

Specifications to describe ELI Datasets

Table of contents

Introduction	2
Objectives of ELI Dataset description	2
Summary of an ELI Dataset	3
ELI dataset “1 star”: plain list	3
The URI list file	3
File structure	3
Dissemination of URI list file	4
The Dataset entry in a national open-data portal	4
Rationale	4
Description template of an ELI Dataset	4
ELI dataset “2 stars”: regular file update	6
The URI list file	6
The Dataset entry update	6
ELI dataset “3 stars”: date of modification of each ELI	6
The URI list file	6
File structure	6
Example of 3-stars URI list with timestamps	7
Dissemination of the file	7
The Dataset entry in the open-data portal	7
ELI dataset “4 stars”: dump exports	7
Annex : disseminating multiple ELI URI Lists	8

Introduction

As ELI progresses toward making legal metadata more reusable, it should lower the barrier for data consumers to retrieve the ELI metadata from an ELI provider. The role of ELI Datasets is to lower this barrier.

This document describes a **specification for the description and dissemination of ELI Datasets**. ELI Datasets can be regarded as the “Pillar 4” of ELI, as it goes one step further for ELI providers in making their legal data more reusable.

In this document we use the terms “ELI providers” and “ELI consumers” to refer respectively to organizations that publish ELI-compatible metadata and to organizations/individuals willing to retrieve and use the published metadata.

This document refers to the notions of “Dataset”, “Distribution” and their descriptive properties from the “DCAT Application Profile for data portals in Europe” available at <https://joinup.ec.europa.eu/solution/dcat-application-profile-data-portals-europe>.

Objectives of ELI Dataset description

The first objective of an ELI Dataset description is to **provide the exhaustive and up-to-date list of the ELI of LegalResources published by a data provider**. This list is a key information for data consumers, as it guarantees that they can access the complete metadata set of a provider.

To retrieve a complete set of metadata from an ELI provider, an ELI data consumer can do the following :

1. Access the ELI Dataset description of that provider from an open-data portal;
2. For each LegalResource ELI listed in the description:
 - a. Follow the ELI to fetch the corresponding HTML page;
 - b. Parse the semantic markup in this page;
 - c. Aggregate the data in a single file / database, depending on its use;

The second objective of the ELI Dataset description is to **make ELI data visible in open data portals, at national and European levels**. Open-data portals rely on this notion of “Dataset” to describe their notices.

In particular, all the ELI dataset descriptions, once published in national open-data portal, will be automatically aggregated in the European Data Portal (<https://www.europeandataportal.eu/data/en/organization>) that will provide a central point of access to all the ELI Dataset descriptions.

Also, ELI Dataset descriptions can serve as a hub in the future to indicate other sources where the data can be found. Typically in the case that an ELI provider also provides downloadable “dump” files.

Summary of an ELI Dataset

To publish its ELI Dataset, an ELI provider must take the following steps:

1. Generate a **URI list file**: a file that lists all the ELI of LegalResources in the Dataset;
2. Usually **publish this URI list file** at a known URL (or upload it to an open-data portal, see next step);
3. Create or update a **Dataset entry in its national open-data portal, following a specific template**: this entry will contain descriptive information about the dataset, and will refer to the URI list file;
4. (Automatic) wait for its dataset to be **harvested by the European Data Portal** (<https://www.europeandataportal.eu>) that aggregates national open-data portals;

The ELI Dataset can be provided with different level of precision:

- ★ 1 star: publish the plain list of URIs, with few descriptive metadata;
- ★★ 2 stars: update the file on a regular basis, and give the date at which the list file was generated;
 - this avoids unnecessary work for clients that want to access the list, if it has not been modified since the last time they fetched it;
- ★★★ 3 stars: give the date at which *each ELI* in the list was last modified;
 - This makes it possible for a client to fetch only the ELIs added / modified since the last time they fetched the list;
- ★★★★ 4 stars: instead of providing the list of URIs, directly provide an RDF “dump” of all the data.
 - This makes it even easier for an ELI consumer since the parsing and retrieval of the ELI metadata becomes unnecessary;

A higher score in this scale means easier data access for consumers. To follow the general philosophy of ELI, providers implement the degree of precision of their choice, based on the feasibility and availability of the data.

