European Commission

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

D05.01 Updated Guidelines

ISA^2 Action 2016.28: Access to Base Registries

Specific Contract n° 380 under Framework Contract n° DI/07624 - ABCIV Lot 3

Date: 17/12/2020

Doc. Version: 1.00

DOCUMENT CONTROL INFORMATION

Settings	Value
Document Title:	D05.01.Updated Guidelines
Project Title:	Specific Contract n° 380 under Framework Contract n° DI/07624 - ABCIV Lot 3: ISA² - Action 2016.28: Access to Base Registries
Document Author(s):	Ksenia Bocharova Patrice-Emmanuel Schmitz Anastasios Kyrezopoulos Anastasia Sofou
Project Owner (European Commission):	Natalia Aristimuno Perez
Project Manager (European Commission):	Peter Burian
Contractor's Project Manager (CPM):	Ksenia Bocharova
Doc. Version:	1.00
Sensitivity:	Public
Date:	17/12/2020

Revision History

Date	Version	Description	Author(s)	Reviewed by
11/08/2020	0.01	Restructure the document according to Framework v2.00	Ksenia Bocharova	-
18/08/2020	0.02	Review outdated content in the document	Ksenia Bocharova	-
21/08/2020	0.03	Re-write according to decision 42 (PPR M2)	Ksenia Bocharova	Petro Dudi
04/09/2020	0.04	Technical Writer review	Petro Dudi	Ksenia Bocharova

14/09/20	0.05	Review of content to re-work the document according to PM requirements	Ksenia Bocharova	Ludovic Mayot
14/09/20	0.06	Elaboration of Legal Section.	Patrice-Emmanuel Schmitz	Ksenia Bocharova
15/09/20	0.07	Submission to PM for review.	Ksenia Bocharova	Peter Burian
06/10/20	0.08	Review of PM' feedback. Assign resources to address technical comments.	Ksenia Bocharova	-
27/10/20	0.09	Elaboration of new API sub-section and re-work of Semantic section.	Anastasios Kyrezopoulos Anastasia Sofou	Peter Burian
16/11/20	0.10	Implementation of PM' comments on API's sub-section and Semantic section.	Anastasios Kyrezopoulos Anastasia Sofou	Ksenia Bocharova
20/11/20	0.11	Update of document to include content from interviews with MS, public webinars, update conclusions.	Ksenia Bocharova	-
20/11/20	0.12	Submission to PM for final review.	Ksenia Bocharova	Peter Burian
14/12/20	0.13	Implementation of comments from PM on Architecture section	Anastasios Kyrezopoulos	Petro Dudi
16/12/20	0.14	Quality review	Petro Dudi	Ludovic Mayot
17/12/20	1.00	Submit to PM for approval	Ksenia Bocharova	Peter Burian

Disclaimer

This deliverable was prepared for the European Commission by Trasys International under SC380 and it is the ownership of the European Commission.
The views expressed in this document are purely those of the authors 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 document, 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 authors to ensure that they have 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.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

This document represents the deliverable under Task-05 in the framework of the specific contract n°380 under ABCIV-Lot 3, regarding the project on the continuation of an Action running under the ISA^2 Programme (Action 2016.28), namely Access to Base Registries (ABR)[1].

The overall purpose of the aforementioned task is the maintenance and update of the ABR Collection on Joinup (ABR Catalogue / Cartography of Solutions[2]), as well as of the Guidelines document.

This deliverable's purpose is to update the Guidelines by replacing outdated sections and adding new information relevant to base registries, if applicable, as well as to restructure and review the document to improve the logical flow of the content in it.

With regard to the elaboration on this deliverable, the project team performed the following activities:

- Restructured the document according to the Framework of Base Registries and Interconnection (BRAIF) v.3.00[3];
- Reviewed, removed/updated the outdated content and referenced materials;
- Reworked the sections according to the new scope:
  - Introduction to the topics and main concepts;
  - The importance of the topics (why it is discussed, based on which challenges, etc.);
  - Main content of the sections;
  - Conclusions and recommendations.
- Elaborated on the 'Legal', 'Blockchain', 'Cross-border collaboration' sections and reviewed the conclusions;
- Had a Technical Writer review and improve the logical flow in the document;
- Inserted new good practice examples, found during various activities (e.g. webinars, working group meetings, feedback from Member States (MS) and related EU initiatives projects, interviews with MS, work on the pilot, etc.).

The work was based on the relevant aspects from the existing documentation in **Action 2016.28**, and is in alignment with the existing similar initiatives at the European Union level, e.g. **Single Digital Market**[4], **Single Digital Gateway** (**SDG**)[5], **European Data Strategy**[6], etc.

The Guidelines document aims to complement the BRAIF with more practical information on good practice examples and recommendations on how to overcome challenges that MS face, serving as guidance to MS on different aspects in the creation of base registries and registries of registries and on how to interconnect them.

[1] Action 2016.28 of ISA^2 Programme: https://ec.europa.eu/isa2/actions/improving-cross-border-access-government-data_en

[2] ABR Collection of Solutions: https://joinup.ec.europa.eu/collection/access-base-registries/abr-catalogue-solutions-0

[3] BRAIF: https://joinup.ec.europa.eu/collection/access-base-registries/document/braif-framework-base-registries-access-and-interconnection

[4] Single Digital Market: https://ec.europa.eu/digital-single-market/

[5] Single Digital Gateway: https://ec.europa.eu/growth/single-market/single-digital-gateway_en

[6] European Data Strategy: https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy

Overview

Information on basic data items, such as people, companies, vehicles, licences, buildings, locations and roads, is stored in authoritative databases, namely, **base registries**, that represent a trusted and authoritative source of information on the aforementioned data items. In their attempt to adopt customer-centric approaches, Public administrations in Member States (MS) are trying to improve their processes that involve the interaction between such base registries, in order to provide efficient and user-friendly public services to citizens and businesses.

The information needed for operating European public services is owned and managed at the MS level. During the period 2010-2015, the European Union (EU) established the ISA Programme to provide common and shared solutions that facilitated interoperability for European public administrations, including local and regional administrations and Community institutions and bodies.

The **ISA^2 Programme**[7], the successor to the ISA Programme, ran between 2016-2020, and supported the development of digital solutions that enable public administrations, businesses and citizens in Europe to benefit from interoperable cross-border and cross-sector public services.

Among other aspects, the ISA^2 Programme aimed at ensuring a common understanding of interoperability through the **European Interoperability Framework**[8] **(EIF)** and its implementation in MS public administrations. The Framework can be seen as one of the main axes permeating all the ISA projects. In particular, it aims at the following:

- To promote and support the delivery of European public services by fostering cross-border and cross-sectoral interoperability;

- To guide public administrations in their work to provide European public services to businesses and citizens;

- To complement and tie together the various National Interoperability Frameworks (NIFs) at the European level.

In order to establish and foster interoperability of European public services, a collaboration at the MS and EU levels already takes place to define interfaces between base registries, and publish and harmonise the data at legal, organisational, semantic and technical levels.

This scope is covered by Action 2016.28 of the ISA^2 Programme, namely, **Access to Base Registries**. In particular, this document aims to complement the high-level Framework on Base Registries Access and Interconnection (BRAIF) with practical guidance, providing MS with real examples on solutions for their work challenges regarding base registries and registries of registries (RoR) from other MS and EU institutions and projects. It aims at facilitating MS in their future work on RoR, supporting them to achieve cross-border access to governmental data and interoperability.

[7] ISA^2 Programme: https://ec.europa.eu/isa2/home_en

[8] The new version of the EIF: https://ec.europa.eu/isa2/sites/isa/files/eif_brochure_final.pdf

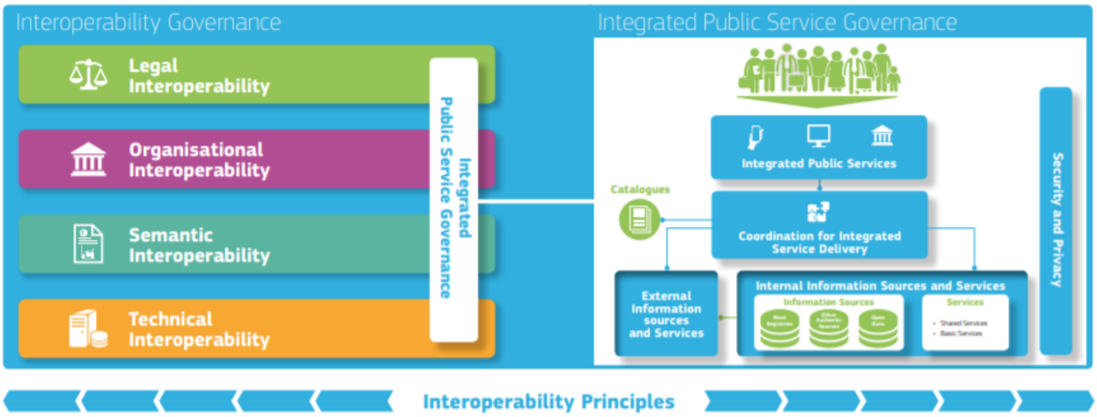
Data Terminology & Definitions

This document provides practical guidance on base registries interconnection and interoperability. Thus, it is essential to define and have a common understanding on the concepts of ‘base registry’ and ‘interoperability’, as well as understand how these concepts are linked to other concepts among various initiatives in the EU.

Base registries are trusted and reliable sources of basic information on data items, such as citizens, corporations, vehicles, driver licences, buildings, and locations. They are the cornerstone of public services and essential entities for public administration management.

The EIF identifies a base registry as a “***trusted and authoritative source of information*** *which can and should be digitally reused by others and in which one organisation is responsible and accountable for the collection, usage, updating and preservation of information”*.[9]

Hence, the EIF depicts a base registry as one of the shared building blocks that make the delivery of integrated public services possible based on interoperability governance, as illustrated in the conceptual model below:



Interoperability Governance
Legal Interoperability
Organisational Interoperability
Semantic Interoperability
Technical Interoperability
Integrated Public Service Governance
Integrated Public Service Governance
Integrated Public Services
Catalogues
Coordination for Integrated Service Delivery
External Information sources and Services
Internal Information Sources and Services
Information Sources
Base Registries
Other Authentic Sources
Open Data
Services
Shared Services
Basic Services
Security and Privacy
Interoperability Principles

Figure 1: The revised EIF conceptual model[10]

In order to succeed with the objective of the Digital Single Market, Member States’ base registries need to be interconnected and exchange data to deliver cross-border and cross-sector public services in the EU.

Interoperability[11] is essential for the effective exchange of data among base registries, public administrations and other authorities, at the Member States and at trans-European level, for single business domains and cross-sector purposes.

[9] The EIF: https://ec.europa.eu/isa2/eif_en

[10] The EIF conceptual model: https://ec.europa.eu/isa2/sites/isa/files/eif_leaflet_final.pdf

[11] Defined in the EIF as the “ability of disparate and diverse organisations to interact towards mutually beneficial and agreed common goals, involving the sharing of information and knowledge between the organisations, through the business processes they support, by means of the exchange of data between their respective ICT systems.”

The interoperability of base registries is key for the development of the **Single Digital Gateway** (**SDG[12])**, a platform that aims to be the single point of access for public services, facilitating digital public services among public administrations and citizens. Its implementation relies on **the once-only principle[13]**, ensuring that data, which are submitted once to at least one MS, could be reused by any public authority across the EU.

An **interoperability-by-design**[14] concept should always drive the development and evolution of public services. As mentioned, new public services should reuse existing information and available services from public administrations, namely, those already available in base registries.

The EIF defines an interoperability model which may be considered as an integral element of the "interoperability-by-design" paradigm. It includes the following elements:

- Four layers of interoperability: **legal, organisational, semantic and technical**[15];
- A cross-cutting component to the four layers called "**Integrated Public Service Governance**";
- A background layer called "**Interoperability Governance**".

This document provides various good practices examples for the Member States' work on base registries and RoR based on the aforementioned layers of interoperability that are the basis of the **European Interoperability Reference Architecture (EIRA**[16]).

Last but not least, in the context of the **European Strategy for Data**[17], BRAIF's complementing guidelines support the implementation of this strategy by provisioning to MS a set of good practices in the area of base registries interconnection, allowing secure cross-border sharing of data, establishing, thus, a base for the future **European Registry of Base Registries (ERBR[18])**.

Scope and target audience

This Guidelines document has the following objectives:

- To share **topics** and note **the challenges** that may hinder the creation of a RoR in Member States or the interoperability of base registries;

- To provide **good practices examples on solutions** and **ABR recommendations** for the creation of a RoR in Member States and for the effective cross-border and cross-sector interoperability with base registries and access to their data.

