European Commission

How Linked Data is transforming eGovernment

… and how the ISA Programme is actively pushing forward this transformation for the benefit of Europe

ISA

What is Linked Data?

The Web is evolving from a "Web of linked documents" into a "Web of linked data".

However, in many cases, **data is still locked** in information systems and databases and is **represented using different, usually not aligned, vocabularies and schemas**.

In Europe, access to government data, and the possibility to freely use it, is seen as an enabler for Open Government and a goldmine of unrealised economic potential. Open Data usually refers to public records (e.g. on transport, infrastructure, education, and environment) that can be freely used and redistributed by anyone - either for free or at marginal cost [1].

But opening-up data, e.g. in Open Data portals, often happens in an **ad-hoc manner**, and in many cases thousands of datasets is published **without adhering to commonly-agreed data and metadata standards** and **without reusing common identifiers**.

Hence, a fragmented data-scape is created, where **finding, reusing, integrating and making sense of data** from different sources is **a real challenge**.

Linked Data can respond to these challenges and can be an enabler of eGovernment transformation, leading to smarter and more efficient government services and applications, and fostering creativity and innovation in the digital economy.

"Linked data is a set of design principles for sharing machine-readable data on the Web for use by public administrations, business and citizens."

The four design principles of Linked Data put forward by Tim Berners-Lee in 2006 [2].

1. Use Uniform Resource Identifiers (URIs) to uniquely identify things (data entities)
2. Use HTTP URLs, corresponding to these URIs, so that information can be retrieved
3. Provide metadata using open standards such as RDF
4. Include links to related URIs, so that people can discover more things

[1] http://europa.eu/rapid/press-release_IP-11-1524_en.htm
[2] http://www.w3.org/DesignIssues/LinkedData.html

How can Linked Data help transform eGovernment?

The Directive on the re-use of public sector information (known as the PSI directive) encourages public administrations to adopt open data policies, allowing broad use of their data records. Linked Data principles enable data to be delivered in both machine- and human-readable formats. Making government data available on the Web enables greater transparency and accountability, and makes possible more informed choices [3].

Linked Data enables public administrations to define links (i.e. relationships) between related datasets of others. Putting data into such a context creates new knowledge and fosters creativity and innovation. Now, governments can come up with smarter and more efficient public services and applications, and businesses and individuals can develop new tools and services in order to work with, analyse, and make sense of the data.

The Linked Data paradigm does not impact the ownership of the original data. Relationships between data are established, but everyone keeps full control of their original data [3].

> "Open data initiatives are expected to promote transparency, foster collaboration across government and beyond, allow the creation of new, innovative, added-value services, and improve the quality of decision-making." [4]

Linked Data enables the flexible virtual integration of government data, through linking, without requiring to redesign information systems and to centralise data in data silos. This will facilitate the collaboration between different public sector agencies in the provision of common services.

Re-using data, commonly-agreed metadata (e.g. vocabularies and taxonomies) and common identifiers (i.e. URIs) ensures semantic interoperability when information systems exchange data, thus making the provision of cross-border public services easier, and creating economies of scale which are likely to result in significant savings.

[3] Bizer, C.; , "The Emerging Web of Linked Data," Intelligent Systems, IEEE , vol.24, no.5, pp.87-92, Sept.-Oct. 2009
[4] Ding, L., Peristeras V., Hausenblas M., "Linked Open Government Data," IEEE Intelligent Systems, vol. 27, no. 3, pp. 11-15, May-June 2012

Looking at the future

At present most information systems store data in relational databases and make its exchange possible according to well defined structures, usually using XML schemas. Sharing data according to some sort of schema has been the technological paradigm of the last decades because it enables computer programmes to process data efficiently. However, when these schemas evolve, information systems using them(as data providers or data consumers) need to be adapted accordingly. Over time maintaining these schemas requires significant effort and can be quite inflexible – especially when the pace of change is high.

This is a key reason for the emergence of a new paradigm for data exchange centred around the Resource Description Framework (RDF). According to its publisher, W3C, *"RDF has features that facilitate data merging even if the underlying schemas differ, and it specifically supports the evolution of schemas over time without requiring all the data consumers to be changed"* [5]. In RDF, data is organised in graphs around subject-predicate-object statements and can be queried using SPARQL. These and other related standards are the foundations of Linked Data.

