

Cooperation OGD Österreich: Abschlussbericht der Arbeitsgruppe Metadaten		White Paper	
		OGD Metadaten – 1.0	
		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.</p> <p>Als Ergebnis liegt ein sogenannter Metadatenkern und ein Kategorienkatalog (siehe Anhang) vor.</p> <p>Der nächste Schritt muss die Ergänzung zum Metadatenkern um (u.a. INSPIRE relevante) weitere Attribute sein.</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)	Projektteam / Arbeitsgruppe	
		Arbeitsgruppe Metadatenstruktur der Cooperation OGD Österreich	
Beiträge von:	Thomas Burg (Statistik Österreich) Stefan Dürauer (Stadt Wien) Gregor Eibl (Bundeskanzleramt) Gerhart Hartmann (Stadt Wien) Johann Höchtl (Donauuniversität Krems) Wolfgang Jörg (Stadt Wien) Dominik Klauser (Bundeskanzleramt) Gustav Lebhart (Stadt Wien) Rudolf Legat (Umweltbundesamt) Brigitte Lutz (Stadt Wien) Stefan Pawel (Stadt Linz) Michael Rederer (Stadt Wien) Thomas Thurner (Semantic Web Company)		

Inhaltsverzeichnis

(1) ALLGEMEINES	3
(2) MANAGEMENT SUMMARY.....	3
(3) WEITERE VORGANGSWEISE DIE FOLGENDEN AKTIVITÄTEN SIND AUS SICHT DER AG METADATEN DURCH DIE COOPERATION OGD ÖSTERREICH ZU ENTSCHEIDEN BZW. ZU BEAUFTRAGEN:.....	4
(4) DETAILS ZUR WEITEREN VORGANGSWEISE	5
(5) ANHANG	6
1. METADATENKERN.....	6
2. GRUNDLAGEN ZUR METADATENSTRUKTUR	9
3. ZUSÄTZLICHE ALTERNATIVE ATTRIBUTE	10
4. VOKABULARE ZUM METADATENKERN	11
5. KATEGORIEN	13

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.

Insgesamt wurden 5 Treffen der Arbeitsgruppe unter der Leitung der Stadt Wien durchgeführt:

1. Treffen der Cooperation OGD Österreich AG - Metadaten 1.8.2011
2. Treffen der Cooperation OGD Österreich AG - Metadaten 17.8.2011
3. Treffen der Cooperation OGD Österreich AG - Metadaten 6.9.2011
4. Treffen der Cooperation OGD Österreich AG - Metadaten 15.9.2011
5. Treffen der Cooperation OGD Österreich AG - Metadaten 28.9.2011

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

- Bundeskanzleramt Österreich, BKA I/11
- Stadt Wien, MD-OS/IKT
- Stadt Wien, MA 14, MA 53, MA 41
- Stadt Salzburg
- Stadt Linz, Open Commons Region Linz
- Statistik Austria
- 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 liegt ein sogenannter **Metadatenkern** und ein **Kategorienkatalog** (siehe Anhang) vor.

Der nächste Schritt muss die **Ergänzung zum Metadatenkern** um (u.a. INSPIRE relevante) weitere Attribute sein.

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:

- Entscheidung über die Veröffentlichung des **Metadatenkerns**.
- Entscheidung über das Verfahren, wie die erforderliche Erweiterung des Katalogs erfolgen soll. Empfohlen wird ein möglichst transparentes Verfahren unter Mitwirkung (vor allem) der INSPIRE relevanten Interessensgruppen.
- Beauftragung eines **Erfassungsleitfadens**.
- Beauftragung der Entwicklung eines **URL** und/oder **URI** Schema für OGD in Österreich.
- Beauftragung der Entwicklung eines Open Government Data / eGovernment **Thesaurus**.
- Beauftragung der Harmonisierung der **finalen (gesamten) Empfehlung zu Metadaten, OGD Thesaurus und des URI Konzepts** auf nationaler und internationaler Ebene.

