GUIDELINES FOR TCS
Metadata Quality

July, 2019

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 676564

TABLE OF CONTENTS

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT N° 676564

1 Introduction

This document is a guideline to help TCS' communities provide their domain specific concepts and related keywords via EPOS-DCAT-AP metadata. Such information will be used to better support discovery and access to TCS resources such as datasets and webservices.

Section 2 describes in detail how to fill in the most relevant metadata elements by illustrating where the metadata values appear in the EPOS portal (2.1) and by providing some basic rules about how some of the metadata values should look like (or shouldn't).

Section 3 points to contacts for getting additional support, together with references for further study.

2 Metadata Quality Control

2.1 Visual presentation of metadata in GUI

The provision of quality metadata aims at: 1) guaranteeing the appropriate functioning of the EPOS Integrated Core Services (ICS) System -- for instance for searching and filtering resources of interest; and 2) providing useful and meaningful information to end-users through the EPOS Graphic User Interface (GUI).

Clear and correct information helps users understand what data they can access, download, and also what services are available.

In Figure 1, a mockup of the EPOS GUI is presented, which illustrates where the actual metadata values appear. This provides a better idea about what metadata values are visualized in what elements of the GUI.

In the following sections, we outline some basic rules about how to fill in information in the EPOS-DCAT-AP files, also by providing examples of what metadata values should or shouldn't be provided in order to furnish useful and clear information to the end users.

Numbers in Figure 1 relates to specific boxes in the GUI. Box 1 is covered in section 2.2, where metadata elements for discovery of resources, search and filters are discussed; Box 2, discussed in section 2.3, is related to metadata elements describing resources details; metadata elements influencing Box 3, i.e. (re)configuration of resources, are discussed in section 2.4.

