

# Semi-automatisierte Metadatenerstellung (profil.AT 3.x) von räumlichen Datensätzen und deren Bereitstellung für INSPIRE und OGD

## *Semi-automated Metadata Creation (profil.AT 3.x) of Spatial Datasets and Making it Available for INSPIRE and OGD*

Bernhard Vockner<sup>1</sup>, Manfred Mittlböck<sup>1</sup>, Victoria Nikmengjaj-Achatz<sup>2</sup>

<sup>1</sup>RSA iSPACE, Salzburg · bernhard.vockner@researchstudio.at

<sup>2</sup>Amt der Salzburger Landesregierung, Abteilung 10: Wohnen und Raumplanung, Salzburg

**Zusammenfassung:** Datensätze für Open und Big Data bedürfen effizienter Ansätze, um einen Mehrwert aus diesen Daten generieren zu können. Dies geschieht mithilfe von strukturierten Metadaten, wodurch Daten zu Information werden. Gerade die Ersterfassung von Metadaten bei sehr vielen Datensätzen ist zeitaufwendig. Demzufolge widmet sich die vorliegende Forschungsarbeit der semi-automatisierten Metadatenerstellung von räumlichen Datensätzen nach profil.AT 3.x sowie deren Bereitstellung für INSPIRE und OGD. Unser Ansatz greift auf das Fachwissen von ExpertInnen in Form von Kurzfassungen zu, erweitert dieses Wissen anhand automatisiert ableitbarer Parameter, wie Umschreibendes Rechteck, und gibt die Inhalte in standardisierten Formaten u. a. für INSPIRE und OGD frei.

**Schlüsselwörter:** Metadaten, Geodateninfrastrukturen, GDI, Geoportal, Offene Daten

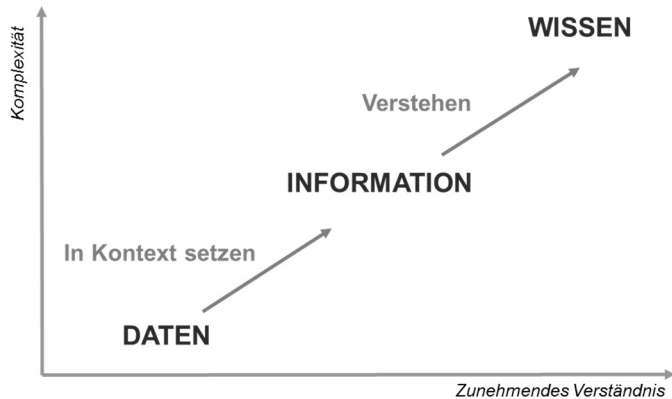
**Abstract:** *Datasets for Open and Big Data need efficient approaches to generate additional value of data. Additional value may be created by structured metadata, which transform data to information. Especially the initial collection of metadata can be time consuming. Consequently, the present research work is devoted to the semi-automated metadata generation of spatial data sets according to profil.AT 3.x and their provision for INSPIRE and OGD. Our approach is based on the knowledge of experts, which is incorporated in the abstracts of datasets. Further, it extends this knowledge by means of automatically extractable parameters such as bounding box and publishes the contents in standardized formats for e. g. INSPIRE and OGD.*

**Keywords:** *Metadata, spatial data infrastructures, SDI, geoportal, open data*

## 1 Einleitung

Im Zuge von Open und Big Data werden Metadaten immer wichtiger, denn erst wenn Daten mithilfe von strukturierten Metadaten in einen Kontext gesetzt werden, werden Daten zu Informationen und erhalten einen Mehrwert (Bellinger 2004; Gantz & Reinsel 2011). Diese Information liefert wiederum die Grundlage für das Verstehen von Inhalten, wodurch Daten letztendlich zu Wissen werden (vgl. hierzu und folgend Abb. 1). Mit einem zunehmenden Verständnis von Inhalten steigt jedoch auch die Komplexität dieser deutlich an. Um einen Überblick über die vorhandenen Daten zu erhalten, hat sich in der Domäne der Geographie das Konzept der Geodateninfrastrukturen (GDIs) etabliert. In GDIs werden Kataloge (Geoportale) zur übersichtlichen Verwaltung und Organisation von räumlichen Metadaten eingesetzt. Allerdings sind noch immer sehr viele Daten nicht zur Nutzung in Katalogen erfasst.

Hierbei gilt es v. a., den Erfassungsprozess so einfach und transparent wie möglich zu gestalten. Dementsprechend schlagen wir einen semi-automatisierten Metadatenerfassungsprozess vor, bei dem ein Brückenschlag zwischen Fachabteilungen und IT stattfindet.