ELI dataset “1 star”: plain list

The URI list file

File structure

To conform to ELI dataset description “1 star”, the ELI URI list:

- MUST be a plain text file;
- MUST be an **exhaustive** list of all the ELIs of LegalResources in the scope of the ELI implementation¹; the ELI provider MUST ensure that all the metadata of its ELI dataset can be fetched by parsing the metadata embedded into the pages listed in this file;
- MUST provide an **up-to-date** list of the available ELIs at the time of its publication;

Example of URI list file

¹ Splitting the list is possible, see section below.

http://country.eu/eli/law/2016/501/001/jo
http://country.eu/eli/law/2016/501/002/jo
http://country.eu/eli/law/2016/501/003/jo
http://country.eu/eli/decree/2005/199/999/jo

Dissemination of URI list file

The URI list file SHOULD be made accessible from a web server of the ELI provider at /eli/uri-list.txt. For example, given a legal web portal at <http://country.eu>, the URI list file will be accessible at <http://country.eu/eli/uri-list.txt>.

This is not a strong requirement, but we think it can make life easier for data consumers if the ELI list files are being published in consistent manner by all ELI partners.

In case the file is split, multiple files will be published (see section below).

Also, depending on national open-data portals, the files may be uploaded to the open-data portal rather than referenced at this URL.

The Dataset entry in a national open-data portal

Rationale

The URI list needs to be advertised in an open-data portal, at national level. ELI providers need to contact the responsible for their national open-data portal to understand how to publish and update a Dataset on their portal.

Description template of an ELI Dataset

For consistency across ELI partners, the description of the ELI Dataset MUST conform to the template given below:

Property/field	URI in DCAT	Description
Title of the dataset	dct:title	This property contains a name given to the Dataset. The property CAN be repeated for parallel language version of the name. The value MUST be expressed in the official language(s) of the country and MUST be repeated in English if the open-data portal allows it, with the value " ELI (European Legislation Identifier) - <country name> ".
Description of the dataset	dct:description	This property contains a free text account of the Dataset. The property CAN be repeated for parallel language version of the description. The value MUST be expressed in the official language(s) of the country and MUST be repeated in

		<p>English if the open-data portal allows it; the English value MUST begin by “This is the ELI dataset of <country name>. ELI (European Legislation Identifier) defines a common metadata model for sharing legislation description on the web of data. <rest of the description>”</p> <p>The description MUST indicate the exact coverage of ELI implementation (e.g. “all acts after 2014”)</p>
Distribution	dcat:distribution	The Dataset MUST have at least one associated distribution, that is the URI list file.
Distribution download URL	dcat:downloadURL (on the distribution entity)	This property MUST points to the URL of the URI list file, typically http://country.eu/eli/uri-list.txt
Distribution media type	dcat:mediaType (on the distribution entity)	The media type of the distribution. The value MUST be “text/plain”
Keyword / tag	dcat:keyword	<p>A keyword or tag describing the Dataset. The value MUST be “ELI”. Another value MUST be “European Legislation Identifier”.</p> <p>The property CAN be repeated for other values.</p>
Theme	dcat:theme	<p>A category of the Dataset. The value MUST be the entry defined in DCAT : “Justice, legal system and public safety” (URI http://publications.europa.eu/resource/authority/data-theme/JUST)</p> <p>The property CAN be repeated for other themes if needed.</p>
Publisher	dct:publisher	<p>The organization responsible for making the Dataset available. This MUST be a reference to a URI identifying the publisher.</p>
documentation	foaf:page	<p>A page or document about the dataset. This CAN refer to the documentation of the specific ELI implementation of the ELI provider if it exists.</p>
Coverage	dct:spatial	<p>The spatial coverage of the Dataset. The value MUST indicate the country being covered, either trough a String or a URI (e.g. “France”).</p>

ELI providers are encouraged to provide detailed description of their dataset with further descriptive metadata allowed by the national open-data portal, typically if their ELI implementation is restricted to a subset of the legislation (“All acts after 20xx”, “all acts of type xxx”, etc.).

ELI dataset “2 stars”: regular file update

The URI list file

No change in the format of the URI list file are implied by the “2 stars” level.

However, the ELI provider **MUST** make sure the file is updated on a regular basis, preferably in an automatic way. As a general rule, we think an update **every 2 to 3 months** is a good idea.

