Seventh Framework Programme
Theme 6
Environment

Project: 603587 SWITCH-ON

Full project title:
Sharing **W**ater-related **I**nformation to **T**ackle **C**hanges in the **H**ydrosphere – for **O**perational **N**eeds

Deliverable: D2.6

Standard Information Model for Meta-Data

FINAL VERSION

Due date of deliverable: 30/06/14

Date accepted for publication/submission to EU: 27/06/14

Actual submission date: 30/06/14

Title	Standard Information Model: Definition and Description
Creator	Frank Keppel (Deltares)
Editor	Frank Keppel (Deltares)
Brief Description	This document provides a detailed and explanatory description of the Standard Information Model for Meta-Data (SIM) which constitutes an intrinsic part of the Spatial Information Platform (SIP).
Publisher	SWITCH-ON Consortium
Contributors	Pascal Dihé (CISMET), Frank Keppel (Deltares)
Type (Deliverable/Milestone)	Software Deliverable
Format	MS Word 97-2003 Document
Creation date	07/05/14
Version number	FINAL
Version date	26/06/14
Last modified by	Pascal Dihé (CISMET)
Rights	Copyright "SWITCH-ON Consortium". During the drafting process, access is generally limited to the SWITCH-ON Partners.
Audience	☐ internal ☒ public ☐ restricted, access granted to: EU Commission
Action requested	☐ to be revised by Partners involved in the preparation of the deliverable ☐ for approval of the WP Manager ☐ for approval of the Internal Reviewer (if required) ☐ for approval of the Project Co-ordinator
Deadline for approval	27/06/14

Version	Date	Modified by	Comments
0.1	08/05/14	Frank Keppel (Deltares)	Draft structure (TOC)
0.2	23/05/14	Pascal Dihé (CIS)	Updated draft
0.3	26/05/14	Pascal Dihé (CIS)	Standard Information Model
0.4	03/06/14	Frank Keppel (Deltares)	Introduction
0.5	06/06/14	Pascal Dihé (CIS)	Mapping, Find, Bind, Transform, Publish, API
0.6	10/06/14	Sascha Schlobinski (CIS)	Use cases description
0.7	11/06/14	Pascal Dihé (CIS)	Revisions merged
0.8	16/06/14	Frank Keppel (Deltares)	Remaining sections
0.9	16/06/14	Pascal Dihé (CIS), Frank Keppel (Deltares)	Pre review comments

1.0	23/06/14	Pascal Dihé (CIS)	Meta-Data Example + review
1.1	26/06/14	Pascal Dihé (CIS)	Update after internal review, + specifications of tag groups
FINAL	27/06/14	Esa Falkenroth (SMHI) Eleanor O'Rourke (SMHI)	Final edit and formatting

Table of Contents

List of Tables

Table of Figures

Abbreviations and Acronyms

Acronym	Description
API	Application Programming Interface
CKAN	Comprehensive Knowledge Archive Network
CSW	Catalogue Service for the Web
DCAT	Data Catalogue Vocabulary
DoW	Description of Work
EU	European Union
FP7	7th Framework Programme (EC)
FTP	File Transfer Protocol
GIS	Geographic Information System
HDF	Hierarchical Data Format
HTML	HyperText Markup Language
HTTP	Hypertext Transfer Protocol
I/O	Input & Output
INSPIRE	Infrastructure for Spatial Information in Europe
ISO	International Organization for Standardization
JSON	JavaScript Object Notation
MIME	Multipurpose Internet Mail Extensions
OGC	Open Geospatial Consortium
OPeNDAP	Open-source Project for a Network Data Access Protocol
PFB	Publish-Find-Bind
PostGIS	Spatial and Geographic Objects for PostgreSQL
RDBMS	Relational Database Management System
REST	Representational State Transfer
RMI	Remote Method Invocation
SIM	Standard Information Model
SIP	Spatial Information Platform
SQL	Structured Query Language
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
UUID	Universally Unique Identifier
W3C	World Wide Web Consortium
WebDAV	Web-based Distributed Authoring and Versioning
WFS	Web Feature Service
WP	Workpackage
WPS	Web Processing Service
XML	eXtensible Markup Language

Executive Summary

This document, deliverable D2.6, has been produced by the consortium of the European Project FP7-603587 “Sharing Water-related Information to Tackle Changes in the Hydrosphere – for Operational Needs” (SWITCH-ON). It is a more detailed and explanatory description of the Standard Information Model for Meta-Data (SIM) which constitutes an intrinsic part of the Spatial Information Platform (SIP). The SIP is defined in Deliverable D2.1 where the SIM is mentioned on a less detailed level.

This document provides insight in how the SIM is structured to be able to meet its requirements by:

- defining the entities and relations that constitutes the relational (standard information) model
- demonstrating the mapping between the SIM and various meta-data profiles
- explaining how the SIM supports find, bind, transform/repurpose and publish

1 Introduction

1.1 Introduction to the Standard Information Model

For an open data repository to be useful to a broad range of potential users it has to supply meta-data describing the data it holds. To store and manage meta-data a Meta-Data Repository (deliverable D2.1, section 3.3.1) as part of such an open data repository is a necessary component. For the open data repository to be conceived in the SWITCH-ON project, this necessary component is the Standard Information Model for Meta-Data (SIM).

1.2 Purpose of this document

This document is a detailed description of the Standard Information Model for Meta-Data which is an intrinsic part of the Spatial Information Platform (SIP) which is being conceived in the SWITCH-ON context. This document should provide enough detail on how the SIM will function in a production environment.

1.3 Intended audience

This document targets foremost the SWITCH-ON consortium members. On the one side the members responsible for implementing the SIP with SIM and on the other side the members in work packages 3 and 4 who are primarily on the usage side of the SIM of find, bind, publish and repurpose.

Otherwise any individual with affinity to open data and meta-data might find material of interest in this document

1.4 Structure of the document

Chapter 1 sketches the purpose, scope and context of the subject (Standard Information Model for Meta-Data). Chapter 2 introduces the SIM and its basic concepts and describes its requirements from SWITCH-ON WP3 and WP4 perspective, defines the technical details of its (relational) model and explains the mapping between the SIM relational model and meta-data profiles. Finally it explains how the SIM supports find, bind, repurpose/transform and publish and how its meta-data content is accessible to exercise find, bind, repurpose/transform and publish.

1.5 Relation to other Documents

Deliverable D2.1, “Spatial Information Platform and Application Programming Interface” defines the overall architecture of the SIP for which the SIM is an intrinsic part.

Deliverable D3.1, “Report on data requirements for a Virtual Water-Science Laboratory” defines data and meta-data requirements for WP3 activities. The SIM should be able to comply with the WP3 meta-data requirements.

Deliverable D4.2, “Functional and technical design of the products” describes the WP4 models and applications to be developed in SWITCH-ON context. These products will interact with the SIP and the SIM by using and/or providing meta-data.

1.6 Relation to other work packages

Work Package 3, the Virtual Water-Science Laboratory, will conduct collaborative hydrological experiments in the course of the SWITCH-ON project. By doing so they will both generate data and meta-data for storage in the SIP and SIM as well as searching and (re)using data from the SIP by means

of meta-data from the SIM. Work Package 3 largely defines the meta-data profiles that the SIM should be able to store and make available for find, bind, transform/repurpose and publish purposes.

Work Package 4 defines and implements products and services on top of the SIP (and SIM). These products and services will operate on the (repurposed) data and meta-data that the SIP/SIM provides. They will be mostly data consumers (search and bind).

2 Standard Information Model

2.1 Introduction

In accordance to the Open Data Repurpose Pattern (see D2.1, chapter 3), the main purpose of the SIM is to provide information about how to find, bind, transform and publish open data (Figure 1). Introducing new data models and data formats for open data is beyond the scope of this model. Although the Standard Information Model is generally not directly concerned with the storage of open data it provides information for accessing, interpreting, processing, etc. these data. Typically, such information about data is called meta-data.

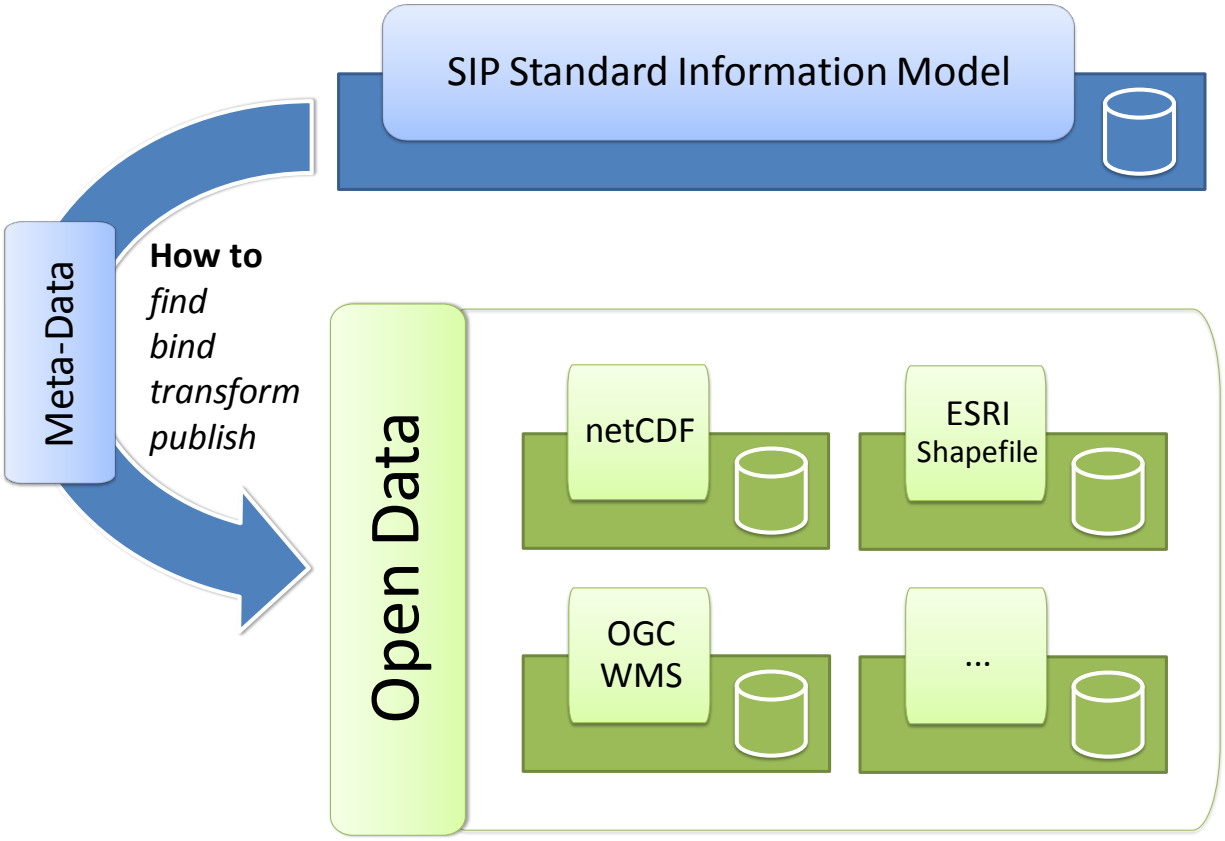


Figure 1: Relation of the Standard Information Model to Open Data

In SWITCH-ON, open data that is described by the Standard Information Model in terms of meta-data refers to the following three main data categories (Figure 2):

- **Public Open Data**
  General Open Data harvested from external data catalogues that can be found (*find*) and made available (*bind*) by the SIP.
- **SWITCH-ON Experiment Input Data**
  Open Data that has been pre-processed (*transform / repurpose*) to become suitable input data for an experiment in the WP3 Virtual Water-Science Laboratory or a WP4 product.
- **SWITCH-ON Experiment Result Data**
  The result of an experiment, including the actual result data and the protocols which are made publicly available through the SIP (*publish*).

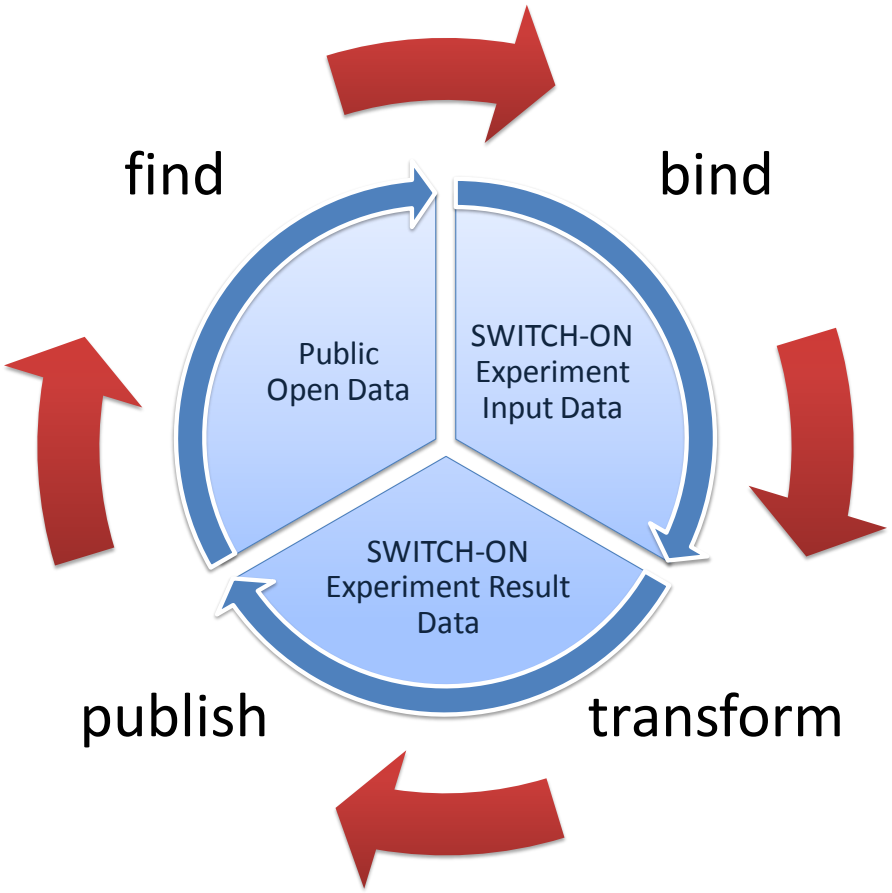


Figure 2: Open Data Repurpose Pattern

Although widely adopted information models for the description of data and services do exist (e.g. ISO19115 (2003) and ISO 19119 (2005), the Standard Information Model of the SIP is not solely based on one of these standards. Instead of defining one fixed information model that is based on a selection of particular meta(data) standards or profiles, the Standard Information Model of the SIP has been tailored to the actual information needs of the SIP, auxiliary services, and tools as well as its end users (product developers and researchers working in the virtual water-science lab). Thereby, the concepts of the CKAN (Comprehensive Knowledge Archive Network) domain model as well as support for meta-(data) standards like Dublin Core, ISO 19115, etc., have been considered in the design of the SIM.

Figure 3 lists the following three main meta-data categories of the SIM. A more detailed description of these meta-data categories and how they relate to find, bind, transform and publish is provided in section 2.6.

- **Basic Metadata**
  Basic or general Metadata refers to a meta-data schema based on an ISO 19115 profile. ISO 19115 has been chosen because it is also used by GEOSS project and it is the standard recommended by the INSPIRE directive. The ISO 19115 profile gives already enough information to provide comprehensive find functionality and to support bind, repurpose and publish. Furthermore, it is compliant to the Dublin Core and CKAN domain model, and Data Catalogue Vocabulary of the W3C.
- **Meta-Data for Data Quality**
  Meta-Data for data quality refers to any type of information that is useful to estimate uncertainty of a dataset. Thereby it has to be noted that quantitative estimate of data uncertainty, if available, is provided jointly with the data. The meta-data schema for data

quality has been inspired by the ISO norms ISO 19113:2002 and ISO/TS 19138:2006. These norms refer to geographic information but are to large extent relevant to other data types.

- **Lineage Meta-Data**
  Lineage meta-data is used to describe data transformation and processing which is in many ways the information that will be available in the protocol of an experiment. The meta-data schema for data quality is currently based on the protocols of the first experiments of WP3.

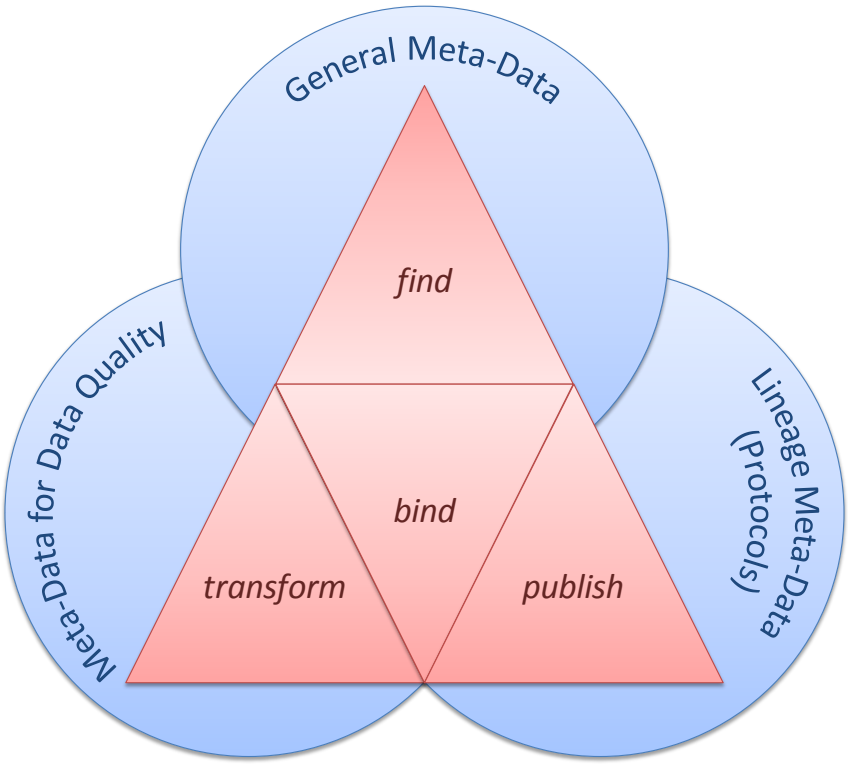


Figure 3: Meta-Data Categories

Since the development of the SIP follows an iterative agile approach, also the definition of meta-data and their relevance and usage for find, bind, transform and publish are subject to change. As a consequence, the SIM has to have the ability to adapt to such changes without causing a redesign of components that rely on the model. The design of the SIM follows therefore a graduated approach with the following three different levels of increasing extensibility and flexibility shown in Figure 4:

- **Relational Model**
  The relational model defines the outline for an object relational database model and supports the core business processes of the SIP.
- **Dynamic Tag Extensions**
  Dynamic tag extensions augment the relational model by user definable code lists and thus provide a simple yet powerful extension mechanism.
- **Dynamic Content Extensions**
  Dynamic Content Extensions form a mechanism to dynamically inject complex structured or semi-structured content in the SIM without the need to change the relational model.

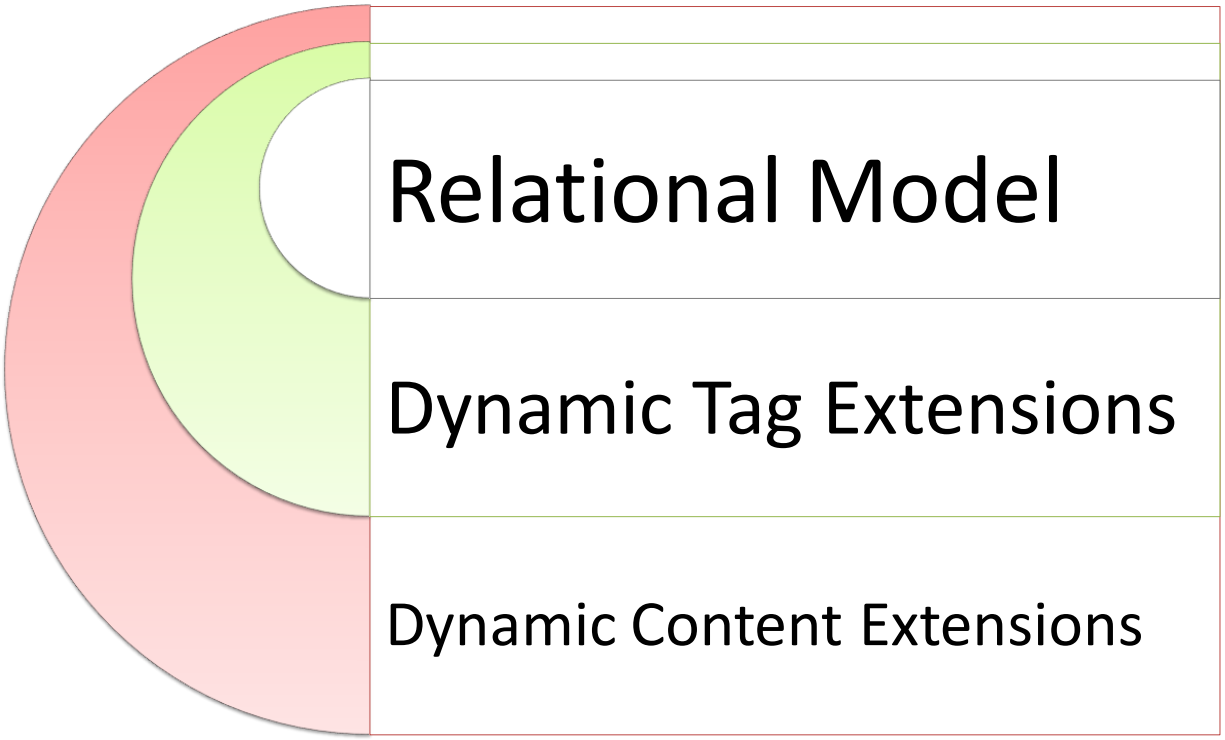


Figure 4: Layers of the Standard Information Model

Those three layers of the Standard Information Model as well as their mapping to three categories of meta-data are described in detail in the subsequent chapters.

2.2 Requirements

In principle, the general information modelling requirements on the SIM are set by the general requirements on the SIP expressed deliverable D2.6. Thereby, support for standards-based meta-data, search and cataloguing and ability to link with existing external catalogues of open data play a predominant role. However, since the development of the SIP follows an iterative agile approach, detailed functional requirements (formulated as “user stories”) on the SIP and the information model, respectively, are subject to change. As a consequence, the Standard Information Model has to have the ability to adapt to such changes without causing a redesign of components that rely on the model.