The solutions and recommendations are of interest to Member States' public administrations and their representatives, in particular, those roles that design strategies and policies, develop from scratch or adapt a base registry data model, interoperate with other systems that provide public administration services, etc. It also may be of interest to those representatives of Member States who are responsible for the integration of public services that require data from base registries, since they need to ensure

[12] Single Digital Gateway: https://ec.europa.eu/growth/single-market/single-digital-gateway_en

[13] Once-Only Principle Project (TOOP): http://www.toop.eu

[14] Interoperability-by-design paradigm outlined in the EIF: https://ec.europa.eu/isa2/sites/isa/files/eif_brochure_final.pdf

[15] Section "3 Interoperability layers" of the EIF for more details about the dimensions of the interoperability: https://ec.europa.eu/isa2/sites/isa/files/eif_brochure_final.pdf

[16] EIRA Collection on Joinup: https://joinup.ec.europa.eu/solution/eira

[17] European Commission, 2020: European Data Strategy

[18] Plan for ERBR: https://joinup.ec.europa.eu/solution/abr-specification-registry-registries/document/plan-registry-registries-released

that such data is used appropriately, i.e. according to the law and the agreements between all involved stakeholders. This includes, but is not limited to, the following roles:

- **Public Agents** who are involved in both the interoperability and public services governance (including e-Government and legal experts);
- **Public Officers** from the base registries and the roles responsible for the execution and monitoring of the processes needed for the exchange of data between stakeholders;
- **Data scientists, analysts and engineers** who take part in the definition, reuse and/or profiling of common data models, vocabularies, reference data and other semantic assets;
- **Technical roles** from the base registries and public administrations that are responsible for the implementation, validation or security of the data exchange processes, interconnecting platforms, etc.

1. Legal aspects of access to base registries

This chapter aims to deliver comprehensive guidelines for supporting decision-makers in EU Member States who are drafting or trying to improve their legislation on base registries/data management/data sharing, and then guide them in what provisions they should consider and include in such legislation.

The first section lists applicable regulations, especially in the field of data protection, and their potential impact on ABR systems (on their design, architecture, restrictions regarding third countries, use of cloud computing services and blockchain technologies).

The second section provides an overview on existing Trans-European Systems experiences, on National Interoperability Strategies (taking as example the Irish Data Sharing and Governance Act 2019), and on a representative cross-border initiative (the Estonian-based X-Road implemented in Nordic countries).

The third section, based on and in relation to the two above sections, provides guidelines or points to consider in Member States' legislations, with the aim of making ABR not just an exception, a simple possibility or a good practice, but a default rule based on existing EU instruments and the cooperation of Member States.

Applicable law and guidelines

EU and national law applicable to Base Registries

Base registries are reliable sources of basic information on items such as persons, companies, vehicles, licenses, buildings, locations, roads etc. Although access to base registries implies the exchange of business data as well, the principal regulations impacting ABR are related to the processing of personal data. Indeed, most base registries contain such type of data, even when their main purpose is focused on other entities, such as real estate, vehicles or businesses – because of the relationship with the owner (of the house, of the car, of the company etc.).

The General Data Protection Regulation (GDPR)

Since 25 May 2018, the GDPR Regulation (EU) 2016/679 protects natural persons with regard to the processing of personal data and on the free movement of such data. As a directly applicable regulation, the GDPR ends most of the fragmentation in different national systems implementing the previous Directive 95/46/EC and its related decisions.

The GDPR regulates personal data protection in the "GDPR zone" (the European Union and EEA[19] countries Norway, Iceland and Lichtenstein) and even abroad in "third countries", since data processing bodies must comply with the GDPR everywhere in the world, if the person (called "data subject") is a resident of the European Union.

In short, the GDPR defines the **principles of lawfulness** and **conditions** of the processing, the rights of the data subject, the responsibility of the data controller (the person who determines the purpose and means of the processing) and the lighter, but real and shared as well, responsibilities of data processors (i.e. contractors or service providers acting on behalf of a controller). It conditions transfers to third

[19] European Economic Area.

countries and defines the competences of national supervisory authorities (NSAs), data protection authoritiesDPAs) and of a European Data Protection Board (EDPB)[20].

Data protection rules beyond the GDPR

There are other rules applicable to data protection, too, such as:

- when the European Union Data Protection Representatives (EUDPR)[21] is applicable due to the data controller being a European institution (under supervision of the European Data Protection Supervisor (EDPS) that is the data protection authority for European institutions);

- when personal data are processed by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties[22];

- when data are processed and exchanged in the framework of EUROPOL[23];

- when focusing on specific aspects, like the ePrivacy Directive (and coming Regulation) i.e. complementing the GDPR regarding on-line/direct marketing or like the privacy aspects of telecommunications.

National laws

The GDPR has not replaced existing national instruments that were adopted for implementing the previous Directive 95/46/EC (or ones that existed even before 1995 in specific countries and were adapted/updated over the time). Knowing that the GDPR is directly applicable, national laws may not enter in contradiction with the GDPR, but may reflect diversity in adapting its principles into national frameworks. In several Member States, the legislator has opted for maintaining existing national laws as far as possible, unless the GDPR disallows this. National provisions continue to set the tone in the remaining policy-making fields as provided by the GDPR (certification bodies, code of conduct, supervisory authority, restrictions for national security, defence etc.). They may clarify specific aspects or add stronger requirements (for example shorter delays for notifying about data breaches, provisions regarding specific technologies, cookies, etc.). Therefore, when a system is implemented, it makes a reference to both the European and the National frameworks[24].

Impact of data protection regulation

Based on the same principles, the GDPR has reinforced (with the possibility of strong penalties) the data protection already provided under Directive 95/46. It would be out of scope here to detail the GDPR principles (lawfulness of processing, legitimate purpose, minimisation, accuracy, limitation in time, integrity, confidentiality and accountability) and conditions for personal data processing where at least one is enough (consent, legal obligation, public interest etc.). We will not detail here the specific GDPR provisions (rights of data subjects, obligations of the controller, role of the Data Protection Officer (DPO)) neither the fact that service providers or sub-contractors acting on behalf of data controllers (as data

[20] The EDPB is an EU body in charge of the application of the General Data Protection Regulation (GDPR). It’s made up of the head of each DPA and of the European Data Protection Supervisor (EDPS) or their representatives. The European Commission takes part in the meetings of the EDPB without voting rights. The secretariat of the EDPB is provided by the EDPS.

[21] Regulation (EU) 2018/1725 of 23 October 2018

[22] Directive (EU) 2016/680 of 27 April 2016 (PDPD – Police data protection Directive

[23] Regulation (EU) 2016/794 of 11 May 2016 on the European Union Agency for Law Enforcement Cooperation

[24] See hereafter the X-Road system implemented by Nordic countries, where it is specified that personal data are processed according to the Personal Data Protection Act of the Republic of Estonia and the EU General Data Protection Regulation (GDPR), unless otherwise provided by the applicable (national) law.

processors) are equally committed to comply with these provisions, but it would be useful to highlight some points that are relevant for implementing ABR sharing systems.

Security by design

The GDPR requires privacy and security by design and by default. What was formerly considered to be a simple best practice is now a mandate that will need to be operationally demonstrable. Therefore, when planning an ABR, project owners must in all cases process an ex-ante assessment of compliance with the GDPR principles and conditions, and determine if a more in-depth Data Protection Impact Assessment (DPIA) and consultation of the supervisory authority is needed (Art. 35 GDPR).

Centralised or decentralised architecture

The GDPR has impacted the choice of IT system architectures: unless there is a specific need for centrally-based functionalities, a decentralised architecture minimises the risks for the individuals whose data are processed because it avoids single points of failure and better complies with the data minimisation principle. As noted by the EDPS[25], it maintains responsibilities at local level (Member States and authorities that are responsible for the civil or commercial records databases and the processing of personal data within these databases).

However, extremely decentralised architectures may become problematic in case the wide dissemination of data controllers makes difficult the allocation of responsibilities (see further below section related to Blockchain technologies). Regarding access to base registries, the allocation of data controller responsibilities to local registries owners is a barrier to opening access, especially, in a cross-border/pan-European environment with multiple stakeholders. This may be compensated by the assignment of a joint or lead data controller at the top system architecture management level.

Cloud computing

Cloud computing is more a marketing concept than a specific technical or legal concept. In general a cloud service provider proposes remote services that are less expensive and said to be "more secure" than keeping ICT infrastructure "in users' homes". The economy is based on the sharing of central manpower, devices and services like servers, storage media, backups, physical security, access right management etc. Large cloud providers also take advantage of the fact that peak hours (for the consumption of services) vary across the world: China, India, Europe and the US do not work at the same time. A consequence of such optimisation is that the location of data (and the applicable legal protection) may vary and, thus, bypass any European protective regulation. The principal potential cause of issues is the prominent role of United States-based cloud providers and the possibility that EU citizens' data are being abused and illegally processed not only by US corporations but also by US Government agencies.

Therefore, **a cloud computing agreement** with a data processor must include awareness of specific points:

- Is the service exclusive (private cloud for the data controller only), shared with a compatible users' group (i.e. other government services), or with anyone in general (public cloud)?

- Are all personal data stored in the GDPR zone, is any being transfered outside it, including backups?

- Does the service comply with all GDPR obligations, including the permanent data deletion when

[25] Opinion 5/2019 of the EDPS on the revision of the EU Regulations on service of documents and taking of evidence in civil or commercial matters: https://edps.europa.eu/sites/edp/files/publication/19-09-13_opinion_service_doc_taking_evidence_civil_matters_en.pdf

not used anymore, and are data accessbile for correction?

Data transfers to third countries

Outside the GDPR zone and territories that ensure an adequate level of protection according to an ad-hoc EC adequacy decision, conditions of data transfer to third countries are strictly limited (GDPR Articles 44 - 50).

The United States of America, which hosts the most important ICT, cloud computing and social network providers, is today the most problematic case. The European Commission attempted to remedy the situation in 2000 with a decision to make US principles comply with the EU Directive. The "Safe Harbour" decision, was an agreement that made possible to work with US business actors, provided they comply with the conditions of a "safe harbour" policy. But in October 2015, The Court of Justice of the European Union (CJEU) declared that the "Safe Harbour" decision was invalid, leading to further talks being held by the Commission with US authorities that produced a new framework for transatlantic data flows, known as the "EU-US Privacy Shield". However, on 16/07/2020, the CJEU invalidated also the Privacy Shield Agreement.

As of this writing (September 2020), the **consequences** of the above are that:

- European Data Controllers cannot use Joint Controllers/Processors/Sub-Processors, located in the US or controlled by a US entity, when the invalidated Privacy Shield Agreement is the legal basis used for data transfer between the GDPR zone and US;

- Data Controllers cannot use Joint Controllers/Processors/Sub-Processors, located in the US or controlled by a US entity, when Standard Contractual Clauses are the basis used to transfer data to the US.

There may be some US/EU Joint Controllers/Processors/Sub-Processors that can be used but the local Controller (EU data exporter) will need to verify on a case by case basis if these business actors could be subject to critical US regulations (i.e. US FISA Section 702 and/or Executive Order 12.333). The case of the United Kingdom's Brexit is still to be clarified as agreements are still under discussion; with the UK having left the EU in January 2020, there is a chance it may be considered a third country after 31 December 2020.

Blockchain

Blockchain technologies share and synchronise a database via a consensus algorithm that stores data on multiple computers (each of them storing a complete local version of the database). Through this wide distributed replication, a strong data integrity (or resilience against alterations) is obtained, due to the difficulty of knowing all storage points and changing content everywhere at the same time. There have been discussions regarding the compatibility between Blockchain and the GDPR[26] because points of tension coexist:

- The GDPR is based on the assumption that in relation to each personal data processing point there is one data controller – whose data subjects can enforce their rights. Using Blockchain, where a unitary actor is replaced by many different players, leaves open the question of responsibility and accountability, unless if a specific body is assigned with the responsibility.

[26] See in particular the Study "Blockchain and the GDPR" – European Parliamentary Research Service, PE 634.445 – July 2019

- The GDPR is based on the assumption that data can be modified or erased where necessary (the famous “right to be forgotten”) to comply with legal requirements, such as Articles 16 and 17 of the GDPR. Blockchain, however, intentionally renders difficult or onerous the unilateral modification of data in order to ensure data integrity and to increase trust in the network.

- The GDPR data minimisation principle requires that the processed personal data is kept to a minimum and only for purposes that have been specified in advance, which can be hard to apply to Blockchain technologies where data is replicated on many different computers.

