European Commission

Directorate - General Informatics
Directorate D – Digital Services
DIGIT D2 - Interoperability

D02.02 Plan for the Registry of Registries

ISA^2 Action 2016.28: Access to Base Registries

Specific Contract n° 157 under Framework Contract n° DI/07624 - ABCIV Lot 3

Date: 14/11/2019

Doc. Version: 2.00

DOCUMENT CONTROL INFORMATION

Settings	Value
Document Title:	SC157_D02.01.Draft Specification
Project Title:	Specific Contract n° 157 under Framework Contract n° DI/07624 - ABCIV Lot 3: ISA² - Action 2016.28: Access to Base Registries
Document Author(s):	Martin Alvarez-Espinar Ksenia Bocharova
Project Owner:	Natalia Aristimuno Perez
Project Manager (European Commission):	Peter Burian
Contractor's Project Manager (CPM):	Ksenia Bocharova
Doc. Version:	2.00
Sensitivity:	Limited
Date:	14/11/2019

Revision History

Date	Version	Description	Author(s)	Reviewed by
21/03/2019	0.01	Initiation of the document	Martin Alvarez-Espinar	Ksenia Bocharova
25/03/2019	0.02	Added section 'Introduction' and quality review of the v0.01	Ksenia Bocharova	Patrick Jacobs
25/03/2019	1.00	Sent to the PM for review, the version to serve for the high-level paper to be shared with ABR WG members prior to the first webinar	Ksenia Bocharova	Peter Burian
28/03/2019	1.01	Added two use cases in the introductory part to illustrate the motivation of the work.	Martin Alvarez-Espinar	Ksenia Bocharova
20/10/2019	1.03	More requirements gathered from the Guidelines and the ABR WG meetings.	Martin Alvarez-Espinar	Ksenia Bocharova
24/10/2019	1.04	Added information to section 'Future activities'. Review of the deliverable.	Ksenia Bocharova	Ludovic Mayot
30/10/2019	1.05	Submission to PM for review.	Ksenia Bocharova	Peter Burian
06/11/19	1.06	Review of PM comments, addressing part of them.	Ksenia Bocharova	-
07/11/19	1.07	Expanding the comments with technical details.	Martin Alvarez-Espinar	Ksenia Bocharova
12/11/19	1.08	Quality review.	Patrick Jacobs.	-
14/11/19	2.00	Submission to PM for approval.	Ksenia Bocharova	Peter Burian

Disclaimer

The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the Commission. The Commission does not guarantee the accuracy of the data included in this study. Neither the Commission nor any person acting on the Commission's behalf may be held responsible for the use which may be made of the information contained therein.

Table of Contents

List of illustrations

Context

This document represents the deliverable under Task-02 in the framework of the specific contract n°157 under ABCIV-Lot 3, regarding the project on the continuation of an Action running under the ISA² programme (Action 2016.28), namely Access to Base Registries (ABR).

The overall purpose of the aforementioned task is to draft a specification of an application profile for the Data Catalogue Vocabulary (DCAT), describing the data in base registries of the Member States (MS), namely BRegDCAT-AP, and define the aspects and activities for the creation of potential Registry of Registries at European level in the future.

This deliverable's purpose is to develop a concept and propose future activities for the facilitation of creation of the European Registry of Registries.

With regard to the elaboration of this deliverable, the project team worked based on the following:

- Relevant aspects from the existing documentation on Action 2016.28;
- Alignment with the existing similar initiatives on the European Union (EU) level;
- Best practices and challenges that MS face in creation of their registries of registries;
- Elaboration of the specification for a registry of registries, namely BRegDCAT-AP, that will serve MS with specification / data model to create their own registries of registries;
- Suggestions, challenges and other feedback on the elaboration of BRegDCAT-AP, shared by the MS representatives and other ABR working group (WG) members;
- Future activities for the ABR Action, defined with the PM.

The outcome of the deliverable, on the long-term, will serve to fulfil one of the goals of a European Registry of Base Registries (ERBR), namely to provide a full interconnection of base registries at the European level.

1. Registry of Registries interoperability

One of the targets of the **Digital Single Market**[1] is to deliver cross-border and cross-sector public services in Europe. So, in order to succeed with this objective, Member State's base registries need to be interconnected. **Base registries** are trusted and reliable sources of basic information on data items such as citizens, corporations, vehicles, driver licenses, buildings, and locations. They are the cornerstone of public services and essential entities for public administration management.