**Abb. 1:** Semi-automatisierte Metadatenextraktion und Bereitstellung

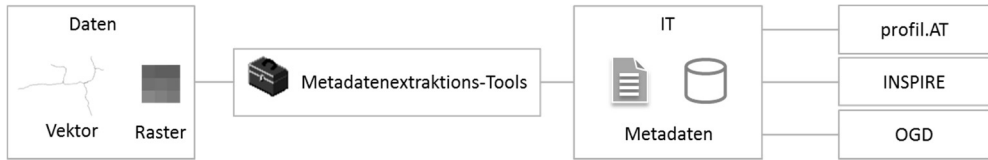
Damit Kataloge für Endnutzer sinnvoll verwendbar sind, benötigen sie einen gewissen Umfang an Metadaten, die durchsucht werden können. Diese gilt es, möglichst einfach und zeiteffizient zu erfassen.

## 2 Herausforderung

Metadaten zu erfassen benötigt gerade bei sehr vielen Datenquellen und großen Datenmengen (Big Data) einen erheblichen Zeitaufwand. V. a. gilt es bei Metadaten auch, implizites Wissen explizit zu machen. Fachabteilungen verfügen über sehr detailliertes Wissen, jedoch liegt dieses Wissen in vielen Fällen nicht als Metadaten direkt an den Daten vor. Andererseits sind Experten in Fachabteilungen in vielen Fällen nicht mit XML-Strukturen zur standardisierten Metadatendokumentation vertraut. Dementsprechend bedarf es inhaltlich sowie technisch einer Brücke, um das Fachwissen aus den Fachabteilungen in die IT zu überführen und dort bei Bedarf um technische Parameter, sowie Erfordernisse kommend aus nationalen und internationalen Standards und Normen, zu erweitern.

## 3 Lösungsansatz

Als Lösungsansatz zum Brückenschlag von Fachabteilungen zu IT schlagen wir eine semi-automatisierte Metadatenextraktion vor. Technisch gesehen wird hierfür ein Tool bereitgestellt, das in Fachabteilungen erhobene Metadaten extrahiert, in der IT gespeichert und standardisiert national (profil.AT 3.x – Mittlböck et al., 2016 / ÖNORM A 2270:2010/2017) und international (INSPIRE) bereitstellt. Zusätzlich können die Metadaten auch für Open Government Data (OGD) geöffnet werden (Abb. 1).



**Abb. 2:** Semi-automatisierte Metadatenextraktion und Bereitstellung

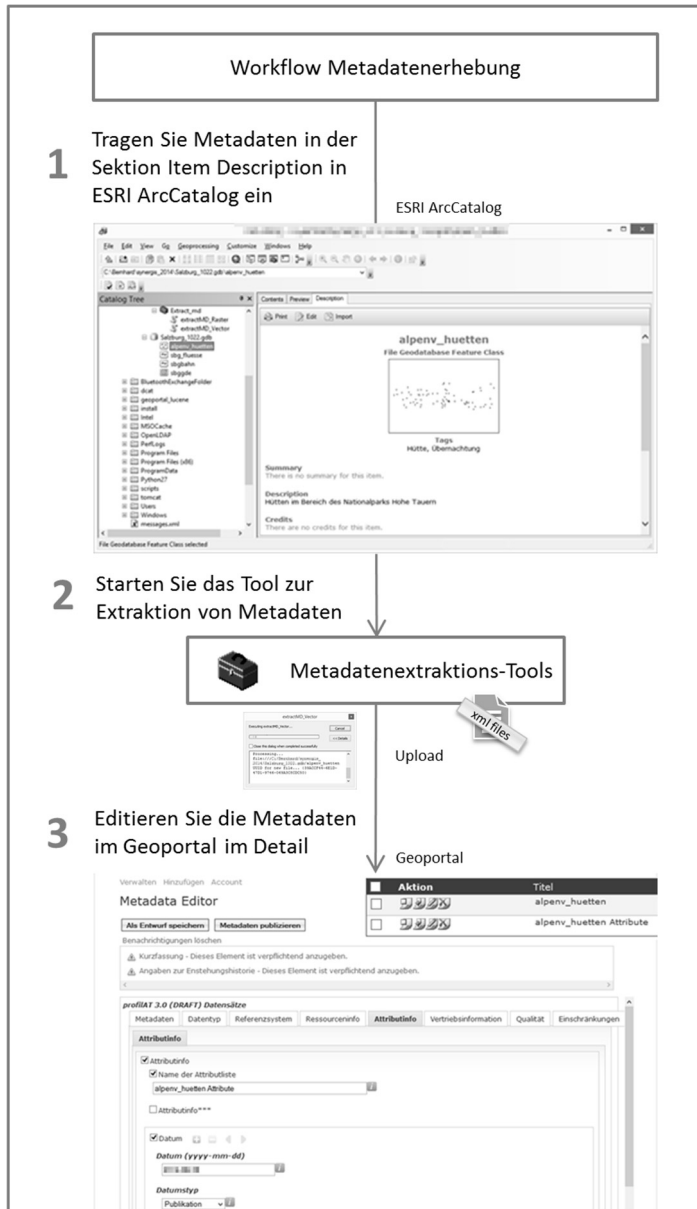
Der Brückenschlag erfolgt in drei Schritten: Schritt 1 beinhaltet Metadatenabschnitte, die direkt von und in den Fachabteilungen erhoben werden können. Dies umfasst z. B. Kurzfassung und Verwendungszweck. Automatisierbar ableitbare Parameter wie z. B. Umschreiben des Rechteck (Bounding Box), Koordinatensystem, Attribute usw. werden automatisiert in Schritt 2 extrahiert, in ein ISO-Schema nach profiL.AT 3.x überführt und in einem Katalog zur Verwaltung und Organisation von räumlichen Daten registriert. In Schritt 3 können die Metadaten dann im Katalog zusätzlich ergänzt werden, um vollständig nationalen (profiL.AT 3.x / ÖNORM A 2270:2010/2017) und internationalen (ISO 19115:2003/2006, ISO 19115-1:2014, INSPIRE) Normen und Standards zu entsprechen. Optional können die Metadaten auch als OGD Metadaten 2.3 (OGD 2015) exportiert werden.