The ISA programme of the European Commission is attentive to this paradigm shift and is running an action on semantic interoperability. This action is putting these standards and technologies together for improving eGovernment, at EU-wide level. The results of this action will then be re-used by Public Administrations in the Member States to implement open data policies, to open up their base registries and to ease data exchange across borders.

Present	Future
Data shared using the eXtensible Markup Language (XML)	Data shared using the Resource Description Framework (RDF)
XML Schema (XSD) used as a means to validate data	Rules (e.g. SWRL) and SPARQL used as a means to validate data and infer facts
Structure is central to reuse i.e. reuse of XSDs	Re-use of vocabularies and reference data is a means of achieving common meaning
Data is stored in a relational database	Data is stored in RDF graphs
Specialised SOAP-based Web Services make access to data possible	Lightweight RESTful Linked Data services make access to data possible
System integration	Semantic-based data inter-linking
Data and schemas are assumed to be complete (Closed World)	Data and vocabularies are assumed to be incomplete (Open World)
Each data entity has a unique identifier at information system level	Each data entity has common and inter-linked identifiers, across information systems, at Web level

[5] http://www.w3.org/RDF/

From Technical to Semantic interoperability

In the last decades, interoperability has been a key topic for both the IT industry and governments. The road leading to effective system-to-system communication has always been paved with many obstacles: proliferation of standards, non-compatible technologies, high costs for data transformation and mapping.

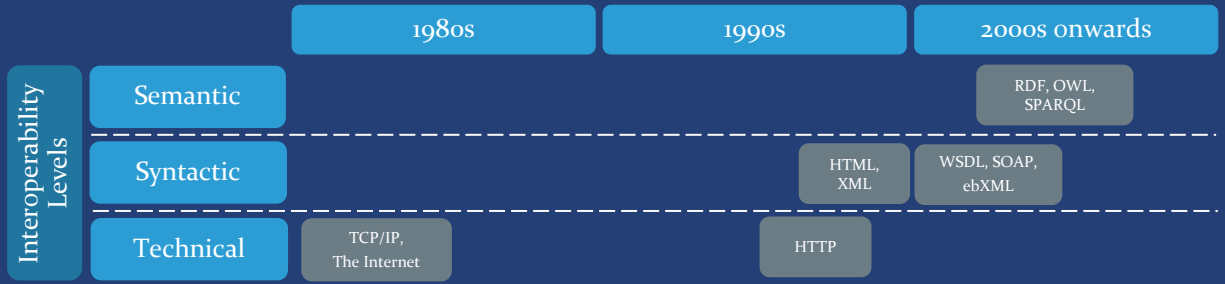
Different technologies and protocols have been created to address the issues above. The advent of the Internet in the 1980s constitutes the first major milestone, with the standardisation of TCP and IP protocols, which made world-wide system-to-system communication and **technical interoperability** [6] a reality.

However, at that time, systems were still "speaking different languages".

To make communication effective, data formats needed to be harmonised, thus improving **syntactic interoperability** [6]. Only in the late 1990s standards like XML became widely adopted. However, XML and related technologies (XSD, SOAP, WSDL, ebXML) left one problem open: the need of sharing "off-line" a common naming scheme and agreeing upfront on the meaning and on strict data typing. Mutual understanding was impossible otherwise.

In the 2000s, thanks to semantic standards like RDF and OWL, a new step has been achieved: **semantic interoperability** [6], which entails reaching consensus on the *meaning of data elements* and the relationships between them. Semantic interoperability requires developing common vocabularies to describe data entities, and ensures that these are understood in the same way by communicating parties **[8]**.

Linked Data is an enabler of semantic interoperability. The evolution of semantic and Linked Data technologies is summarised in the following slide.



[6] As defined by EC's European Interoperability Framework: http://ec.europa.eu/isa/documents/isa_annex_ii_eif_en.pdf

Key milestones for Linked Data, particularly for eGovernment

2004

RDF published as W3C recommendation
http://www.w3.org/RDF

2006

Tim Berners-Lee publishes the 4 Linked Data principles

http://www.w3.org/DesignIssues/LinkedData.html
http://5stardata.info/

2008

SPARQL, the query language for RDF, published as W3C recommendation

http://www.w3.org/TR/rdf-sparql-query/

2009

The UK Government released data.gov.uk. Pioneer in linked government data they contributed to set the example world-wide.