- If open without restrictions, to store data “anywhere in the world”, without any restrictions posed by a closed number of localised partners, then Blockchain technologies are not compatible with the principle of not transferring personal data to non-GDPR zones or to countries that are not in line with the GDPR protection.

However, Blockchain is a technology class with many flavours or versions. The relationship between the technology and the legal framework cannot be determined in a general manner but, rather, must be determined on a case by case basis. It seems that no Blockchain technology is used so far in ABR projects, like the Nordic countries’ X-Road project (see further below for more details).

Other relevant EU legal principles

Other relevant EU legal principles or guidelines also have legal impact on organising access to base registries, in particular:

- **Availability of PSI**: the Directive on the reuse of public sector information (PSI) encourages Member States to make public information available for access and reuse as open data;

- For processed/shared spatial or geographical information, the **INSPIRE Directive** requires sharing of spatial datasets and services between public authorities with no or as few as possible restrictions or practical obstacles for its reuse;

- **The European Interoperability Framework** recommends the use of open licences for both data and software (EIF recommendations 2 & 3).

ABR Practices

The experience of Trans-European Systems (TES)

Trans-European Systems or solutions (TES) are operational interoperable European solutions owned by the European Commission or other bodies (in some cases by private initiatives, foundations or co-funded by Member States in support of the implementation and advancement of EU policies).

TES generally include the cross-border exchange of data between public administrations’ base registries and, in some cases, between citizens or businesses.

Examples of running (or planned) projects are presented in the table below:

Table 1. EU Projects

Abbreviation	**Project name**
ECRIS	European Criminal Records Information System (restricted access; for relevant authorities only)
EUCARIS	European car and driving licence information system (restricted access; for relevant authorities only)
IRI	Insolvency Registers Interconnection (public access available through the e-Justice portal)
BRIS	Business Registers Interconnection System (public access available through the e-Justice portal)
EPRIS	European Police Records Index System (restricted access; for relevant authorities only)
LRI	Land Registers Interconnection (public access; advanced features only for authenticated legal professionals)
EVIDENCE2 e-CODEX	Exchanges in counterterrorism operations and in the fight against global crimes (restricted access; for relevant authorities only)
EBOCS	European Business Ownership and Control Structures (restricted access; for relevant authorities only)
eu-LISA managed systems	SIS-II (Schengen system alerts regarding various cases / criminal matters); VIS (visa applications from third country nationals to the Schengen area); EURODAC (fingerprint of asylum seekers and irregular border-crossers); EES Entry-Exit System (planned monitoring of border-crossing third-country nationals)

The above TES implementations often require the negotiation and adoption of a specific legal instrument (legal basis). A project can also be built on voluntary participation where members can opt-in (like for LRI, the Land Register Interconnection initiative). In most cases it requires consensus-building between multiple stakeholders and multiple years between the inception, the building of a pilot project and its operation.

A central legal question when implementing such systems is the data controllership: who is the responsible body, playing the role of data controller, facing the applicable data protection regulations and in charge of interaction with data subjects? According to Article 23 of the GDPR, it is the natural or legal person who, alone or jointly with others, determines the purposes and means of the processing of personal data, and therefore bears the responsibility for the processing.

A usual method for assessing this question is to differentiate the collection, storage and provision of data that characterises data controllership, from the provision of technical means (maintaining the network and the IT system), which implies a lighter but real – according to the GDPR – data processor responsibility.

When the common platform is interconnecting decentralised databases, when all the data displayed from the central platform are “transient data” that are not stored in any central component and where no tracing or logs from user-run queries through the service are centrally recorded or kept, the general practice is that the data controllers are the various (i.e. national) providers, each one for the data they collect, own and store. This is the applied solution for systems like EPRIS, EUCARIS or for the systems managed by the eu-LISA Agency.

On the other hand, when a system stores core data in a central component (i.e. at the European Commission level which the national base registries keep updating in regular intervals), and when queries are run against this central database, the owner of this central system looks as the most convenient data controller. This is the case for BRIS, where the Commission provides central storage, and for other shared services provided by the e-Justice Portal (i.e. IRI) where the legal notice privacy statement reports that "Although the responsibility for the Portal's content and its management is a responsibility shared between the European Commission and the individual European Union Member States, the data controller for the European e-Justice Portal is the European Commission".

These general guidelines are indicative, to the extent that an agreement between the various stakeholders in a joint project – complemented by information published in a privacy statement – may allow assigning the responsibility of data controller to a specific body. This is the case for projects such as the e-CODEX or EVIDENCE2 (counterterrorism information), where a foundation is assigned with the role of data controller for all personal data processed through the system.

National Interoperability Strategies (NIS)

In order to facilitate, accelerate and provide a comprehensive legal framework for sharing data between base registries – and, in some cases, to make this sharing mandatory –, Member States may implement a general purpose NIS. Such a framework has been implemented by Ireland through the ***Data Sharing and Governance Act 2019***[27].

The Irish regulation (hereafter "the Act") may be the most recent and complete, but we could also take inspiration from other countries such as Spain which due to the existence of multiple authorities at central, regional and local levels – combined with different local languages – presents a complexity similar to the EU. Spain's legal efforts resulted in the Laws 39 and 40/2015, which force public bodies to share data without requesting paper certificates from citizens. Such instruments specify that in relation to base registries data, sharing can be done, unless data subjects (citizens) explicitly specify the opposite[28].

The Irish Act regulates the sharing of information, including personal data, between Irish public bodies and supports the exclusion of special categories of personal data: all kind of "criminal data" prevention, prosecution (some are processed in the framework of specific TES above), state security, espionage, and defence related data. Noting that the Act breaches the EU GDPR protection (referred in the Act as the fundamental legal instrument).

Under the governance of a minister, public bodies that must comply with the Act are defined and listed by name (i.e. the Attorney General, the Guardia Siochána), or by category (i.e. recognised school, bodies delivering services to the public under an agreement with a public body) with exceptions (Act – 10.5). Specific public bodies may be exempted by the government, or on the contrary, some non-public bodies may be included (Act – 10.4) i.e. contractors or private entities delivering services to the public under an agreement with public bodies – such as companies providing legal insurance for driving cars, etc.

The principle (when no other EU rules apply already) is that a relevant body may and must disclose personal data to another, when this is needed for (and is proportionate to) the performance, the verification, avoidance of burden, facilitation or assessment of the public service.

The implementation of data sharing must be documented in a standard, written agreement that is concluded voluntarily between relevant bodies, knowing that entering into such an agreement could be an obligation in case no other enactment or law of the European Union is requiring (or on the contrary, prohibiting) such data-sharing. Therefore the conclusion of an agreement could also be directed to the relevant public bodies by the government (the minister), after appropriate assessment (and knowing that the directed body may exercise recourses or refuse sharing in case it seems in breach of EU/MS laws).

The agreement must specify in writing the stakeholders, the shared information, the purpose and functional link, the legal basis, and who will or may use/disclose data. It must specify if the disclosure is

[27] http://www.irishstatutebook.ie/eli/2019/act/5/enacted/en/html

[28] Joinup ABR Factsheet, Access to base registries in Spain (2017)

about a specific data subject or classes of data subjects (i.e. statistics), is on a one-off or an ongoing basis, what processing will result, any restrictions, compliance with Article 5 of the GDPR (with all six conditions of transparency, legitimate purpose, data minimisation, accuracy, limited duration and integrity), if a DPIA was carried out (Art. 35 GDPR), the security measures when sharing/transmitting, the duration of the retention, the deletion process, the procedure in case of withdrawal, and any other related aspects (Act – 19.2 & 3).

The agreement must specify a Lead Agency (which is, in general, the responsible data controller or a joint controller), and may be open to new members accession. It must be reviewed on a regular basis (< 5 years). When the information collection is qualified as a "base registry", the registry owner must ensure that the personal data are accurate and up to date.

Moreover, the Act defines the roles and the categories of specific public service information, especially regarding the administration of the public service pension scheme.

The minister role includes the designation of base registries for use by public bodies, so that they can access BR personal data without having to collect it directly from service users.

Data subject rights to assess their own data are provided through a personal data access portal, which allows individuals to view their personal data, as well as, information in relation to any data breaches affecting their personal data, along with the data sharing agreements under which their personal data is processed.

The Act is not confined to the sharing of personal data: business information can be shared between public bodies in the performance of their functions. Part 6 of the Act provides the allocation of a "unique business identifier number" for the purpose of uniquely identifying any undertaking that has a transaction with a public body.

Cross-border Interoperability Strategy (the X-Road case)

While the Irish Act and other national NIS organise data transfers between bodies located in the same country, the purpose of a cross-border framework is to organise access to base registries in different countries, i.e. "trans-European access" as the case may be.

The X-Road[29]® data exchange layer is continuously developed and managed by the Nordic Institute for Interoperability Solutions (NIIS). X-Road is used nationwide in the Estonian data exchange layer *X-tee* and in the *Suomi.fi Data Exchange Layer* service in Finland. Moreover, X-Road is used in the Faroe Islands in their *Heldin* environment, and Iceland is currently implementing its national X-Road environment *Straumurinn*. Two X-Road ecosystems can be connected together, which enables easy and secure cross-border data exchange between countries using X-Road.

Although not a member of the European Union, Iceland is part of the GDPR zone as an EEA state (along with Norway and Liechtenstein). The Icelandic Parliament passed Act 90/2018 ("the Data Protection Act") to implement the GDPR in July 2018. The case of the Faroe Islands is special: they are not part of the European Union, but part of the Nordic Passport Union. There are no border checks between the Nordic countries and the rest of the Schengen Area. In addition, the European Commission has recognised Faroe Islands as a country providing adequate GDPR protection.

X-Road is based on a distributed architecture, however an X-Road ecosystem is managed and operated centrally. The owner of the ecosystem, the X-Road Operator, controls who are allowed to join the ecosystem, and the owner defines regulations and practices that the ecosystem must follow. Each member organisation of the ecosystem manages its own data and controls who is allowed to access it.

[29] X-Road: https://x-road.global/

This applies to the operations of NIIS as the development organisation of the X-Road software. Nevertheless, NIIS does not operate any X-Road ecosystem itself and has no direct connection with base registries or any other X-Road user organisations.

NIIS members – Estonia and Finland – have their own X-Road ecosystems and they are responsible for operating them. Any legal constraints that are related to X-Road are regulated on the NIIS member level and they apply to the X-Road ecosystem of the country in question.

For example, both Estonia and Finland have their own laws and regulations that regulate the use of X-Road. Therefore, when we talk about X-Road – the open source software – there are no laws or regulations directly related to it. However, when we talk about the X-Road ecosystem in Estonia or Finland, the use of the national X-Road ecosystem is regulated on the national level, and each country has its own laws and regulations.

X-Road implements the below set of standard features to support and facilitate data exchange and ensures confidentiality, integrity, and interoperability between data exchange parties:

- address management;
- message routing;
- access rights management;
- organisation-level authentication;
- machine-level authentication;
- transport-level encryption;
- time-stamping;
- monitoring and reporting;
- digital signature of messages;
- logging;
- error handling.

Regarding the copyright, the X-Road software is open source and provided for free under the MIT licence. This means that any individual or organisation can copy the source code of the software, adapt it to their own needs as far as necessary, and use it for developing their own service.

Like all open source licences, the MIT licence specifies that the software is provided "as is", without warranty of any kind, and that the authors will never be liable for any claim, damages or other liability. To help new X-Road users get started, NIIS provides a set of online resources that are all available free of charge. However, NIIS does not provide technical support or consultation services. There is an X-Road Technology Partner programme in which members are companies providing X-Road consultation services. It is recommended to contact one of the partner companies for more extensive support.

The MIT licence is simple, very permissive (recipients can do what they want) and not reciprocal: a recipient can develop their own version, keep all improvements secret and make their whole version proprietary. This is one of the principal differences with the European Union Public Licence (EUPL[30]) that is reciprocal: in the case the software is re-distributed (and providing access online via a network is a form of distribution) recipients must disclose and provide the source code back, under the same EUPL licence. The two licences are compatible, meaning that source code covered by the MIT licence can be re-licensed under the EUPL (the reverse is not true, though).

While the source code is licensed under the MIT, the same is not the case for the X-Road name and logo that are registered trademarks of the Estonian Information System Authority (RIA) and therefore, their use is forbidden without permission from the trademark owner. NIIS is responsible for sublicensing the trademark.