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT N° 676564

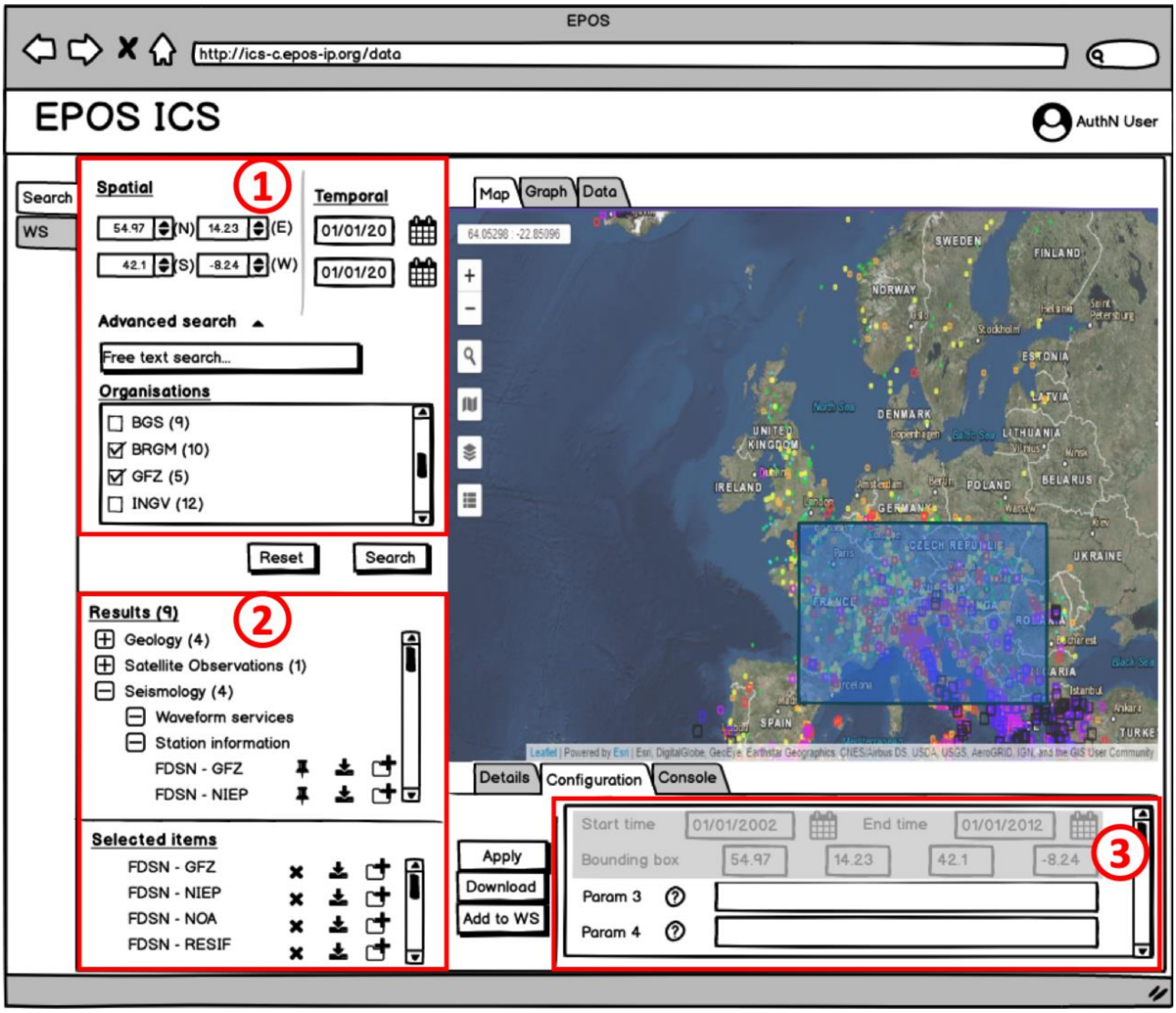

Figure 1 - A mockup of the EPOS GUI. Red numbered boxes mark three main tasks that can be achieved on the GUI's "search"-panel: (1) Discovery, search, and filtering; (2) listing and display of resource details; (3) (re-)configuration of resources.

2.2 Resource discovery

The EPOS portal will support a number of mechanisms to search for and to discover community resources such as Dataset and Webservice. In particular, it will enable users to:

- look for specific terms or keywords associated with a resource;
- navigate through facets that combine groups of resources;
- filter resources provided by a specific Institution and/or Organisation;
- apply spatio-temporal filters.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 676564

As such functionalities are enabled by the underpinning information maintained in the metadata catalogue, it is important to verify that this information is as correct and complete as possible.

Keywords

Keywords need to be associated with **Dataset** and **Webservice**. The values of such keywords should come from a controlled list which is available here. We recommend to use the values in the column "*proposed*". This activity is fundamental for the proper and correct functioning of the search features, therefore this task has the highest priority.

In addition, terms and keywords should be organised and represented in the EPOS-DCAT-AP notation as *Simple Knowledge Organization System* (SKOS) Concepts and Concept Schemes. This activity is desirable but it has a lower priority. The idea is that each TCS would create a file named `[TCSname]_Concepts.ttl`. Such a file would contain one or more `skos:ConceptScheme` defining the category(-ies) of the corresponding TCS, and several `skos:Concept` elements. An example can be found here.
The values assigned to `skos:prefLabel` are the "*proposed*" keywords, whereas one or many `skos:altLabel` can be used to represent the synonyms indicated here.

Dataset			
Metadata element	**Rules**	**Example**	**Wrong values**
`dcat:keyword`	Multiple entries in a keywords list are typically delimited by commas.	`dcat:keyword "underground mining", "anthropogenic seismicity";`	

WebService			
Metadata element	**Rules**	**Example**	**Wrong values**
`schema:keywords`	Multiple entries in a keywords list are typically delimited by commas.	`schema:keywords "anthropogenic hazards", "episodes";`	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 676564

Facets

The facets organise related resources into groups of elements structured by domain. At the moment (until the September 2019 release) facets are static and defined in a JSON file according to the descriptions available here. These have been provided by TCS WP leaders who are responsible for their definition and maintenance.

When the Concept Schemes are in place facets can be dynamically built based on the information about categories maintained therein (this will happen after September 2019).

Institutions/organizations

Organizations and Institutions should be described by the **Organization** entity; `schema:legalName` property will be used by the GUI in the facets search. Please ensure that you only use organization names that agree with the controlled list of organizations available at the following link (if you discover errors or missing institutions in the controlled list, please correct them or leave a comment): https://docs.google.com/spreadsheets/d/1xcYHyh5_cS3sCevSmieZD5oTvPMi7s8xQZDP6RRtA_8/edit#gid=0

Dataset			
Metadata element	**Rules**	**Example**	**Wrong values**
`schema:legalName`	Full English language name. Optionally full native language name. Do not use abbreviations.	`schema:legalName` "European-Mediterranean Seismological Centre"**@en** ; `schema:legalName` "Istituto Nazionale di Geofisica e Vulcanologia"**@it** ;	`schema:legalName` "EMSC"

After defining the Organization entities, it is possible to link each of them to one or more Dataset (by using `dct:publisher`) and Webservice (by using `schema:provider`). It is possible to define multiple DataProviders institutions for the same ServiceProvider.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 676564

Spatio-temporal information

The spatio-temporal coverage needs to be specified by using `dct:spatial` and `dct:temporal` properties in **Dataset** and **WebService** entities, as shown in the following tables.
More details about spatio-temporal aspects have been included in the previous guidelines available on GitHub (https://github.com/epos-eu/EPOS-DCAT-AP/blob/EPOS-DCAT-AP-shapes/docs/GUIDELINES%20FOR%20TCS%20Metadata%20Mapping%20-%20Spatio-temporal%20information.pdf).