2010

The large research initiative on Linked Open Data (LOD2) and the LOD Around The Clock (LATC) project are launched under FP7

http://lod2.eu/
http://latc-project.eu/

2011

W3C establishes the Government Linked Data (GLD) Working Group

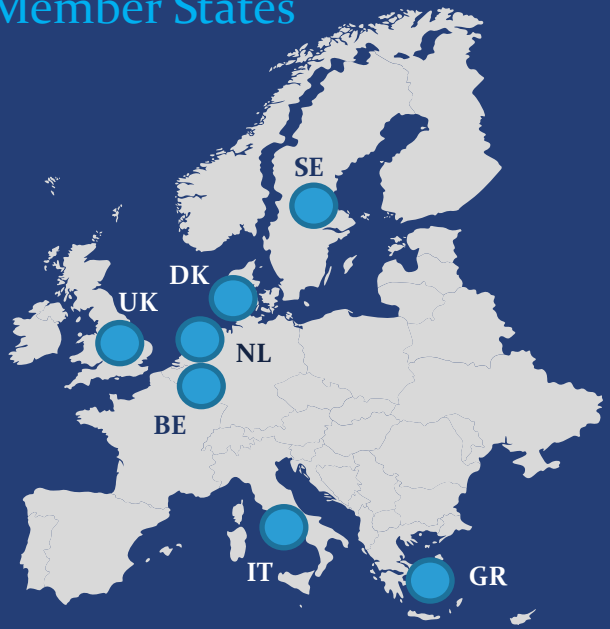
http://www.w3.org/2011/gld

2012

ISA Member State representatives endorse key Linked Government Data specifications

http://joinup.ec.europa.eu/news/isa-member-state-representatives-endorse-key-specifications-e-government-interoperability

Key Linked Government Data initiatives in the Member States



BE - Interconnecting Belgian National and Regional Address Data
http://location.testproject.eu/BEL/

DK – Good basic data for everyone
http://uk.fm.dk/publications/2012/good-basic-data-for-everyone

EC – Open Data Portal
http://open-data.europa.eu/open-data/linked-data

EU – Financial Transparency System
http://fts.publicdata.eu/page/startpage

GR – Legal Entities & Public Spending
http://linkeddata.ihu.edu.gr:8080/rdf-browser/
http://publicspending.medialab.ntua.gr/

IT – Linked Data Guidelines
http://www.digitpa.gov.it/notizie/linee-guida-open-data-interoperabili

SE – Linked Open Data Sweden
http://www.ida.liu.se/~evabl45/lodsweden.sv.shtml

UK – Government Linked Data Working Group
http://data.gov.uk/linked-data
http://education.data.gov.uk/
http://location.data.gov.uk/
http://legislation.gov.uk/

NL - Geonovum.nl
http://www.geonovum.nl/

7 Good Practices for publishing Linked Data [7]

- Model the data
- Re-use vocabularies whenever possible
- Name things with persistent URIs
- Publish human and machine readable descriptions
- Convert data to RDF
- Specify an appropriate license
- Host the Linked Dataset publicly and announce it!

[7] W3C's Cookbook for Open Government Linked Data http://www.w3.org/2011/gld/wiki/Linked_Data_Cookbook

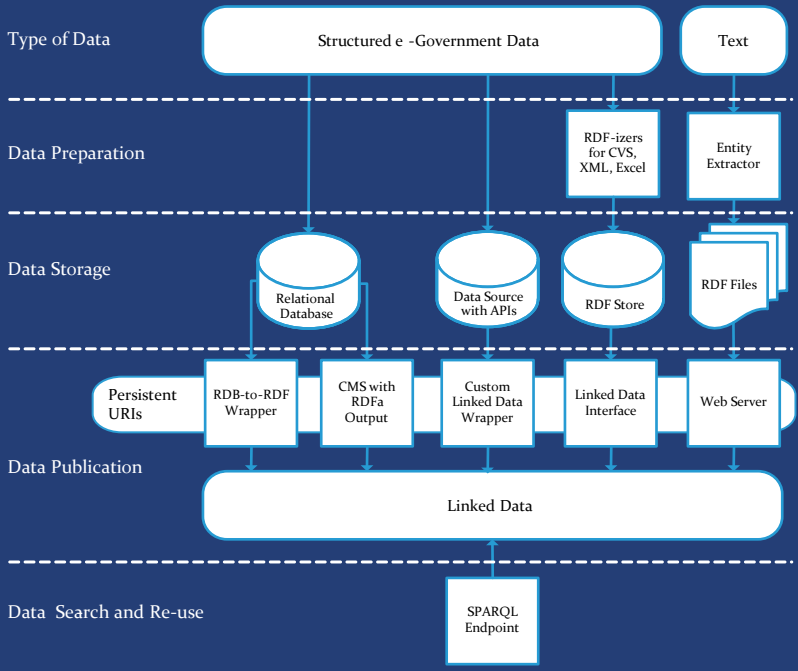
A Linked Data architecture for data integration