The most important requirements on the Standard Information Model and the respective section of this document which explain how those requirements have been addressed by the SIM are:

- Consume and possibly to feed-back meta-data from existing catalogues: sections 2.6.2.4, and 2.6.3.1.
- Support for publishing meta-data in a web crawler friendly format so that it can be indexed by major search engines: section 2.6.2.4 and 2.6.4.
- A flexible information modelling approach which support different (meta-)data formats and standards: sections 2.6 and 2.3.4.
- Support for the documentation of scientific analyses, the description of software (tools), aspects related to re-repurposing open data: section 2.3.3, 2.6.2.3 and 2.6.3.3.
- Support for inherent features for cataloguing: sections 2.6.3.4 and 2.6.4.4.
- Store and provide “basic” meta-data, sections 2.3.1, 2.6.1 and 2.6.2.1.
- Store and provide “quality” meta-data: section 2.3.4, 2.6.1 and 2.6.2.2.
- Store and provide “lineage” meta-data: sections 2.6.2.3, 2.6.1 and 2.6.2.3.

2.2.1 Use cases supporting WP3 and WP4

The current status of WP3 requirements on the SIM can be capture in the following three core user stories:

1. As researcher doing collaborative hydrology, I want to look at (find and inspect) open data sets relevant for our experiment.
   a. As researcher doing collaborative hydrology, I want to look at (find and inspect) daily precipitation amount and daily mean temperature from open data-sets so I can decide if I can use the data in our experiment.
   b. As researcher doing collaborative hydrology, I want to look at (find and inspect) water shed behaviour (aggregated area, Mean elevation, mean temperature, CORINE Land Cover[1], etc from open data-sets so I can decide if I can use the data in our experiment.
2. As researcher doing collaborative hydrology, I want to download open data relevant for my research.
3. As researcher doing collaborative hydrology, I want to upload/publish the result of my collaborative research as open data so it can be found by others.

In the context of the SIM this can be expressed as:

(1) Identify relevant data sets by representative aggregated variables such as mean values.
(2) Directly access the identified data.
(3) Publish research result in way so it can be identified as in (1).

From this we can derive the requirement on the SIM to be able to represent dynamic content in the form of aggregated information and highlight the actual type of this information so a specific search and/or indexing algorithm can rely on the existence of proper information (as described in a content type). As the type of aggregated information and the corresponding purpose is heavily use case/open data dependent it is admissible to use a combination of the dynamic content extension concept as described in section 2.5 and the concept for dynamic tag extension as described in 2.4.

Figure 5 shows how the SIM supports alternate representations of an open data resource. Besides the main representation which refers to the actual resource data as described in section 2.3.2, several alternate representations which enable advanced preview and search functionality can be associated with resources.

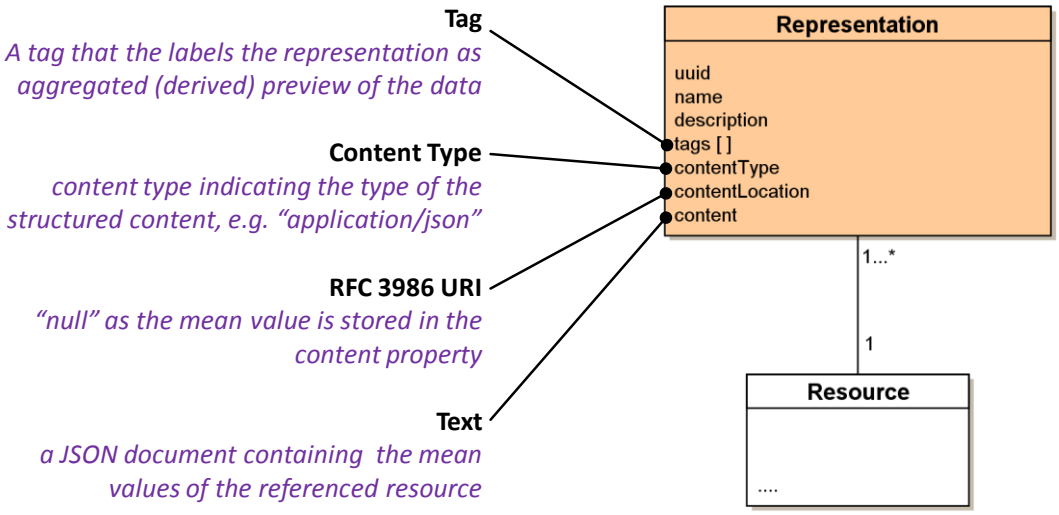


Figure 5: Relational Model Overview

[1] http://www.eea.europa.eu/publications/COR0-landcover

For a specific type of search on aggregated data we can create a content type that is described by the triple (contentType, contentLocation, content) and a tag so that a e.g. a search algorithm can rely on the corresponding syntax (e.g. structure and type of the information).

A corresponding use case for finding open data that takes structured meta-data into account is described in section 2.6.3.1.

2.3 Relational Model

The first layer of the Standard Information Model is an object relational database model which simultaneously defines the outline of the two subordinate layers. Besides basic classes for resources, their relationships and representations, the model also defines a set of attributes needed to describe those classes. The relational model is also the basis for both the internal (SIP Expert Catalogue Management APIs) and external (Published Catalogue Access APIs) catalogues of SIP.

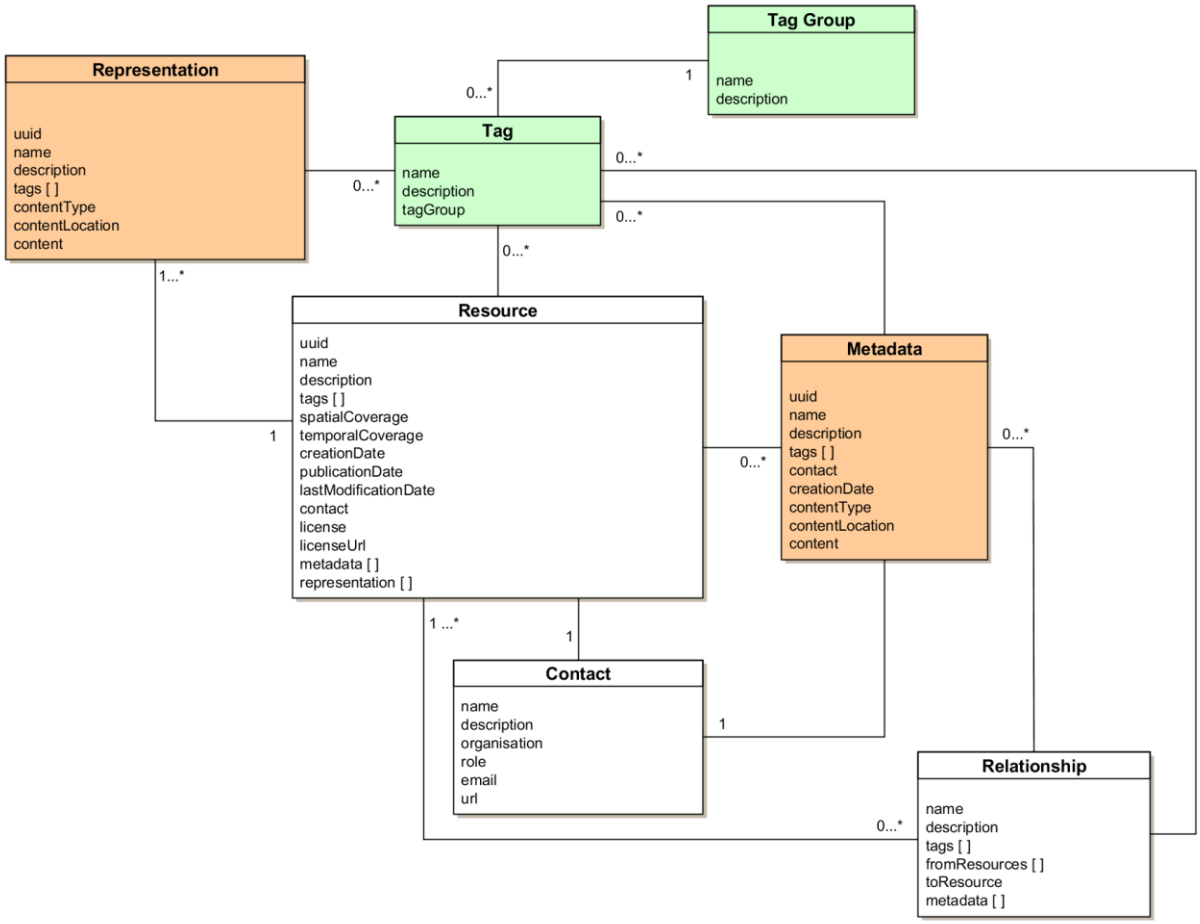


Figure 6: Relational Model Overview

Figure 6 shows a general overview of the relational model. As anticipated in the SWITCH-ON DoW (DoW, 2013), the model is implemented as schema of a RDBMS which is part of the Meta-Data Repository. In the following, the classes of the model, their conceptual background and their realisation as part of a relational data base model are briefly explained:

2.3.1 Resource

Resources are the central entity of the data model and the catalogue, respectively. They logically describe a dataset, a service, a tool, etc. Resources have a set of core attributes that, among other

things, define the spatial and temporal coverage of the resource. Further attributes cover information about the description, owner, version, etc., of the resource. Besides the basic attributes of a resource which have been derived from the basic metadata categories as defined in section 2.6, arbitrary additional metadata can be associated with a resource.

Table 1 provides an overview of the definition of the Resource class in a relational database schema. It contains information on

- primary keys (PK),
- the name of the table columns (Attribute Name),
- the data type of the respective values of the columns (Attribute Type), and
- the name of referenced table (FK Reference Table) in case values of the column are foreign keys.

Table 1: Resource Table Definition

PK	Attribute Name	Attribute Type	FK Reference Table
True	id	integer	
False	uuid	varchar	
False	name	varchar	
False	description	text	
False	tags	integer	jt_resource_tag
False	spatialcoverage	integer	geom
False	fromdate	integer	
False	todate	timestamp	
False	creationdate	timestamp	
False	publicationdate	timestamp	
False	lastmodificationdate	timestamp	
False	contact	integer	contact
False	representation	integer	jt_resource_representation
False	license	integer	tag
False	metadata	integer	jt_metadata_resource

Figure 7 shows a detailed diagram of the Resource table which includes also all relevant reference tables (prefixed with JT_).

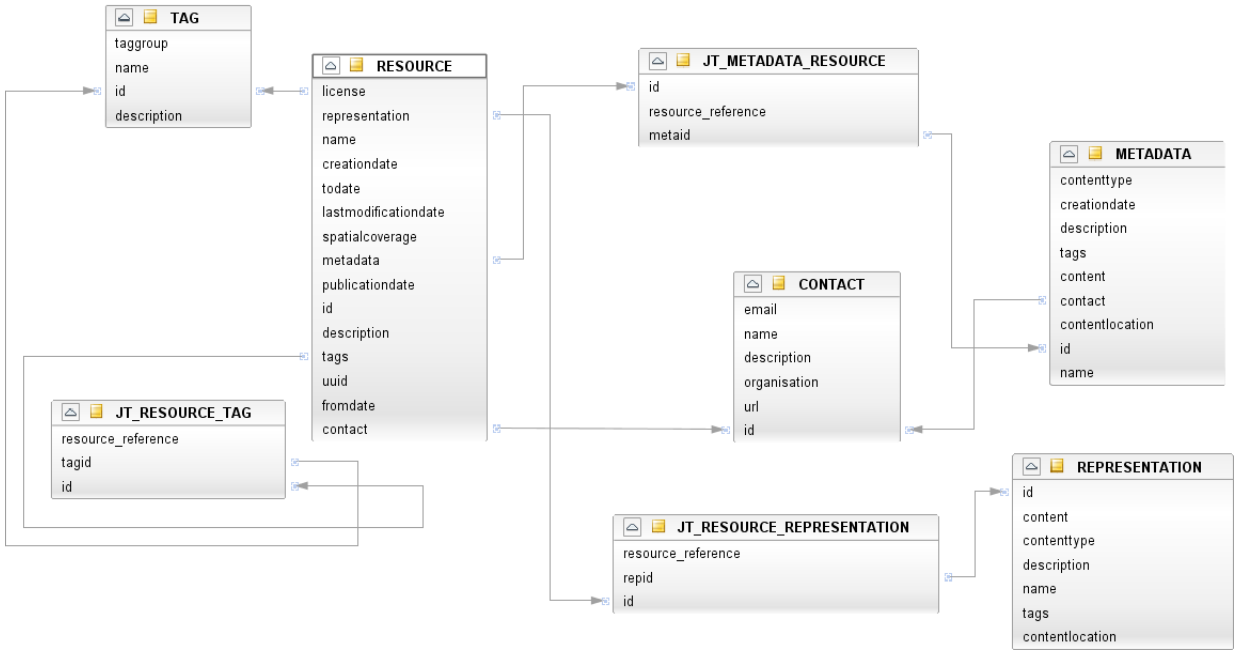


Figure 7: Resource Diagram with Relationships

Please note that the mandatory attribute “id” (primary key) is automatically generated by RDMBS and uniquely identifies a resource in the context of the SIP only. The optional “uuid” attribute on the other hand is a globally unique identifier which has been assigned by the original provider of the resource.

2.3.2 Representation

This is a description of the “physical” representation of a resource. A resource may have different representations. For example, a dataset may be available in different formats or an aggregated representation of the data for enabling preview and search functionality has been derived (see section 2.2.1).

Among other roles, attributes of the representation define how to actually access a particular representation or instance of a logical resource (*bind*). As will be explained later, the representation class is part of the dynamic content extension layer of the Standard Information Model.

Table 2: Representation Table Definition

PK	Name	Type	FK Reference Table
True	id	integer	
False	name	varchar	
False	description	text	
False	tags	integer	jt_representation_tag
False	contenttype	varchar	
False	contentlocation	varchar	
False	content	text	

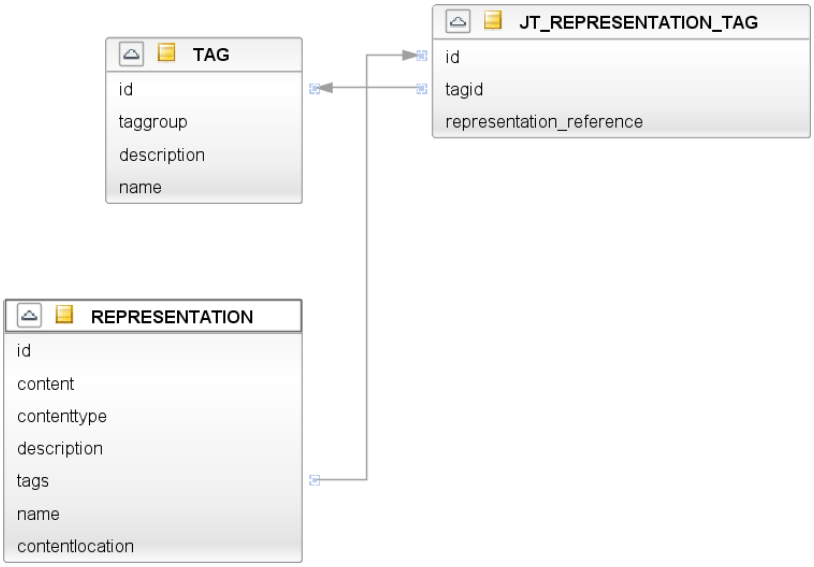


Figure 8: Representation Diagram with Relationships

2.3.3 Relationship

This class is used to specify relationships (e.g. “derived from”) between resources, e.g. to track and document data transformations and processing (lineage) of resources. The type of the relationship can be identified by a respective tag and arbitrary additional meta-data (e.g. protocols of an experiment) can be associated with a relationship.

Table 3: Relationship Table Definition

PK	Name	Type	FK Reference Table
True	id	integer	
False	name	varchar	
False	description	varchar	
False	tags	integer	jt_relationship_tag
False	fromresource	integer	jt_fromresource_relationship
False	toresource	integer	jt_toresource_relationship
False	metadata	integer	jt_metadata_relationship

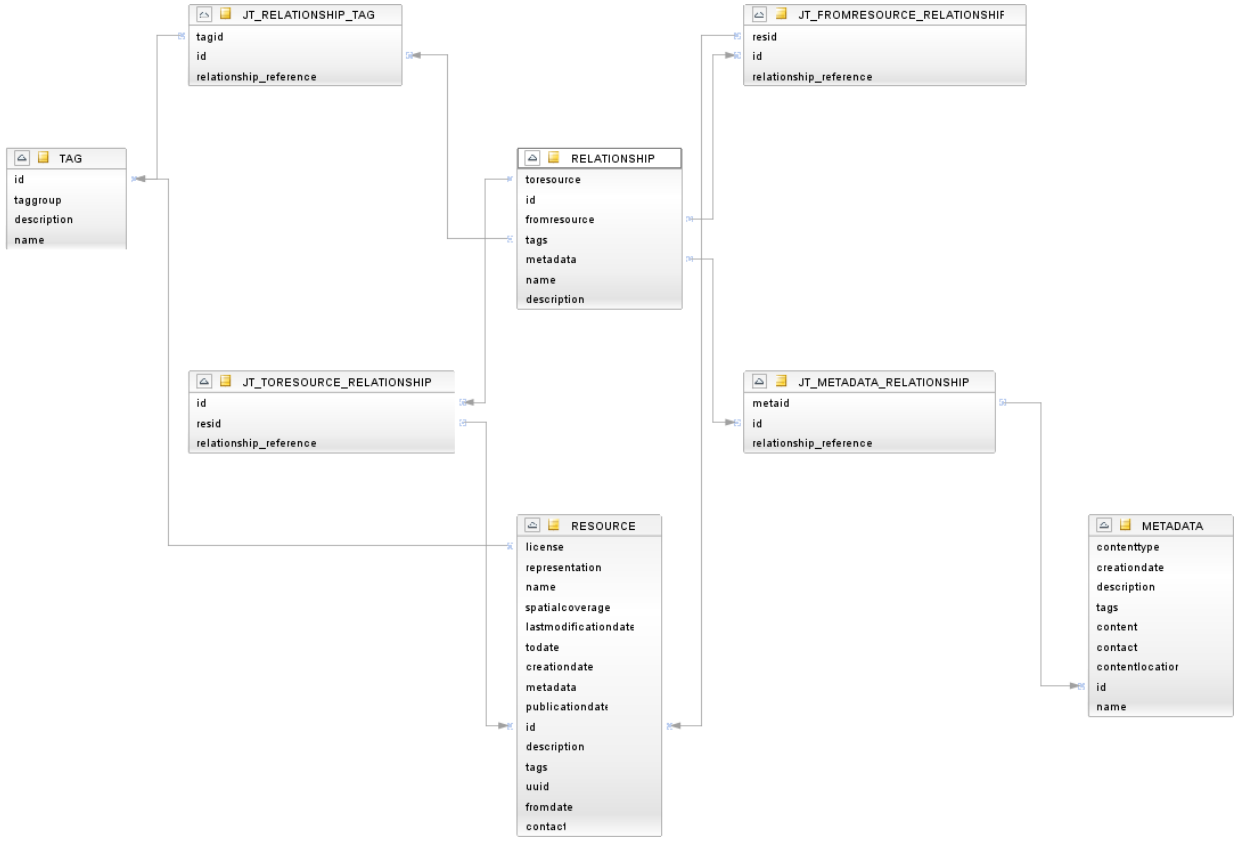


Figure 9: Relationship Diagram with Relationships

2.3.4 Metadata

The metadata class represents additional meta-data about a resource or a relationship in a structured or semi-structured format (e.g. meta-data for data quality). As will be explained later, the metadata class is also part of the dynamic content extension layer of the Standard Information Model. Additional meta-data can be associated with resources and with relationships.

Table 4: Metadata Table Definition

PK	Name	Type	FK Reference Table
True	id	integer	
False	name	varchar	
False	tags	integer	jt_metadata_tag
False	description	text	
False	contact	integer	
False	creationdate	timestamp	
False	contenttype	varchar	
False	contentlocation	varchar	
False	content	text	

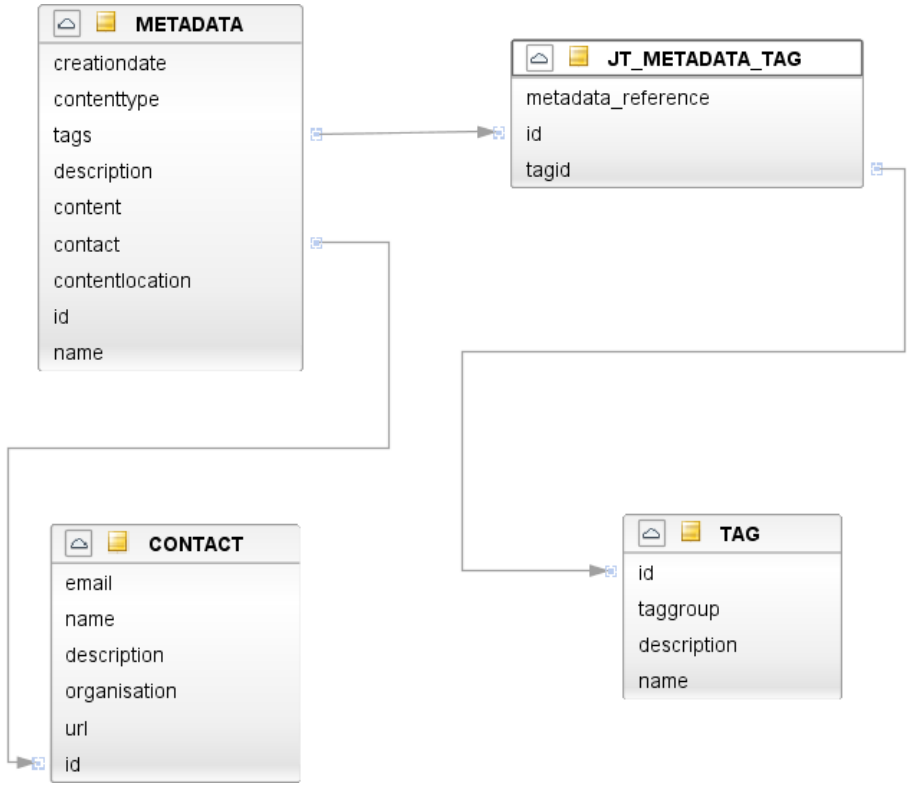


Figure 10: Metadata Diagram with Relationships

2.3.5 Tag and Taggroup

Tag and Taggroup represent the dynamic tag extension layer of the Standard Information Model. TagGroups define a general classification for tags and thus can be used to create lists of predefined tags (e.g. code lists).