The interoperability of base registries is key for the development of the **EU Single Digital Gateway**[2] (or just Gateway), a platform that aims to be the single point of access to public EU Member State's services, facilitating digital public services among public administrations and citizens. The implementation of the Gateway relies on **the once-only principle**, ensuring that data, which is submitted once to at least one EU Member State, could be reused by any public authority across the EU.

The development of a **European Registry of Base Registries (ERBR)**, a pan-European registry of base registries, will improve the interoperability of individual base registries and harmonise the existing registries of base registries, enabling a one-stop-platform for citizens, business and public bodies to access and manage base registries across the European Union and across different domains. The ERBR development initiative works in close liaison with **The Once-Only Principle Project (TOOP)**[3], launched by the European Commission in 2017 with the objective of exploring and demonstrating the once-only principle across borders. In particular, TOOP is focused on creating an innovative pan-European federated architecture for interacting with existing national infrastructures, connecting base registries and eGovernment platforms in different countries. As TOOP is based on the reuse of existing EU interoperability Frameworks, such as the **European Interoperability Reference Architecture (EIRA)**[4], the **European Interoperability Framework (EIF)**[5], and **Connecting Europe Facility (CEF)**[6], an **ERBR** may contribute to this project of interoperability development.

[1] https://ec.europa.eu/commission/priorities/digital-single-market_en
[2] https://ec.europa.eu/growth/single-market/single-digital-gateway_en
[3] http://www.toop.eu
[4] https://joinup.ec.europa.eu/solution/eira
[5] https://ec.europa.eu/isa2/eif
[6] https://ec.europa.eu/inea/en/connecting-europe-facility

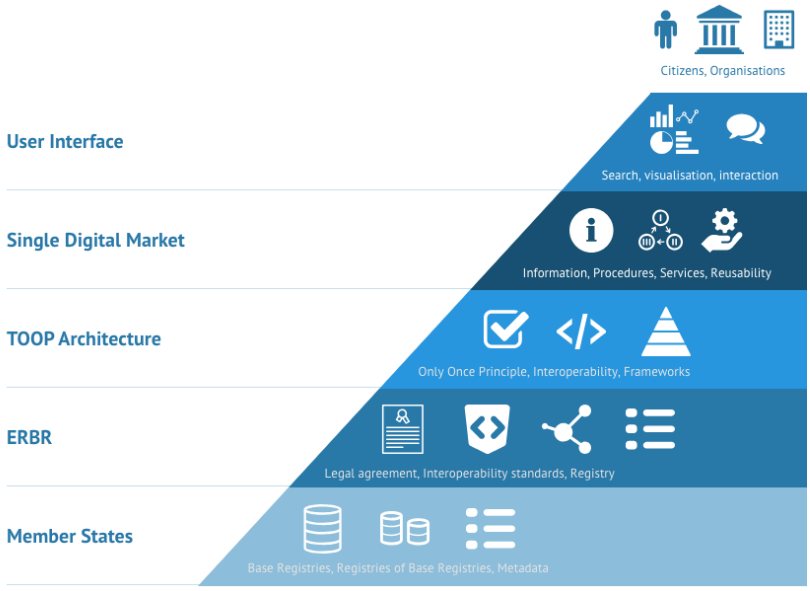

Fig. 1. Stack of ERBR components

Use Cases

The development of the ERBR is motivated by a need of interoperability for Base Registries at European Level. In the ecosystem of the ERBR there are two main types of users: providers and consumers. The former are mainly Base registries' holders, managers or the competent authorities managing Base Registries (i.e., registering, updating information, and maintaining registries); the later are third parties, including individuals, businesses, public authorities, public administrations, and any other stakeholder, benefiting from the ERBR (collecting, processing, reusing data, etc.).

The following use cases illustrate the motivation of the idea for the development of the ERBR.

Use Case 1. Aggregation of Base Registries

Although some Member States have developed infrastructures to connect Base Registries —indeed, some of them have National Registers that integrates several Base Registries—, they are isolated and there is no interoperability among them. The future ERBR will facilitate an efficient cooperation for achieving interoperability and common standards.

Using the ERBR tools and infrastructure, Member States will be able to publish information about their existing National Registers, enabling a subsequent homogeneous integration with the ERBR.

The ERBR will aggregate metadata about National Registers and their underlying catalogues of datasets. This aggregation will enhance the visibility of Base Registries at European level.

Use Case 2. Analysis, Reports and Decision Making

The ERBR will act as a one-stop platform for the base registries' information. So, European public authorities will be able to have access to the overall picture of official data on specific topics across Europe. In this regard, the ERBR would provide overall and centralised information on the existing European base registries and any relevant national registries of registries.