[30] https://ec.europa.eu/isa2/solutions/european-union-public-licence-eupl_en

Points to consider in ABR legislation

Providing guidelines for the elaboration of a better legislation related to the ABR is a difficult exercise, knowing that the BR organisation and their access conditions are specific to each national framework. And, by staying on the European Union level, no global specific regulation is to be expected in the short term.

We have to consider the applicable laws and guidelines, the requirements related to the processing of personal data in order to take advantage of existing ABR practices: TES, NIS, specific cross-border initiatives like X-Road. The Irish Data Sharing and Governance Act 2019 has been especially useful as a source of inspiration. However, this Act only regulates the sharing between national public bodies, while in multiple domains (education, pensions, medicine etc.) the need of cross-border data sharing at the European level is strongly growing.

Overall Governance

Each (national) instrument is to be placed under a national authority. We use the term “the Minister” hereafter, being a national authority in charge of public administration / digital agenda. The Minister will be assisted by a committee or “Board” in charge of advising the authority, supervising data sharing agreements and keeping them in a repository. Depending on the national framework, the Board could be a specific institution or – in order to avoid duplication when the sharing is related to personal data – a section of the independent supervisory authority implemented by Member States (GDPR Chapter VI). The Minister has extended powers for determining the list of BR and the scope of data sharing (inclusion or exclusion from the list of relevant public bodies), for mandating specific data transfers after consulting with stakeholders, for defining transfer conditions or code of conduct (this will cover the question of access fees that can cause potential issues), for processing various inquiries and impact assessments.

Hierarchy of norms and relationships

The agreement will refer to higher (EU) regulations and to other acts that are impacted or are in relation to the specific ABR regulation. An ex-ante inventory is needed as it may be that other specific acts contain ABR restrictions or conditions.

Definitions and pre-requisites

All terms already defined in other general instruments (like the GDPR data controller, processor, data subject, personal data, special categories of data etc.) will be defined by reference to this instrument.

All specific terms will be defined into the specific legislation. For example:

- **Base registry**. A frequently reported issue is that the concept is not yet legally defined in national frameworks. It must be defined as a database designated as such by the authority as a primary and trusted source of information. This must be reported in a “Registry of registries” (RoR) containing relevant information (name, owner, purpose, detailed content of each information field).

- **The RoR** can be considered as a cornerstone and a prerequisite for any global sharing service. The RoR will highlight benefits resulting from access sharing, i.e. creation of ecosystems to facilitate ABR when it is beneficial to businesses or citizens for performing their administrative requests.

- **Data sharing**, in this context, means the disclosure of base registries’ information by a public body to another public body (or its assimilation).

- **Business information**, is a list of all fields that are relevant to the ABR of enterprises.

Delimiting the scope

The legislation (the “legal instrument” hereafter) will:

- **Define the categories of exchanged data** or, on the contrary, exclude special categories of data (for which specific TES may exist already, like exchanges in criminal matters).

- **List targeted public bodies** by name, category (i.e. schools, for students’ data), or by extension: in a “registry of relevant public bodies” the Minister can add organisations acting on behalf of public bodies or providing services to the public under an agreement with a public body. However, the Minister or the instrument may list administrations whose definition of “public body” does not apply in the framework of the ABR.

Define the sharing principles

- **Need and proportionality**: data sharing with another public body must be needed, useful and proportionate for the performance of a public service (i.e. reducing burden, avoiding one administration asking for the data subjects’ information owned by a another administration – according to the “Once Only Principle”).

- **Transparency**: based on written ex-ante data sharing agreements between public bodies, submitted to the Board for validation with details and conditions of the agreed access and information on the architecture (decentralised, with all data and copies located in the GDPR zone), on data controller and processors, if any, and open to upcoming accessions of more stakeholders. An agreement template will be provided. The Minister will publish agreements on a public site. The instrument may specify that the Minister will implement a specific personal data access portal in order to facilitate the exercise of the data subjects’ rights.

- **Direction**: under conditions, the Minister can designate a specific database as a “base registry” and assign it as the unique source for a category of information. When it appears to be needed and useful after consulting with stakeholders, data sharing can be directed by the Minister (becoming an obligation for the relevant public body, with possible recourse/escalation).

- **Agreement governance**: assign a lead agency and review/monitor the agreement application on a regular basis (i.e. every 5 years).

- **Data controllership on several levels**:

    - Each registry “owner” ensures the information is accurate, up to date, and complete. They implement, monitor and document accesses;
    - The lead agency governs the agreement (with *n* parties);
    - The GDPR personal data controller, set by the agreement as the lead agency, who knows that the agreement could assign another body or the technical infrastructure provider with the task;
    - The Minister who will have the power to audit/control the FRAND (fair, reasonable and non-discriminatory) access fee management, when applicable.

Clarify Intellectual property and standards

Place data under an open (possibly free and non-commercial) licence. Encourage the publication of specifications and software under a recognised open source licence, convenient for the distribution of public sector assets (such as the EUPL in case a reciprocal licence is needed for keeping track of further distributed improvements, or an Apache or MIT permissive licence in other cases).

Facilitate data sharing beyond national borders

- **Implementation and reciprocity**: this aspect is missing (not forbidden according to the GDPR, but not specifically regulated) in national instruments like the Irish Data Sharing and Governance Act 2019 or the Spanish Law 39/2015. It may result from cross-border initiatives/projects like X-Road or from the implementation of interoperable regulations or agreements between various

Member States' public authorities.

- **The principle of written agreement** according to a convenient template and the role of a lead agency will be reproduced in the case of cross-border sharing. However, the agreement (lawfulness, utility, proportionality etc.) will be submitted to all the relevant boards or supervisory authorities and to the relevant ministers. The agreements will be open to the accession of new members (i.e. public bodies from other States delivering services related to the same ecosystem).

- Here also, **a cross-border agreement** could be directed by the competent ministers. The simple fact that data sharing is cross-border should not justify refusal or discrimination inside the GDPR zone or between Member States' public bodies.

- When it comes to **personal data**, sharing will be restricted to the GDPR zone or, on a case by case basis, to countries recognised by the European Commission as providing adequate GDPR protection[31]. The agreement will assign the role of data controller, i.e. to the lead agency or – in case of a multiple-stakeholder sharing – to an expert organisation with appropriate skills (like the NIIS – the technical infrastructure owner), who will provide and monitor the service and deliver functionalities without keeping copies of transient requests or data.

Promotion of reliable technical infrastructures and their certification

Multiple works come to the conclusion that if data must stay decentralised in their respective base registries, the technical infrastructure needed for data sharing should be centralised and based on commonly agreed, interoperable standards[32]. This is also the main lesson learnt from existing TES and the X-Road experience: an ABR, and certainly a cross-border ABR, requires skills that go far beyond the normal competences of base register owners. It necessitates cross-border normalisation and central management of a range of services.

As a prerequisite, a specific level of certification such as that of "Trans-European Base Register Access Provider" may be useful, to assess the capacity of the technical infrastructure owner/provider who controls cross-border exchanges, and to certify their efficiency in delivering needed services related to: multiple-nodes network address management, message routing, access rights management and authentication, encryption, time-stamping, monitoring of infrastructure traffic and reporting statistics, checking digital signature of messages, anonymised logging and tracing, issues' error handling. (This list of functionalities provides examples and criteria for assessing skills in the framework of this specific level of certification. It must be checked and validated with ICT experts).

In addition to the above "technical" infrastructure services, the certification will include cross-border data controllership that necessitates competences in various other fields (communication and multilingual policy information to the public, in-depth knowledge of data subjects' rights and ability to apply high quality standards in that field[33], other relevant GDPR matters, ICT architecture, technology and security, risk assessment, human resources training, contacts with relevant supervisory authorities etc.). Implementation of cross-border data controllership will substantially reduce barriers for the multiple base

[31] Andorra, Argentina, Canada, Faeroe Islands, Guernsey, Israel, Isle of Man, Jersey, New Zealand, Switzerland, Uruguay and Japan. The decision on Canada applies only to private entities falling under the scope of the Canadian Personal Information Protection and Electronic Documents Act. The decision on the United States (EU-U.S. Privacy Shield) was invalidated by the CJEU on 16/07/2020. The case of UK after Brexit is still to clarify, especially after the CJEU ruling: after the end of the transition period, any transfer of personal data to the United Kingdom other than that governed by Article 71(1) of the Withdrawal Agreement will not be treated as sharing of data within the Union. It will need to comply with the relevant Union rules applicable to transfers of personal data to third countries.

[32] Arno BENS and Stefan Schukraft, „Register Modernisierung und Verwaltungsdatennutzung in der amtlichen Statistik, Statistisches Bundesamt – WISTA – 4 - 2018

[33] In this line, previous study in this project mentioned the original approach undertaken in the Netherlands for defining an open "standard" (QiY - www.qiyfoundation.org)to give back to individuals the control of their data in the digital world.

register owners who may fear of real or hypothetical risks and penalties in the case of data transfers in a cross-border environment. This role of common data controller is important. It does not signify the concentration of controllership in one entity, and could be organised based on voluntary initiatives and/or on a sectoral base, taking into consideration the specific aspects of each relevant sector or ecosystem (i.e. pensions, medical data, finance etc.). The controllership will be extended with the admission of stakeholders/members from the "community", the preparation, collection/storage and periodic monitoring of transfer agreements between the various stakeholders, ex-ante assessment of each ABR implementation (including, when applicable, a DPIA) and the relation to supervisory or governance authorities.

Certification is a specific domain in which European coordination has strong utility to ensure mutual recognition across all GDPR zone states. A common legal basis is already provided by Article 42 of the GDPR which encourages voluntary and transparent certification mechanisms, and by Article 43 of the GDPR which regulates the role of certification bodies. Therefore, no new instrument is necessary to the extent that the conditions or criteria for such a certification are circulated and agreed between the relevant supervisory authorities (according to 43.3) in the framework of the European Data Protection Board (EDPB), which groups the national DPAs and considers any possible advice from the EDPS.

Economic capacity and funding

Another point (which may be the most important one) to be clarified at central management level is the question of funding and fee collection: beyond the business plan of the infrastructure provider, some BR provide access free of charge in one country, while in other countries equivalent BRs may require payment for accessing specific data. Such a question, combined with the covering of the operational costs of infrastructure, is likely to become the main talking point in the definition of cooperations. The infrastructure provider will promote advantages of cutting the costs for submitting requests for all stakeholders, by increasing their numbers and speed and by eliminating the unnecessary paperwork that would have been required in country-to-country transactions. The infrastructure economy justifies an additional set of skills in the field of accounting and financial audit/transparency. Funding and fee management in cross-border exchanges are typically questions that are problematic to regulate through national regulations only, except when trying to harmonise BR fee practices across the various Member States and checking for FRAND conditions. Promising projects can apply for support from European funds in initial phases[34]. Choices have to be made between concentration (public funding/monopoly) and competition between infrastructures. The perennial implementation of services will be a need-driven process where experience, economic skills and strong negotiation capacity will matter**.**

[34] The European Regional Development Fund in the case of X-Road.

2. Common governance and strategy

This chapter introduces activities that can assist public administrations in creating a common ground on governing data in base registries, along with ideas on how to unify such actions in an effective strategy.

Data Governance

A **definition of a data strategy** and the establishment of **a common data governance model** for base registries are the first important steps for the successful interconnection of base registries and their interoperability.

Currently, some Member States tackle various challenges[35] related to data governance. For example, there is a challenge of having many units of command and establishing the flows of data across them. Thus, even if legally each actor has its own smaller scope on data management (e.g. a municipality has a scope on its level, Member States have a wider scope, and so on) it is important to define how all actors collaborate with each other and how a **common data governance model** fulfils this scope.

A concept of **data governance** can be defined as:

> *"[…] a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods"[36].*

Thus, a **data governance model** focuses on the management of data throughout its whole lifecycle.

In the context of **access and interconnection of base registries**, **BRAIF** proposes to implement the following for an efficient common data governance model:

- organisational structures, clear roles and defined responsibilities for the management of data, its access and interconnection;

- common standards, rules and data policies to formalise data management across the integrated public administration;

- simplified processes for data management by the organisation.

In a wider context of interoperability models for data governance, the new EIF enters the picture with support for an interoperability model that includes four (4) layers of interoperability and two (2) distinguished types of governance: **the interoperability governance** and the **integrated public service governance**:

- The **interoperability governance** refers to *"decisions on interoperability frameworks, institutional arrangements, organisational structures, roles and responsibilities, policies,*

[35] See Annex 1.

[36] Data Governance Institute: http://www.datagovernance.com/adg_data_governance_definition/

agreements and other aspects of ensuring and monitoring interoperability at national and EU levels".[37] It can be considered as a more holistic approach to interoperability.