## 4 Implementierung & Validierung

Zur prototypischen Implementierung und Validierung des vorgestellten Ansatzes wird ESRI ArcGIS for Desktop als GIS und Geoportal Server als Katalogdienst verwendet. Im Detail präsentiert sich die Vorgehensweise wie folgt (siehe auch Abb. 2):

In einem ersten Schritt werden die Metadaten von Fachexperten in der Item Description in ArcCatalog eingetragen. Dies umfasst einige wenige Elemente wie eine Kurzfassung, Stichwörter und Verwendungszweck, die vom jeweiligen Domänenfachpersonal (z. B. Wasserwirtschaft, Raumplanung) eingetragen werden können. In diesem Fall ist kein spezielles Know-how über Metadaten, XML-Schemas und Kataloge vonnöten.

Im zweiten Schritt wird das Tool zur semi-automatisierten Metadatenextraktion (SaMe) gestartet, welches die Inhalte in ein XML-Format nach profiL.AT 3.x transformiert und in einem Katalogdienst (im vorliegenden Fall Geoportal Server) hochlädt. Das hierzu verwendete Python Script nützt ausschließlich Python-Standardbibliotheken und zusätzlich arcpy (Esri ArcGIS Bibliothek für Python). Für räumliche Ressourcen finden folgende Metadatenstandards Anwendung: ISO 19110:2005/2016 (Feature Catalogue zur Attributdokumentation), ISO 19115:2003/2006 (Geographic information – Metadata) bzw. ISO 19115-1:2016 (Geographic information – Metadata – Part 1: Fundamentals), ISO 19119:2005/2016 (Geographic information – Services) und ISO 19139:2007 (Geographic information – Metadata – XML schema implementation).



**Abb. 3:** Workflow zur semi-automatisierten Extraktion von Metadaten

Durch das SaMe-Tool können derzeit jegliche Metadaten von Raster- und Vektordaten (File Geodatabase, Enterprise Database/SDE, Shapefiles) aus Esri ArcGIS for Desktop extrahiert werden. Die extrahierten Metadaten werden von Esri ISO zu dem ISO 19110:2005/2016 und ISO 19115:2003/2006 bzw. der ISO 19115-1:2016 umfassenden ÖNORM A 2270:2017 transformiert und in einen Katalogdienst (Geoportal Server) hochgeladen.

ISO 19115:2003/2006 und ISO 19110:2005/2016 Dokumente werden mithilfe von IDs miteinander verknüpft und können im Geoportal als zueinander zugehörig in den Suchergebnissen leicht aufgefunden werden. Zusätzlich zu den Attributen lassen sich auch codierte Wertelisten (Coded Value Domains) extrahieren. Durch die Speicherung des Datensatz-Pfades ist es u. a. durch ein ArcMap Add-in möglich, Datensätze direkt im Katalog zu suchen und via Klick darauf in ArcMap zum Table of Contents hinzuzufügen. Zur Registrierung mit Geoportal Server wird die CSW-Schnittstelle verwendet. Die Verwendung der standardisierten CSW-Schnittstelle hat den Vorteil, dass bei dem von uns präsentierten Ansatz auch alternative Kataloge wie z. B. Geonetwork als Katalogdienst genutzt werden können.