Table 5: Tag Table Definition

PK	Name	Type	FK Reference Table
True	id	integer	
False	name	varchar	
False	description	varchar	
False	taggroup	integer	jt_tag_taggroup

Table 6: Taggroup Table Definition

PK	Name	Type	FK Reference Table
True	id	integer	
False	name	varchar	
False	description	varchar	

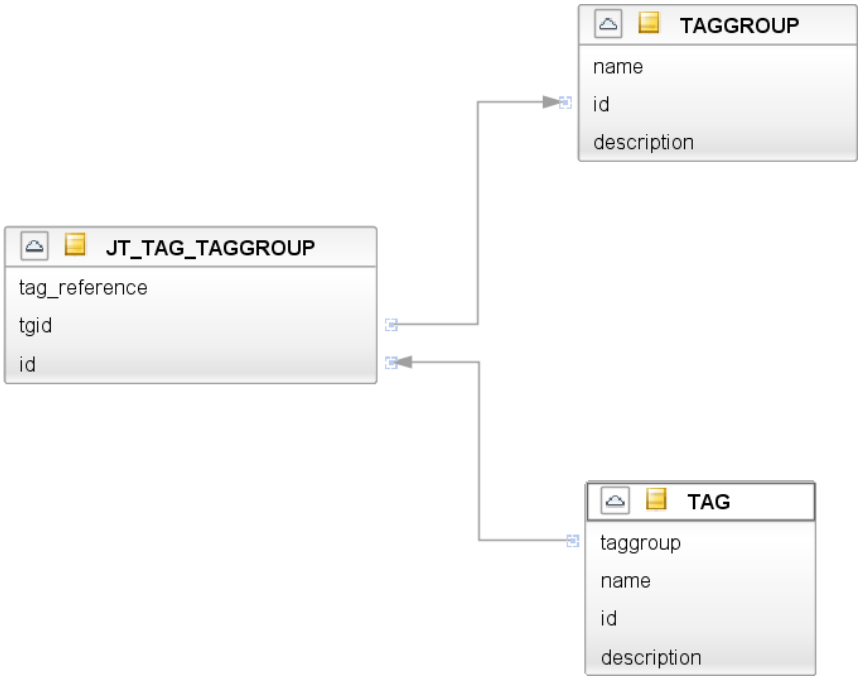


Figure 11: Tag and Taggroup Diagram with Relationships

2.4 Dynamic Tag Extension

Dynamic tag extensions provide the possibility to extend the Standard Information Model at runtime without causing changes to the relational model itself. Thus, dynamic tag extensions fit perfectly into the iterative and agile development approach followed in SWITCH-ON. Accordingly, new information-needs of SIP client applications (external catalogues, tools, etc.) can quickly be fulfilled without the need to change the internal database structure of the SIP. As described in section 2.3.5, dynamic tag extensions are represented in the relational model by tags and their associated tag groups.

Besides the possibility for extending the relational model by introducing new tag groups, tags and tag groups are mainly used to define fixed value lists like standardised topic categories, INSPIRE compliant keywords lists and so on. Thereby, an initial set of fixed tag groups and tags has been defined on basis of the default meta-data schema presented in chapter 2.6. Those lists can be used during search, e.g. allowing a user to select a specific topic category. Of course, also open tag lists are supported, e.g. to support user-defined keywords.

Figure 12 gives an overview on some important Tag Groups that have been defined and populated with tags in the current version of the SIM.

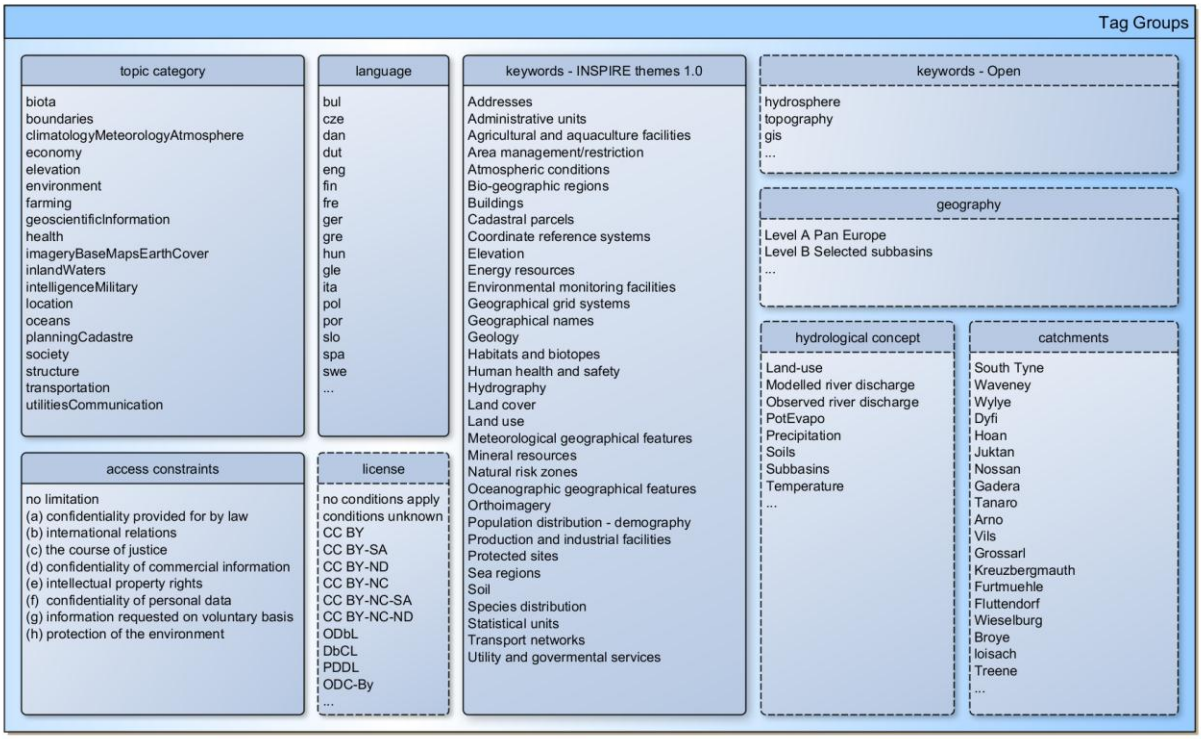


Figure 12: Examples of important Tag Groups of the SIM

In the following, a complete list of tag groups have been defined by the time of writing this document is provided:

- **access constraints**
  Limitations on public access in accordance to Article 13 of Directive 2007/2/EC (fixed group with tags from a standard code list).
- **application profile**
  Application that can be used to open or process the resource representation (open group with some predefined tags).
- **catchments**
  Catchment of a Level B dataset (open group with some predefined tags).
- **collection**
  Assigns the resource to a collection of resources for cataloguing purposes, e.g. SWITCH-ON Experiment Results, etc. (open group with some predefined tags).
- **content type**
  MIME Type of the representation (open group which several predefined tags).
- **function**
  Function that can be performed following the "contentLocation" link to the resource representation (fixed group, standard code list).
- **geography**
  Geographical classification of the resource, currently Level A and Level B are defined (open group with some predefined tags).
- **hydrological concept**
  Hydrological concept of a Level A Dataset (open group with some predefined tags).
- **keywords - INSPIRE themes 1.0**
  Keywords from Gemet – INSPIREthemes, version 1.0, publication, 2008-06-01 Hydrography (fixed group).

- **keywords - open**
  User defined to describe the subject (open group with several predefined tags).
- **language**
  The language(s) used within the resource or in which the metadata elements are expressed. The value domain of this tag is limited to the languages defined in ISO 639-2.
- **license**
  License regulating the conditions for access and use of the data (open group with some predefined tags).
- **location**
  Geographic Location of the resource (covered by the data). E.g. Continent, Country, Region, City, etc. Open Group. (open group with some predefined tags)
- **meta-data standard**
  Official standard on which the meta-data record is based (open group with some predefined tags).
- **meta-data type**
  SIP internal type of the Meta-Data Record (fixed group).
- **protocol**
  Protocol of the service that can be accessed at the "contentLocation" of the resource representation (open group with several predefined tags from standard code lists).
- **realtionship type**
  SIP internal type of the relationship between resources (fixed group).
- **representation type**
  SIP internal type of the representation of the resource (fixed group).
- **role**
  Function performed by the responsible party (fixed group, standard code list).
- **scope**
  Scope of the resource. (fixed group with tags from a standard code list)
- **srid**
  The Spatial Reference System Identifier (SRID) of the spatial coverage of the resource EPSG format (open group with some predefined tags).
- **topic category**
  High-level classification of resources in accordance with ISO 19115 for grouping and topic-based search (fixed group).

The specifications of the respective tag groups can be found in Appendix I: Specification of Tag Groups.

Another field of application for dynamic tag extensions is the creation of (dynamically constructed) catalogue structures. For example, when a specific catalogue is structured according to hydrological concepts or geography represented by tags belonging to the respective tag groups, introducing new hydrological concepts or geographical classification is just a matter of adding a new tag to these groups. An example is provided along with the find use case in section 2.6.3.1.

Further possibilities for using dynamic tag extensions include, for example, specifying the type of relationships between resources, the type of the resource representation, and so on. The definition of such additional tag groups is work in progress and will continue as the development of the SIP and related products and services progresses.

2.5 Dynamic Content Extension

Dynamic content extension is a simple mechanism for further extending the SIM by either dynamically injecting arbitrary content encoded in plain text into the model or by providing references (URIs) to externally stored content.

In the relational model (Figure 6), dynamic content extensions are represented by triples of the attributes “contentType”, “contentLocation” and “content” which are described in the following:

- **contentType**
  defines the (MIME) type of the content that is injected or referenced. Examples of content types are “text/uri” and “application/json”.
- **contentLocation**
  defines the location of the content when the content isn’t dynamically injected but is referenced. In this case, the value of the attribute “content” is set to null and the value of the attribute “contentType” is set to a content type that properly indentifies the content. The type of this attribute is a URI (RFC 3986).
- **content**
  contains the actual content encoded in plain text when the content isn’t referenced but is dynamically injected. In this case, the value of the attribute “contentLocation” should be set to null and the value of the attribute “contentType” must be set to a content type tag that properly indentifies the content. The type of this attribute is string.

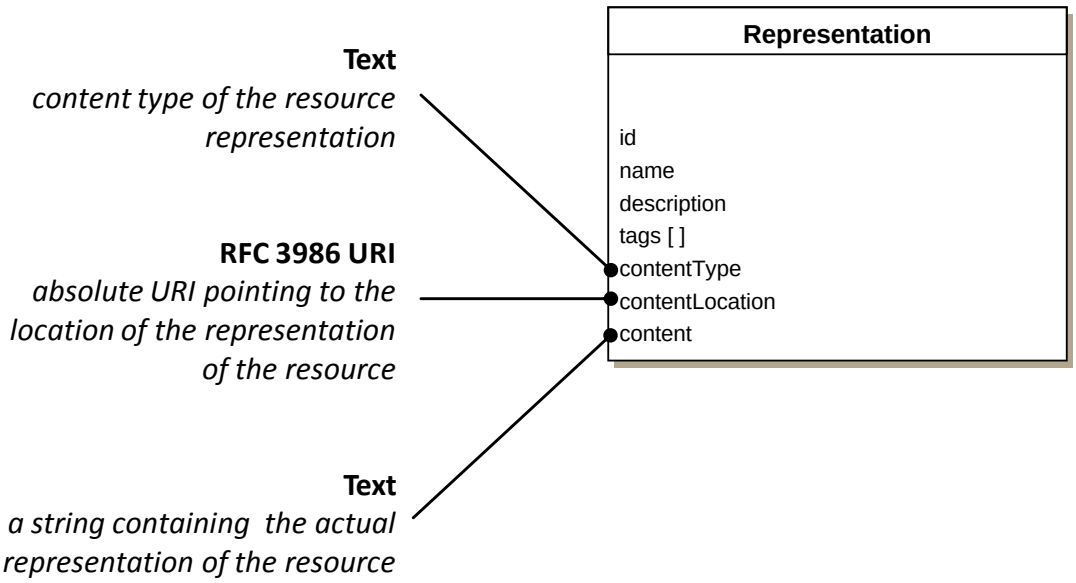


Figure 13: Dynamic Content Extension

Figure 13 provides an overview of the concept of dynamic content extension. Currently, classes of the relational model with support for dynamic content extension, and thus containing the three aforementioned attributes, are “Representation” and “Metadata”.

In the case of a resource representation (“Representation” class of the relational model), content is generally provided by reference. Thus the “contentLocation” attribute contains a URI that points to actual resource data. This may, for example, be the URI of a file on a FTP or WebDAV server, the URI of a WMS layer as well as an URI pointing to an OPeNDAP data server (e.g. a THREDDS[2] or ERDDAP[3] instance). Figure 14 shows an example of a resource that is represented by different data sets (JSON document and WMS layer) which are indentified by a URI (“content location”).

[2] https://www.unidata.ucar.edu/software/thredds/current/tds/
[3] http://coastwatch.pfeg.noaa.gov/erddap/download/setup.html

Expectations may apply for representations that contain highly aggregated data for search and visualisation purposes. In this case, the aggregated data should be stored in the “content” property in a structured format (e.g. a JSON document), so that is directly available for search and indexing (see also section 2.6.4.5).

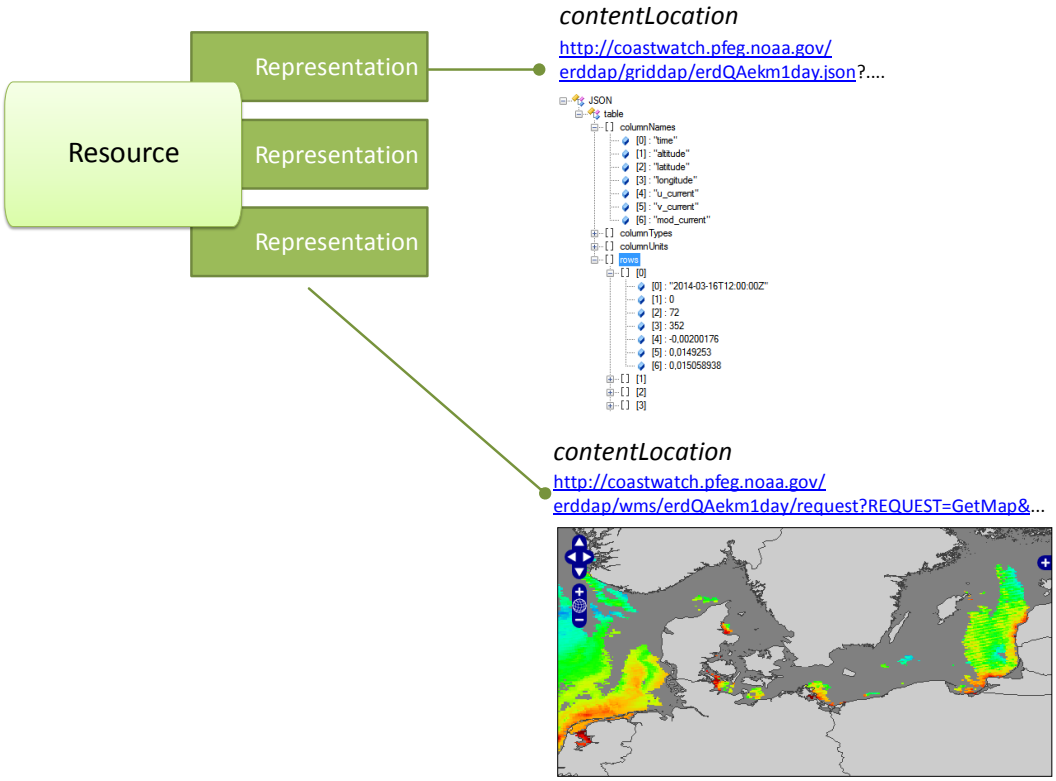


Figure 14: Different Resource Representations

In the case of resource meta-data (“Metadata” class of the relational model), content is generally provided by value. Thus the “content” attribute contains the actual meta-data encoded in some textual format, e.g. XML. This may be, for example, an ISO19115 meta-data XML document that can be directly stored in the SIP Standard Information Model as shown in Figure 15.

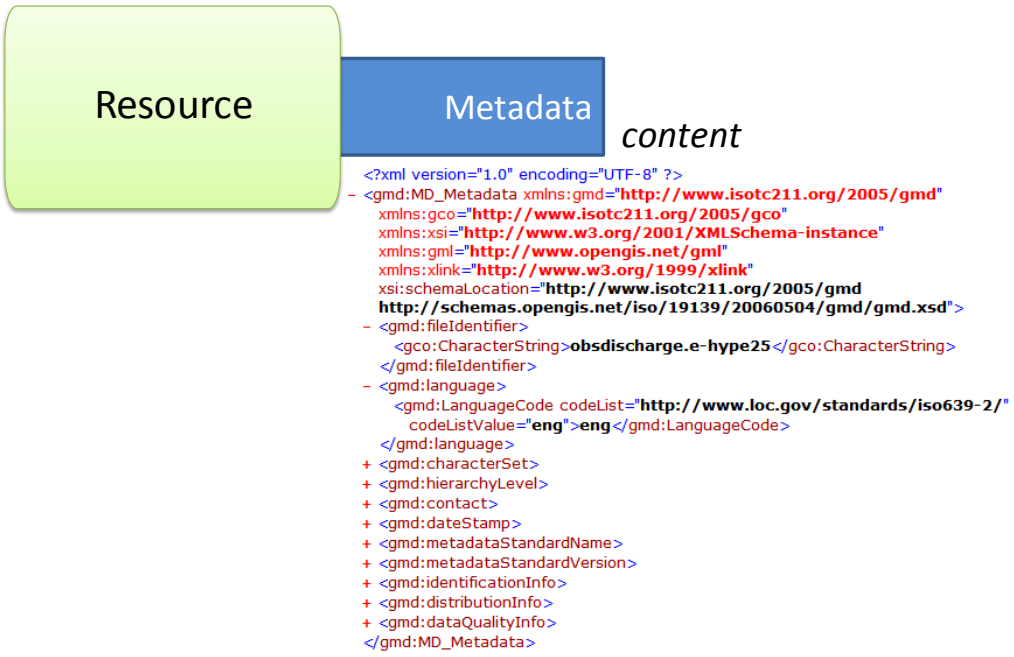


Figure 15: Meta-data Example

The contentType attribute refers to standardised Internet Assigned Numbers Authority (IANA) media types[4] (image/png, text/plain, ...) as well as to custom industry standard media types (application/x-netcdf, application/gml+xml, ...). While the content type is in general sufficient to identify the data stored in the content field of the representation, a URL that is stored in the contentLocation field may however not directly point to actual data. Instead, the link may lead to an online form for accessing the data (in different formats), a service endpoint and so on. Therefore the fixed tag groups “function”, “protocol” and “application profile” have been defined. They allow giving additional information about the handling, access and processing of content. Tags of these groups can be assigned to Representation and Resource classes which support also Dynamic Tag Extensions (2.4).

A comprehensive example is provided by means of the tags of the “function” tag group:

- **information**
  The link provides further information about resource representation, e.g. a capabilities document.
- **template**
  The link provides a template to access resource representation, e.g. an URL with request parameters.
- **download**
  The link points directly to the resource representation (e.g. a SHP file).
- **service** l
  The link points to a service endpoint, e.g. an OGC WMS service.
- **order**
  The link points to a website requiring user interaction to order/request access to the resource representation.
- **search**
  The points to a website requiring user interaction to search/browse/subset the resource.
- **offlineAccess**
  The link points to the resource representation in local files system.

[4] http://www.iana.org/assignments/media-types/media-types.xhtml

2.6 Meta-Data

2.6.1 Relevant meta-data profiles

Basic Meta-Data

Work package 3 considered the data and meta-data requirements for the Virtual Water-Science Laboratory. Deliverable D3.1 provides the outline: INSPIRE/ISO19115 for 'basic' meta-data.
Table 7 lists the meta-data properties as they are set out in deliverable D3.1:

Table 7: Basic meta-data profile

Category	Subcategory	Value
Point of contact	Organisation	Name of organization
	E-mail	e-mail for contact person
Metadata date		Date when meta-data was created
Metadata language		Language for meta-data
Resource title		Title of dataset
Identifier	Code	Dataset name, any string consisting a-z, 0-9 and -
	Codespace	urn:x-waterswitchon:dataset:<organisation>:
Resource abstract		Abstract text, must contain unit of measurement
Topic Category		Inland Waters
Keyword value		(Gemet – INSPIREthemes, version 1.0, publication, 2008-06-01) Hydrography
Geographic bounding box		Coordinates for the data
Temporal extent	Starting date	Starting data for time series
	Ending date	Ending data for time series
Date of Publication		Date for publication of data
Lineage		Information on how the data was created. For resulting data form experiments this must contain the protocol.
Conditions applying to access and use		Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) http://creativecommons.org/licenses/by-sa/4.0/legalcode
Limitations on		No limitation

public access		
Responsible party	Organisation name	Name of organisation
	E-mail	e-mail for contact person
	Responsible party Role	Role of contact person

This set of meta-data properties currently doesn't hold the linkage (location of the actual open data resource). The linkage information is nevertheless considered in the Standard Information Model. Since the metadata entry tools of the SIP are not yet available, it is proposed that meta-data for external open data sources will be collected with the help of the EUOSME Metadata Editor[5]. Consequently, the linkage must be included in the ISO 19115 profile in order to be able to collect this important information. Therefore the WP2 recommends extending the ISO 19115 profile as follows:

Table 8: Extension of the Basic Meta-Data Profile

Category	**Subcategory**	**Value**
Resource locator	Linkage	The resource locator defines the link(s) to the resource and/or the link to additional information about the resource. The value domain of this metadata element is a character string, commonly expressed as uniform resource locator (URL)
Others?		*Additional categories to be identified*

Whether additional categories have to be added or not depends on the actual requirements of WP3 regarding *find*, *bind*, *repurpose* and *publish*. Specifically, it has to be discussed with WP3 which metadata and which types of queries are required for the search for open data (*find*), which metadata is needed for the interpretation of open data (e.g. whether found data fits a specific purpose or not) (*bind*), etc..

For example, the basic meta-data schema doesn't formally define (e.g. as value of a code list) observed properties (e.g. temperature) and units of measurement (e.g. °C). Although this information can be provided as free text in "Resource abstract", only a full text search could be performed which may not deliver satisfactory results in some cases.

Quality Meta-Data

WP3 provided 2 examples (D3.2) for meta-data on data quality, of which one is shown in Table 9. These sets of meta-data properties are under discussion at the time of writing. The results of this discussion will not be available for this deliverable (D2.6). However, the SIM can basically store any meta-data profile by means of tag and content extension features. This flexibility of the SIM could potentially result in non-uniform meta-data on Quality or Lineage which might be more challenging for find and bind purposes. It could be concluded that meta-data properties should be moved from the

[5] http://inspire-geoportal.ec.europa.eu/editor/

dynamic storage to (static) attribute level in the SIM hence the (relational model definition of the) SIM could be subject to (minor) change.