- The **integrated public services governance**, on the other hand, covers all the interoperability layers. According to EIF, it requires *"organisational structures and roles and responsibilities for the delivery and operation of public services, service level agreements, establishment and management of interoperability agreements, change management procedures, and plans for business continuity and data quality".*[38]

The interoperability governance of base registries is a specific form that takes into account interoperability in public services, and hence the effective deployment of generic interoperability enablers and artefacts in end-to-end public service provisioning.

Thus, to support public administrations in their work on interoperability governance, the new EIF offers public administrations **47 concrete recommendations** on how to improve governance of their interoperability activities, establish cross-organisational relationships, streamline processes supporting end-to-end digital services, and ensure that both existing and new legislation do not compromise interoperability efforts[39].

Data sharing at national and European levels could be extremely challenging due to the fact that the exchange of data occurs in a complex and changing environment that requires strong and effective cooperation. In order to achieve good results, there should be political support and stakeholders' agreement over a common vision and the objectives of a **data strategy**.

In conclusion, public officers should foster the **coordination** with Member States and with the European Commission (EC) in order to avoid redundancy and inconsistency (non-interoperable solutions serving the same objectives in public service provisioning).

Structure: Define and establish a common data governance model

As has been mentioned, the interoperability of public services based on the data of base registries could be ensured by using a **common data governance model** for the tight management and sharing of the data over time.

When setting up a common data governance model, it is important to distinguish between **upper level governance, e.g. at national level**, and **lower level governance – at individual base registries level**. Thus, there should be organisational bodies that will take decisions when reuniting different stakeholders on upper levels, e.g. national ones that establish a data security policy, and decisions made on a lower level, e.g. base registries owners who choose the way of access for their base registry from the options available in the established data security policy.

The common data governance model should be established on **the higher level** (e.g. national level) and should aim at governing the interoperability of base registries and that of public services to which they provide their data. Additionally, the governance model should be initiated taking into account that the assets of any base registry are its data and its main goal is the governance of each data-related aspect (data security, data quality, etc.).

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

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

[39] EIF: https://ec.europa.eu/isa2/eif

Thus, **a data governance model** is established by governance bodies on the upper level. The decisions which data governance bodies take at the beginning of the data governance cycle should be maintained (via change management) throughout the whole cycle[40]:

- **Definition of data policies** (e.g. data access authorisation, data quality, security);
- **Definition of organisational structure, roles and responsibilities** (e.g. on the level of the interoperability of base registries in one country, national level);
- **Definition of standards, principles and rules** (e.g. common concepts on data).

For instance, the governance bodies need to identify if common definitions of different concepts for base registries already exist, and whether they can reuse them. Otherwise, governance bodies need to reunite and define the single definition for each concept (e.g. what should fall under ‘property data’ or ‘location data’?).

A suitable moment to define and establish the common data governance model is at the very beginning of the development of any interoperability initiative, e.g. during the integration of base registries into the national interoperability framework. This allows all stakeholders to agree on the collaborative means and participate in the implementation and fine-tuning of the model.

Regarding the **lower level governance**, decision-makers at the individual base registry level should establish their own governance model and maintenance plans for their base registry, cascading the higher-level decisions (from the common data governance model) and putting them into practice.

Base registries governance and maintenance plans should cover at least three main aspects:

- The charting of **governance bodies** for the base registry, roles and responsibilities;
- A policy for the **sustainability** of the base registry and services;
- **A data maintenance plan** and **a standards management model**, including description on how the data policies and standards defined at the upper level will be implemented at the base registry’s level.

The current state of affairs on the establishment of data governance models differ in Member States.

In the complex **Belgian**[41] landscape, where the data governance is divided across autonomous regions, it represents a challenge for federal authorities to introduce standards and create registries. However, BOSA, and more specifically, the **Directorate-General for Digital Transformation (DGDT)**, is successfully addressing the challenge and has positioned itself as an organisation where other institutions and citizens go to when interested in either exposing their data or expressing their needs for establishing certain standards or data source interconnections.

The **Data Strategy vision** is to provide maximum support for the re-use of government data by offering simplified secure services and secure access to the data, respecting applicable regulations, with transparency to data managers and to natural and legal persons who are the subject of the data. Belgium is trying to reach this through:

- Unlocking and promoting authentic sources;
- Providing digital data exchange services;
- Developing common standards with a great deal of attention to security and the GDPR;

[40] These topics are being discussed more in detail in next sections of the document.

[41] https://joinup.ec.europa.eu/collection/access-base-registries/news/highlights-interview-belgium

- Promoting digital transformation by sharing best practices.

The Federal government wishes to focus on the **maximum digitalisation** of its **customer interactions** and the underlying operational processes, with the ultimate goal of improving the satisfaction of these 'customers', as well as realising efficiency benefits.

Instead, in **Spain**, there is no overall legislation for base registries, and every competent authority regulates the access to their own data. The governance is based on a protocol and platform, which act as a central point for the services offered to other competent authorities for their procedures. These procedures are registered across the governance's functionality, where data services are consumed and provided by the platform, as well as offered to these competent authorities.

One good practice example of how a common data governance model at the national level was set up and implemented, is that from the **Danish Basic Data Programme**[42].

As background information, the basic types of data (such as cadastres, buildings, road systems, watercourses and lakes), are of a sufficient quality level in Denmark – as was stated by representatives of Denmark in a relating webinar. However, they could be used in a more harmonised way in both public and private sectors. Moreover, some challenges existed in terms of data management in different base registries. Therefore, the Programme was triggered by the willingness of public administrations to overcome the challenges, such as duplications, differences in definitions, data silos not connected to each other (resulting in shadow registers)[43], and the need to improve the data management aspects.

Denmark approached that goal by setting up the Programme with the relevant governance principles and policies, managed by relevant governance bodies. Thus, the Danish government, the Local Government Denmark (KL)[44], and the Danish Regions signed the agreement 'Good basic data for everyone – a source of growth and efficiency', as a practical outcome of the objectives contained in the eGovernment Strategy 2011-2015 (of the Danish government, 2011).

The main focus of the Programme is on basic data, which was established as rudimentary information that public authorities should register about citizens, companies, property, addresses, etc., and the way it should be used and reused in the whole public sector – allowing public authorities to perform their activities efficiently across units, administrations and sectors.

There are 5 categories of basic data, selected by the Programme, which cover more than 10 authoritative registries:

1. Spatial Data;
2. Address Data;
3. Property Data;
4. Company Data;
5. Personal Data.

Moreover, the basic data governance model targeted the following rules:

- Basic data needs to be correct, complete and up to date as much as possible;
- All public authorities must use public sector basic data;
- As far as possible, basic data (excluding sensitive personal data) must be made freely available to businesses, as well as to the public;

[42] The slides 13 and 14 of this presentation about the Basic Data Programme in the context of INSPIRE: http://inspire.ec.europa.eu/events/conferences/inspire_2014/pdfs/plenaries/Grunddata_INSPIRE_JRO5.pdf

[43] The Basic Data Programme (Denmark): https://www.digst.dk/Servicemenu/English/Digitisation/Basic-Data

[44] https://www.kl.dk/english/kl-local-government-denmark/

- Basic data must be distributed efficiently, accommodating the needs of the users.

With the approval of the Basic Data Programme, all basic data are, as a rule, to be placed at the free disposal of all public authorities, private companies and citizens[45].

Denmark also identified five (5) processes to be implemented[46], as a "way to open easy-to-access high-quality basic data":

Process 1: In order to ensure the reuse of data and to prevent double registration and shadow registries, map data, cadastral maps, Central Business Register data, and company data, will be financed by the government and released to the public and the private sectors, as is already the case with address and real property data.

Process 2: In order to enhance the quality of data, the registries of map data, real property data, address data, as well as business registries, will be expanded to include other necessary data.

Process 3: In order to make it possible to link data, efforts will be made to ensure that all data conforms to the same technical requirements.

Process 4: In order to improve the distribution of common public sector data, a common infrastructure is to be established providing for stable and efficient distribution of data; a data distributor.

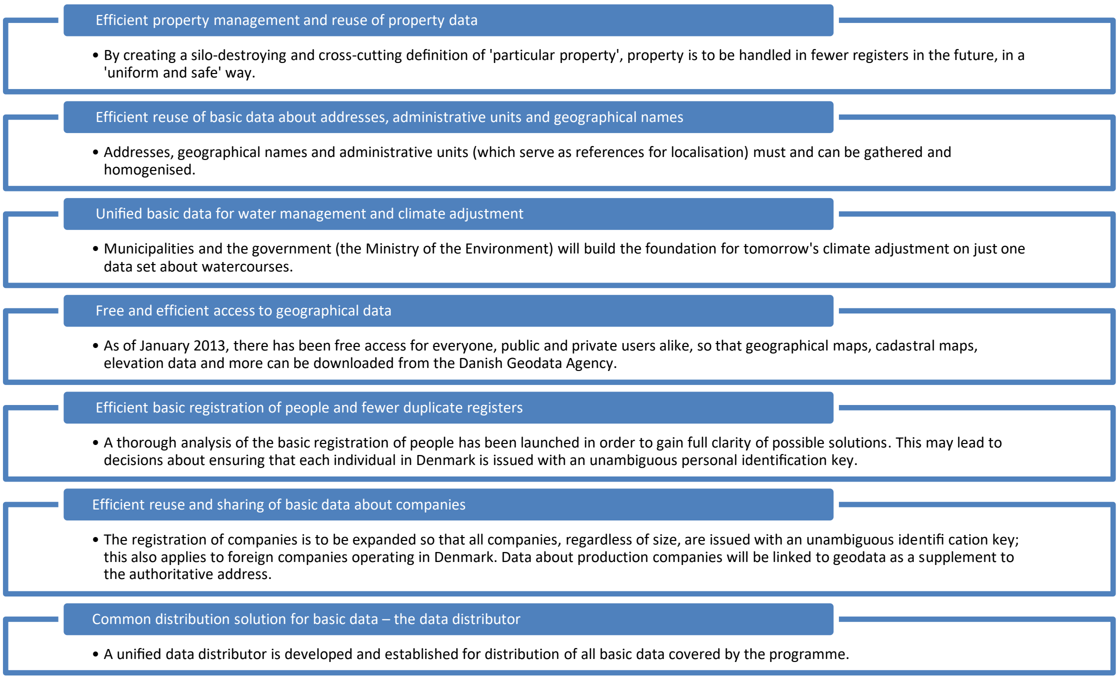
Process 5: In order to ensure efficient, effective and coordinated development and use of basic data, a cross-institutional basic data committee is to be established.

In order to implement the established model, Denmark identified the data silos and then defined a sub-programme for each of them, as described in the following diagram[47].

[45] The Basic Data Programme – A Danish Infrastructure Model for Public Data (Denmark): https://www.academia.edu/29858925/The_Basic_Data_Programme_A_Danish_Infrastructure_Model_for_Public_Data

[46] Good basic data for everyone – A driver for growth and efficiency, The Danish Government, Denmark October 2012, presentation available on ABR Collection on Joinup here: ABR_2019-04-08_Denmark_Data_Distributor.pptx

[47] Excerpt from aforementioned The Basic Data Programme – A Danish Infrastructure Model for Public Data.



Efficient property management and reuse of property data

- By creating a silo-destroying and cross-cutting definition of 'particular property', property is to be handled in fewer registers in the future, in a 'uniform and safe' way.

Efficient reuse of basic data about addresses, administrative units and geographical names

- Addresses, geographical names and administrative units (which serve as references for localisation) must and can be gathered and homogenised.

Unified basic data for water management and climate adjustment

- Municipalities and the government (the Ministry of the Environment) will build the foundation for tomorrow's climate adjustment on just one data set about watercourses.

Free and efficient access to geographical data

- As of January 2013, there has been free access for everyone, public and private users alike, so that geographical maps, cadastral maps, elevation data and more can be downloaded from the Danish Geodata Agency.

Efficient basic registration of people and fewer duplicate registers

- A thorough analysis of the basic registration of people has been launched in order to gain full clarity of possible solutions. This may lead to decisions about ensuring that each individual in Denmark is issued with an unambiguous personal identification key.

Efficient reuse and sharing of basic data about companies

- The registration of companies is to be expanded so that all companies, regardless of size, are issued with an unambiguous identification key; this also applies to foreign companies operating in Denmark. Data about production companies will be linked to geodata as a supplement to the authoritative address.

Common distribution solution for basic data – the data distributor

- A unified data distributor is developed and established for distribution of all basic data covered by the programme.