To succeed in this goal, the ERBR would implement some features that will enable:

- consumers to select datasets from specific domains or areas of interest;
- consumers to select datasets from specific countries or administrative areas;
- consumers to receive information about the competent authority responsible for the operation of the Base Registries related to the filters included;
- consumers to access and process the data in machine-readable formats;
- consumers to process data, analysing information to elaborate reports with trustworthy official data.

Design

The design of ERBR will enable interoperability through the application of technical standards and the implementation of the **EIF core interoperability principles and recommendations**[7]:

- **Openness**: The ERBR should be designed following the *open government data* principles, making public sector information freely available for use and reuse by third parties, unless restrictions apply;
- **Transparency**: The ERBR should be a common mechanism to enable visibility inside the administrative environment of public administrations (i.e., rules, processes, data, services, and decision-making). It also should provide and ensure the availability of interfaces that facilitate reuse of systems and data, always securing the right to the protection of personal data;
- **Reusability**: The ERBR should reuse existing solutions (e.g. TOOP, CEF, DCAT-AP, MDR, and ELI), and share its interoperability solutions, concepts, specifications, frameworks, specifications, tools, and components with others fostering subsequent use and reuse of resources;
- **Technological neutrality and data portability**: The ERBR should be designed and implemented using minimum technological dependencies, to avoid imposing specific restrictive technical implementations or products. All the data and metadata produced by the ERBR should be delivered in open standard formats.

[7] https://ec.europa.eu/isa2/sites/isa/files/eif_brochure_final.pdf

2. Semantic Model and BRegDCAT-AP

The development of the ERBR requires defining a **data model** to support the description of Base Registries and Registries of Base Registries. This abstract model will drive the creation of a subsequent vocabulary to represent registries, registries of registries and the ERBR itself. This vocabulary, along with a proper set of taxonomies and value schemas, would be a key component to share information between national registries and the ERBR.

Since the assets of any base registry are its data (master data) and this data is essential for the delivery of public services so the data model should enable the definition of quality, validity, authenticity, preservation and availability of the information and repositories.

Since the ERBR will manage registries —i.e, catalogues of data and catalogues of catalogues—, the development of the vocabulary will be based on the W3C DCAT specification, a standard to describe data catalogues and their content. More specifically, the ERBR will extend **DCAT-AP** (DCAT Application Profile for Data Portals in Europe)[8], a technical specification that ISA[2] developed for describing public sector datasets in order to achieve a successful exchange of metadata among data portals in Europe. Thus, a new **DCAT Application profile for base registries in Europe** (**BRegDCAT-AP**) will be created, as a DCAT-AP extension for describing base registries, their contents, and the services they provide.

BRegDCAT-AP will enable cross-border interoperability between Base Registries and Registries of Base Registries, defining a semantic model to describe registries and their contents, facilitating data discovery and exchange of data, reducing redundancy by supporting the Once-Only principle. This set of recommendations will enable a mechanism for the update of EU base registries and their content, reducing technical, organizational and multilingual barriers.

The model and the vocabulary should cover the requirements gathered on the Guidelines document (D04.01):

Identification of registries and master data

Instances of registries and master data should be identified unique and unambiguously. BRegDCAT-AP is based on the Semantic Web, promoting the use of uniform resource identifiers (or URIs) to identify all single entity of the model and descriptions. Those identifiers are based on Web standards so they facilitate the access to them from anywhere, using Web protocols such as HTTP(S). These identifiers are unique and they should not change. In case of any change, there are standard mechanisms to inform about the updates, so machines may be aware of those changes.

[8] https://joinup.ec.europa.eu/solution/dcat-application-profile-data-portals-europe

Management of the quality of data

The quality of data is essential so datasets and data services should include annotations about the accuracy, accessibility, completeness and consistency of data. BRegDCAT-AP includes mechanisms to indicate the level of compliancy of quality standards, to track the provenance of data and define global and specific metrics about data quality.

Definition of registries

Each registry and its contents should be described. Registries are considered as catalogues that include information resources (i.e., data sets, data services, or other catalogues). Catalogue descriptions include information about the managing entities, theme of the catalogue, spatial and temporal scope, relations with the public services running the registry, and other useful information to guarantee the universal interoperability.

The DCAT-AP, created for this specific purpose, will make possible the automatic discovery of semantic assets, the federation of registries and the sharing of datasets between data portals and services connected to registries.

Definition of master data

In cross-border initiatives, but especially in cross-sector and multilingual ones, stakeholders do not always interpret the master data the same way. Master data attributes and identifiers may differ largely in number and nature. The model and vocabulary should include mechanisms to include standard metadata to describe, explain, locate, or otherwise makes it easier to retrieve, use, or manage the information resource of registries.