Dataset and WebService			
Metadata element	**Rules**	**Example**	**Wrong values**
dct:temporal	The date format is “YYYY-MM-DDThh:mm:ssZ”	dct:temporal [a dct:PeriodOfTime ; schema:startDate "1992-03-23T02:15:00Z"^^xsd:dateTime ; schema:endDate "2010-12-20T23:00:00Z"^^xsd:dateTime ;];	
dct:spatial	POLYGON format is: POLYGON((lon1 lat1, lon2 lat2,...)) POINT format is: POINT (lon lat elevation)	dct:spatial [a dct:Location ; locn:geometry "POLYGON((3.053 47.975, 7.24 47.975, 7.24 53.504, 3.053 53.504, 3.053 47.975))"^^gsp:wktLiteral ; locn:geometry "POINT(3.053, 47.975)"^^gsp:wktLiteral ;];	usage of single round bracket for POLYGON(3.053 47.975, 7.24 47.975, 7.24 53.504, 3.053 53.504, 3.053 47.975) OR usage of double round bracket for POINT((3.053 47.975))

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 676564

2.3 Resource details

The values of the following properties are used to show the details of the resource. In particular, dct:title property is used as the name of the resource into the results list (Figure 1 - box 2), while the other properties are used to populate the Details pane (Figure 1 - box 3).

We ask service providers to take special care in reviewing the Resource names (`dct:title`, `schema:name`) and their descriptions (`dct:description`, `schema:description`). Names and descriptions should briefly but clearly describe what a service is about. Domain-experts who know the service should be able to identify it, but also geoscientists who have never used the service should be able to understand its purpose. The tables below list some guidelines and examples on the choice of titles/names and descriptions.

Distribution			
Metadata element	**Rules**	**Example**	**Wrong values**
`dct:title`	1. Should succinctly describe the Distribution using common terminology, i.e., understandable by geoscientists. 2. Should not include institution/organization information (exception: if multiple resources have the same name, the institution name can be appended. 3. Should not use any acronyms.	~~NIEP~~ Radon counts and ~~local~~ air temperature ~~Distribution~~ Seismic waveforms from the National Observatory of Athens	
`dct:license`	Should list the licence under which the Distribution is made available	dct:license "https://creativecommons.org/licenses/by-nc/4.0/"^^xsd:anyURI ;	
`dct:description`	1. Should be written in complete sentences. 2. Should provide information on (a) what the distribution provides, (b) any standard format that it may adhere to (e.g.,CSV, NetCDF, XML), (c) what important coverage limits apply to the underlying data (e.g., specific region).	Suggested standard formulation: *"This distribution provides access to dataset XYZ, including x, y, and z. It is encoded in the xy standard format, version X. Its related dataset covers the time period since YYYY and is limited to countryX."*	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 676564

	3. Should not include institution/organization information.	*E.g.:* This distribution provides access to time series of temperatures and radon gas concentrations at sampling sites. ….	

WebService			
Metadata element	**Rules**	**Example**	**Wrong values**
`schema:description`	1. Should be written in complete sentences. Start with a capital letter. 2. Should provide information on (a) what the service provides, (b) any implementation standards that it may adhere to (e.g., FDSN station web service 1.0), (c) whether any limits apply to the response (e.g., maximum 1000 events), and (d) what important coverage limits apply to the underlying data (e.g., specific region or a catalog's start time). 3. Should not include institution/organization information.	Access to seismic station, instrumentation, and response data ~~maintained by SED~~. Service implements FDSN station web service 1.0 standard Suggested standard formulation: *"This web service provides access to XYZ, including x, y, and z. It is implemented according to the xy web service standard, version X. It returns no more than NNN measurements per request. The catalog/dataset covers the time period since YYYY and is limited to countryX."* *E.g.:* This web service provides access to seismic station metadata, including locations, operation periods, instrumentation, and response data. It is implemented according to the standard 1.0 for seismic station web services set by the Federation of Digital Seismograph Networks (FDSN).	
`schema:name`	1. Should succinctly describe the Resource using common terminology (i.e., understandable by geoscientists). 2. Should not include institution/organization information (exception: if multiple resources have the same basic name, then the	Waveform Catalog of the European Integrated Data Archive	`EIDA WFCatalog ~~- Swiss Seismological Service at ETH Zurich~~`

	institution can added behind the basic name). 3. Should not use any acronyms.		