Für den optionalen Upload in CKAN (Portal für Open Data) wird die Bibliothek ckanapi in Python verwendet. Durch einen integrierten Scheduler mit einem frei definierbaren Intervall (z. B. jede Stunde, einmal am Tag), können die Metadaten von Geoportal Server als ISO 19115-1:2016/19139:2007-XML in Form von profil.AT 3.x ausgelesen und in einem CKAN-Format ausgegeben werden. Alternativ könnte auch die DCAT-Schnittstelle für den Informationsaustausch genutzt werden, welche jedoch den Nachteil hat, dass die Elemente nicht so gezielt zugematcht werden können, wie mit dem SaMe-Tool.

Personen, die speziell mit Metadatenstandards vertraut sind, können im dritten Schritt durch Nacheditieren einiger weniger Metadaten eine vollständige ISO- und profil.AT-Konformität erreichen.

Bei nachträglichen Updates dient die Metadaten-ID dazu, die Metadaten direkt an den Raster- und Vektordaten mit jenen im Katalog zu synchronisieren. Der österreichische Metadatenstandard profil.AT 3.x ist semantisch mit OGD Metadaten 2.3 abgestimmt, wodurch die Metadaten direkt nach OGD transformiert werden können. Damit lassen sich die Metadaten auch direkt z. B. für data.gv.at bzw. opendataportal.at freischalten, wodurch sie sich einerseits von den Fachabteilungen selbst, andererseits von einer deutlichen breiteren Nutzergruppe auffinden lassen.

## 5 Fazit & Ausblick

Mithilfe des vorgestellten Ansatzes können Metadaten semi-automatisiert im Zusammenspiel von Fachabteilungen und IT-Standard-Experten erfasst werden. Das Verfahren in drei Schritten ermöglicht den Miteinbezug der Expertisen beider Fachdomänen und ermöglicht den Brückenschlag von Fachabteilung hin zu IT. Die in der IT katalogisierten Metadaten können wiederum von Fachabteilungen dazu genutzt werden, einen Überblick über die Datenbestände zu behalten und diese effizient wiederaufzufinden.

In weiterer Folge gilt es, unseren Ansatz hinsichtlich Effizienz und Zeitersparnis zu validieren. Erste Ergebnisse aus Befragungen lassen bereits darauf schließen, dass sich der Ansatz sehr gut in die tägliche Praxis in Verwaltungsprozessen integrieren lässt.

## Danksagung

Das Tool SaMe wurde in Zusammenarbeit von RSA iSPACE und dem Amt der Salzburger Landesregierung (Abteilung 10: Wohnen und Raumplanung, Referat 10/04: Geodateninfrastruktur und Landesamtsdirektion – Fachgruppe 0/2: Landesinformatik) entwickelt.

## Literatur

- Bellinger, G. (2004). *Knowledge Management – Emerging Perspectives*. Retrieved Jan 11, from <http://www.systems-thinking.org/kmgmt/kmgmt.htm>.
- Gantz, J., & Reinsel, D. (2011). *Extracting Value from Chaos*. Retrieved Oct 17, 2011, from <http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf>.
- ISO (2003/2006). *ISO 19115: Geographic information – Metadata*.
- ISO (2005/2016a). *ISO 19119: Geographic information – Services*.
- ISO (2005/2016b). *ISO 19110: Geographic information – Methodology for feature cataloguing*.
- ISO (2007). *ISO 19139: Geographic information – Metadata – XML schema implementation*.
- ISO (2014). *ISO 19115-1: Geographic information – Metadata – Part 1: Fundamentals*.
- Mittlböck, M., Grillmayer, R., & Vockner, B. (2016). *profil.AT 3.02*. Retrieved Jan 11, 2017, from <http://data.opendataportal.at/dataset/e6ed78b0-aaaf-4035-bc87-f38083db03be/resource/bf6408b7-0b11-43b2-be54-391d7609a2a3/download/profilat302160811.pdf> (last update: Aug 08, 2016).
- OGD (2015). *OGD Metadaten – 2.3*. Retrieved Jan 11, 2017, from [https://www.ref.gv.at/fileadmin/\\_migrated/content\\_uploads/OGD-Metadaten\\_2\\_3\\_2015\\_02\\_19.pdf](https://www.ref.gv.at/fileadmin/_migrated/content_uploads/OGD-Metadaten_2_3_2015_02_19.pdf) (last update: Feb 19, 2015).
- ÖNORM (2010/2017). *ÖNORM A 2270:2010/2017 – Geoinformation – profil.AT – Metadatenprofil für Geoinformation – Regeln zur Umsetzung der ÖNORM EN ISO 19115*.