Table 9: Example meta-data related to data reliability for time series

Meta data property	Value based on example data set (mean areal precipitation for the Arno river at Subbiano)
Data Origin	Repurposing of observation
Time resolution	Daily
Observation period	1.1.1992 - 31.12.2013
Description of the observing instrument/experiment	The Mean Areal Precipitation is estimated through an Inverse Squared Distance weighting of daily raingauge observations recorded by the SIR (Regional Hydrologic Service) of Regione Toscana.
Information on data collection	Tipping-bucket raingauges, resolution 0.2 mm.
Data quality measures	No quantitative information is available.
Coherency of the measurement method and instrument	No quantitative information is available.
Any other information on data quality	A description of the procedures for the data prevalidation and validation applied by the SIR (Regional Hydrologic Service) of Regione Toscana is available at the web page: http://www.sir.toscana.it/index.php?IDS=2&IDSS=51 (in Italian) In particular, as far as the observation period is concerned, the data from 1.1.2013 to 31.12.2013 have been “prevalidated”, whereas the majority of the remaining data have been “validated”.
Notes	No notes

Lineage Meta-Data

In the context of WP3 data lineage is described through the protocols defined for the experiments in the Virtual Water-Science Laboratory. As suggested in D3.1 these protocols can be represented in the meta-data by a link to the document describing the used protocol. This is already foreseen in the SIM; chapter 2.6.2.3 explains this feature.

For the moment this seems an adequate solution. However the pending meta-data discussion mentioned could result in additional but minor measures for the SIM.

2.6.2 Mapping

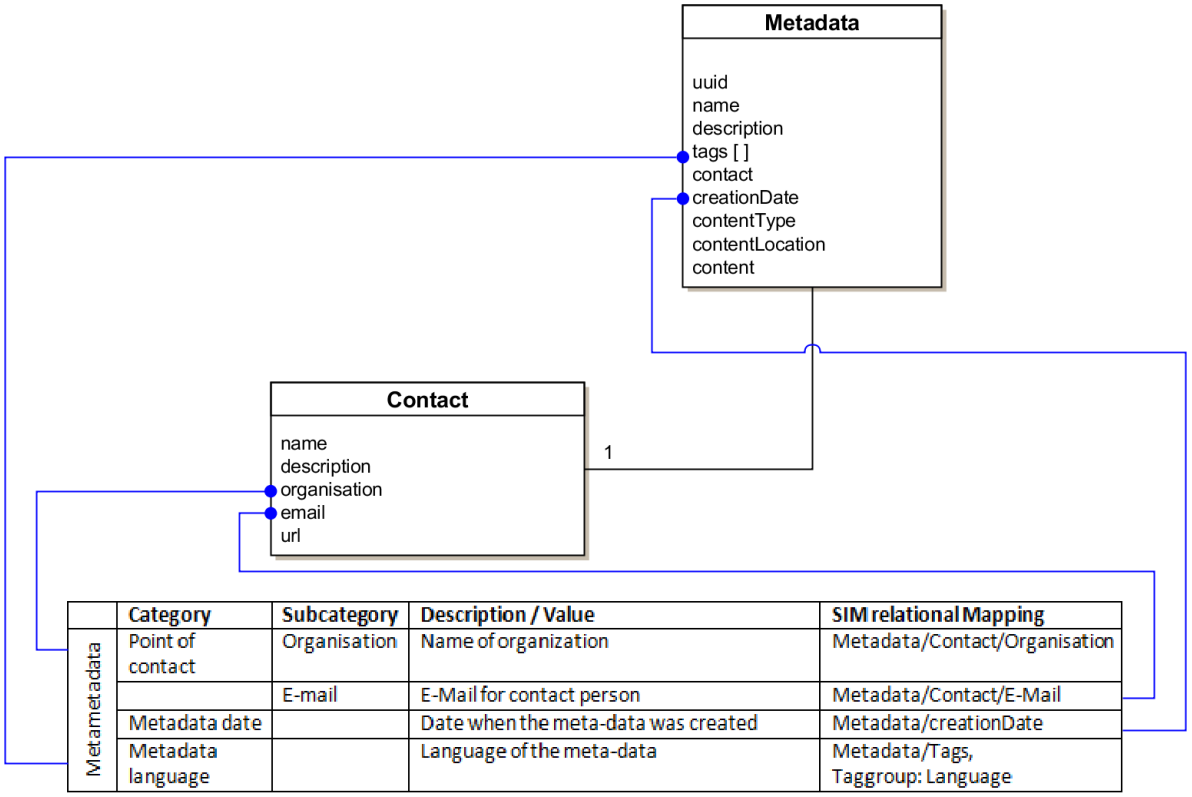
As described in section 2.6.4, access to the meta-data on resources stored in accordance with the Standard Information Model in the Meta-Data Repositories is preformed through the SIP Expert Catalogue Management APIs and the Published Catalogue Access APIs.

The following section explains therefore not only how the relevant meta-data schemas presented in the previous section are mapped to the Standard Information Model but also how the Standard Information Model can be mapped to other meta-data schemas of standards-based Published Catalogue Access APIs like OGC CSW.

2.6.2.1 Mapping from General Meta-Data (ISO 19115 Profile)

The Standard Information Model has been aligned to the current basic meta-data schema of SWITCH-ON which is based on a profile of the ISO 19115 meta-data standard. Thus, the ISO 19115 meta-data profile can be expressed completely by the relational model of the Standard Information Model. This means that categories and subcategories of the ISO 19115 meta-data profile can be mapped directly to tables and columns of a relational data base model described in section 2.3.

The following four figures give an overview on those mappings. The tables in Figure 16 represent an excerpt of the ISO 19115 meta-data profile (section 2.6.1), the boxes represent the respective tables of the relation model (section 2.3) and the blue lines the actual mapping.



	Category	Subcategory	Description / Value	SIM relational Mapping
Metametadata	Point of contact	Organisation	Name of organization	Metadata/Contact/Organisation
		E-mail	E-Mail for contact person	Metadata/Contact/E-Mail
	Metadata date		Date when the meta-data was created	Metadata/creationDate
	Metadata language		Language of the meta-data	Metadata/Tags, Taggroup: Language

Figure 16: Mapping from ISO 19115 Meta-Meta-Data

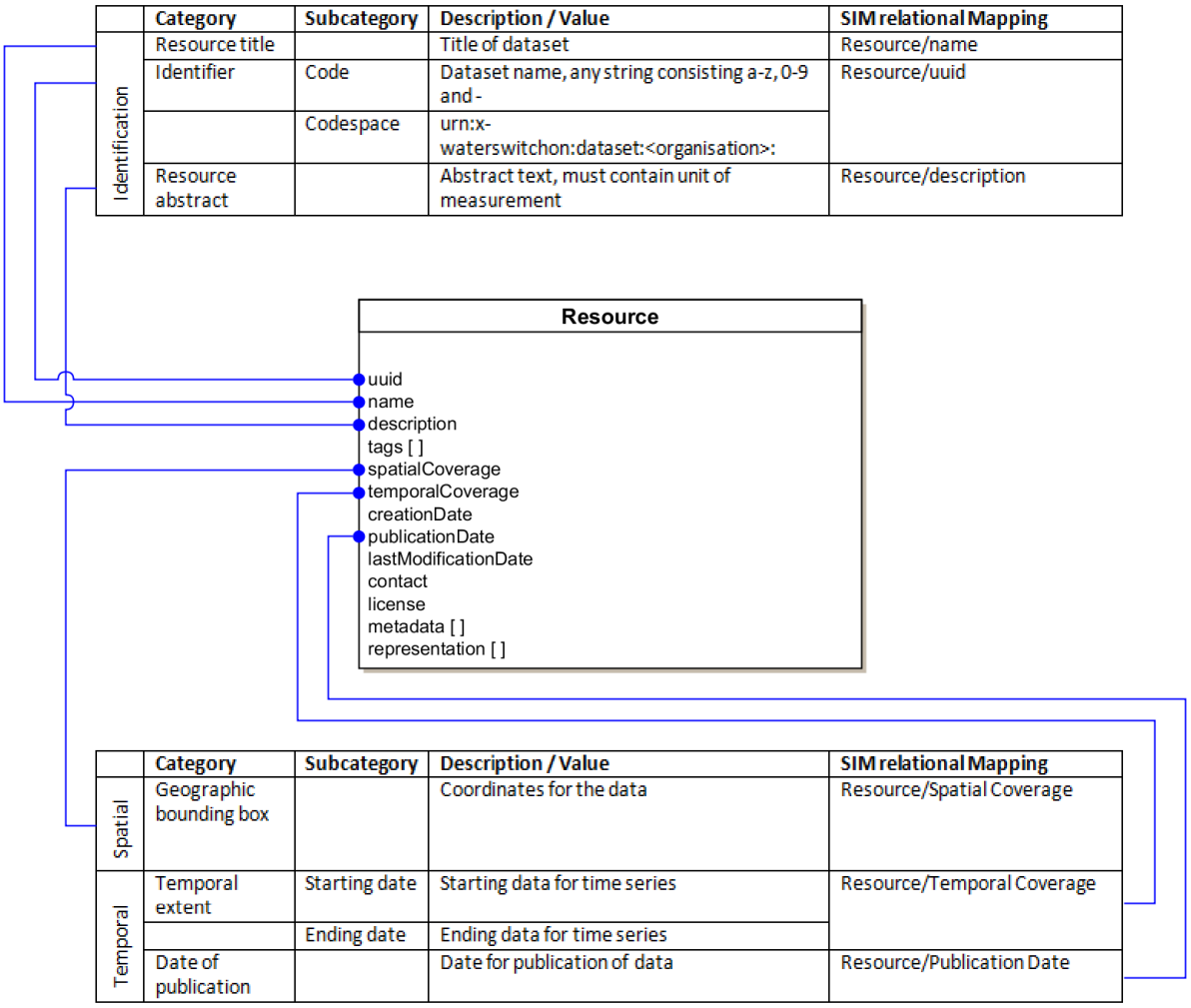
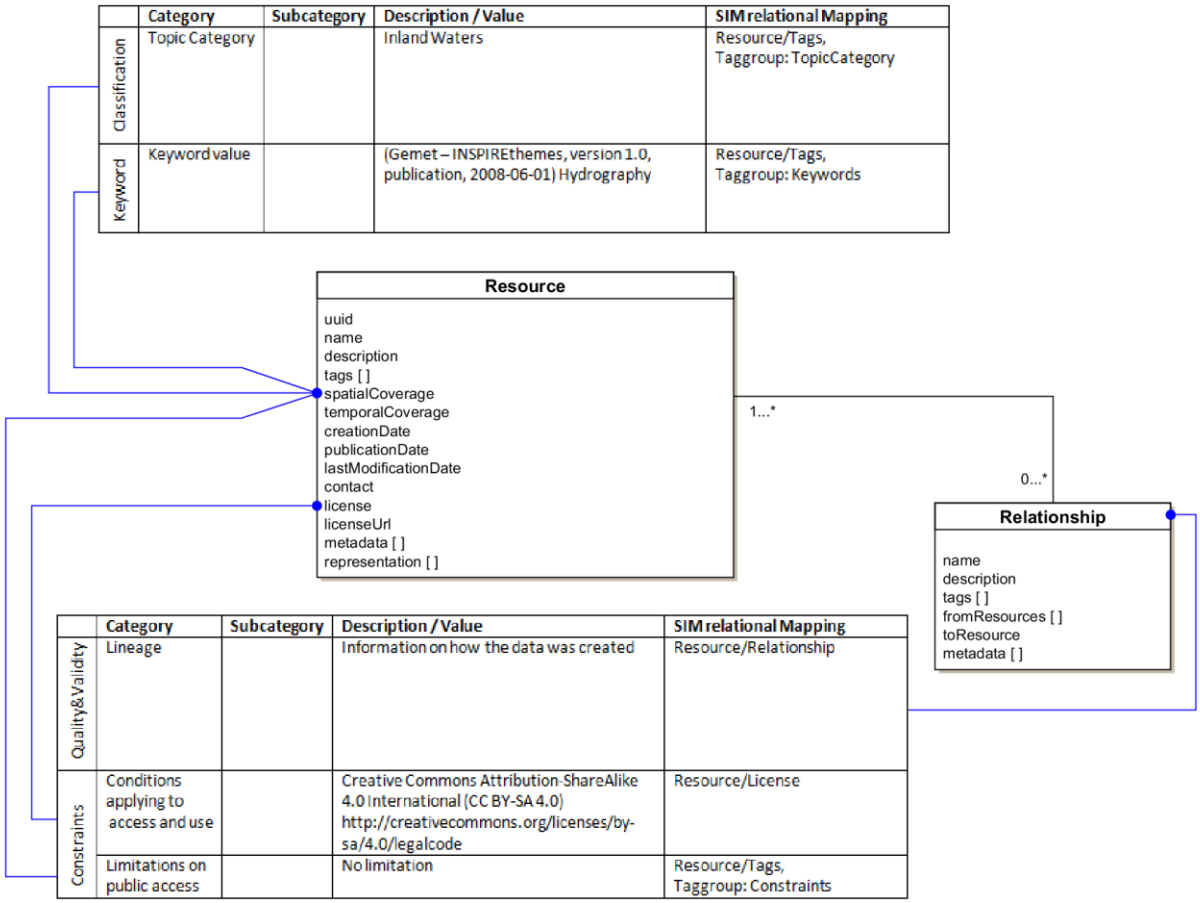


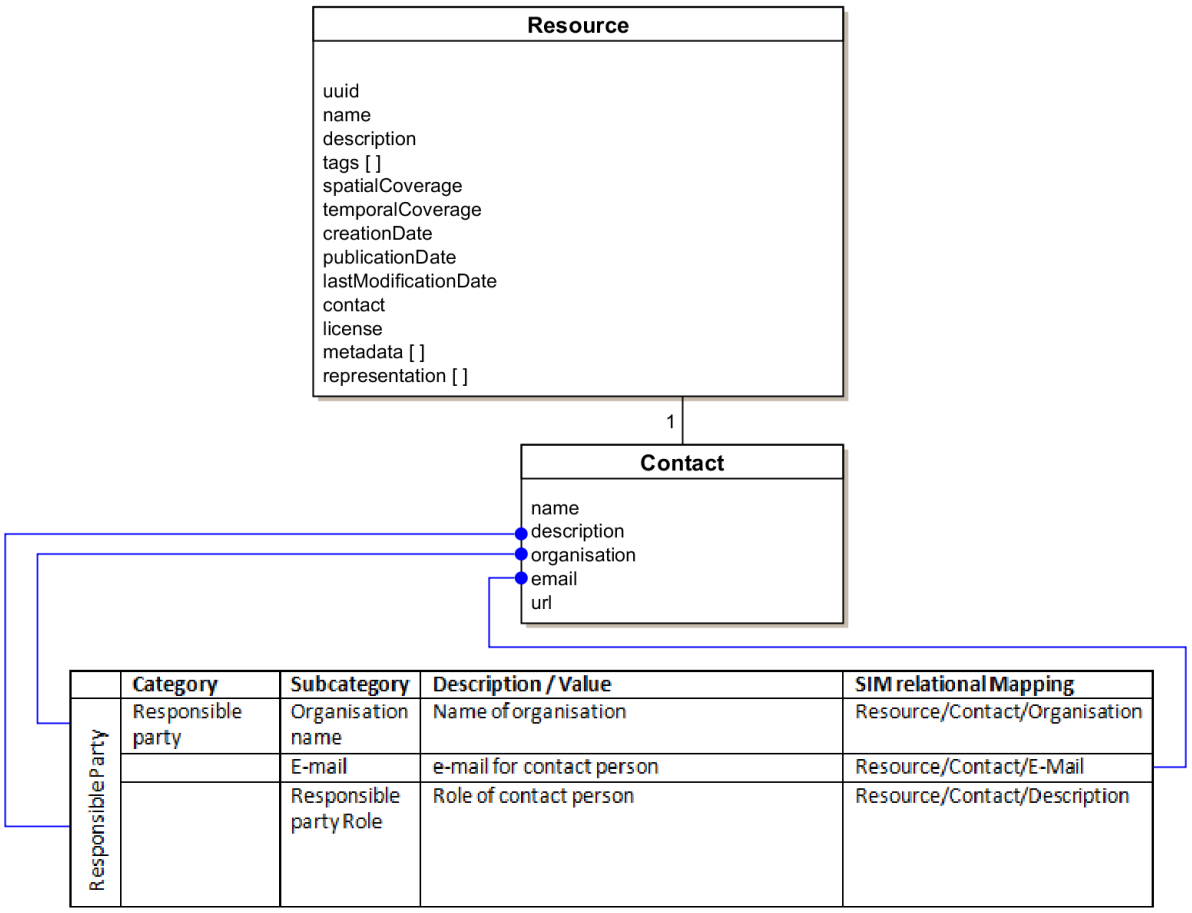
Figure 17: Mapping from ISO 19115 Spatial and Temporal Meta-Data



	Category	Subcategory	Description / Value	SIM relational Mapping
Classification	Topic Category		Inland Waters	Resource/Tags, Taggroup: TopicCategory
Keyword	Keyword value		(Gemet – INSPIREthemes, version 1.0, publication, 2008-06-01) Hydrography	Resource/Tags, Taggroup: Keywords

	Category	Subcategory	Description / Value	SIM relational Mapping
Quality&Validity	Lineage		Information on how the data was created	Resource/Relationship
Constraints	Conditions applying to access and use		Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) http://creativecommons.org/licenses/by-sa/4.0/legalcode	Resource/License
	Limitations on public access		No limitation	Resource/Tags, Taggroup: Constraints

Figure 18: Mapping from ISO 19115 Constraints and Quality&Validity Meta-Data



	Category	Subcategory	Description / Value	SIM relational Mapping
ResponsibleParty	Responsible party	Organisation name	Name of organisation	Resource/Contact/Organisation
		E-mail	e-mail for contact person	Resource/Contact/E-Mail
		Responsible party Role	Role of contact person	Resource/Contact/Description

Figure 19: Mapping from ISO 19115 ResponsibleParty Meta-Data

2.6.2.2 Mapping from Meta-Data for Data Quality

Meta-Data for Data Quality as currently defined in section 2.6.1 of this document is associated with resources with help of the Dynamic Content Extensions mechanism (section 2.5) of the Standard Information Model. The figure below shows the relevant excerpt of the Standard Information Model:

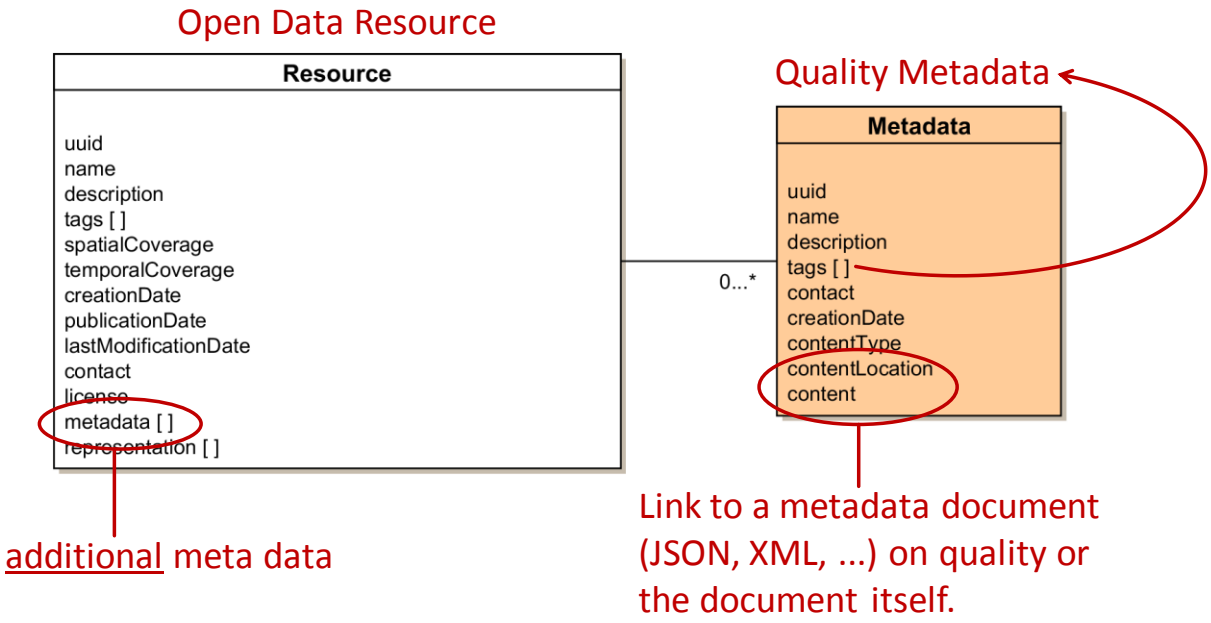


Figure 20: Mapping Meta-Data for Data Quality

The type of the meta-data is identified by a respective tag, e.g. “Quality Metadata”. The representation (encoding) of the metadata which is identified by the “contentType” attribute (e.g. application/json or text/xml) is JSON (Java Script Object Notation).

Figure 21 gives an example of a JSON representation of meta-data on data quality based on the initial specification of meta-data for data quality in deliverable D3.1.

```
{
  "dataDescription":           "Mean Areal Precipitation data for the Arno River at Subbiano",
  "dataOrigin":                "Repurposing of observation",
  "timeResolution":            "daily",
  "observationPeriod":         [
                                "2013-12-31T00:00:00+00:00",
                                "2014-04-01T00:00:00+00:00"
                               ],
  "instrumentDescription":     "The Mean Areal Precipitation is estimated through an Inverse
                                Squared Distance weighting of daily raingauge observations
                                recorded by the SIR (Regional Hydrologic Service) of Regione
                                Toscana.",
  "dataCollectionInformation": "Tipping-bucket raingauges, resolution 0.2 mm.",
  "dataQualityMeasures ":      "No quantitative information is available.",
  "measurementCoherency":      "No quantitative information is available.",
  "dataQualityInformation":    "51 (in Italian). In particular, as far as the observation
                                period is concerned, the data from 1.1.2013 to 31.12.2013 have
                                been “prevalidated”, whereas the majority of the remaining data
                                have been “validated”.",
  "notes":                     "No notes"
}
```

Figure 21: Structured Meta-Data for Data Quality (JSON Encoding)

Meta-data for data quality is mainly useful to estimate data uncertainty and thus is interpreted by the user. If the metadata for data quality is stored with help of the Dynamic Content Extensions as described above it can be easily visualized in the SIP Expert GUI and presented to the user. Although this type of meta-data is currently not part of the relational model, complex queries on meta-data for data quality are still possible thanks to the advanced query mechanisms of the cids Domain Server (Meta-Data Repository).

2.6.2.3 Mapping from Lineage Meta-Data

Since lineage information is included in the ISO 19115 profile it has been considered also in the relational model of Standard Information Model. However, the value of the "Lineage" meta-data element in the ISO 19115 standard is defined as free text. Therefore also the Dynamic content Extensions are used to represent the actual lineage meta-data. In contrast to meta-data on data quality, which is associated directly with the resource class as described in section 2.6.2.2., a new "Relationship" table to describe the relationship (lineage) between resources has been introduced in the relation model as shown in Figure 22.