2.4 Resource (re)configuration

When the resource is a web service, it needs to be described properly by specifying and defining the parameters that allow to query it.
The configuration pane (Figure 1 - box 3), is created by using **Operation** entity and its properties.

In particular: **rdfs:label** is used to describe the meaning of the parameter; **hydra:required**, indicates whether the parameter is mandatory or not; **hydra:property**, contains the semantic description of the parameter (e.g., epos:westernmostLongitude). It is also used by the GUI to organize parameters in the configuration pane in three groups: "spatio-temporal", "mandatory" and "optional".

The output format of the web service payload needs to be specified by defining **hydra:returns** property. In case the web service supports multiple output formats, then they need to be specified by defining: a) multiple occurrences of **hydra:returns** property (one for each output format); b) a **hydra:property** property with a "**schema:encodingFormat**" value, as shown below.

```
hydra:mapping [ a hydra:IriTemplateMapping;
     hydra:variable "format"^^xsd:string;
     rdfs:label "Output format";
     hydra:property "schema:encodingFormat";
     rdfs:range "xsd:string";
     http:paramValue "JSON"^^xsd:string;
     http:paramValue "CSV"^^xsd:string;
     schema:defaultValue "JSON"^^xsd:string;
     hydra:required "false"^^xsd:boolean;
];
```

In order to help users understand how to use the web service parameters, it is possible to link **APIDocumentation** entity to the **WebService** entity. The documentation URL, specified by using **hydra:entrypoint** property, will appear in the GUI configuration pane.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 676564

Operation			
Metadata element	**Rules**	**Example**	**Wrong values**

`rdfs:label`	This string describes the meaning of the parameter.
	1. Use short (ca. 10-50 characters), but clear descriptions.
	2. Follow standard capitalization rules (first word, proper nouns, etc.) for better readability.
	2. Denote the physical units
	3. Use standard wording for the following labels (if applicable, adjust only if needed):
	Minimum longitude (deg)
	Maximum longitude (deg)
	Minimum latitude (deg)
	Maximum latitude (deg)
	Start time
	End time
		`rdfs:label "Network code" ;`
		`rdfs:label` “Date after start time (in days)” :
		`rdfs:label “Maximum radius (deg)” :`	

`hydra:returns`	The possible values are listed here: https://www.iana.org/assignments/media-types/media-types.xhtml	`hydra:returns "application/vnd.fdsn.mseed" ;`	

`hydra:property`	`schema:startDate;`
	`schema:endDate;`
	`epos:southernmostLatitude;`
	`epos:northernmostLatitude;`
	`epos:westernmostLongitude;`
	`epos:easternmostLongitude;`
	`schema:encodingFormat;`
		`hydra:property "schema:endDate" ;`	

`hydra:required`	`“true” or “false”;`	`hydra:required “true”;`	

WebService			
Metadata element	**Rules**	**Example**	**Wrong values**
`dct:conformsTo`		`dct:conformsTo <https://www.epos-eu.org/epos-dcat-ap/Seismology/WebService/001/APIDocumentation> ;`	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 676564

APIDocumentation			
Metadata element	**Rules**	**Example**	**Wrong values**
`hydra:entrypoint`		`hydra:entrypoint "https://www.emidius.eu/AHEAD/services/"^^xsd:anyURI;`	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 676564

3 Contacts and references

Contacts:

- **Rossana Paciello**, rossana.paciello@ingv.it, for Metadata description issues;
- **Jan Michalek**, jan.michalek@uib.no, for DDSS-ID and other issue;
- **Felix Halpaap,** felix.halpaap@uib.no, for DDSS descriptions;
- **Luca Trani,** trani@knmi.nl, for Concepts related issues

References:

- **EPOS-DCAT-AP documentation**: https://github.com/epos-eu/EPOS-DCAT-AP/blob/EPOS-DCAT-AP-shapes/docs/EPOS-DCAT-AP_extension_v0.1%20-%202019.07.10.pdf
- **EPOS-DCAT-AP RDF/Turtle example:** https://github.com/epos-eu/EPOS-DCAT-AP/blob/EPOS-DCAT-AP-shapes/examples/EPOS-DCAT-AP_example.ttl

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 676564