datensatzes mit der Definition eine umrahmenden Rechtecks	Documentation of the geographical expansion of the definition of a data set with a framing rectangle	Angabe im Koordinatensystem EPSG:4326 (WGS84) mit einer Mindestgenauigkeit von zwei Dezimalstellen in "Well Known Text String" Repräsentation. POLYGON (Rechtswert_min Hochwert_min, Rechtswert_max Hochwert_max)
25	Zeitliche Ausdehnung (Ende) <i>end_datetime</i>	1	Ende der Gültigkeit einer Ressource	End of validity of a resource	DATUM in Form JJJJ-MM-TT (hh:mm:ss) Zumindest JJJJ
26	Aktualisierungszyklus <i>update_frequency</i>	1	Menschenlesbare Frequenz der Datenaktualisierung	Human readable frequency with which modifications and deletions are made	Eintrag basiert auf der Codeliste, die in der ON EN ISO 19115:2003, Abschnitt B5.18 definiert ist (siehe Vokabulare)
27	Datenqualität <i>lineage_quality</i>	1	Menschenlesbare Beschreibung von Qualitäts- und/oder Entstehungsgenese der Ressource z.B. Methode der Erhebung	Human readable description of metering method, survey methodology, statistic parameters, etc.	

ID	Bezeichner <i>OGD-Schema Kurzname</i>	Anz	Beschreibung Deutsch	Beschreibung englisch	Kommentar
28	Titel und Beschreibung Englisch	1	Englische Angabe von Titel und Beschreibung der Ressource	Name and description of resource in english	
29	Grösse der Ressource	1	Filegröße	Size of resource	Angabe in k
30	Datenquelle	1			Datenquelle laut Nutzungsbedingungen, z.B. "Datenquelle: Stadt Wien - data.wien.gv.at"

(6) Grundlagen zur Metadatenstruktur

- Der Metadatenkern beinhaltet ausschließlich Pflichtfelder
- Als Designrichtlinie wurde festgesetzt, die Anzahl der Attribute des Metadatenkerns, also jene Attribute die jedenfalls zu befüllen sind, so gering als möglich zu halten.
- Ein Metadatenblatt muss sowohl für den Menschen, als auch für die „Maschine“ gut lesbar und interpretierbar sein.
- Es werden in der Empfehlung 'zusätzliche alternative Felder angeboten' (kein Bestandteil des Metadatenkerns und auch keine Pflichtfelder).
- Der Metadatenkern wurde in englische Sprache übersetzt (zwecks internationaler Überprüfbarkeit, bietet eine erste Empfehlungen für verwendete Vokabulare & eine erste Evaluierung von 'Big Lists' zur Hinterlegung von z.B.: Kategorien etc.
- Die Metadatenstruktur muss erweiterbar sein. D.h. sie muss ermöglichen Attribute zu ergänzen, was bedeutet, dass mehrere Versionen der empfohlenen Metadatenstruktur zulässig sind. Dies bedeutet, die Versionen müssen voneinander eindeutig unterscheidbar sein.
- Die Metadatenstruktur soll in Hinsicht auf international vorhandene Metadatenstrukturen im Bereich Open (Government) Data kompatibel sein.
- Die Pflichtfelder des Metadatenkerns müssen angeführt werden, ansonsten kann der OGD- Metadatenatz nicht als gültig betrachtet werden. Dies ist deswegen notwendig, da bei Nichtangabe die Integrität der verschiedenen Kataloge bei einer (künftigen) Vernetzung sehr stark gefährdet, wenn nicht gar unmöglich wäre.
- ALLE OGD Metadaten-Elemente sollten für die Dokumentation österreichischer OGD Ressourcen (Dokumente, Datensätze, Services, Medien etc.), wenn möglich angegeben werden.
- Jedes optionalen OGD- Metadatenelemente sollte unbedingt angeführt und dokumentiert werden, wenn ausreichend Information für das jeweiligen Metadatenelement vorhanden sind.
- Jede OGD Community ist darüber hinaus eingeladen, eigene zusätzliche Metadatenelemente frei zu definieren und in das jeweilige Community Profil zu integrieren. Diese Metadatenelemente stellen aber keinen Bestandteil der Metadatenstruktur für Österreich dar.
- Bei INSPIRE – Daten übernimmt bzw. harvestet ein OGD-Metadatenatz die entsprechenden INSPIRE-Metadaten. Es ist nicht vorgesehen, dass INSPIRE OGD-Metadaten harvestet.

(7) Vokabular zur Metadatenstruktur

a. *RDF Property*

Um die Metadaten maschinell verarbeitbar / maschinenlesbar zur Verfügung zu stellen wird jedes Attribut mit einer Property aus allgemein bekannten und oft verwendeten (RDF) Vokabularen versehen – dies dient u.a. der Interoperabilität mit anderen Systemen.