(4) Details zur weiteren Vorgangsweise

In Folge die oben genannte Vorgehensweise im Detail:

- **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>
- Evaluierung und Empfehlung für den Einsatz eines **Open Government Data / eGovernment Thesaurus** 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)
- und schließlich die **Prüfung der finalen (gesamten) Empfehlung zu Metadaten, OGD Thesaurus und URI Konzept** 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.
- Ein weiterführendes Projekt für die AG Metadaten, wie oben empfohlen, muss hinsichtlich der Ressourcen der AG Mitglieder geklärt werden, da das angeführte Projekt vom Umfang her grösser ist als das bis dato durchgeführte.

(5) Anhang

1. Metadatenkern

Bezeichner	Beschreibung Deutsch	Anz.	Kommentar	Englische Beschreibung
Titel	Titel der Daten	1		Identifier; A name given to the resource.
Beschreibung	Kurze Beschreibung für BenutzerInnen, keine Wiederholung anderer Datenfelder, Kontext des Datensatzes zusammenfassend beschrieben	1		Data abstract giving a short, human-readable description of the data set including domain contextualisation.
URI zu den Metadaten	Permanente, eindeutige URL auf die Metadaten eines Datensatzes (das Datenblatt selbst). Evtl. ein Zeiger auf sich selbst	1		Permanent and unique URL to meta data, possibly a referrer to self
Typ	Kollektion, Events, Tonmaterial, Bildmaterial, Text, reale physische Objekte, interaktive Ressource, interaktives Dokument, Programm, Dataset, Service, API	N		Collection, Event, Sound, Video ("MovingImage"), Text, physical Object, interactive Resource, Software, API
Format	Dateiformat, Schreibweise wird übernommen aus den Cooperation OGD Rahmenbedingungen: ohne Punkt, frei eingebbar in Kleinbuchstaben (z.B.: csv)	N		File, transport or encapsulation format without path delimiters, all in lower case as specified by OGD Austria Framework
Erzeuger	vgl. DC:Creator der Daten (Datenverantwortliche Stelle), auch im Sinn von Autor	1		Data accountable, possibly the author
Kontaktseite zum Erzeuger	Link auf Ansprechpartner/-stelle, Link auf den Erzeuger	1		Link to contact oft the author

Kategorien	Merkmal zur Einordnung des Datensatzes in ein vorgegebenes Kategorienschema	N	Ziel: harmonisierte Ontologie. Erster Schritt: an Lebenslagen-Kategorien anlehnen und veröffentlichen, sofern es Abweichung gibt	Categories to categorise a dataset
Schlagworte	Schlagworte die den Inhalt des Datensatzes beschreiben (aber nicht zwingend im Datensatz vorkommen müssen, aber durchaus vorkommen können)	N	Freitext, solange kein OGD Thesaurus verwendbar ist	Free keywords that describe the dataset until harmonized using an Austrian OGD / eGovernment thesaurus
Lizenz	Lizenz unter welcher der Datensatz verwendet, verändert und weitergegeben werden kann.	1	In Österreich durchgehend CC-BY 3.0 AT bei kostenfreien Datensätzen im Sinne von Open Data. Link auf die Lizenz.	Link to licence; in Austria CC BY 3.0 AT is recommended for open data
Aktualisierung	menschenlesbare Frequenz, zu dem eine Datenänderung stattfinden sollte.	1		Human readable period at which data will be re-published.
Räumliche Bezeichnung	Menschenlesbare, räumliche Begrenzung des (geografischen) Bezugs der Daten, z.B. „Österreich“, „Wien“	1		Human readable, possibly habituated naming of spatial coverage
Zeit von	Zeitliche Ausdehnung (Einzeldatum, Datumsintervall bzgl. Ressource) und Erstellungsdatum Ressource	1		Temporal coverage FROM
Zeit bis	Zeitliche Ausdehnung (Einzeldatum, Datumsintervall bzgl. Ressource) und Erstellungsdatum Ressource	1		Temporal coverage TO

Datenqualität	menschenlesbar, z.B. Methode der Erhebung	1		Human readable description of metering method, survey methodology, statistic parameters, etc.
----------------------	---	---	--	---