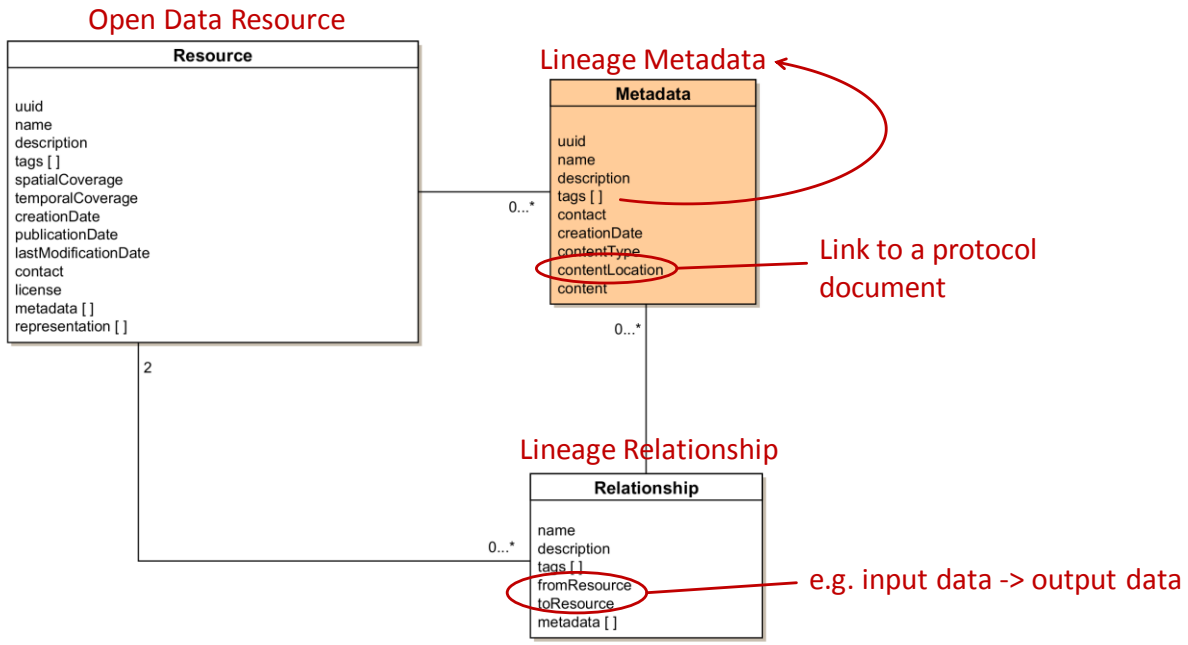


Figure 22: Mapping Lineage Meta-Data

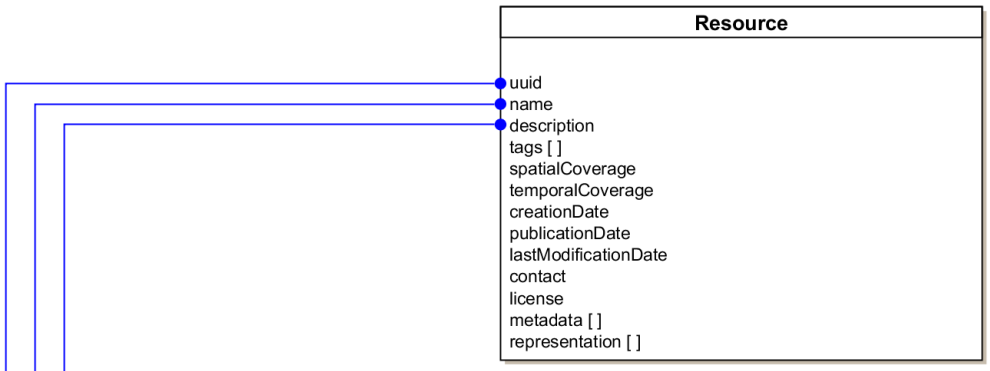
This simple relationship (fromResource / toResource) can be used to describe the I/O of an experiment (model run), a script for repurposing, etc. The type of the relationship can be identified by a respective tag. Interestingly, arbitrary additional meta-data can be associated with a relationship. This is where the actual lineage meta-data, e.g. a link to the protocols of the experiment, can be stored.

2.6.2.4 Mapping to CSW Core Metadata

As observed in deliverable D2.1, the SIP also has to interface with other types of catalogues like OGC Web Catalogue Service (CSW) or Open Knowledge Foundation for the Comprehensive Knowledge Archive Network (CKAN) catalogues. For this purpose, the Published Catalogue Access APIs have been introduced which provide public and standards based access to meta-data stored in the SWITCH-ON Meta-Data Repositories.

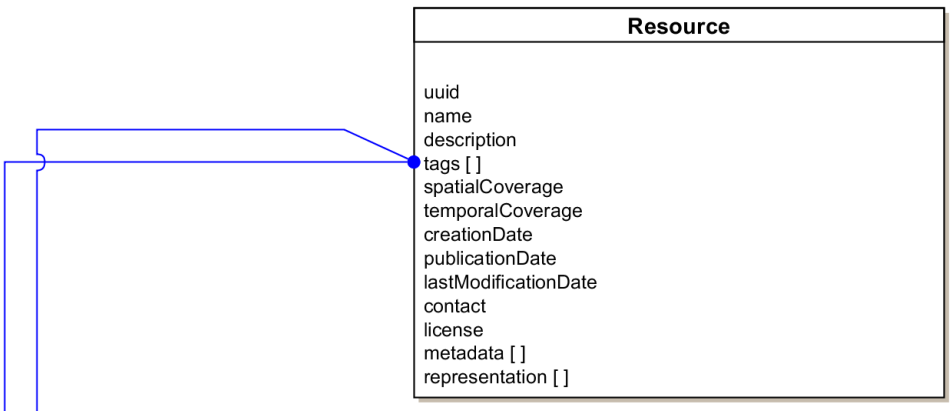
Because OGC CSW is one of the most commonly used data catalogues, the open source CSW implementation pyCSW[6] which is the reference implementation of CSW has been selected as one realisation of a Published Catalogue Access API. Since pyCSW retrieves its complete data from a relational data base model, a mapping of the relational model of the SIM to the Dublin Core encoding of CSW Core Metadata schema has been defined.

This mapping is described in the following six Figures. The tables in Figure 23 represent an excerpt of the Dublin Core meta-data profile that has been adopted by the CSW standard, the boxes represent the respective tables of the relation model (section 2.3) and the blue lines the actual mapping.



	Category	Subcategory	Description / Value	SIM relational Mapping
Identification	dc:title		Title of dataset	Resource/name
	dc:identifier		Catalog Unique Identifier	Resource/uuid
	dc:description	dct:abstract	Abstract text, must contain unit of measurement	Resource/description

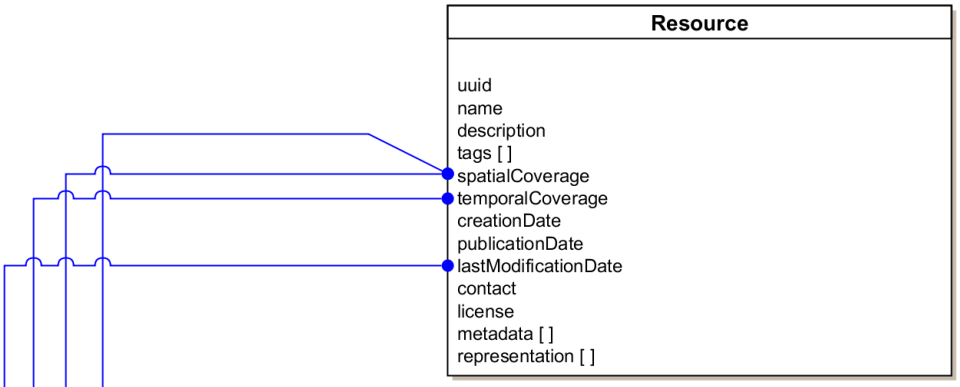
Figure 23: Mapping to Dublin Core Identification Meta-Data



	Category	Subcategory	Description / Value	SIM relational Mapping
Classification	dc:subject		Topic Category and Keywords	Resource/Tags, Taggroup: TopicCategory AND Taggroup: Keywords
	dc:language		Language of the resource	Resource/Tags, Taggroup:Language

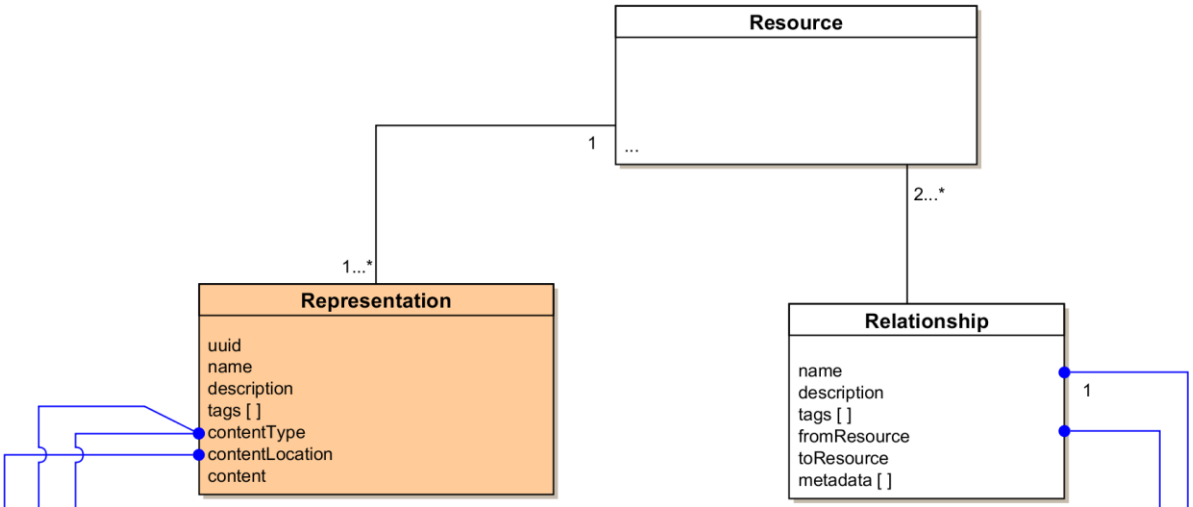
Figure 24: Mapping to Dublin Core Classification Meta-Data

[6] http://pycsw.org



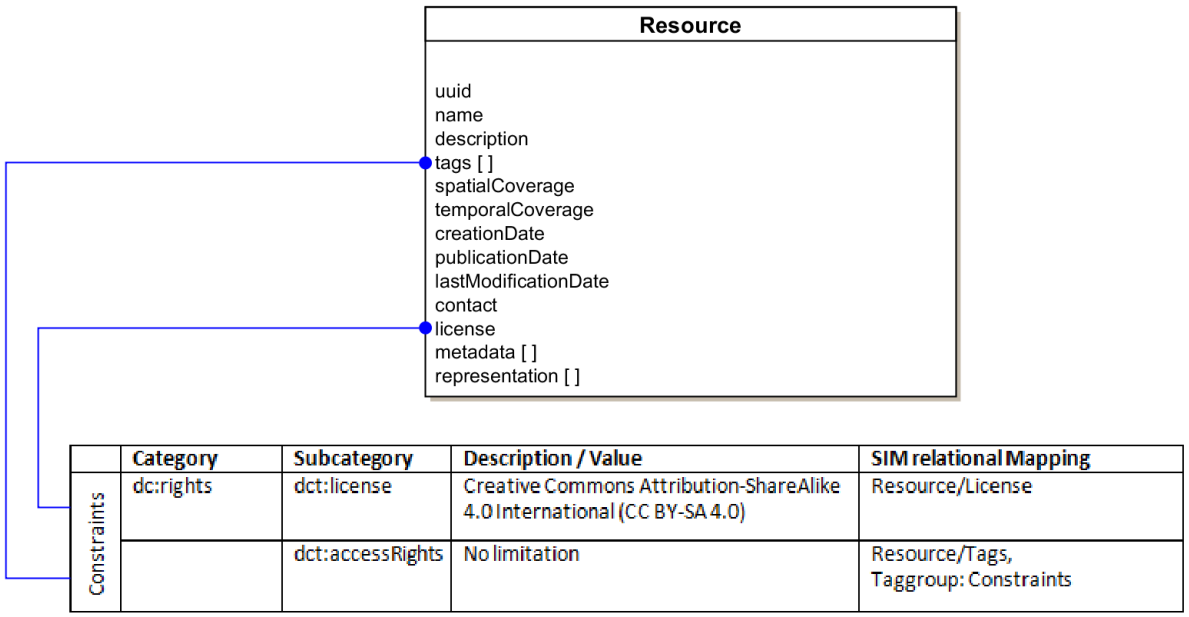
	Category	Subcategory	Description / Value	SIM relational Mapping
Temporal & Spatial	dc:coverage	dct:spatial	Geographic Coordinate Referencesystem	Resource/Spatial Coverage
		ows:Bounding Box	Spatial Coverage as bounding box	Resource/Spatial Coverage
		dct:temporal	Temporal coverage	Resource/Temporal coverage
	dc:date	dct:modified	Date of last modification of catalog record	Resource/lastModificationDate

Figure 25: Mapping to Dublin Core Temporal&Spatial Meta-Data



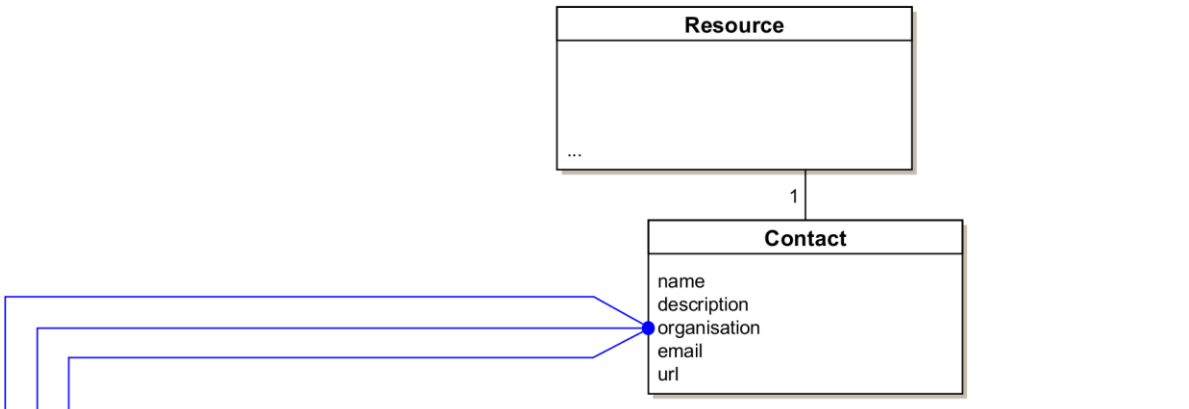
	Category	Subcategory	Description / Value	SIM relational Mapping
Relation&Representation	dc:source		A reference to a resource from which the present resource is derived.	Relationship/fromResource
	dc:relation		Name of an Relationship	Relationship/name
		dct:references	Link to the resource Data	Representation/contentLocation
	dc:type		Nature or Genre of the content.	Representation/contentType
	dc:format		Physical or Digital manifestation.	

Figure 26: Mapping to Dublin Core Relation&Representation Meta-Data



	Category	Subcategory	Description / Value	SIM relational Mapping
Constraints	dc:rights	dct:license	Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)	Resource/License
		dct:accessRights	No limitation	Resource/Tags, Taggroup: Constraints

Figure 27: Mapping to Dublin Core Constraints Meta-Data



	Category	Subcategory	Description / Value	SIM relational Mapping
Responsible Party	dc:creator		Creator of the resource	Resource/Contact/Organisation
	dc:publisher		Publisher of the resource	Resource/Contact/Organisation
	dc:contributor		Contributor to the resource	Resource/Contact/Organisation

Figure 28: Mapping to Dublin Core Responsible Party Meta-Data

Due to the fact, that the Meta-Data Repository and the internal database of the pyCSW can share the same PostgreSQL RDBMS instance, the mapping can be implemented as a combination of SQL Views and Triggers. Figure 27 shows the relevant components and meta-data flows: pyCSW (Published Catalogue API) is directly connected to the PostgreSQL database service. The Expert Catalogue Management API of the SIP which is realised by cids Broker (see deliverable D2.1, section 4.2.1 for more information) does not communicate directly with the database. Instead, it relies on the cids Domain Service which offers advanced capabilities beyond the possibilities of plain SQL Queries or

PL/SQL functions. This includes for example search on structured meta-data stored according to the Dynamic Content Extension Concept (section 2.5).

The mapping is furthermore defined in a bidirectional manner. Thus, the *Harvest* and *Transaction* operations of the pyCSW can be used to feed the SWITCH-ON Meta-Data Repository with new meta-data that is entered into pyCSW or is harvested by pyCSW from external CSW catalogues.

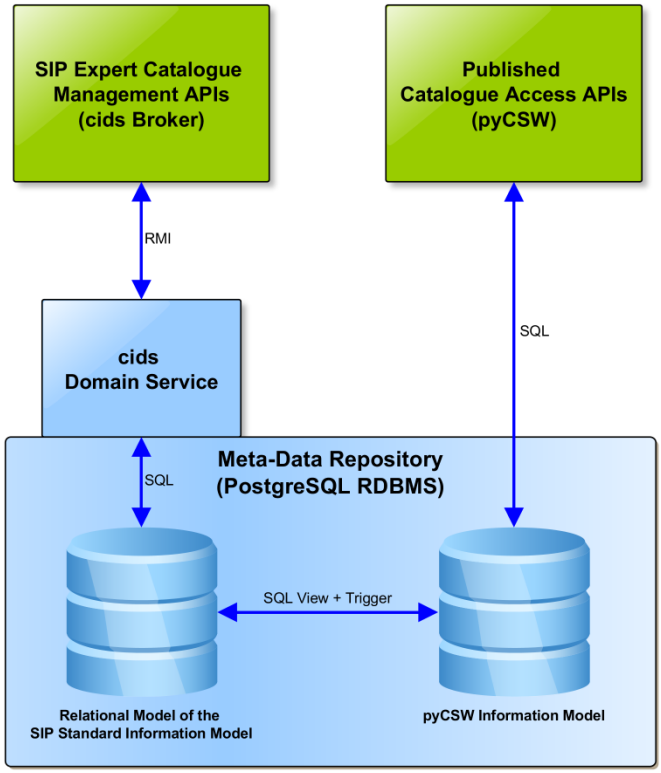


Figure 29: Mapping to CSW Core Metadata

If needed, mappings to other types of meta-data schemas like the Data Catalogue Vocabulary (DCAT[7]) specified by the W3C or the domain model of the Comprehensive Knowledge Archive Network (CKAN) specified by the Open Knowledge Foundation[8] can be implemented to support additional types of catalogues.

2.6.3 Find, Bind, Transform, Publish

The primary objective of the Standard Information Model is to support the open data repurpose pattern (Figure 30) that has been introduced in deliverable D2.1 (D2.1, 2014).

[7] http://www.w3.org/TR/vocab-dcat/

[8] http://okfn.org/

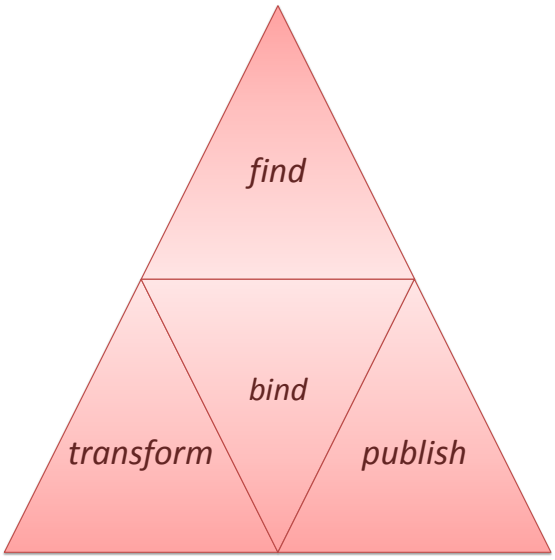


Figure 30: Open Data Repurpose Pattern

The next four sections describe by what meta-data and by what means the SIM support finding, binding, transforming and publish of open data.

2.6.3.1 Find

Find refers to supporting discovery and investigation of open data to assess its relevance including the ability to locate relevant open data and assess the suitability for the task at hand.

The general meta-data profile based ISO 19115 (see section 2.6.1) gives already enough information to provide comprehensive geospatial and temporal discovery functionality which is often demanded by experts seeking for open data. Moreover, since the general meta-data profile is mapped to a relational database model as explained in 2.6.2.1, any meta-data item of this profile can become subject of a SQL query which opens a wide range of possibilities for finding open data that is catalogued in the SIP.

Some default queries to be supported by SIP GUI and the respective APIs as well as the according meta-data item of the query are listed below:

- Geospatial Search (“Geographic bounding box”) by providing coordinates (bounding box) or drawing a bounding / polygon on a map.
- Temporal Search (“Temporal extent”, “Date of publication”) by entering start/end dates.
- Keyword Search (“Keyword value”) by specifying free keywords or selecting a keyword from a list of predefined keywords.
- Category Search (“Topic Category”) by selecting one or more categories from a list of known categories.
- Full-text Search (“Resource Title”, “Resource abstract” and others) by specifying a search string.

However, also search for structured meta-data which is not part of the relational model and thus the ISO 19115 profile has to be supported. The Meta-Data Repositories of the SIP are implemented by the cids Kernel which consists of the cids Integration Base, a standard Relational Database Management System (PostgreSQL), and the cids Domain Service, which translates the generic meta-data structure into concrete meta-objects and classes. The Domain Service is also responsible for the construction of the dynamic catalogue structure at runtime and offers a highly customisable interface to the dynamic search capabilities.

Figure 31 shows how a search for relational as well as for structured meta-data can be performed on basis of the Standard Information Model:

(4) A SWITCH-ON client, for example a WP4 application or the SIP Expert GUI (cids Navigator), issues a query for meta-data to the Search API (see section 2.6.4.5) of the SIP Expert Catalogue Management API (cids Broker). The concrete query is no relevant in the example below, since any query that can be satisfied by the meta-data stored in a SWITCH-ON Meta-Data Repository can theoretically be supported. Please note also that the Expert Catalogue Management API currently supports only the RMI protocol. A HTTP REST interface that can easily be used by non-Java based clients (WP4 applications, web clients, etc.) will be developed in the context the project.

(5) The cids Broker service, which actually offers the Expert Catalogue Management API, delegates the query to an appropriate cids Domain Server instance. There can be many different cids Domain Server instances each representing the interface to distinct Integration Bases (PostgreSQL RDBMS). The cids Broker hides this distribution aspect and offers on entry point for clients. Clients never communicate with the Domain Server or the Integration Base directly.

(6) The cids Domain Server processes the query and translates it into SQL statements that are executed on the data base. If necessary, it fetches structured meta-data stored in accordance to the Dynamic Content Extension Mechanism of the Standard Information Model (see section 2.5). Structured meta-data can then be queried efficiently with appropriate query languages like JSON Query Language (JSONiq)[9] for JSON-encoded meta-data or XML Path Language (XPath)[10] for XML-encoded meta-data.