Linked Data can be used for integrating data housed in different types of sources, in different formats, both structured and unstructured. We put forward an architecture for that, inspired by the work of Heath & Bizer [8].

The first step is to design and assign persistent URIs to the data and describe it in RDF using commonly-agreed metadata. In the case of structured data (e.g. stored in relational databases, CSV, Excel or XML files), RDF wrappers, exporters and APIs can be adapted and used for transforming it to RDF and publishing it on the Web or any other shared environment data stored in structured sources. Whereas for unstructured data (e.g. a series of news stories or business reports), entity extractors and text mining techniques have to be employed for discovering the data entities to be published as Linked Data.

While data is being represented in RDF, it should be linked to other data coming from trusted sources, in order to provide context and enrich its meaning.

After Linked Data is published, any system capable of running semantic queries (using the SPARQL query language) can extract and re-use the relevant information.



No need for complex system-to-system integration

Semantic queries can be optimised for large sets of data

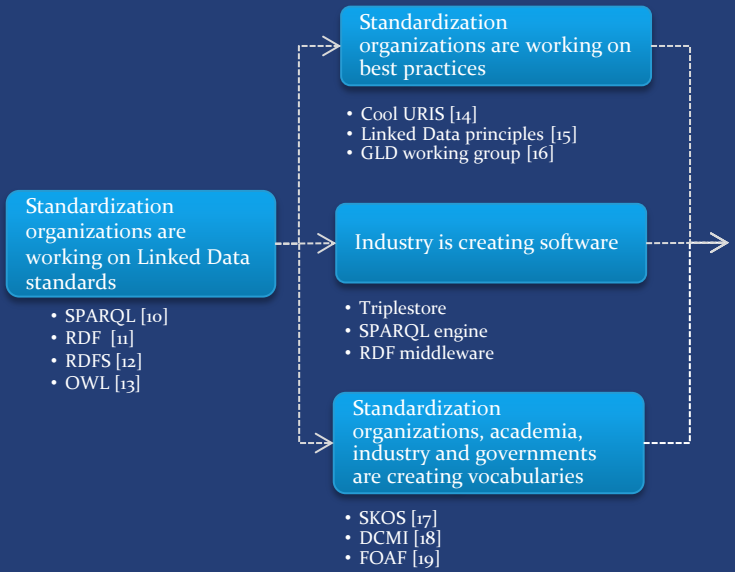
No need for strict data typing

Use of good practises on persistent URI design is key for re-use [9]

[8] http://linkeddatabook.com/editions/1.0/
[9] https://joinup.ec.europa.eu/community/semic/document/10-rules-persistent-uris

What is the role of the ISA programme?

ISA action 1.1 promotes semantic interoperability through Linked Data.



[10] http://www.w3.org/TR/rdf-sparql-query/
[11] http://www.w3.org/RDF/
[12] http://www.w3.org/RDF/
[13] http://www.w3.org/TR/owl-features/
[14] http://www.w3.org/TR/cooluris/

[15] [http://www.w3.org/standards/semanticweb/data
[16] http://www.w3.org/2011/gld/wiki/Main_Page
[17] http://www.w3.org/2004/02/skos/
[18] http://dublincore.org/
[19] http://xmlns.com/foaf/spec/

ISA produces good practices

ISA provides good practices and practical examples to help public administrations apply Linked Data technologies to eGovernment. The following are some studies already available on Joinup:

Study on URI design [20]

This study explores good practices on the publication of Uniform Resource Identifiers (URI) sets, both in terms of format and of their design rules and management.