2. Grundlagen zur Metadatenstruktur

- Grundsätzlich können im Bereich OGD verschiedene 'Datenblatt Typen' unterschieden werden, wie z.B.: OGD, INSPIRE et al.
- Alle diese 'Datenblatt Typen' haben einen CORE (Metadatenkern) an Attributen (denselben kleinsten gemeinsamen Nenner von Metadaten Feldern)
- 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.
- Mindestens die INSPIRE Attribute (zumindest die Pflichtfelder der INSPIRE Attribute) sollen vorhanden sein.
- Je nach 'Datenblatt Typ' werden diesem Metadatenkern weitere spezifische Felder hinzugefügt. Es werden daher 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.

3. Zusätzliche alternative Attribute

Welche empfohlen sind, aber kein Bestandteil des Metadatenkerns sind, sind wie folgt:

- **Veröffentlicht am:** [Datum + Uhrzeit] nach ISO 8601 / RFC 3339 der Veröffentlichung des Metadatenblattes.
- **Attribute:** menschenlesbare Beschreibung der Bedeutung der Datenfelder in einem Datensatz (also zB die Spaltennamen und deren Bedeutung in einer CSV Datei oder ähnliches).
- Bezeichnung des **Metadatenstandards** inklusive Versionsnummer
- **Größe des Datensatzes** (File Größe)
- **Englische Angabe** von Titel und Beschreibung des Datenblattes
- Vergabe von **Schlagworten durch UserInnen** – sogenannte ‚user generated tags‘.

4. Vokabulare zum Metadatenkern

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.

Bezeichner	RDF Property	Kommentar
Titel	dcterms:title	
Beschreibung	dcterms:abstract	
URI zu den Metadaten	dcat:record oder owl:sameAs	owl:sameAs nur dann benutzen, wenn beide URIs eindeutig (oder identisch) sind und immer den exakt selben Metadatensatz identifizieren
Typ	dcterms:type	Mit diesem Property können die bei Kommentar 1 angeführten DC Types (und auch selbst erstellte) verwendet werden
Format	dcterms:format	Recommended best practice is to use a controlled vocabulary such as the list of Internet Media Types: http://www.iana.org/assignments/media-types/
Erzeuger	Noch ungeklärt: Evtl. eigene Property a la "creator_name"	Noch offen und zu überprüfen
Kontaktseite zum Erzeuger	dcterms:creator	dcterms:creator benötigt einen URI des Ersteller, deshalb für obiges Feld "Erzeuger" wegen dem dortigem Freitext nicht möglich
Kategorien	dcat:theme oder tags:associatedTag	Beides nur bei Nutzung von kontrollierten Tags sinnvoll, also URIs statt Strings. Tags:associatedTag wird bereits sehr häufig verwendet Zusätzliche Alternative: dcat:theme
Schlagworte	dcterms:subject	dcterms:subject weil Freitext erlaubt - sonst wäre es wie bei Kategorien möglich

Lizenz	dcterms:license	Um automatisierte Analysen von Datensätzen zu ermöglichen, ist es wichtig ,canonical identifiers' für bekannte Lizenzen zu verwenden.
Aktualisierung	dcterms:accrualPeriodicity	Wenn die Werte aus einem kontrollierten Vokabular kommen , wie beispielsweise aus einer vordefinierten Liste von Orten. Es können auch zB Intervalle von http://www.placetime.com/ verwendet werden.
Räumliche Bezeichnung	dcterms:spatial	Räumliche Bezeichnung / Begrenzung. Sollte, wenn möglich, auch ein URI sein
Zeit von	dcterms:temporal	„Zeit von und Zeit bis“ kann in einem einzelnen RDF Value ausgedrückt werden, welcher Beginn und Zeitdauer enthält, dazu z.B.: verwendbar: http://www.placetime.com/
Zeit bis	dcterms:temporal	siehe oben
Datenqualität	dcat:dataQuality	dcat:dataQuality verwendet einen Freitext als Wert

5. 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, Groß-Britannien, 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.