Figure 2: The 7 sub-programmes of the Danish Basic Data Programme

Each sub-programme targets different governance bodies and data governance aspects, based on the rules described in previous paragraphs.

In conclusion, with the implementation of the Basic Data Programme, Denmark defined the basic data as a common digital resource to be freely used for commercial and non-commercial purposes, and established a common data governance model through data governance rules, and specific governance bodies with roles and responsibilities for different categories of basic data.

Well-defined related roles and responsibilities play an essential role in the successful implementation of the aforementioned model.

Roles: Clearly define responsibilities and liabilities

The definition of the data governance model requires a compromise between providing stakeholders with adequate means to channel their requirements, needs and/or complains, and a flexible decision-making process, allowing to cope with changes in a timely fashion. During the description phase, different **organisational roles** should be identified, such as steering committee member, service owner, etc., including **specific data-related roles**, such as data owner and data steward. **BRAIF** provides a detailed overview on typical data governance bodies and their roles and responsibilities in Chapter 2.1.1. Data Governance[48].

Thus, specific responsibilities and liabilities regarding the management of data shall be clearly defined for each role. The difference between the terms ‘Responsibility’ and ‘Liability’ lies in the following:

- **Responsibility** is used to define who must do what, when, and what for, under which circumstances;
- **Liability** is used to define what are the consequences, who must face them – and how – when something goes wrong.

[48] BRAIF: https://joinup.ec.europa.eu/collection/access-base-registries/document/braif-framework-base-registries-access-and-interconnection

Sources of liability can be crucial when managing data (e.g. reduced data quality, data loss), along with failure in determining who is responsible for disclosing and sharing the data (e.g. lack of defining who, when and how the data can be disclosed, or transmitting the data through insecure means).

It is extremely important that public services relying on base registries should clearly define both aspects:

- In the case of responsibilities, the design of the service should include **a clear workflow** in which each actor, process, inputs and outputs are depicted. The result is a clear **end-to-end vision of the data lifecycle**, which should be used to draw and document how the data should be maintained and shared, and by whom;

- Concerning liabilities, the officers responsible for the development and maintenance of base registries and public services should define and keep **a list of liabilities** which each actor may incur.

Examples in which the definitions of responsibilities and liabilities are frequently used are the initiatives interconnecting business registries. For instance, the registration of the insolvency or bankruptcy of a company. If the company has branches in other Member States, the registry has the obligation of notifying the event to each business registry in any Member State where the company had a registered branch. An error in the name or address of a company may end up in a request for striking-off the wrong branch. If that happens, the company could experience real damage and claim liability.

To overcome these challenges, during the development of the **Business Registers Interconnection System (BRIS)** Trans-European System, the Member States identified the need for a clear definition of the liabilities[49] and requested to document them.

On a national level, for example, in **Italy,** there is a national code for digital administrations that establishes which databases are of national interest and who are the administrations in charge of each such database. Hence, there are base registries with information of national interest, e.g. a national residents population registry, the registry of tax administration etc., and each administration is responsible for the maintenance and provision of legal value for the data they maintain.

One good practice example on how different roles, responsibilies and liabilities are implemented is **X-Road[50]** (for Estonia and Finland). The X-Road ecosystem consists of an X-Road Operator, Member organisations, and Trust Service Provider(s). As the owner of the X-Road ecosystem, the Operator is responsible for all the aspects of the operations. The responsibilities include defining regulations and practices, accepting new members, providing support to Members, and operating the central components of the X-Road software. X-Road Members are organisations that have joined the ecosystem and produce and/or consume services with other Members. Thus, a Member organisation can be a service provider, a service consumer, or both. A functioning X-Road ecosystem requires two types of trust services: 1) a time-stamping authority (TSA) and 2) a certification authority (CA). Trust Service Providers are organisations providing these services. Hence, each organisation has its own clear role and responsibilities in the data management process.

[49] [BR22] System Wide Requirements in the ABR Catalogue of Solutions on Joinup.

[50] X-Road: https://x-road.global/

Data Policies

Data policies help ensure that all data and information assets are properly – and consistently – handled and, thus, should be considered a fundamental aspect of any data governance model. By setting up proper data policies early on in a data strategy, public administrations can build on available information assets effectively and efficiently.

Thus, this section is dedicated to the aspects of implementation of data policies practices that represent the practices required to ensure a correct management of the data throughout all the processes.

There are traditional data policies that need to be elaborated and implemented, as listed below:

- **Data policy on authorisation and accessibility** is based on the national legal frameworks, and it defines the legitimate users that can be authorised to access the data in base registries, the types of access rights, etc.;

- **Data protection policy** is based on data protection-related legal frameworks[51], and it defines how these legal requirements apply to the interoperability of the base registries;

- **Data security policy** is based on the data policy on authorisation and accessibility, and it defines how the channels that transmit data from one registry to another (or to a registry of registries) are protected, which security protocols are being used, etc.;

- **Data quality policy** defines procedures, roles and responsibilities, liabilities to ensure the data provided in base registries is accurate, complete and consistent.

It is important to take into account that data policies should be aligned with an overall IT security policy, too.

Data Policy on Authorisation and Accesibility: Ensure the right users access original and authentic data

The data kept in the base registry is, by definition, public data, but that does not mean that anyone or any system can access it at any time or at all. Only **legitimate** users are able to access the related data in base registries. In general, these users are as follows:

- Persons to whom the data relates to;
- Persons involved in an administrative procedure (or empowered to represent these procedures);
- Authorities that need to access the data for the execution of their duties and in accordance with the laws.

In practice, such users, as well as aspects of their access rights and responsibilities, should be defined and specified in each national legal framework. Moreover, the process of identifying, tracking and logging of who is requesting access needs not only an operational algorithmic solution, but also the implementation of well-designed data and metadata models that will manage the **authentication**, **authorisation** and **annotation / logging** of the operations.

The EU has **Regulation (EU) N°910/2014**[52] that concerns the electronic identification and trust services for electronic transactions in the internal market[53], to which all new solutions should comply with. In a

[51] E.g. EU Regulation 2016/679 (repealing Directive 95/46/EC), known as the General Data Protection Regulation (GDPR).

[52] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.L_.2014.257.01.0073.01.ENG

[53] EUR Lex: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.L_.2014.257.01.0073.01.ENG

nutshell, the regulation ensures that people and businesses can use their own national electronic identification schemes to access public services in other EU countries where electronic IDs are available.

To understand how the aforementioned Regulation can be technically implemented, public administrations can check the solutions developed jointly by the European Commission and the Member States, namely the ones promoted by DG CNECT through the CEF[54] funding. Thus, **CEF Digital** provides a **catalogue of reusable building blocks** developed by the EC and MS along these types of initiatives. This way, service providers and national eID infrastructure owners could study **eID**[55] and check with CEF regarding guidelines on how a Member State could implement a needed eIDAS-Node.

Member States use different approaches to manage the access and authorisation of data in their BRs or RoRs, and some of them offer good practice examples for inspiration.

MyGuichet[56], in Luxembourg, is connected to more than 15 **authentic sources** and can reuse data from them. The authentic sources are not directly interconnected and no possibility exists for a civil servant to view, in one glance, all the data kept on a user in the different authentic sources. **Access rights** are always managed at the level of each authentic source, and a civil servant can each time – and only if they have the right to do so – access one specific authentic source by authenticating with it. However, users can view in their MyGuichet personal space (after authenticating with their **eID**) all the data kept on them. In order to guarantee a high level of data protection and, as far as possible, privacy by design, Luxembourg is not striving to create a RoR. Instead, its trying to make reuse of data more efficient, access to their data easier for users, and access to citizens' data by civil servants transparent via **tracking and logging** of such accesses.

In **Denmark**[57]**,** the access to centralised data is managed by the responsible base registries owners, and not by the authority operating as the **Data Distributor**[58].

Currently, three types of access rights are in place, via username with password or eID:

- Anonymous;
- Known User;
- Individual Permission.

The Data Distributor is a fairly new solution that it is growing quickly, and it will be the primary source of access to data collected from base registries in the country. A standardised management of access is being considered to a certain extent, but for the moment the rules of access are distinct to each responsible base registry.

In **Estonia** and **Finland**, **X-Road** provides an organisational framework which defines an onboarding process that an organisation must complete to become an X-Road member, and to be able to exchange data with other X-Road members. The X-Road Operator controls who are allowed to join the ecosystem. For example, in Estonia and Finland the ecosystems are open for all kinds of organisations (public, private, non-profit etc.) and joining them is free.

First, a new member organisation must be registered to the X-Road ecosystem by the X-Road Operator. The organisation's identity is verified during the onboarding process by a trusted certificate authority. Once the onboarding process has been completed, the new member organisation must contact other

[54] Connecting Europe Facility: https://ec.europa.eu/digital-single-market/en/connecting-europe-facility

[55] https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eID

[56] https://guichet.public.lu/en/myguichet.html

[57] Presentation from ABR webinar: ABR_2019-04-08_Denmark_Data_Distributor.pptx

[58] Danish Data Distributor: http://datafordeler.dk/ (available only in Danish).

member organisations that provide services that the new member wants to access. Getting access to services requires a service agreement between the service consumer and service provider. The service agreement is done with each service provider separately, and the service provider defines the content of the service agreement. Once the service agreement has been concluded, the service provider grants the service consumer access to the services[59].

Regarding the access to data, one should ensure that the data is provided by an original and authentic source, defined in the MS as being a trustful source of data via a related data strategy or other legal ways (see Legal aspects of access to base registries). Thus, additional challenges might arise while the data is shared across different public administrations, especially cross-border, and solutions should be implemented to:

- Ensure that the data was actually provided by the legal entity who claims its issuance (**non-repudiation at the origin**);
- Justify that a public service providing data from a base registry cannot deny the reception of the data (**non-repudiation at the destination**);
- Ensure (and demonstrate, if necessary) that the content shared by a base registry has not been altered before it reaches its recipient (**data integrity**).

In this sense, one technical development that system developers are recommended to assess is the **provenance ontology** developed by **W3C**. Among other interesting features, the **PROV Ontology**[60] uses the OWL2 web ontology language to express the PROV Data Model, and can be used to represent and expose control information generated in different systems and under different contexts. It also proposes a scalable and yet efficient storage model by exploiting structures of provenance logs and separating metadata from the generating process. Moreover, it not only addresses the creation of data, but also proposes a capturing provenance model for information from web-based data access, as well as information regarding the creation of data.

One example of a trans-European service that exports electronic documents with references to service providers and base registries is the **European Single Procurement Document** (**ESPD**) **Service**[61]. These references are provided by service providers that aggregate data from the base registries or redirect data to a specific base registry[62].

An option which ensures a strong certainty of non-repudiation consists of signing electronically the shared data by using **qualified signatures**[63]. From the legal perspective, the eIDAS Regulation facilitates this option, as it promotes the use of Secure Signature Creation Devices (SSCD) to create qualified signatures. Another option to consider is the use of **Hardware Secure Modules**[64] (**HSM**) solutions for strong and efficient authentication and non-repudiation purposes, especially in automated processes. Another common method used to ensure non-repudiation consists of **logging**, i.e. creating and storing electronic evidence of who shared the data, at what exact date and time, from which system, to which address, etc. One of the publicly available logging and monitoring solutions is the **European Criminal Records Interconnection System (ECRIS)**[65], in which the monitoring process is carried out

[59] https://www.niis.org/history

[60] PROV Ontology: https://www.w3.org/TR/prov-o/

[61] ESPD: https://ec.europa.eu/growth/tools-databases/espd/filter?lang=en

[62] E.g e-Attestation in France (https://www.e-attestations.com/fr/) or ROLECE in Spain (http://www.minhap.gob.es/es-ES/Areas%20Tematicas/Patrimonio%20del%20Estado/Contratacion%20del%20Sector%20Publico/Paginas/ROLECE.aspx)

[63] Article 3 of the eIDAS Regulation for the normative definitions of the terms (EU): http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014R0910&from=EN

[64] HSM are commercial products, physical computing devices that safeguard and manage digital keys for strong authentication and provide cryptoprocessing. These modules traditionally come in the form of a plug-in card or an external device that attaches directly to a computer or network server.

[65] ECRIS: http://data.consilium.europa.eu/doc/document/ST-11274-2011-INIT/en/pdf

through continuous evaluation and correlation of non-personal statistical data produced by the logging systems and procedures, throughout all message exchanges performed via ECRIS (i.e. data that is not protected by specific European legislation regarding free movement of data and privacy, such as the existing general rules on the protection of personal data processed in the framework of police and judicial cooperation in criminal matters).