For this purpose, BRegDCAT-AP will be based on DCAT-AP as well as on recognised schemes and ontologies such as the **ISA[2] Core Vocabularies**[9], **EUROVOC**[10], Publication Office's **MDR**[11] (Metadata Registry), **NUTS**[12] (Nomenclature of Territorial Units for Statistics) and **ELI**[13] (European Legislation Identifier).

Definition of authenticity and origin of data

When data is shared across different administrations, especially cross-borders, there is a need for guaranteeing that the data was actually provided by a specific legal entity that content shared by a base registry has not been altered (integrity). The model and vocabulary should include mechanisms to

[9] https://joinup.ec.europa.eu/page/core-vocabularies
[10] http://eurovoc.europa.eu
[11] http://publications.europa.eu/mdr/authority/
[12] http://ec.europa.eu/eurostat/web/nuts/background
[13] https://eur-lex.europa.eu/eli-register/about.html

describe the provenance of the data, indicate who is the creator and/or publisher and add additional information about trustfulness.

Definition of the legal basis

National legislation plays an important role in the provision of the public service running the registry. Thus, the model and vocabulary may include information about the rules and resources that conforms the legal basis to run the registries.

The specification will be based on the **Core Public Service Vocabulary Application Profile (CPSV-AP)**[14] and the **European Legislation Identifier (ELI)**[15], guaranteeing the full compatibility with the existing implementations across Europe.

[14] https://ec.europa.eu/isa2/solutions/core-public-service-vocabulary-application-profile-cpsv-ap_en
[15] https://publications.europa.eu/en/web/eu-vocabularies/eli

3. Future activities for Registry of Registries

Alignment of the specification with DCAT-AP v2

After the publication of BRegDCAT-AP as a stable draft before the major release of DCAT (v2) was presented as standard, it is recommended to perform a revision of the specification to ensure its alignment with the new version of the original W3C DCAT and with the European DCAT Application Profile[16].

The specification should be revised looking for alignment with other on-going initiatives, such as TOOP Project, Single Digital Governance (SDG), also with other Core Public Vocabularies, and well-recognised schemas and vocabularies such as the ISA2 Core Vocabularies, the ELI (European Legislation Identifier), and taxonomies like EUROVOC, Publication Office's MDR (Metadata Registry) and NUTS (Nomenclature of Territorial Units for Statistics). Using these vocabularies, the specification enables the definition of public services, the legal basis of the registries and other relevant information about management of registries.

Definition of new terms and classification taxonomies

In order to complement the classification schemas to describe registries, content and all the entities that will enable interoperability across Europe, it is needed to perform desk research on the existing classification taxonomies, such as Eurovoc and EU Publication Office's authoritative lists, complementing them with terms required in the BRegDCAT-AP specification. If it is identified that new terms are needed, they should be proposed to EU Publication Office to extend the mentioned taxonomies.

Testing and validation of the specification/ data model

BRegDCAT-AP needs to be validated by the existing use cases, demonstrating that the model and vocabulary fit the existing implementations of registries. The refinement of BRegDCAT-AP should follow the unexpected requirements from real-world use cases.

BRegDCAT-AP needs to be mapped with the existing specifications, already implemented by Member States. It is important to collect use cases from Member States, apply the BRegDCAT-AP vocabulary to describe the existing reality. Following this activity, a proof of concept to manage data should be implemented, at least to cover discoverability, federation and visualisation.

[16] Both standards are expected to be released in December 2019.

Existing Member State initiatives with specific semantic representations and models should be mapped in an exercise of alignment of vocabularies and taxonomies. This alignment map should be exposed publicly, and a specific tool will help MS in this exercise, as described in section 'Mapping Editor'.

Member States should be guided to implement and validate the model of BRegDCAT-AP. In order to help Member States to use the model, they should be provided with tools, such as the existing ISA Interoperability Testbed, for validation of the model of BRegDCAT-AP. A SHACL (Shapes Constraint Language) schema should be created to validate BRegDCAT-AP documents. This language aims at creating schemas (shapes) for the specific RDF model. These schemas will serve to validate the semantic descriptions automatically. This validation tool would inform Member States about errors, inconsistencies, missing data, allowing them to correct errors and improve the quality of the information.

Another recommended tool to help MS in the process of understanding and using BRegDCAT-AP is a visual editor, as described in section '**Error! Reference source not found.**'.