Diese Definition wird in der nächsten Version des Dokuments publiziert

Bezeichner	RDF Property	Hinweise
Eindeutiger Identifikator <i>identifier</i>		
Datum des Metadaten-satzes <i>metadata_date</i>		
Titel <i>title</i>		
Beschreibung <i>identifier</i>		
Kategorie <i>categorization</i>		
Schlagworte <i>keyword</i>		
Ressourcen-Link <i>resource_url</i>		
Ressourcen-Format <i>resource_format</i>		

Ressourcen-Titel <i>resource_</i> <i>description</i>		
Datenverant- wortliche Stelle <i>maintainer</i>		
Lizenz <i>licence</i>		
Zeitliche Ausdehnung (Anfang) <i>begin_datetime</i>		
Eindeutiger Identifikator <i>identifier</i>		
Datum des Metadaten-satzes <i>metadata_date</i>		
Titel <i>title</i>		

b. Aktualisierungszyklus

Eintrag basiert auf der Codeliste, die in der ON EN ISO 19115:2003, Abschnitt B5.18 definiert ist: „MaintFreqCd frequency with which modifications and deletions are made to the data after it is first produced“

Deutsch	Name	Domain code	Definition
kontinuierlich	continual	001	data is repeatedly and frequently updated
taglich	daily	002	data is updated each day
wochentlich	weekly	003	data is updated on a weekly basis
14-tagig	fortnightly	004	data is updated every two weeks
monatlich	monthly	005	data is updated each month
quartalsweise	quarterly	006	data is updated every three months
halbjahrlich	biannually	007	data is updated twice each year
jahrlich	annually	008	data is updated every year
nach Bedarf	asNeeded	009	data is updated as deemed necessary
unregelmaig	irregular	010	data is updated in intervals that are uneven in duration
nicht geplant	notPlanned	011	there are no plans to update the data
unbekannt	unknown	012	frequency of maintenance for the data is not known

c. OGD-Formate

Text-, Tabellen- und Bildformate	Endung
Klassische Textdateien	txt
Comma Separated Value	csv
Hypertext Markup Language	html
Extensible Markup Language	xml
Resource Description Framework	rdf
Open Document Formats	odt, ods...
Newsfeed/Webfeed Syndication	rss
JSON (JavaScript Object Notation)	json
Geodatenformate	Endung
Geography Markup Language	gml
GPS Exchange Format	gpx
Keyhole Markup Language	kml
ESRI Shapefile	shp, shx, dbf, prj
Georss:	rss
Schnittstellen	
Web Map Service	
Web Feature Service	
Web Map Tile Service	

d. **Kategorien**

Kategorien Österreich	englisch
Arbeit	employment
Bevölkerung	population
Bildung und Forschung	education and science
Geographie und Planung	geography and planning
Gesellschaft und Soziales	society
Gesundheit	health
Kunst und Kultur	culture
Sport und Freizeit	recreation
Umwelt	environment
Verkehr und Technik	transport
Verwaltung und Politik	government and politics
Wirtschaft und Tourismus	economy and tourism

Die vorgeschlagenen 12 Kategorien beruhen auf den der BLSG-Konvention "E-Government Verfahrens-/Leistungsbereiche" (v1b 1.3) (<http://reference.e-government.gv.at/EP-VV-v1b-1-3-0-Version-vom.563.0.html>) und der Bereichsabgrenzungsverordnung (E-Gov-BerAbgrV, StF: BGBl. II Nr. 289/2004, Anlage zu § 3 Abs. 1, Teil 1) (in der Tabelle gelb bezeichnet).

Verglichen wurden diese mit den Kategorien der Open Data Portalen von Wien, Linz, Berlin, London, Seattle, Großbritannien, Belgien, Canada und Kenia und der Menü-Seite der Statistik Austria.

Es wurden die Kategorien mit einer Übereinstimmung zwischen 50% und 100% ausgewählt.

Ergänzt wurden diese um die Kategorien, die aus der Praxis praktikabel erscheinen bzw. die bei der Wiener Umfrage zu Open Government als die interessantesten Bereiche gewertet worden sind (in der Tabelle grün bezeichnet).

Die aus diesen Quellen erstellte thematische Klassifikation wird als Normierungsgrundlage eingebracht, die aus der best practice Perspektive Cooperation OGD Österreich erweitert werden kann.

Die englischen Ausdrücke können für die URL-Konvention verwendet werden. Die Erstellung eines Thesaurus ist aus der vorliegenden Tabelle ebenfalls möglich.