Data Protection Policy: Ensure that data protection laws are enforced[66]

Protecting the **privacy** of **personal data** while ensuring the principle of public availability of the base registries may constitute a challenge.

In legislation, it is recommended to always refer to the whole set of data protection-related legal frameworks currently in force, and on how they will be applied or complemented for the interoperability of base registries and public services. Especially relevant in this sense is EU Regulation 2016/679 (repealing Directive 95/46/EC), also known as the **General Data Protection Regulation (GDPR)**.

In order to implement this regulation, MS and other countries defined clear guidelines on how to ensure citizens' rights in relation to their personal data protection.

As stated in the chapter Legal aspects of access to base registries, when aligning with the legal framework, one should rethink public services and the interoperability with base registries and consider whether they could become more user-centric. To do so, policy makers, jurists and IT experts should assist the e-Government authorities in the configuration of a legal framework, to define the responsibilities regarding the management of the data and to provide citizens with "**proportionated" legal instruments for self-control of personal data**.

Along these lines, one original approach undertaken in the **Netherlands** has been the definition of a "standard" to give back to individuals the control of their data in the digital world. This open standard, named **QiY**[67], defines a trust framework for individual users, companies and governmental organisations to obtain full, secure and private control of their personal data, and the possibility to share their data of choice with people, companies and governments they are dealing with.

Related to the "proportionated" instruments, one example could be the objective of unburdening citizens, reflected in the legislation in **Spain** (Laws 39 and 40 of 2015) that amend the previous legal regime on public administrations and common administrative procedures. Now the situation is more flexible and, similarly to other EU countries, Article 28 of the new Law 39 allows the sharing of data, unless citizens expresses specifically their non-consent.

Interesting examples of how the control of data can be legally and safely given back to citizens (when accessing base registries) are the solutions **eBoks[68]**, **Finland's MyData**[69] and the **Czech Republic's Datove Schranky**[70], that have strong roots in their national legal frameworks. Many other MS have also worked in this regard, such as Sweden and Belgium. Sweden and Denmark have also considered user-centric approaches for the compulsory relationship between legal entities and public administrations.

[66] See Legal section for detailed information.

[67] QYI Foundation: www.qiyfoundation.org

[68] eBoks: https://www.e-boks.com/danmark/da

[69] The proposal of a framework, principles, and a model for a human-centric approach to the managing and processing of personal information (Finland): http://www.lvm.fi/publication/4440204/mydata-a-nordic-model-for-human-centered-personal-data-management-and-processing

[70] An example from the Czech Republic, as well as other ties of data boxes to base registries on this ePractice article, published on Joinup: https://joinup.ec.europa.eu/community/epractice/case/data-boxes-%E2%80%93-easy-economic-and-environmentally-friendly-delivery-official-d

Data Security Policy: Ensure the security of the data access and its communication

This policy aims at securing the communication channels and the means used to transmit the data from one base registry to others, and it defines that a strong security environment is needed to ensure the data exchange in a secure manner.

The previously mentioned EU Regulation also covers the possibility of using certificates for website authentication, which may be used to assure users (in particular, citizens and Small and Medium-Sized Enterprises (SMEs)) that behind the website there is a legal person identifiable by trustworthy information. The Regulation sets clear requirements for website authentication certificates to be considered trustworthy together with minimal obligations for providers of such certificates, with regard to the security of their operations, their liability and their (light-touch) supervision regime.

Many MS have also developed and implemented strong security measures, and among good practice examples to follow are **X-Road**[71] **(X-Tee)** in **Estonia** and **Finland**, the **SARA network**[72] in **Spain**, etc.

Taking the example of X-Road, the Security Server is the entry point to X-Road. It represents an OSI Level 7 Application Gateway, and it mediates service calls and service responses between Information Systems. The Security Server encapsulates the security aspects of the X-Road infrastructure: managing keys for signing and authentication, sending messages over a secure channel, creating the proof value for messages with digital signatures, time-stamping and logging. The requests sent over X-Road are protected from eavesdropping, unauthorised change, loss and duplication. Message routing is based on organisation- and service-level identifiers that are mapped to physical network locations of the services by X-Road. All the evidence regarding the data exchange is stored locally by the data exchange parties, and no third parties have access to the data. The Security Server manages two types of keys. The authentication keys are assigned to a Security Server and used for establishing cryptographically secure communication channels with other Security Servers. The signing keys are assigned to the Security Server's clients and used for signing the exchanged messages. A trusted certification authority issues certificates for the keys, and certificates issued by other certification authorities are considered invalid.

Data Quality Policy: Ensure and control the quality of the data by all means

As has been already stated, the quality of data is **essential** and if not ensured public authorities will probably face loss of trust and administrative actions that may have legal consequences, as well as result in data inconsistencies, etc. Thus, this policy aims to identify how to provide accurate, understandable, complete, and consistent data and define the technical support for them.

In order to implement this policy, Master Data Management (MDM) approaches and solutions can be taken into account, but base registries owners should also define and implement their own **Data Quality Assurance Plan** to identify the procedures, roles, responsibilities, liabilities and workflows in the correct manner. While drafting this plan, any occasion and means to check the quality of the data should be used, not only during the registration of new master data, but especially while registering modifications and, if possible, before sharing the data.

The importance of data quality is reflected in the legislations and strategies, and translated into practical solutions that many Member States and EU institutions have been developing for the past few years. As is well known, without a legal mandate, the success of any type of service is jeopardised. Thus, the first step is to define the data quality in national legislations, then to have it as an essential part of the national strategy related to data management, which when implemented brings many tangible benefits.

[71] Article and video on the X-Road infrastructure (Estonia and Finland): "A secure and Scalable Infrastructure for Inter-Organizational Data Exchange and e-Government applications"; X-Tee User Guide: https://moodle.ria.ee/mod/page/view.php?id=420

[72] Red Sara (Spain): http://administracionelectronica.gob.es/ctt/redsara#.V4QfRvmLTIU (In Spanish)

Regarding the legal part, one good example is the Spanish data protection **Law 15/1999**[73], **Article 4 on Data Quality**, that is also mentioned in the **EU Regulation 2016/679**[74] (General Data Protection Regulation, Article 47, Binding corporate rules; paragraph D referring to Section 63 on Consistency).

Regarding the implementation of a defined strategy on data management, one way of facilitating the control of the data quality is to involve citizens, businesses and service providers in the process: for this, **user-centred or business-centred solutions** (such as Data Boxes and e-Government personal workspaces) should be considered. The advantages of these types of solutions is that they offer a view of the data to users and allow them to check and validate their data through secured electronic channels.

Different MS already implemented solutions supporting the above, and among good practice examples to study are **eEsti**[75] **in Estonia**, '**services for citizens and business'**[76] **in Austria**, '**Mi Carpeta'**[77] in **Spain**, '**My Data'**[78] in **Finland**, and **MyGuichet.lu**[79] in **Luxembourg**.

This has also been the case with the **Danish e-Boks project**[80]. The idea behind this approach to allow citizens to monitor their data as it arrives directly from the key national registries, and to notify the authorities about any possible quality issues related to their data. Currently, only Denmark has been able to back up by law the use of this solution for the relation between citizens and public administrations (and it is being used by 100% of the population, with the next step being the enforcement of a similar use of data-boxes by companies).

For example, with regard to real **use cases** on data quality in Denmark[81], the update of data occurs on a base registry level, and a **Data Distributor** joins the process when the data are ready for publication and distribution to data consumers. In the case the data is incorrect, then the update follows a complex set of rules. In simple cases, citizens can correct mistakes, if stated in the law, by updating their own data, e.g. on the municipality level. In such cases, the responsible entity to fill-in the corrected information is the first contacted registry, i.e. the municipality, and other registries that reuse these data are responsible to approve the data inputted by the first registry. Challenges might occur when a specific case is regulated differently among the laws. However, Denmark is continuously improving the legislation, requesting feedback from the authorities that are applying the laws.

When it comes to data quality in **Luxembourg**, **citizens** can contribute to raise the accuracy of data for some authentic sources via possibilities given to them in **their personal space on MyGuichet.lu**. This permits them to directly view the data kept on them in authentic sources, and request online corrections of inaccurate data. Additionally, what is interesting is that the Luxembourgish government also sends an extract from the registry once per year by postal mail, allowing citizens to notify the authorities in case a correction is needed.

In the **Netherlands**, and other countries as well, there is also a responsibility of 'data subject' to ensure that the provided data is correct, e.g. there is an obligation to report a move to another address to the

[73] https://www.boe.es/diario_boe/txt.php?id=BOE-A-1999-23750

[74] http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN

[75] E-Services (Estonia): https://www.eesti.ee/est/teenused

[76] Service for Foreign Citizens (Austria): https://www.help.gv.at/Portal.Node/hlpd/public/en, The service platform for business: https://www.usp.gv.at/Portal.Node/usp/public

[77] Citizen Folder (Spain): https://sede.administracion.gob.es/carpeta/clave.htm

[78] My Data (Finland): http://www.lvm.fi/publication/4440204/mydata-a-nordic-model-for-human-centered-personal-data-management-and-processing

[79] https://guichet.public.lu/en/myguichet.html

[80] E-Boks (Denmark): http://www.e-boks.dk/

[81] Information on how Denmark approaches data quality improvements can be also consulted here: https://economie.fgov.be/en/themes/enterprises/crossroads-bank-enterprises/data-quality

relevant authorities and base registries. What some Member States are missing from national frameworks regarding data quality is the functionality or service that enables people to report errors in the base registries (i.e. the absence of a feature to highlight errors).

However, one good example is in **Flanders, Belgium,** where in order to resolve this issue they created the **OSLO Framework**[82] standard, which allows notifications and feedback to be provided for a topic, and to the right organisation, thus it is a general data standard. If someone has a comment on a topic, then they can notify the relevant distributor who has to implement any applicable changes. This process guarantees that the comment will reach the correct base registry or responsible entity, disengaging thus the user from dealing directly with the issue.

Norway also invests great efforts and means to ensure that the data collected (i.e. captured or created) are of the best quality, and executes a tight control on their quality from that moment on until they are no longer used or legally disposed of[83]. Norway established the platform **Altinn**[84], where the data to and from the government are shared (not stored). Data from the public sector can be shared through the portal, allowing pre-filling, when reporting to the government. About 95% of reporting from businesses to the government in Altinn is machine-structured data, and with regard to data quality, the data are validated as entered, and their quality is assured by frequent and various usage.

In **Slovakia**, it has taken several years to come up with significant progress in the domain of linked government data. Slovakia has established the 'Strategic Priorities – Enhanced Data and Open Data', and launched several working groups online, dedicated to discussions in different related domains, inviting interested users to participate in debates.

Thus, the working group 'IPA Working Group K9.4 Better Data'[85] was established, aiming to improve the use and quality of data in public administrations. This working group addressed the following topics, among others:

- Data quality;
- Reference data, codebooks and URIs;
- Open data;
- Linked data;
- "One Time and Enough";
- My Data service;
- Big data in public administration.

One example from this working group's discussions was structured around the main challenges that people wanted to resolve such as:

- What should be addressed at the central level and what at the local level, what are the combinations?
- What licence models will be used?
- How will privacy be set?
- How will the credibility of published data be assured?
- What do we want when saying 'better data'?

The discussions touched such topics as data quality target withinopen data interoperability rules, disclosure of datasets vs hiding them behind APIs, etc. The working group debated to establish data quality levels, where all new and innovated vendors publish open data of a central nature (reference data, central registries, metadata or priority datasets, data and electronic services to towns and

[82] https://data.vlaanderen.be/doc/applicatieprofiel/notificatie-basis/

[83] Information received during interview with MS: https://joinup.ec.europa.eu/collection/access-base-registries/news/interview-highlights-1

[84] Altinn: https://www.altinn.no/en/

[85] https://platforma.slovensko.digital/t/upvii-pracovna-skupina-k9-4-lepsie-data/2736

municipalities) must be published at quality level 5, meaning they must use the URI (Uniform Reference Identifier/Uniform Resource Identifier) to identify ISVs' data, and the Central Model of Public Administration Data to describe them.

The discussions in this working group lasted almost 2 years, catching the attention of lots of users, who actively participated in debates, and ended by launching the '**Strategic Priority Open Data**[86]' document, that went through many iterations based on the working group members' contributions.

What is worth mentioning is how the challenges were solved and what solutions were agreed to be used by key data controllers, which can be summarised below:

- The central registry will not store the data;
- Each public office has its own reference for master data management, and is responsible for the data quality;
- The central component for Government to government (G2G) data access will be a central software, a unified API, and a unified data model;
- The access control will be based on the basic level of "what authority can see, what type of data, for what purpose", on the central, data provider detailed level;
- "Who can have what data" is established by what the law allows.