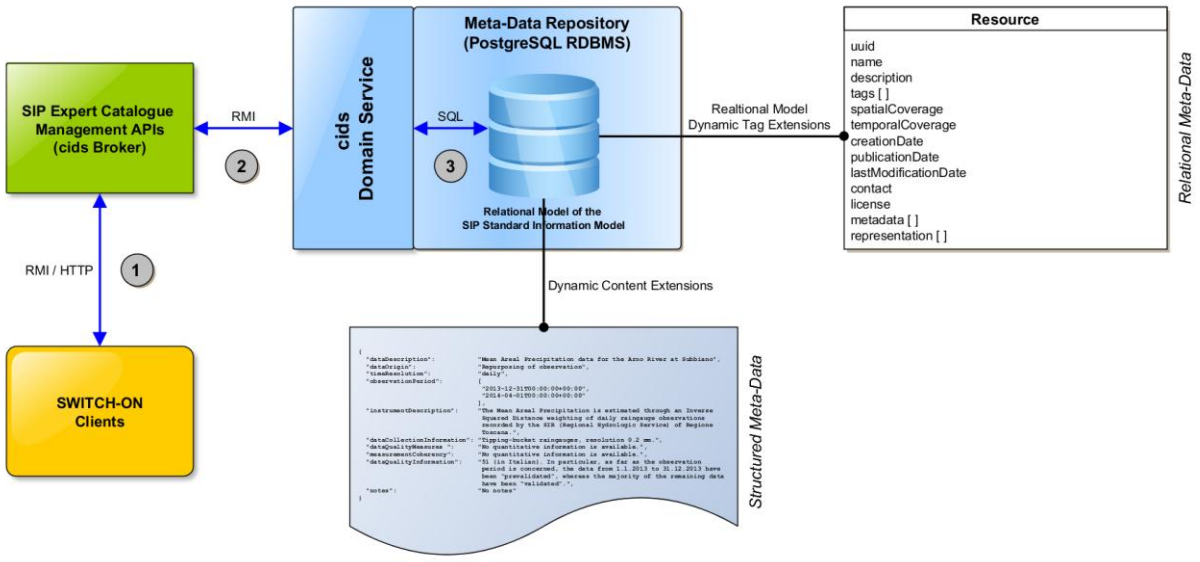


Figure 31: Search on Relational and Structured Meta-Data

Besides the possibilities to take a combination relational and structured meta-data into account to provide highly customised search capabilities, the SIM leverages also the creation of dynamic catalogues. The overall approach for creating catalogue structures based on Dynamic Tag Extensions is briefly in section 2.4. An example that uses the tag groups “Geography” and “Hydrological Concept” to create a custom catalogue in the SIP Expert GUI (cids Navigator) is illustrated in Figure 32. Of course,

[9] http://www.jsoniq.org/

[10] http://www.w3.org/TR/xpath/

any other tag group and possibly even any meta-data item can be used to categorise open data in such a catalogue structure.

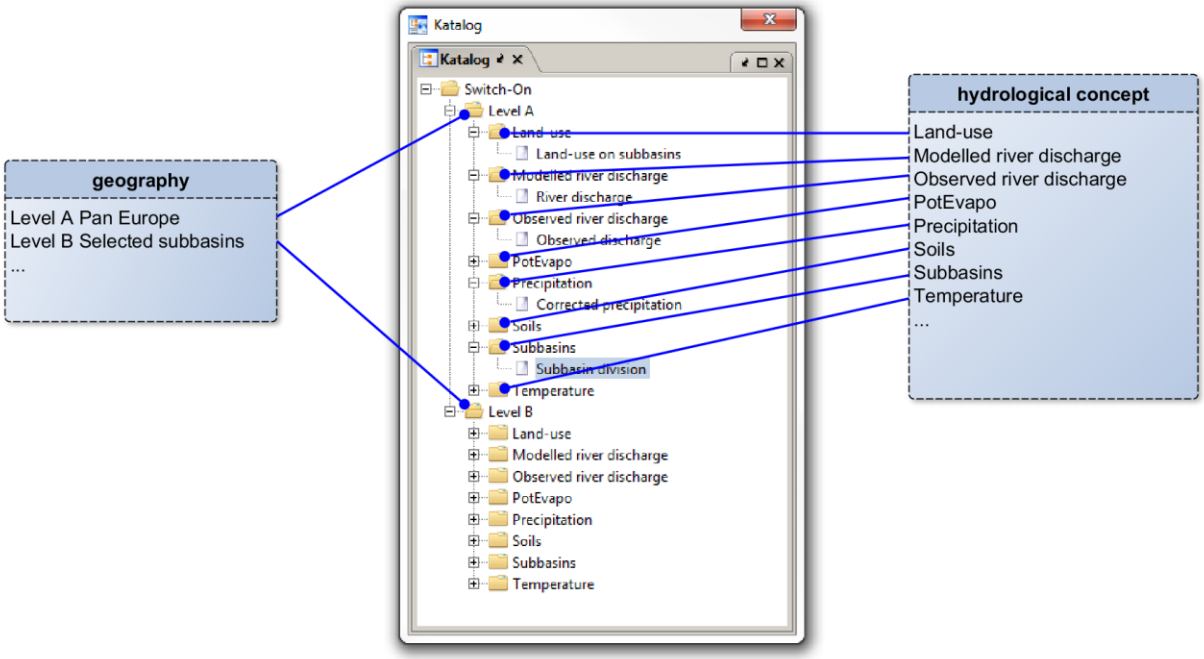


Figure 32: Creating Dynamic Catalogue Structures with Tag Groups

The assessment of the suitability of the open data found for a specific purpose is supported by Standard information Model only on a more general level. Especially the common description of the data set ("Resource abstract"), meta-data on data quality and the meta-data on lineage (e.g. protocols of an experiment) are valuable for this assessment. The SIP, in particular the SIP Expert GUI, can leverage this assessment by presenting the relevant meta-data in an appealing and comparable manner. However, it has to be admitted that quality and lineage meta-data is often not available for public open data found in external catalogues.

2.6.3.2 Bind

Bind refers to technically enable access to and use of open data including support for putting it to use, which requires descriptive information about syntax and semantics.

The actual technical access to open data is leveraged by the "Representation" class of the Standard Information Model which can be seen as an extension of the "Linkage" category of the ISO 19115 profile. More information and concrete examples on how the technical access information is represented is described in detail in section 2.5 in the context of Dynamic Content Extensions.

Figure 33 provides an example how the Standard Information Model supports technical access to open data:

(1) A SWITCH-ON client, for example a WP4 application or the SIP Expert GUI (cids Navigator), retrieves meta-information about a resource from the Meta-Data Repository through one of the Catalogue Access APIs. The meta-information encompasses also information about the representations (the actual data) of the resource and the type of the representation (e.g. aggregated data for preview).

(2) The client can then evaluate the “contentLocation” property of the “Representation” class. In general, it contains a URL which points to the data. The technical access to the data is initiated by the client. Please note, that highly aggregated data for preview might be stored directly in the “content” property as mentioned in section 2.2.1.

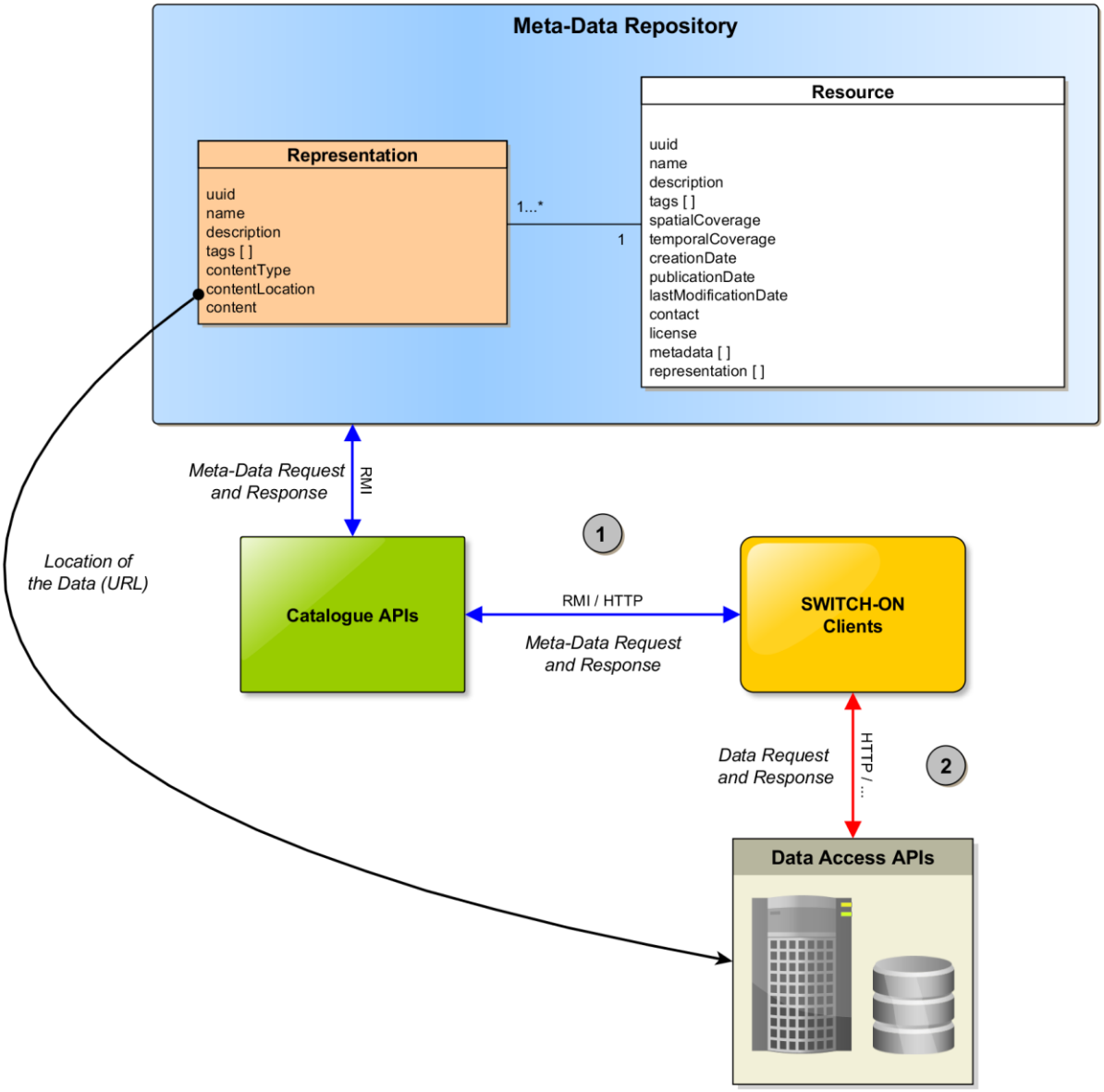


Figure 33: Data Access leveraged by Meta-Data

2.6.3.3 Transform

Transform refers to setting the data into the desired context, either through structural changes (format conversion, filtering, extraction, selection) and/or by changing the actual data content, e.g. algorithmic conversion, aggregation or by producing different representations.

Prerequisite for transform is the technical access to the data, thus bind as described in the previous section (2.6.3.2). The actual transformation is performed by Tools, in particular Conversion and Schema Mapping Tools (deliverable D2.1, section3.3.4). Complex rules and processing instructions on how data has to be transformed are not described by meta-data. However, meta-data for transform

supports the description of result of a transformation including the description of the transformation process (e.g. a protocol of an experiment).

This kind of meta-data is modelled in the "Relationship" class which encompasses also lineage information (see section 2.3.3) of the resource. Thus, in the context of the Standard Information Model meta-data for data transformation refers both to repurposed data as well as to data resulting from an experiment. On the level of the information model, this distinction is only made by the type of the relationship ("tags" of the "Relationship" class) between resources.

Figure 34 provides a simple example on what kind of meta-data for transform is supported by the SIM:

(1) A SWITCH-ON Processing Tool, for example a tool for open data repurposing, retrieves data from a SWITCH-ON internal or external Data Repository.
(2) The Tool performs a repurposing algorithm which results in a new data set. The new data is stored in the SWITCH-ON Data Repository.
(3) The Tool creates also an appropriate set of meta-data ("Resource" and "Representation") including a description of lineage of the resource ("Relationship") and stores it in the Meta-Data Repository. Depending on whether the new data has to be published (see section 2.6.3.4), the respective meta-data of the resource is also made available through the Published Catalogue Access APIs.

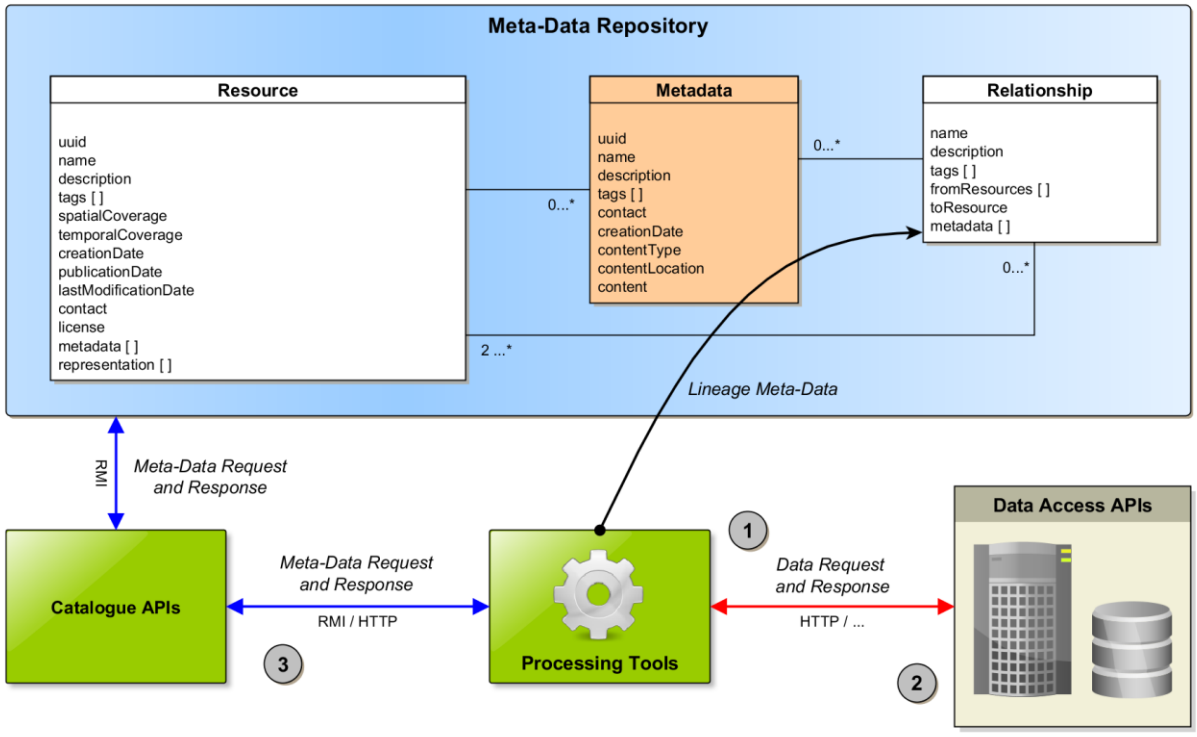


Figure 34: Lineage Meta-Data resulting from Data Transformation

2.6.3.4 Publish

Publish refers to making repurposed data available to others and adding descriptive information (meta-data) to support (find, bind, transform) them.

Thus, publish is leveraged by the SIM by supporting all relevant types of meta-data needed for find, bind, transform open data. Coming back to the transform example shown in Figure 34, the last step in the open data repurpose chain would be to publish the generated meta-data via the Published Catalogue Access APIs.

The meta-data stored in accordance with the Standard information Model can furthermore be published in many different standards and formats. Section 2.6.2.4 provides an example on how meta-data in the SWITCH-ON Meta-Data Repository can be published via a CSW service interface. Support for other standards and formats for meta-data publication will be investigated and implemented if needed.

2.6.4 Access

Access to meta-data stored in accordance with the Standard Information Model is provided through the SIP Expert Catalogue Management APIs and the Published Catalogue Access APIs. Published Catalogue Access APIs are web services that provide (in general read only) access to meta-data through a standardised service interface, mainly for the purpose of finding open data. They can be used by any (external) client that supports the respective standard. The SIP Expert Catalogue Management APIs provide advanced functionalities to SIP Expert Catalogue GUIs for manipulating the meta-data catalogue, including import of meta-data and the (semi-automatic) creation of meta-data from open-data, and advanced search functionalities not supported by the catalogue or respective meta-data standards.

The general usage of an Expert Catalogue Management API (cids Broker) by a SWITCH-ON client is briefly explained in section 2.6.3.1 and Figure 31. An example of a Published Catalogue Access API (pyCSW) is given in 2.6.2.4 and Figure 29.

Since access and usage of Published Catalogue Access API is described in the respective standards, e.g. the OpenGIS Catalogue Service Implementation Specification[11], the focus of this section is on the specification of the Expert Catalogue Management REST HTTP API. Please note that the REST HTTP API is currently not yet available and will be developed in the context of the project.

Figure 35 gives an overview on the different types of APIs for Expert Catalogue Management. The purpose of the different APIs is explained in the following subsections.

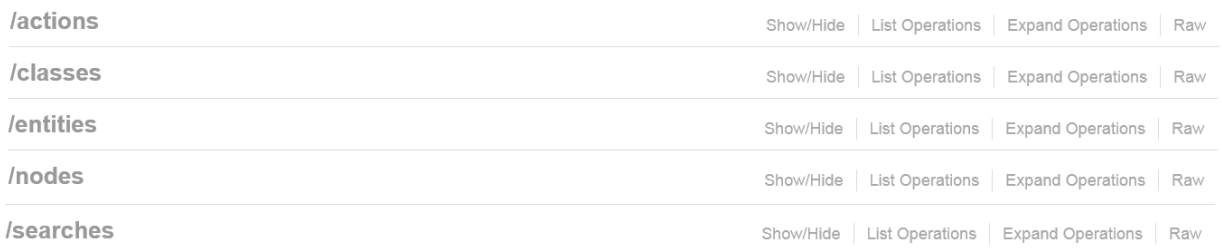


Figure 35: SIP Expert Catalogue Management REST HTTP APIs Overview

Please note that administrative APIs like user and permission management are currently excluded from this specification since they are not relevant for clients other than the SIP Administration GUIs.

2.6.4.1 Actions

Actions are predefined tasks that are executed within the execution context of the SIP Backend, e.g. on the server or the virtual machine the SIP server components are deployed. Actions are implemented as Java classes in the cids Domain Server component and thus can perform long lasting and complex operations. Actions provide a simple mechanism to delegate business logic from the client to the server and can for example be used to directly call a local transformation script or initiate a data

[11] http://portal.opengeospatial.org/files/?artifact_id=20555

transformation on a remote server (e.g. WPS). Figure 36 gives an overview about the specification of the Actions API.

/actions

GET	/actions	Get all actions.
GET	/actions/{domain}.{actionkey}/tasks/{taskkey}	Get task status.
DELETE	/actions/{domain}.{actionkey}/tasks/{taskkey}	Cancel task.
GET	/actions/{domain}.{actionkey}/tasks/{taskkey}/results/{resultkey}	Get task result.
GET	/actions/{domain}.{actionkey}	Show and describe an action.
GET	/actions/{domain}.{actionkey}/tasks	Get all running tasks.
POST	/actions/{domain}.{actionkey}/tasks	Create a new task of this action.
GET	/actions/{domain}.{actionkey}/tasks/{taskkey}/results	Get task result.

Figure 36: Actions API Specification

2.6.4.2 Classes

The Classes API provides information on the entity types (classes) supported by the Expert Catalogue APIs. The entity types are defined by relational model of the SIM as described in section 2.3. The Classes API delivers a JSON representation of the respective class. Instances of those classes can be retrieved via the Entities API (section 2.6.4.3). Figure 37 gives an overview about the specification of the Classes API.

/classes

GET	/classes/{domain}.{classkey}	Get a certain class.
GET	/classes	Get all classes.
GET	/classes/{domain}.{classkey}/{attributekey}	Get a certain class.

Figure 37: Classes API Specification

A complete example of the JSON Class Definition of the “Representation” Class is shown below:


```
{
            "$self": "/SWITCHON.REPRESENTATION",
            "configuration": {
                "name": "representation",
                "policy": "STANDARD",
                "attributePolicy": "STANDARD",
                "pK Field": "id"
            },
            "attributes": {
                "id": {
                    "$self": "/SWITCHON.REPRESENTATION/id",
                    "name": "id",
                    "position": 0,
                    "javaclassname": "java.lang.Integer"
                },
                "name": {
```

```
                "$self": "/SWITCHON.REPRESENTATION/name",
                "visible": true,
                "name": "name",
                "position": 1,
                "javaclassname": "java.lang.String"
            },
            "description": {
                "$self": "/SWITCHON.REPRESENTATION/description",
                "visible": true,
                "optional": true,
                "name": "description",
                "position": 2,
                "javaclassname": "java.lang.String"
            },
            "tags": {
                "$self": "/SWITCHON.REPRESENTATION/tags",
                "visible": true,
                "array": true,
                "optional": true,
                "name": "tags",
                "position": 3,
                "referenceType": "/SWITCHON.TAG"
            },
            "contenttype": {
                "$self": "/SWITCHON.REPRESENTATION/contenttype",
                "visible": true,
                "optional": true,
                "name": "contenttype",
                "position": 4,
                "javaclassname": "java.lang.String"
            },
            "contentlocation": {
                "$self": "/SWITCHON.REPRESENTATION/contentlocation",
                "visible": true,
                "optional": true,
                "name": "contentlocation",
                "position": 5,
                "javaclassname": "java.lang.String"
            },
            "content": {
                "$self": "/SWITCHON.REPRESENTATION/content",
                "visible": true,
                "optional": true,
                "name": "content",
                "position": 6,
                "javaclassname": "java.lang.String"
            }
        }
    }
```

The most interesting part of this JSON Class definition shown in the example above is the list of attributes of the class. The type of the attribute is determined by the “referenceType” and by “javaclassname” attributes as shown in Figure 38.