The Dataset entry update

To conform with the “2 stars” level, the dataset entry **MUST** have an additional property stating the date at which the URI list file was last generated. This is helpful for data consumers as it avoids them to fetch and process the list of URIs if it hasn’t change since the last time it was fetched.

Property	URI	Description
Update / modification date	dct:modified (on the distribution entity)	The most recent date at which the distribution was modified. This MUST correspond to the date the file was regenerated. The file MUST not be older than 2 or 3 month.

ELI dataset “3 stars”: date of modification of each ELI

The URI list file

File structure

To conform with the “3 stars” level of the ELI dataset description, the URI list file:

- **MUST** conform to all the requirements of the “2 stars” level (and as such **MUST** be regenerated on a regular basis);
- **MUST** be expressed in a second file; that second file:
 - **MUST** be a valid RDF Turtle file;
 - **MUST** contain one “dct:modified” statement for each ELI (“date of modification”), indicating the date at which the ELI was last modified. This **MUST** correspond to the date the ELI was first published if it was never modified;
 - **MUST** contain the same list of ELIs as the plain text list, and **MUST** be generated at the same time;

The date of modification in this file gives an indication to a data consumer on whether it is necessary to update the data it has previously fetched for a given ELI. Note that this information is not part of the “official” ELI metadata and has no legal meaning. It **CAN** correspond to one of the date in the ELI metadata (eli:date_publication for example), but it can also follow completely different rules depending on the publication process.

Example of 3-stars URI list with timestamps

```
@prefix dct: <http://purl.org/dc/terms/> .
@prefix eli: <http://data.europa.eu/eli/ontology#LegalResource> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

<http://country.eu/eli/law/2016/501/001/jo> dct:modified "2016-03-22"^^xsd:date .
<http://country.eu/eli/law/2016/501/002/jo> dct:modified "2016-03-23"^^xsd:date .
<http://country.eu/eli/law/2016/501/003/jo> dct:modified "2016-03-24"^^xsd:date .
<http://country.eu/eli/decreet/2005/199/999/jo> dct:modified "2005-08-18"^^xsd:date .
```

Dissemination of the file

In the “3 stars” level:

- The plain text “1 star” URI list file MUST remain accessible and MUST follow the requirements of the “1 star” level;
- The new Turtle URI list file, with modification dates, SHOULD be made accessible at /eli/uri-list.ttl. (note the different file extension). For example, given a legal web portal at <http://country.eu>, the new URI list file will be accessible at <http://country.eu/eli/uri-list.ttl>.

Again, this is not a hard requirement, and other ways of disseminating the file are possible if it remains accessible through the open-data portal.

The Dataset entry in the open-data portal

To conform with the “3 stars” level of the ELI dataset description, the dataset entry:

- MUST conform to all the requirements of the “2 stars” level;
- MUST contain an additional distribution, pointing to the URI list with timestamps. This Distribution:
 - MUST have “text/turtle” as a value for its “media type” field;
 - MUST point to the URL of the new URI list with timestamps;
 - MUST have a date of modification with the same value as the plain text file (since both files are supposed to be generated at the same time);

ELI providers should get in touch with the support team from their national open-data portal to understand whether it is possible to update this dataset entry on a regular, automated basis.

ELI dataset “4 stars”: dump exports

The ELI Dataset CAN be used to indicate other Distributions of the same data. Typically, if a complete downloadable dump of the ELI metadata is available somewhere, it can be indicated as another Distribution of the dataset. Or if the full textual content of the legislation is available, it can also be a Distribution of the Dataset.

The new Distribution(s) should be described with proper metadata, and readers can refer to the documentation of their national open-data portal.

Annex: disseminating multiple ELI URI Lists

In some situations, it may be easier to split the complete list of ELI URIs in multiple files (potential use-case: in Italy with multiple series of the *GazzettaUfficiale*, it might make sense to generate one list per serie).

If this is the case, then:

- Separate Datasets **MUST** be declared in the open-data portal;
- The distribution of each Dataset will point to a distinct URI list file;
- Each Dataset **MUST** follow the description guidelines given above, notably on the keywords, title, description, etc.
- Each dataset **SHOULD** be described with proper metadata from to indicate its coverage: which OJ serie, which time spans, etc.