In addition to these tools, a set of implementation guidelines with useful technical information is recommended. It would include concrete semantic recommendations to serve MS with information on how to adopt and implement the specification. This set should serve as a set of technical recipes to implement and represent the data model of base registries.

Creation of reusable tools

It is recommended to create a set of tools that MS will be able to reuse to facilitate their work on creation of their registry of registries and align their registry of registries on the EU level in the future, using common standards.

Mapping Editor

A mapping tool would be beneficial for MS, so a BRegDCAT-AP Mapping Editor is recommended. This tool would serve as a facilitator for comparison of data models in an intuitive way. This tool should allow MS to visualise, save and export their mappings, and compare different data models, which are used by public administrations to describe base registries on different levels (local, regional, national), and even across the borders in the future.

On a long term, this tool should be useful for MS to harmonise different data models that they created inside the country to comply with BRegDCAT-AP data model.

Creator

A tool, namely, BRegDCAT-AP Creator, would be recommended, allowing MS to create their descriptions in an easy, user-friendly manner. This tool should serve as a 'translator from a human to a machine',

producing the results of MS inputs from a pre-defined web-template into a machine-readable format. This tool should allow a MS to create a specification and export it. In conclusion, this tool should enable the creation of machine-readable RDF documents through human-friendly interfaces.

Visual Editor

An editing tool, namely, BRegDCAT-AP Visual Editor, should serve as a content management application, supporting MS in creation and management of their descriptions. This tool should allow MS to create, edit, remove a specification, as well as export them, make a search, save and visualise saved specifications.

Data Validator

As described in section 'Testing and validation of the specification/ data model', the data validator through SHACL schema would be beneficial to MS to test the application of BRegDCAT-AP data model. It will be based on the existing ISA Interoperability Testbed, adding a new schema ready to validate the implementations.

Data Harvester

A BRegDCAT-AP tool would be recommended to allow MS to achieve up-to-date information in a central registry. This tool should serve as an automatic collector of the data from the registries on multiple levels (local, regional, national). This tool aims to allow a MS to reduce the effort on manual update of information in different registries, and facilitate automatic data collection on a central level into a central catalogue.

On a long term, this tool should facilitate to achieve an integrated view on the data in a MS, and further cross-border.

Pilot of reusable tools

It is highly recommended to implement the set of reusable tools, mentioned in section above, in several MS, to test, and upon testing results, to improve the tools, resolving potentially identified issues. A pilot to gather all the information from the existing registries may be implemented. The registries collected (their metadata) should be shown on a website, both in human and machine-readable formats.

This tool will help the general public to understand the purpose of these activities and check the status of implementation and deployment of the specification. Also, it will prove the feasibility of the model and the specification itself.

Member States Case Studies

During this phase of the Action, an exchange of experiences and best practices between MS during different activities (interviews, webinars, working group meetings) proved to be a source of valuable information to inspire other MS in their work on their registry of registries. Also, this action will include the gathering of information related to the existing registries. The shared information from MS will be harvested and exposed and promoted to showcase the early adoptions and gain interest from MS.

In order to expand current sources, such as ABR Catalogue of Solutions and Guidelines, serving MS with different type of information on challenges and best practice examples from other MS or EU initiatives, it is recommended to continue engaging MS in sharing their experiences on creation of their registry of registries and related aspects. This could be done in a format of case studies on different topics, which are of high interest for MS, such as preparation activities and actual steps taken to create a central registry or even start interconnection and exchange of data with another MS.

Alignment with EU Initiatives

Continuous follow-up on ongoing work of other ISA² Programme and EU initiatives, related to ABR Action, is needed to align the Action's work accordingly. Collaboration with relevant stakeholders to find synergies, identify if their or ABR inputs can be shared and re-used, is an important and continuous activity. It should include accommodating feedback and inputs, received from other initiatives, but also providing feedback and inputs from ABR Action that will be relevant for those initiatives.

One example could be to follow-up on Core Vocabularies and ELI that might impact the specification, and update the specification accordingly. Another example would be to provide relevant input for other European programmes such as the Single Digital Governance (SDG) initiative, aiming at aligning the different semantic models and offering interesting information to illustrate the benefits of the SDG initiative such as the publication of the list of base registries and their content, the underlying model and other potential re-usable pieces of information (e.g., taxonomies, pilots, implementations, etc.). Alignment and bi-directional discussions with other working groups, like the CPSV-AP specification will be also encouraged, publishing the alignments and mapping of differences, in case there is any.

Member States and individual users will be also encouraged to use the model and specification. As a result of this, reports with the alignments of models and implementations should be published.