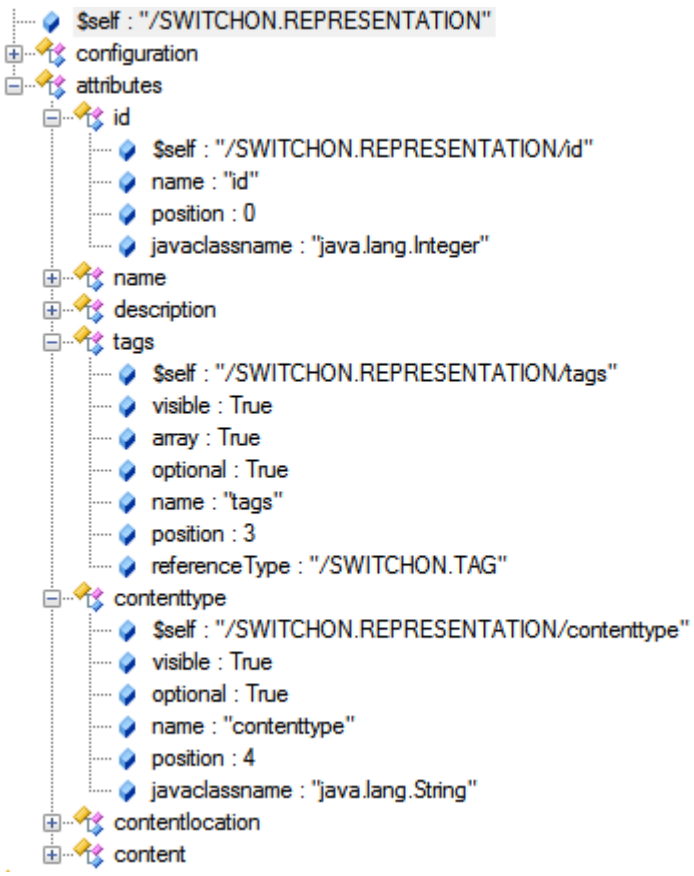


Figure 38: JSON Structure of the “Representation” Class

As comparison, the Relational Structure of the “Representation” Class as defined in the relational model (section 2.3) is shown in Figure 39 below.

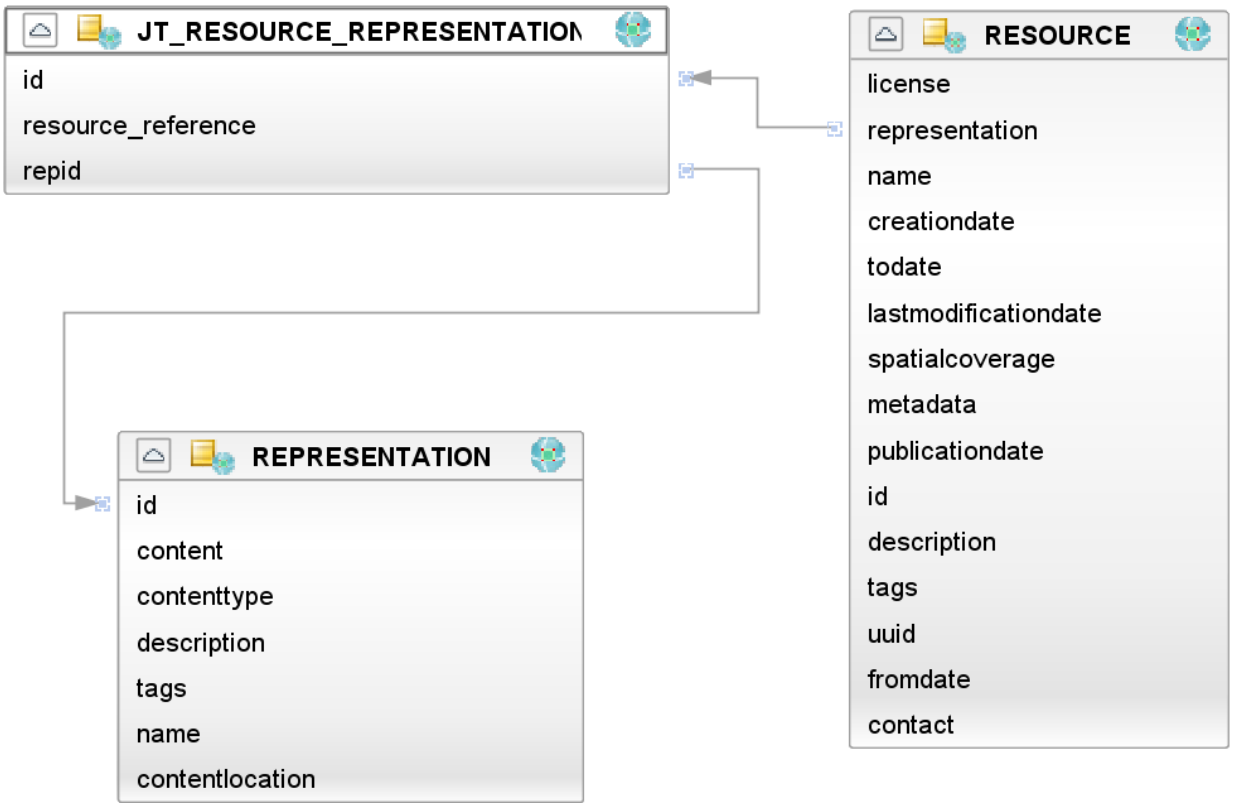


Figure 39: Relational Structure of the "Representation" Class

2.6.4.3 Entities

The Entities API is the most important API. It provides access to the actual object instances stored in the Meta-Data Repository. Thereby, the API offers basic CRUD functionality as shown in Figure 38.

/entities

GET	/{domain}.{class}/{objectid}	Get a certain object by its id.
PUT	/{domain}.{class}/{objectid}	Update or creates an object.
DELETE	/{domain}.{class}/{objectid}	Delete a certain object.
POST	/{domain}.{class}	Create a new object.
GET	/{domain}.{class}	Get all objects of a certain class.
GET	/{domain}.{class}/emptyInstance	Get an empty instance of a certain class.

Figure 40: Entities API Specification

The JSON Class definitions of all supported entities can be obtained via the Classes API that is described in section 2.6.4.2. Below is an example of the structure (Figure 41) and the JSON representation of a resource.

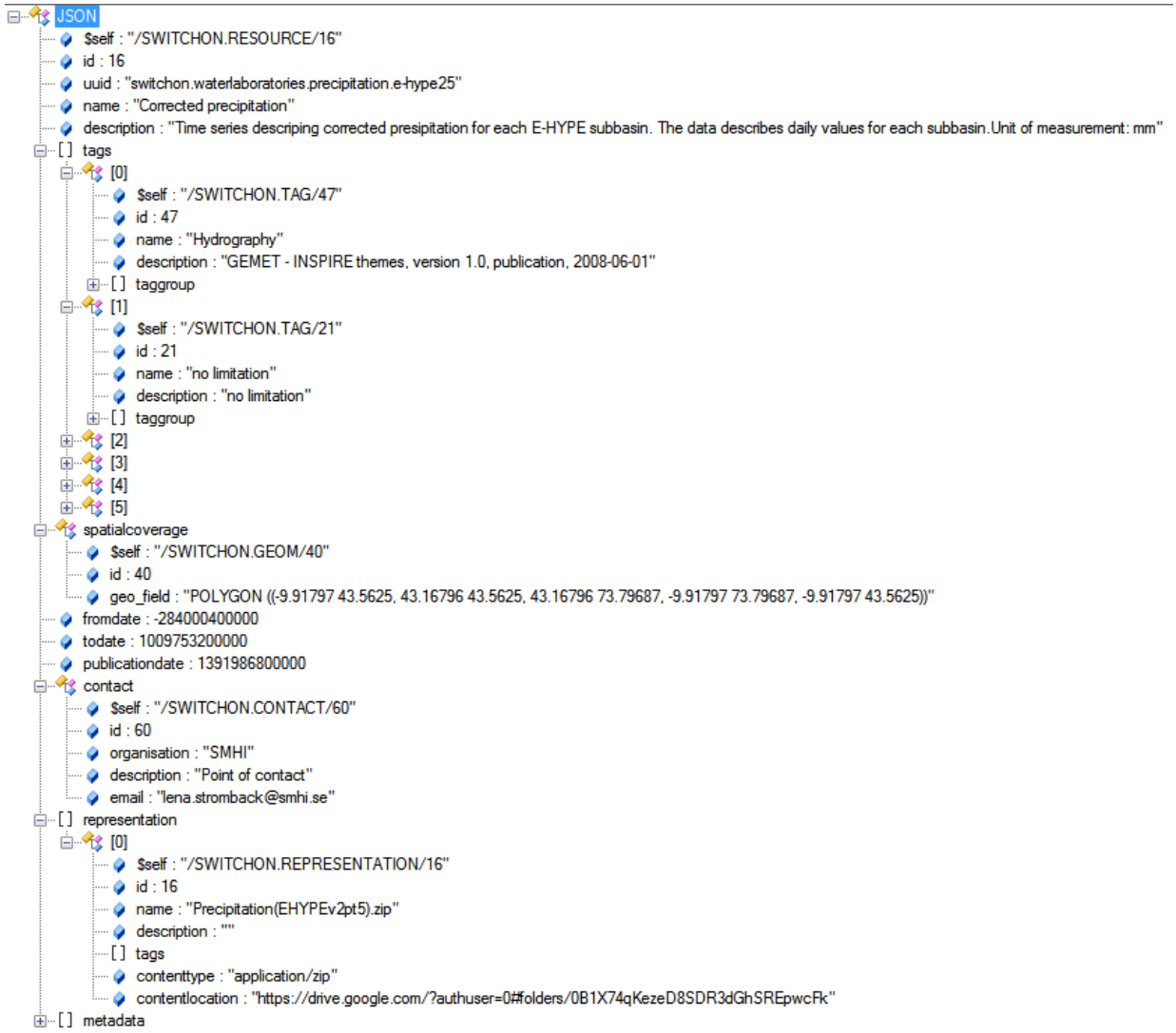


Figure 41: JSON Structure of a “Resource” Object


```
{
    "$self": "/SWITCHON.RESOURCE/16",
    "id": 16,
    "uuid": "switchon.waterlaboratories.precipitation.e-hype25",
    "name": "Corrected precipitation",
    "description": "Time series descriping corrected presipitation for each E-HYPE
subbasin. The data describes daily values for each subbasin.Unit of measurement: mm",
    "tags": [
        {
            "$ref": "/SWITCHON.TAG/47"
        },
        {
            "$ref": "/SWITCHON.TAG/21"
        },
        {
            "$ref": "/SWITCHON.TAG/12"
        },
        {
```

```
            "$ref": "/SWITCHON.TAG/68"
        },
        {
            "$ref": "/SWITCHON.TAG/95"
        },
        {
            "$ref": "/SWITCHON.TAG/118"
        }
    ],
    "spatialcoverage": {
        "$ref": "/SWITCHON.GEOM/40"
    },
    "fromdate": -284000400000,
    "todate": 1009753200000,
    "publicationdate": 1391986800000,
    "contact": {
        "$ref": "/SWITCHON.CONTACT/60"
    },
    "representation": [
        {
            "$ref": "/SWITCHON.REPRESENTATION/16"
        }
    ],
    "metadata": [
        {
            "$ref": "/SWITCHON.METADATA/40"
        }
    ]
}
```


As can be seen in the JSON document above, properties referring to complex objects like tag, representation, metadata, etc. are only provided by reference. The behaviour whether JSON object obtained via the Entities API shall be expanded (contain all object values instead of object references) or not is controllable by “expand” and “level” parameters of API. Figure 42 gives an overview on all parameters of the GET Entities operation of the Entities API.

Parameter	Description	Parameter Type	Data Type
domain	**identifier (domainname) of the domain.**	path	string
class	**identifier (classkey) of the class.**	path	string
objectid	**identifier (objectkey) of the object.**	path	string
version	version of the object, 'current' version when not submitted	query	string
role	role of the user, 'default' role when not submitted	query	string
expand	a list of properties in the resulting objects that should be expanded	query	string
level	the level of expansion	query	string
fields	the fields of the resulting object, all fields when not submitted	query	string
profile	profile of the object, 'full' profile when not submitted and no fields are present	query	string
omitNullValues	Omit properties that have 'null' as value	query	boolean
deduplicate	if you don't want already expanded properties to be expanded again, set this parameter to true	query	boolean
Authorization	Basic Auth Authorization String	header	string

Figure 42: Parameters of the GET Entities Operation of the Entities API

2.6.4.4 Nodes

The Nodes API can be used to obtain a catalogue structure whereby the catalogue entries (nodes) can point to entities. Section 2.4 explains how tags can be used to define such a dynamic catalogue structure. Figure 43 gives an overview about the specification of the Nodes API.

/nodes

GET	/nodes/{domain}.{nodekey}	Get a certain node.
POST	/nodes/{domain}/children	Get the children of a certain node from the dynamicchildren section of the node.
GET	/nodes/{domain}.{nodekey}/children	Get the children of a certain node.
GET	/nodes	Get all Rootnodes.

Figure 43: Nodes API Specification

2.6.4.5 Searches

The Search API provides a convenient mechanism to search for specific entities. Similar to the Actions API (section 2.6.4.1), a custom search is implemented as a Java Class in SIP backend. Such a custom search can be parameterised and executed via the Search API. As described in section 2.6.3.1, search can be performed on relational (object entities) as well as on structured (value of "content" properties) meta-data. Apart from the default searches mentioned in section 2.6.3.1., other types of user defined searches, e.g. for aggregated mean values, can be implemented as required. Figure 43 gives an overview about the specification of the Nodes API.

/searches

GET	/searches	Get all custom searches.
GET	/searches/{domain}.{searchkey}	Get a certain custom search.
GET	/searches/{domain}.{searchkey}/results	Execute a custom search.

Figure 44: Searches API Specification

2.6.5 Validation

The SIM does not employ XML explicitly to store meta-data although XML formatted content can be stored using the dynamic content extensions. Meta data validation will not be performed by means of XML-against-XSD-validation. A meta-data entry tool will be built as part of the Expert GUI of the SIP and meta-data validation functionality will be part of this tool and performed during the process of meta-data entry.

3 Conclusions

The Standard Information Model for Meta-Data (SIM) defines the storage of meta-data in the Spatial Information Platform (SIP). The flexible and extendible nature of the SIM has been introduced in deliverable D2.1. More details and specially the applicability of the SIM are demonstrated in this deliverable D2.6.

In this document the SIM is described, first in more technical detail. This provides the detailing for the actual implementation of the SIM in a RDBMS.

The meta-data requirements in SWITCH-ON context have been extensively researched by WP3 which resulted in 2 relevant deliverables (D3.1 and D3.2). The INSPIRE/ISO 19115 meta-data standard is designated as the core standard for SWITCH-ON. The meta-data is split in three categories: basic, quality and lineage meta-data. The way the SIM stores these basic, quality and lineage meta-data and how it supports the find, bind, repurpose and publish concepts is explained in detail.

The meta-data profiles for quality and lineage are subject of a pending discussion between WP2 and WP3 at the time of writing and the outcome will not be available before the delivery of D2.6. It is expected that the outcome results in additions to the profiles. This could result in minor changes to the SIM. The setup of the SIM allows for storage of any meta-data profile. However for more efficient search or to diminish full text searches it might be desirable to move meta-data properties in the SIM from the dynamic content extension to single attribute level.

The find, bind, repurpose/transform and publish support of the SIM and how SIP clients access meta-data to do find, bind, repurpose/transform and publish is described and explained.

A full example of meta-data for data set (HydroSHEDS) in Appendix II: Meta-Data Example and a specification of tags groups indentified so far in Appendix I: Specification of Tag Groups.

4 References

SWITCH-ON Consortium. 2013-08-05. Description of Work (DoW).

SWITCH-ON Consortium. 2014-04-30. D2.1 Spatial Information and Application Programming Interface version 1.0

SWITCH-ON Consortium. 2014-04-30. D3.1 Report on data requirements for a Virtual Water-Science Laboratory.

SWITCH-ON Consortium. 2014-04-30. D3.2 Report on protocols and procedures for quality assurance of experiment procedure

SWITCH-ON Consortium. 2014-04-30. D4.2 Functional and technical design of the products

ISO/TC. 2003. ISO 19115:2003 GEOGRAPHIC INFORMATION - METADATA.
http://www.isotc211.org/Outreach/ISO_TC_211_Standards_Guide.pdf

ISO/TC. 2003. ISO 19119:2005 GEOGRAPHIC INFORMATION – SERVICES.
http://www.isotc211.org/Outreach/ISO_TC_211_Standards_Guide.pdf

Appendix I: Specification of Tag Groups

Access Constraints

Limitations on public access in accordance to Article 13 of Directive 2007/2/EC (fixed group with tags from a standard code list).

access constrains	
name	**description**
no limitation	no limitation
(a) confidentiality provided for by law	(a) the confidentiality of the proceedings of public authorities, where such confidentiality is provided for by law
(b) international relations	(b) international relations, public security or national defence
(c) the course of justice	(c) the course of justice, the ability of any person to receive a fair trial or the ability of a public authority to conduct an enquiry of a criminal or disciplinary nature
(d) confidentiality of commercial information	(d) the confidentiality of commercial or industrial information, where such confidentiality is provided for by national or Community law to protect a legitimate economic interest, including the public interest in maintaining statistical confidentiality and tax secrecy
(e) intellectual property rights	(e) intellectual property rights
(f) confidentiality of personal data	(f) the confidentiality of personal data and/or files relating to a natural person where that person has not consented to the disclosure of the information to the public, where such confidentiality is provided for by national or Community law
(g) information requested on voluntary basis	(g) the interests or protection of any person who supplied the information requested on a voluntary basis without being under, or capable of being put under, a legal obligation to do so, unless that person has consented to the release of the information concerned
(h) protection of the environment	(h) the protection of the environment to which such information relates, such as the location of rare species.

Catchments

Catchment of a Level B dataset (open group with some predefined tags).

catchment	
name	**description**
Arno	Arno
Broye	Broye
Dyfi	Dyfi
Fluttendorf	Fluttendorf
Furtmuehle	Furtmuehle
Gadera	Gadera
Grossarl	Grossarl
Hoan	Hoan
Ioisach	Ioisach
Juktan	Juktan

Kreuzbergmauth	Kreuzbergmauth
Nossan	Nossan
South Tyne	South Tyne
Tanaro	Tanaro
Treene	Treene
Vils	Vils
Waveney	Waveney
Wieselburg	Wieselburg
Wylye	Wylye
...	

Collection

Assigns the resource to a collection of resources for cataloguing purposes, e.g. SWITCH-ON Experiment Results, etc. (open group with some predefined tags).

collection	
name	**description**
HydroSHEDS	n/a
...	

Content Type

MIME Type of the representation (open group which several predefined tags).

content type	
name	**description**
application/adrg	ADRG/ARC Digitilized Raster Graphics
application/agr	ESRI ArcGIS Ascii Grid
application/aib	ARC/INFO Coverages
application/dem	USGS DEM/USGS Digital Elevation Model
application/dlg	DLG/Digital Line Graph
application/drg	DRG/Digital raster graphic
application/gml+xml	Geography Markup Language
application/hdf	hdf5/hierachical data format
application/json	Java Script Object Notation
application/netcdf	netCDF-CF/Network Common Data Form
application/ntf	National Imagery Transmission Format
application/octec-stream	General Binary Data
application/sdf	Spatial Data File
application/shp	ESRI Shapefile
application/vnd.geo+json	GeoJSON/JavaScript Object Notation
application/vnd.google-earth.kml+xml	KML/Keyhole Markup Language
application/wkb	WKB/Well-known binary
application/wkt	WKT/Well-known text
application/xhtml+xml	Extensible HyperText Markup Language
application/xml	eXtensible Markup Language
application/x-netcdf	Network Common Data Format
application/zip	zip archive file format

image/ecw	ECW/Enhanced Compressed Wavelet
image/geotiff	GeoTIFF/Tagged Image File Format
image/gif	Graphigs Interchange Format
image/jp2	JPEG2000
image/jpeg	Joint Photographic Experts Group
image/png	Portable Network Graphics
image/tiff	Tagged Image File Format
image/vnd.dwg	AutoCAD DWG
image/vnd.dxf	AutoCAD DXF
image/x-mrsid	MrSID/Multi-Resolution raster format
text/csv	Comma Separated Values
text/html	Hypertext Markup Language
text/plain	raw, ascii text
text/rtf	Rich Text Format
text/tab-separated-values	Tab Separated Values
...	

Function

Function that can be perfomred following the contentLocation link to the resource representation (fixed group, standard codelist).

function	
name	**description**
information	link provides information about resource
template	link provides template to access resource
download	link will get resource. function of link it to get a representation of the resource
service	link is service endpoint
order	link value is URL of web application requiring user interaction to order/request access to the resource
search	link value is URL of web application requiring user interaction to search/browse/subset the resource.
offlineAccess	link points to the local filesystem

Geography

Geographical classification of the resource, currently Level A and Level B are defined (open group with some predefined tags).

geography	
name	**description**
Level A	Datasets covering Pan-Europe
Level B	Datasets covering a specific Subcatchment

Hydrological Concept

Hydrological concept of a Level A Dataset (open group with some predefined tags).

hydrological concept	
name	**description**
Land-use	Land-use
Modelled river discharge	Modelled river discharge
Observed river discharge	Observed river discharge
PotEvapo	PotEvapo
Precipitation	Precipitation
Soils	Soils
Subbasins	Subbasins
Temperature	Temperature
...	