Case study - Interconnecting business registers with the Core Business Vocabulary [21]

The business case underscores the potential of the Core Business Vocabulary as the foundation for a common semantic standard for interconnecting business registers in Europe and beyond.

Case study - How to describe organizations in RDF using the Core Business Vocabulary and the Organization Ontology? [22]

This case study explains how organizations can be described in RDF using the Core Business Vocabulary and the Organization Ontology. To do this, it provides an example of a real organization, PricewaterhouseCoopers Enterprise Advisory a legal entity registered in the Belgian company register, which is then described in RDF using these specifications.

"ISA provides good practices and practical examples to help public administrations apply Linked Data technologies to eGovernment."

SEMIC SEMANTIC INTEROPERABILITY COMMUNITY

All studies produced by ISA's semantic interoperability initiative are available for download on Joinup through the Semantic Interoperability Community (SEMIC) [23]

[20] https://joinup.ec.europa.eu/community/semic/document/10-rules-persistent-uris
[21] https://joinup.ec.europa.eu/asset/core_business/document/interconnecting-business-registers-core-business-vocabulary
[22] https://joinup.ec.europa.eu/asset/core_business/document/describe-organizations-rdf-core-business-vocabulary-and-org-ontology
[23] https://joinup.ec.europa.eu/community/semic/description

ISA produces Core Vocabularies

What is a Core Vocabulary?

A Core Vocabulary is a simplified, reusable, and extensible data model that captures the fundamental characteristics of a data entity in a context-neutral fashion. Core Vocabularies can be used in interoperability agreements for enabling information systems, developed independently, to exchange information, thus making it possible for information to travel across borders and domains.

The first three Core Vocabularies developed by ISA, namely Core Person, Core Location and Core Business (renamed to Registered Organization) [24], have been undertaken by the Government Linked Data Working Group (GLD WG) of W3C [25]. A fourth Core Vocabulary, the Core Public Service, is currently under development.

Why is it useful?

Although there are many cross-border public services in the EU, citizens and businesses are still not receiving them in a seamless and efficient way. When citizens of one Member State receive services in another, the necessary exchange of data is often hampered by incompatible data standards and specifications. Core Vocabularies help describe data entities by defining their components. They are designed to become the foundation of new, context-specific vocabularies to be used for exchanging public sector information. They can also be used as pivotal vocabularies for mapping and aligning other vocabularies.

The Core Vocabularies abide by the Linked Data principles

- They promote the use of common identifiers for organizations in the form of URIs.
- They can be easily combined with other Linked Data vocabularies.
- They can easily be extended with new classes and attributes to fulfil new domain requirements.

Find more about the Core Vocabularies on Joinup

CORE BUSINESS VOCABULARY

CORE PERSON VOCABULARY

CORE LOCATION VOCABULARY

CORE PUBLIC SERVICE VOCABULARY

[24] https://joinup.ec.europa.eu/community/core_vocabularies/description
[25] http://www.w3.org/2011/gld/

ISA is carrying out Linked Government Data pilots

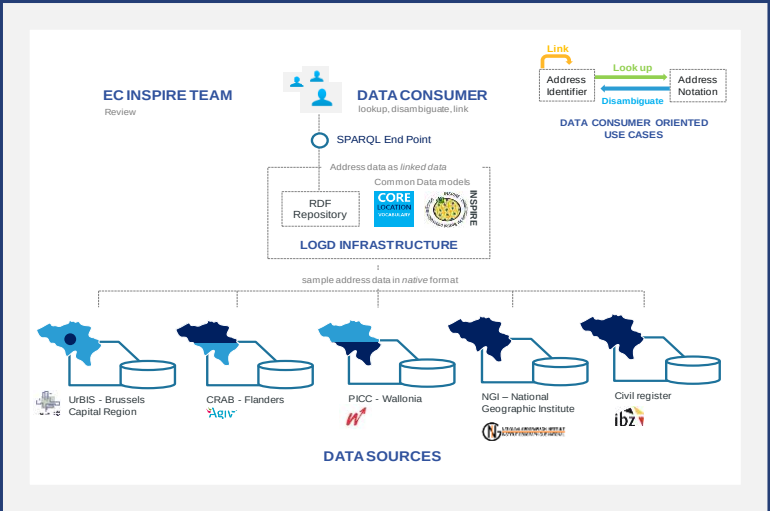
Interconnecting Belgian National and Regional Address Registers

This pilot uses the Core Location Vocabulary [26] to publish and inter-link data from the address registers of the three Belgian regions, namely UrBIS in the Brussels Region, CRAB in the Flanders and PICC in Wallonia, and the Belgian National Geographic Institute (NGI).