Keywords - INSPIRE themes 1.0

Keywords from Gemet – INSPIREthemes, version 1.0, publication, 2008-06-01 Hydrography (fixed group).

keywords - INSPIRE themes 1.0	
name	**description**
Addresses	Location of properties based on address identifiers, usually by road name, house number, postal code.
Administrative units	Units of administration, dividing areas where Member States have and/or exercise jurisdictional rights, for local, regional and national governance, separated by administrative boundaries.
Agricultural and aquaculture facilities	Farming equipment and production facilities (including irrigation systems, greenhouses and stables).
Area management/restriction/regulation zones and reporting units	Areas managed, regulated or used for reporting at international, European, national, regional and local levels. Includes dumping sites, restricted areas around drinking water sources, nitrate-vulnerable zones, regulated fairways at sea or large inland waters, areas for the dumping of waste, noise restriction zones, prospecting and mining permit areas, river basin districts, relevant reporting units and coastal zone management areas.
Atmospheric conditions	Physical conditions in the atmosphere. Includes spatial data based on measurements, on models or on a combination thereof and includes measurement locations.
Bio-geographic regions	Areas of relatively homogeneous ecological conditions with common characteristics.
Buildings	Geographical location of buildings.Geographical location of buildings.
Cadastral parcels	Areas defined by cadastral registers or equivalent.
Coordinate reference systems	Systems for uniquely referencing spatial information in space as a set of coordinates (x, y, z) and/or latitude and longitude and height, based on a geodetic horizontal and vertical datum.
Elevation	Digital elevation models for land, ice and ocean surface. Includes terrestrial elevation, bathymetry and shoreline.
Energy resources	Energy resources including hydrocarbons, hydropower, bio-energy, solar,

	wind, etc., where relevant including depth/height information on the extent of the resource.
Environmental monitoring facilities	Location and operation of environmental monitoring facilities includes observation and measurement of emissions, of the state of environmental media and of other ecosystem parameters (biodiversity, ecological conditions of vegetation, etc.) by or on behalf of public authorities.
Geographical grid systems	Harmonised multi-resolution grid with a common point of origin and standardised location and size of grid cells.
Geographical names	Names of areas, regions, localities, cities, suburbs, towns or settlements, or any geographical or topographical feature of public or historical interest.
Geology	Geology characterised according to composition and structure. Includes bedrock, aquifers and geomorphology.
Habitats and biotopes	Geographical areas characterised by specific ecological conditions, processes, structure, and (life support) functions that physically support the organisms that live there. Includes terrestrial and aquatic areas distinguished by geographical, abiotic and biotic features, whether entirely natural or semi-natural.
Human health and safety	Geographical distribution of dominance of pathologies (allergies, cancers, respiratory diseases, etc.), information indicating the effect on health (biomarkers, decline of fertility, epidemics) or well-being of humans (fatigue, stress, etc.) linked directly (air pollution, chemicals, depletion of the ozone layer, noise, etc.) or indirectly (food, genetically modified organisms, etc.) to the quality of the environment.
Hydrography	Hydrographic elements, including marine areas and all other water bodies and items related to them, including river basins and sub-basins. Where appropriate, according to the definitions set out in Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy (2) and in the form of networks.
Land cover	Physical and biological cover of the earth's surface including artificial surfaces, agricultural areas, forests, (semi-)natural areas, wetlands, water bodies.
Land use	Territory characterised according to its current and future planned functional dimension or socio-economic purpose (e.g. residential, industrial, commercial, agricultural, forestry, recreational).
Meteorological geographical features	Weather conditions and their measurements; precipitation, temperature, evapotranspiration, wind speed and direction.
Mineral resources	Mineral resources including metal ores, industrial minerals, etc., where relevant including depth/height information on the extent of the resource.
Natural risk zones	Vulnerable areas characterised according to natural hazards (all atmospheric, hydrologic, seismic, volcanic and wildfire phenomena that, because of their location, severity, and frequency, have the potential to seriously affect society), e.g. floods, landslides and subsidence, avalanches, forest fires, earthquakes, volcanic eruptions.
Oceanographic geographical features	Physical conditions of oceans (currents, salinity, wave heights, etc.).
Orthoimagery	Geo-referenced image data of the Earth's surface, from either satellite or airborne sensors.

Population distribution - demography	Geographical distribution of people, including population characteristics and activity levels, aggregated by grid, region, administrative unit or other analytical unit.
Production and industrial facilities	Industrial production sites, including installations covered by Council Directive 96/61/EC of 24 September 1996 concerning integrated pollution prevention and control (1) and water abstraction facilities, mining, storage sites.
Protected sites	Area designated or managed within a framework of international, Community and Member States' legislation to achieve specific conservation objectives.
Sea regions	Pysical conditions of seas and saline water bodies divided into regions and sub-regions with common characteristics.
Soil	Soils and subsoil characterised according to depth, texture, structure and content of particles and organic material, stoniness, erosion, where appropriate mean slope and anticipated water storage capacity.
Species distribution	Geographical distribution of occurrence of animal and plant species aggregated by grid, region, administrative unit or other analytical unit.
Statistical units	Units for dissemination or use of statistical information.
Transport networks	Road, rail, air and water transport networks and related infrastructure. Includes links between different networks. Also includes the trans-European transport network as defined in Decision No 1692/96/EC of the European Parliament and of the Council of 23 July 1996 on Community Guidelines for the development of the trans-European transport network (1) and future revisions of that Decision.
Utility and governmental services	Includes utility facilities such as sewage, waste management, energy supply and water supply, administrative and social governmental services such as public administrations, civil protection sites, schools and hospitals.

Keywords - Open

User defined to describe the subject (open group with several predefined tags).

keywords - open	
name	**description**
edc	edc
hydrosphere	hydrosphere
topography	topography
gis	gis
hydropattern	hydropattern
digital elevation model	digital elevation model
rivers	rivers
mapping	mapping
terrain elevation	terrain elevation
land surface	land surface
cartography	cartography
streams	streams
usgs	usgs
drainage	drainage
eros	eros

surface water	surface water
3-arc-second dem	3-arc-second dem
watershed characteristics	watershed characteristics
earth science	earth science
...	

Language

The language(s) used within the resource or in which the metadata elements are expressed. The value domain of this tag is limited to the languages defined in ISO 639-2.

language	
name	**description**
bul	Bulgarian
cze	Czech
dan	Danish
dut	Dutch
eng	English
fin	Finnish
fre	French
ger	German
gre	Greek
hun	Hungarian
gle	Irish
ita	Italian
pol	Polish
por	Portuguese
slo	Slovak
spa	Spanish
swe	Swedish

License

License regulating the conditions for access and use of the data (open group with some predefined tags).

license	
name	**description**
CC BY	Creative Commons Attribution 4.0 International (CC BY 4.0) http://creativecommons.org/licenses/by/4.0/legalcode
CC BY-SA	Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) http://creativecommons.org/licenses/by-sa/4.0/legalcode
CC BY-ND	Creative Commons Attribution-NoDerivatives 4.0 International (CC BY-ND 4.0) http://creativecommons.org/licenses/by-nd/4.0/legalcode
CC BY-NC	Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) http://creativecommons.org/licenses/by-nc/4.0/
CC BY-NC-SA	Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) http://creativecommons.org/licenses/by-nc-sa/4.0/legalcode
CC BY-NC-ND	Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) http://creativecommons.org/licenses/by-nc-nd/4.0/legalcode

ODbL	Open Database License (ODbL) v1.0 http://opendatacommons.org/licenses/odbl/1.0/
DbCL	Database Contents License (DbCL) v1.0 http://opendatacommons.org/licenses/dbcl/1.0/
PDDL	ODC Public Domain Dedication and Licence (PDDL) http://opendatacommons.org/licenses/pddl/1.0/
ODC-By	Open Data Commons Attribution License (ODC-By) v1.0 http://opendatacommons.org/licenses/by/1.0/
proprietary	Proprietary License.
...	

Location

Geographic Location of the resource (covered by the data). E.g. Continent, Country, Region, City, etc. Open Group. (open group with some predefined tags)

location	
name	**description**
Africa	Africa
Europe	Europe
...	

Meta-Data Standard

Official standard on which the meta-data record is based (open group with some predefined tags).

meta-data standard	
name	**description**
ISO19115	Geographic information -- Metadata
ISO19119	Geographic information -- Services
FGDC-STD-001-1998	Standard for Digital Geospatial Metadata
DublinCore	Dublin Core Metadata Element Set
...	

Meta-Data Type

SIP internal type of the Meta-Data Record (fixed group).

meta-data type	
name	**description**
SIM meta-data	SIM meta-data
origin meta-data	origin meta-data
quality meta-data	quality meta-data
lineage meta-data	lineage meta-data

Protocol

Protocol of the service that can be accessed at the contentLocation of the resource representation (open group with several predefined tags from standard codelists).

protocol	
name	**description**
OGC:CSW	OGC Catalogue Service for the Web
OGC:SOS	OGC Sensor Observation Service
OGC:WCS	OGC Web Coverage Service
OGC:WFS	OGC Web Feature Service
OGC:WMS	OGC Web Map Service
OGC:WPS	OGC Web Processing Service
ESRI:ArcIMS	ESRI ArcIMS Service
ESRI:ArcGIS	ESRI ArcGIS Service
ESRI:MPK	ArcGIS Map Package
OPeNDAP:OPeNDAP	OPeNDAP root URL
UNIDATA:NCSS	NetCDF Subset Service
UNIDATA:CDM	Common Data Model Remote Web Service
UNIDATA:CdmRemote	Common Data Model index subsetting
UNIDATA:CdmrFeature	Common Data Model coordinate subsetting
OGC:GML	OGC Geography Markup Language
WWW:LINK	Web Address (URL)
WWW:WSDL	Web Service Description Language XML document describing service operation
OpenSearch1.1	OpenSearch template
OpenSearch1.1:Description	OpenSearch description document

Representation Type

SIP internal type of the representation of the resource (fixed group).

representation type	
name	**description**
raw data	raw data
processed data	processed data
preview data	Preview data

Role

Function performed by the responsible party (fixed group, standard codelist).

role	
name	**description**
resourceProvider	Party that supplies the resource. Person or organisation responsible for the availability of the data resource. Different form data distributor, who actively distributes the data resource at user's request.
custodian	Party that accepts accountability for the data and ensures appropriate care and maintenance of the resource. Person or organisation responsible for care and maintenance of the data resource.

owner	Party that owns the resource. Person or organization with the title to intellectual property rights.
user	Party that uses the resource. Person or organization that is, or can be, the key user of the resource.
distributor	Party that distributes the resource. Person or organisation responsible for the distribution of the data resource. Data distributor is not necessarily the owner of data.
originator	Party that created the resource. Person of organization that created the data resource. Can be the same as author, but in cases when a data resource is based on other resources, the creator cannot be the author.
pointOfContact	Party that can be contacted for acquiring knowledge about or acquisition of the data resource. Person or organization that can be contacted to acquire data on the resource.
principalInvestigator	Key party responsible for gathering information and conducting research. Key person responsible for gathering information and conducting research resulting in the data resource. Appointed principal investigator or project manager or leading researcher.
processor	Party that has processed the data in a manner that the resource has been modified. Person or organization processing the data in the described form. Applicable only if the data has been subsequently processed or modified.
publisher	Party that publishes the resource. Person of organization that published the data resource.
author	Party that authorized the resource. Party that created the resource. More often, the party that published the data resource is listed than the party author of "raw" data. For instance, the person or group of persons or the organization is listed that created the dataset (collected data from multiple resources and created a data resource) or published the reviewing service.

Scope

Scope of the resource. A codelist used in hierarchyLevel, DQ_Scope.level, and updateScope (fixed group with tags from a standard codelist)

scope	
name	**description**
attribute	information applies to the attribute class
attributeType	information applies to the characteristic of a feature
collectionHardware	information applies to the collection hardware class
collectionSession	information applies to the collection session
dataset	information applies to the dataset
series	information applies to the series
nonGeographicDataset	information applies to non-geographic data
dimensionGroup	information applies to a dimension group
feature	information applies to a feature
featureType	information applies to a feature type
propertyType	information applies to a property type
fieldSession	information applies to a field session
software	information applies to a computer program or routine
service	information applies to a capability which a service provider entity makes available to a service user entity through a set of interfaces that define a behaviour, such as a use case

model	information applies to a copy or imitation of an existing or hypothetical object
tile	information applies to a tile, a spatial subset of geographic data

Srid

The Spatial Reference System Identifier (SRID) of the spatial coverage of the resource EPSG format (open group with some predefined tags).

srid	
name	**description**
EPSG:4326	WGS 84 / World Geodetic System 1984
...	

Topic Category

High-level classification of resources in accordance with ISO 19115 for grouping and topic-based search (fixed group).

topic category	
name	**description**
farming	rearing of animals and/or cultivation of plants Examples: agriculture, irrigation, aquaculture, plantations, herding, pests and diseases affecting crops and livestock
biota	flora and/or fauna in natural environment Examples: wildlife, vegetation, biological sciences, ecology, wilderness, sealife, wetlands, habitat
boundaries	legal land descriptions Examples: political and administrative boundaries
climatologyMeteorology Atmosphere	processes and phenomena of the atmosphere. Examples: cloud cover, weather, climate, atmospheric conditions, climate change, precipitation
economy	economic activities, conditions and employment. Examples: production, labour, revenue, commerce, industry, tourism and ecotourism, forestry, fisheries, commercial or subsistence hunting, exploration and exploitation of resources such as minerals, oil and gas
elevation	height above or below sea level Examples: altitude, bathymetry, digital elevation models, slope, derived products
environment	environmental resources, protection and conservation Examples: environmental pollution, waste storage and treatment, environmental impact assessment, monitoring environmental risk, nature reserves, landscape
geoscientificInformation	information pertaining to earth sciences Examples: geophysical features and processes, geology, minerals, sciences dealing with the composition, structure and origin of the earth's rocks, risks of earthquakes, volcanic activity, landslides, gravity information, soils, permafrost, hydrogeology, erosion
health	health, health services, human ecology, and safety Examples: disease and illness, factors affecting health, hygiene, substance abuse, mental and physical health, health services

imageryBaseMapsEarthCover	base maps Examples: land cover, topographic maps, imagery, unclassified images, annotations
intelligenceMilitary	military bases, structures, activities Examples: barracks, training grounds, military transportation, information collection
inlandWaters	inland water features, drainage systems and their characteristics Examples: rivers and glaciers, salt lakes, water utilization plans, dams, currents, floods, water quality, hydrographic charts
location	positional information and services Examples: addresses, geodetic networks, control points, postal zones and services, place names
oceans	features and characteristics of salt water bodies (excluding inland waters) Examples: tides, tidal waves, coastal information, reefs
planningCadastre	information used for appropriate actions for future use of the land Examples: land use maps, zoning maps, cadastral surveys, land ownership
society	characteristics of society and cultures Examples: settlements, anthropology, archaeology, education, traditional beliefs, manners and customs, demographic data, recreational areas and activities, social impact assessments, crime and justice, census information
structure	man-made construction Examples: buildings, museums, churches, factories, housing, monuments, shops, towers
transportation	means and aids for conveying persons and/or goods Examples: roads, airports/airstrips, shipping routes, tunnels, nautical charts, vehicle or vessel location, aeronautical charts, railways
utilitiesCommunication	energy, water and waste systems and communications infrastructure and services Examples: hydroelectricity, geothermal, solar and nuclear sources of energy, water purification and distribution, sewage collection and disposal, electricity and gas distribution, data communication, telecommunication, radio, communication networks

Appendix II: Meta-Data Example

The following tables show a complete example of a SWITCH-ON meta-data record for a HydroSHEDS[12] resource. This meta-data record is based on different sources and has been created according to the specification of the SIM (chapter 2).

Descriptive meta-data of the resource:

Resource	uuid	537f6c6be4b021317a87279c
	name	HydroSHEDS (ACC) - Flow accumulation [cells] at 15s resolution for Africa
	description	HydroSHEDS (Hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales) provides hydrographic information in a consistent and comprehensive format for regional and global-scale applications. HydroSHEDS offers a suite of geo-referenced data sets (vector and raster), including stream networks, watershed boundaries, drainage directions, and ancillary data layers such as flow accumulations, distances, and river topology information. HydroSHEDS is derived from elevation data of the Shuttle Radar Topography Mission (SRTM) at 3 arc-second resolution. Available HydroSHEDS resolutions range from 3 arc-second (approx. 90 meters at the equator) to 5 minute (approx. 10 km at the equator) with seamless near-global extent.

Tags of the resource (Dynamic Tag Extensions):

Resource	tags	topic category	geoscientificInformation
		keywords -	Hydrography, Land cover
		keywords - open	edc, hydrosphere, topography, gis, hydropattern, digital elevation model, rivers, mapping, terrain elevation, land surface, cartography, streams, usgs, drainage, eros, surface water, 3-arc-second dem, watershed, characteristics, earth science
		scope	data
		srid	EPSG:4326
		location	Africa
		collection	HydroSHEDS
		license	non-commercial proprietary
		access constraints	no limitation
		geography	n/a
		hydrological concept	n/a
		catchments	n/a
		language	en

Spatial and temporal meta-data of the resource:

Resource	spatialCoverage	POLYGON((55.0004 38,55.0004 -35.0003,-19 -35.0003,-19 38,55.0004
	fromDate	2002-01-01
	toDate	2014-06-18
	creationDate	2010-10-13 10:54:25
	publicationDate	2014-05-23 09:41:37 MDT
	lastModificationDate	2010-10-28 07:01:19
	licenseUrl	http://www.hydrosheds.org/page/license

Contact information of the resource: Owner or data provider of the resource.

[12] http://hydrosheds.org

Resource	contact	name		n/a
		description		HydroSHEDS has been developed by the WWF Conservation Science Program in partnership with the U.S. Geological Survey, the International Centre for Tropical Agriculture, The Nature Conservancy, and the Center for Environmental Systems Research of the University of Kassel, Germany.
		organisation		World Wildlife Fund, Inc.
		email		hydrosheds@wwfus.org
		url		http://www.hydrosheds.org/
	contact	name		n/a
		description		
		organisation		U.S. Geological Survey EROS Data Center
		email		hydrosheds@wwfus.org
		url		http://hydrosheds.cr.usgs.gov/

SWITCH-ON meta-data on meta-data about the resource: Description when an and by whom the SWITCH-ON meta-data record has been created.

Resource	metadata	uuid		n/a
		tags	meta-data type	SIM Meta-Data
			language	en
			meta-data standard	n/a
		name:		SWITCH-ON SIM Meta-Data
		description		Meta-Data record created, derived or imported by the SWITCH-ON project according to the SWITCH-ON Standard Information Model (SIM) for Meta-Data for the SWITCH-ON Spatial Information Platform (SIP).
		+ contact	name	Emma Lagerbäck Adolphi
			description	Sharing Water-related Information to Tackle Changes in the Hydrosphere – for Operational Needs (SWITCH-ON).
			organisation	SWITCH-ON
			email	emma.l.adolphi@gmail.com
			url	http://www.water-switch-on.eu/
		creationDate		2014-06-11
		contentType		n/a
		contentLocation		n/a
		content		n/a

Meta-data available at the original resource location: A link to a catalogue entry which provides original meta-data about the resource in ISO 19119 format.

Resource	metadata	uuid		537f6c31e4b021317a8720e4
		tags	meta-data type	origin meta-data
			language	en
			meta-data standard	ISO19115
		name		HydroSHEDS (ACC) - Flow accumulation [cells] (Africa) at 15s resolution ISO 19115 Meta-Data
		description		ISO 19115 Meta-Data for "HydroSHEDS (ACC) - Flow accumulation [cells] at 15s resolution (Africa)" data from the ScienseBase Catalogue.
		+ contact	name	n/a
			description	n/a
			organisation	Conservation Biology Institute
			email	anovacek@usgs.gov
			url	https://www.sciencebase.gov/
		creationDate		May 23 09:41:37 MDT 2014
		contentType		text/xml
		contentLocation		https://www.sciencebase.gov/catalog/item/537f6c31e4b021317a8720e4?format=iso
		content		n/a

Quality meta-data about the resource: A link to a website describing the quality and lineage of the data.

Resource	metadata	uuid		n/a
		tags	meta-data type	quality meta-data
		name		Meta-Data for Data Quality
		description		Meta-Data for Data Quality collected from the official HydroSHEDS website.
		+ contact	name	n/a
			description	The leading organization in wildlife conservation and endangered species.
			organisation	World Wildlife Fund, Inc.
			email	n/a
			url	http://www.hydrosheds.org/
		creationDate		2014-06-20
		contentType		text/html
		contentLocation		http://www.hydrosheds.org/page/development
		content		n/a

A representation of the resource: Download link to the original data set of the resource.

Resource	representation	uuid		537f6c31e4b021317a8720e4
		name		15sec ESRI ArcInfo Grid Flow Accumulation for Africa
		description		ZIP Package containing the data in ESRI ArcInfo Grid Format. The ESRI Grid format is supported by the ESRI GIS software environment, but it can also be imported into various independent GIS packages. Each raster coverage is provided in a root folder (named after the raster layer) which contains two subfolders: the grid data folder (again named after the raster layer) and a corresponding "info" folder. In the grid folder, basic metadata information is provided both in XML format (.xml) and in HTML format (.htm) following the FGDC standard. Projection information is provided in an ASCII text file (prj.adf). All raster data are in geographic (latitude/longitude) projection, referenced to datum WGS84.
		tags	protocol	WWW:LINK-1.0-http--link
			representation type	raw data
			application profile	Web Browser
			function	download
		contentType		application/zip
		contentLocation		http://earlywarning.usgs.gov/hydrodata/sa_15s_zip_grid/af_acc_15s_grid.zip
		content		n/a

Another representation of the resource: Link to a Web Map Service (WMS).

Resource	representation	uuid		537f6c31e4b021317a8720e4
		name		WMS Service for HydroSHEDS data
		description		HydroSHEDS (ACC) - Flow accumulation [cells] at 15s resolution for Africa available as OGC WMS Layer.
		tags	protocol	OGC:WMS
			representation type	service
			application profile	WMS Client
			function	online access
		contentType		application/xml
		contentLocation		https://www.sciencebase.gov/catalogMaps/mapping/ows/537f6c31e4b021317a8720e4?service=wms&request=getcapabilities&version=1.3.0
		content		n/a

Another representation of the resource: Link to a Web Feature Service (WFS).

Resource	representation	uuid		537f6c31e4b021317a8720e4
		name		WFS Service for HydroSHEDS data
		description		HydroSHEDS (ACC) - Flow accumulation [cells] at 15s resolution for Africa available as OGC WFS Layer.
		tags	protocol	OGC:WFS
			representation type	service
			application profile	WFS Client
			function	online access
		contentType		application/xml
		contentLocation		https://www.sciencebase.gov/catalogMaps/mapping/ows/537f6c31e4b021317a8720e4?service=wfs&request=getcapabilities&version=1.0.0
		content		n/a

Another representation of the resource: Link to a preview image of the resource.

Resource	representation	uuid		
		name		Preview image
		description		Thumbnail Preview Image of the data.
		tags	protocol	WWW:LINK-1.0-http--link
			representation type	preview
			application profile	Image Viewer
			function	download
		contentType		preview/png
		contentLocation		http://databasin2-filestore.s3.amazonaws.com/c418bccf7af8416dadd3f63f0e2865cb/images/preview.png?v=1288270879
		content		n/a

Relationship of the Resource: Describing the lineage of the resource.

Realtionship	name			HydroSHEDS Data Sources
	description			HydroSHEDS is derived primarily from elevation data of the Shuttle Radar Topography Mission (SRTM) at 3 arc-second resolution. SRTM data were used in different versions and processing stages, including SRTM-3 unfinished data; DTED-1 finished data; and CGIAR void-filled SRTMv2 data. A variety of auxiliary datasets were used for reference and quality control, including the SRTM Water Body Data (SWBD); the river network of the Digital Chart of the World (also known as VMAP0); the Global Lakes and Wetlands Database (GLWD); and several other global and regional digital river maps.
	tags	realtionship type		derived
	fromResources			n/a

Meta-data about the relationship of the Resource: Link to the detailed technical documentation about collection and processing of data.

Realtionship	metadata	uuid		n/a
		tags	meta-data type	lineage meta-data
			language	en
			meta-data standard	n/a
		name:		Technical Documentation
		description		Technical Documentation of HydroSHEDS Data Sources.
		+ contact	name	Bernhard Lehner
			description	The leading organization in wildlife conservation and endangered species.
			organisation	World Wildlife Fund, Inc.
			email	n/a
			url	http://www.hydrosheds.org/
		creationDate		2013-06-01
		contentType		application/pdf
		contentLocation		http://hydrosheds.org/images/inpages/HydroSHEDS_TechDoc_v1_2.pdf
		content		n/a