A demo of the pilot is available at:
http://location.testproject.eu/BEL/

ADMS SW Pilot - Sample AMDS.SW Software Descriptions from national and regional forges

This pilot aims at publishing sample descriptions of reusable, open-source software from national and regional forges in Europe in HTML+RDFa, RDF-XML, and Turtle formats using the AMDS.SW vocabulary [27].



"The pilot demonstrates how to link address data coming from the 3 Belgian address registers and NGI using the Core Location Vocabulary"

[26] https://joinup.ec.europa.eu/asset/core_location/description
[27] https://joinup.ec.europa.eu/asset/adms_foss/news/sample-amdssw-software-descriptions-cenatic-and-apache-published

ISA online search service based on Linked Data technologies

ISA has launched a new online service to make it easier for public administrations to find and reuse IT specifications (i.e. semantic assets), based on Linked Data technologies [28]. More than 1300 assets from fifteen organisations, shown in the figure, including several Member States and standardization bodies, can be found via the European Commission Joinup Portal [29].

By increasing the visibility and promoting the reuse of semantic assets, the European Commission fosters semantic interoperability among information systems developed in different Member States.

What is ADMS?

This service is powered by the Asset Description Metadata Schema (ADMS) [30]. This is a standardised metadata vocabulary that helps public administrations, standardisation bodies and other stakeholders to document their semantic assets in a uniformed and structured manner (e.g. name, status, version, where they can be found on the Web, etc). In other words, ADMS defines a common way to describe semantic assets.

What are semantic assets?

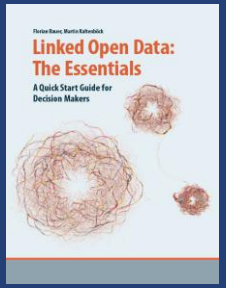
Semantic assets are highly reusable metadata (e.g. xml schema, generic data models) and reference data (e.g. code lists, taxonomies, dictionaries, vocabularies) that are used by public administrations, in their information systems, to share information.

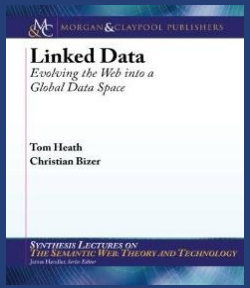
[28] https://joinup.ec.europa.eu/community/cesar/news/european-commission-launches-online-service-promote-semantic-interoperability
[29] http://joinup.ec.europa.eu/
[310 https://joinup.ec.europa.eu/asset/adms/description

Contact us

https://joinup.ec.europa.eu/contact

Further reading about Linked Data





http://www.semantic-web.at/LOD-TheEssentials.pdf
http://linkeddatabook.com/editions/1.0/
http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6237454

Visit our initiatives

CESAR – COMMUNITY OF EUROPEAN SEMANTIC ASSETS REPOSITORIES
SOFTWARE FORGES COMMUNITY
eGOVERNMENT CORE VOCABULARIES
ADMS – ASSET DESCRIPTION METADATA SCHEMA
ADMS.SW
CORE BUSINESS VOCABULARY
CORE PERSON VOCABULARY
CORE LOCATION VOCABULARY
CORE PUBLIC SERVICE VOCABULARY

Get involved

Follow @SEMICeu on Twitter

Join SEMIC group on LinkedIn

Join SEMIC community on Joinup

Disclaimer

The views expressed in this document are purely those of the writer and may not, in any circumstances, be interpreted as stating an official position of the European Commission. The European Commission does not guarantee the accuracy of the information included in this study, nor does it accept any responsibility for any use thereof. Reference herein to any specific products, specifications, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favouring by the European Commission. All care has been taken by the author to ensure that s/he has obtained, where necessary, permission to use any parts of manuscripts including illustrations, maps, and graphs, on which intellectual property rights already exist from the titular holder(s) of such rights or from her/his or their legal representative.

This Case Study was prepared for the ISA programme by PwC EU Services EESV