Another important part of the data policy is to ensure that data served through the public services of one public administration should not be duplicated at the business level, since this may result in poor data quality (e.g. inconsistency, obsolete or non-updated data) and would require a duplicated investment in governance and sustainability.

Occasionally, the reasons behind data duplication lie in **the purpose of the data**. For example, real estate registries and land registries can register similar (if not identical) information, but for different purposes: real estate registries ensure the legal entities' rights, while land registries identify their taxation obligations. For example, in **Spain**, the Real Estate Registry (Ministry of Justice) and the Cadastre (Ministry of Finance and Public Administrations) came up with a coordinated resolution[87] to harmonically identify the location, shape and area of a real estate: the Real Estate Registry uses the geo-referenced graphical description of its real estate through the specifications and resources defined by the Cadastre. This aims to provide consistency and quality to the data about the location, delimitation and areas of the real estate in legal procedures. The technical solution is based on the specification developed by **INSPIRE**[88].

[86] Strategic priority: Open Data

[87] http://www.catastro.meh.es/esp/CoordinacionCatastroRegistro.asp (in Spanish)

[88] http://www.catastro.minhap.es/documentos/formatos_intercambio/Formato%20GML%20parcela%20catastral.pdf (in Spanish)

3. Standards and lean processes

This chapter aims to facilitate public administrations with information on how to redesign and simplify business processes, introducing the conceptual part and going into further details.

Organisational Interoperability

Organisational interoperability, being part of the EIF interoperability model's governance, explains how public administrations align their business processes, responsibilities and expectations to achieve commonly agreed and mutually beneficial goals. It implies documenting, integrating and aligning business processes and associated exchanged information.

The organisational interoperability in the case of base registries represents the area where high attention is needed. It is both challenging at national and EU level:

- **National level**: different authorities manage similar or identical information, frequently organised and represented in heterogeneous ways, and sometimes very fragmented (distributed into many different systems);

- **EU level**: the registry has to interoperate with other registries, where each MS has its own rules, legal framework, business processes, language and technologies.

Moreover, putting in place flexible – and possibly complex – architectures requires an organisational effort mainly addressed to **aligning all the processes** executed in each of the different stakeholders' systems. Its achievment is feasible by gathering the stakeholders together and making them agree on using **common standards**, while allowing them to continue performing their usual processes. And if the alignment of processes is not possible, or extremely complex, the stakeholders should reach **agreements** and make decisions on the development of **common processes** for them to follow.

A good governance plan should take this organisational challenge into account and devise the right governance structure for the stakeholders. The plan should also clearly detail the responsibilities of each actor being part of the system, as well as the liabilities these actors may incur (see section 'Data Governance'), and it should also envision how common processes – or different processes from various stakeholders (base registries owners, other actors) – will be linked, and how interactions will take place. Therefore,

> *"organisational interoperability is concerned with setting the foundations for collaboration between organisations, such as public administrations in different Member States, in order to achieve their mutually agreed goals in providing interoperable public services that reflect the users' needs".* [89]

During interviews, some Member States emphasised the need for drafting **realistic and implementable roadmaps**, in which priorities should be well-defined and phased, for interoperability initiatives that may span over long periods of time. The recommendation would be that governance bodies should rely on

[89] The study mandated by ISA "D02.03 – European organisational interoperability vision".

the experience of small multidisciplinary teams formed by legal, business, and technical specialists, and on the inspiration from known and largely used approaches (for States mentioned below) that facilitate risk measurement and leverage the lessons learnt, while controlling the development and long-term maintenance of the initiative.

Envision the global (holistic) organisational picture

In many cases of Member States, the same data exist in different base registries and sometimes there is no visibility on what type of data are contained in which base registry, thus these Member States experience issues with data quality and data duplication, following heavy processing of data collection. In order to overcome these challenges, rationalised processes are needed, based on master data management (see section 'Master Data Management'). This includes the mapping of the data and their location, how they are used, what is the data quality, what types of policies exist etc.

Together with setting the data governance model and data policies, it is crucial to organise stakeholders and have them meet to define a **programme** and **set up common processes** or **simplify existing ones**, with the aim to rationalise the activities related to base registries. Thus, many countries have started their national programmes or are already progressing with their implementation.

A good example on overcoming organisational challenges is **Denmark,** with its **Basic Data Programme**. The Danish government and Local Government Denmark (KL) entered into an agreement on basic data in 2012, and started rationalising their base registries, establishing the following targets[90]:

- In the area of real properties, the information about real properties and buildings, as well as their owners, will be registered uniformly in the relevant authentic registries[91];
- In the area of addresses, a coherent infrastructure will be set up and the databasess will be improved to ensure that data on addresses, place names and administrative units are made efficiently available to everyone;
- In the area of utilities management, national common public sector basic datasets will be established for watercourses;
- With regard to geodata, access to such data will be made open for all and the need for a more efficient and binding model for geodata maintenance will be discussed based on the finance agreement of 2014;
- With regard to business data, the databases in the Central Business Register will be improved and there will be open access to business and company data.

The **Stelselcatalogus**[92], in the **Netherlands,** can be used as reference for a data governance model on the interoperability governance. It establishes a clear governance structure, data quality maintenance guidelines and a practical organisational model for the interoperability of base registries with concrete public services. One of the advantages of this catalogue is its integration with the government central portal and tools, which allows public administrations and service providers to orchestrate administrative processes and execute public works in a coordinated way.

As has been mentioned, the setting of the foundations for collaboration between organisations refers to the alignment of cross-organisational business processes and smart service orchestration, ensuring this way the seamless interaction and data exchange among distinct systems using standards and common interoperability interfaces (see chapter 4).

To support the implementation of the programmes in public administrations, the EU provides different useful methodologies, elaborated by the European Commission (EC). For example, **PM^2 Project**

[90] Presentation of Danish authorities in ABR webinar: ABR_2019-04-08_Denmark_Data_Distributor.pptx

[91] See section '1. Common governance and strategy', sub-section 'Data Governance', 'Structure: Data Governance, Define and Establish a governance model'.

[92] System Catalog (Netherlands): https://www.logius.nl/diensten/stelselcatalogus

Management Methodology[93] would support the organisational part of the implementation and **Agile@EC[94]** would provide details on how to realise technical implementations. The **ABR Catalogue of Solutions**[95] also collected some examples of solutions, structured around these methodologies.

There are also other European initiatives, such as the **European Business Registers Interconnection System (BRIS)** and the **European Criminal Records Information System (ECRIS),** that offer good examples on how to set up a project governance on the EU level cross-border governance model.

Establish interoperability agreements to ensure base registries and public services sustainability

Ensuring the sustainability of base registries and public services based on their data does not end with their development and the automated provision of the data. Base registries will continue registering new data and changes, and should ensure their quality, trustworthiness, and permanent accessibility.

Legal enforcement is essential in this situation, as the regulatory framework supporting a base registry should specify how the maintenance and evolution of the base registry are to be financed, by whom, and should cover these aspects independently of whether the registry is managed directly by the government or through a private organisation. Organisational planning, process alignment, and well-orchestrated workflows are also important.

One way of enforcing the sustainability of the services, and consequently of the underlying base registries, is through the formalisation of **interoperability agreements**, which should cover all the dimensions of the interoperability.

There are several types of generic interoperability agreements, such as:

- **Interconnection Security Agreements (ISA)**, which specify technical and security requirements for managing a secure connection between two or more entities. For example, it may stipulate certain types of encryption for all data in transit;

- **Service Level Agreements (SLA)**, that define the parties involved in the system, their roles and obligations as well as the organisational and technical conditions and different ways to use the system. Sometimes they even define the governance or the coordination policies for involved parties;

- **Memorandum of understanding (MOU)**, which expresses an understanding between two or more parties indicating their intention to work together towards a common goal. It is similar but less formal than an SLA and does not include monetary penalties;

- **Business partners agreement (BPA)**, a written agreement that details the relationships between business partners including their obligations, the share of profits or losses each partner will take, their responsibilities to each other, and what to do if a partner chooses to leave the partnership.

A rich and varied range of approaches, models, initiatives and examples on interoperability agreements can be found at the European and national levels.

[93] Entry on ABR Collection on Joinup: [EU12] PM2 - Governance, PM^2 community: http://europa.eu/!gb87FF via contact EC-PM2@ec.europa.eu

[94] For details contact EC-Agile@ec.europa.eu

[95] ABR Catalogue of Solutions: https://joinup.ec.europa.eu/collection/access-base-registries/abr-catalogue-solutions-0

On the Member State level, the existing cross-border **bilateral agreement between Estonia and Finland** illustrates the development of a **joint data exchange platform**[96] in order to make digital services mutually accessible for inhabitants, by reusing existing national infrastructure (i.e. the Estonian X-Road[97]). In practical terms, this means that the data kept in base registries (tax boards and social insurance agencies, for a starter) are made accessible to citizens and authorities of both countries, allowing them to avoid repeatedly submitting data when operating in either country, if they have already filed in one country already. For instance, entrepreneurs will no longer have to prove, in both States, the absence of tax arrears. Also, those wishing to officially work in Finland will no longer have to submit there, every year, the paper copy of the pension insurance certificate.

An interesting research project with an approach to bilateral agreements was conducted within the European Land Registry Association (ELRA), namely, the **Cross Border e-Conveyancing**[98] **(CROBECO)** project. The framework for an alternative conveyancing process for foreign buyers of real estate was based on a process with tools to support foreign conveyancers, as described in a Cross Border Conveyancing Reference Framework (CCRF). Drafts of the CCRF were discussed with ELRA members and other stakeholders. The final version of the CCRF was approved by the General Assembly of ELRA in May 2012. A follow-up project named CROBECO II implemented the tools proposed in the CCRF.

Interoperability agreements are usually modelled on **templates**. One project that is providing simple, but practical templates, is the **Centre of Excellence for Information Sharing**[99] **(CEIS)**. This **Information Sharing Agreement (ISA)** defines the arrangements for processing data between different partners and sits underneath the overarching Information Sharing Protocol (ISP) / Partnership Agreement.

In the domain of business registries, we have the European Business Registers (EBR) organisation (and platform) which is based on the cooperation between the participating registries on an **Information Sharing Agreement**[100], in which the contracting parties undertook the duty to give each other access to information stored in their business registers. The experience and developments accumulated from the EBR were taken into account during the inception of the **Business Registers Interconnection System (BRIS)**[101].

Other initiatives go beyond the definition of templates and the design of **interoperability agreement models**. That is the case of the **Model Interoperability Agreement** (**MIA)**[102] developed by the European E-invoicing Service Providers Association (EESPA) for the transmission and processing of electronic invoices and other business documents.

[96] Article: http://news.postimees.ee/2627590/estonian-x-road-e-services-expand-into-finland

[97] X-Road (Global): https://x-road.global/

[98] Crobeco project: https://www.elra.eu/crobeco/

[99] CEIS: http://informationsharing.org.uk/

[100] For real examples of EBR's ISAs please use this contact point: http://www.ebr.org/index.php/contact/

[101] BRIS: https://ec.europa.eu/inea/en/connecting-europe-facility/cef-telecom/projects-by-dsi/business-registers-interconnection-system-%28bris%29

[102] MIA: https://joinup.ec.europa.eu/solution/electronic-invoicing-part-2-model-interoperability-agreement-transmission-and-processing-electronic/about

Change Management

One of the main challenges to building successful interconnections of base registries is to ensure that the improvement, viability and sustainability of public services are breaking the resistance for change. For that, a well-thought change management plan is necessary in order to define the procedures and processes needed to deal with and control all the implemented changes. This would ensure awareness, accuracy, reliability, continuity and evolution of the service delivered to other public administrations, businesses and citizens.

Draft a change management plan

Closely related to essential aspects of interoperability, there are at least three relevant areas of the assurance of viability and sustainability challenges that should be addressed through a well-planned change management plan:

1. The approach to the structuring and digitisation of the entire corpus of data of the base registries. In some cases, it can be hard to break the resistance to the digitisation or even to the structuring and representation of the data.

2. How to break the barriers hampering the alignment of base registries to the free reuse of public information. This is especially important in case the data are appealing for large public and private initiatives that aggregate them in order to offer added value services.

3. The re-organisation of processes both at the base registry's internal level, and when sharing data with public administrations' services. This process should result in a minimum set of modular and coordinated simplified processes that reuse, as much as possible, core functional services and share well-identified core entities (master data).

To manage situations like these, the recommendation is to define a **change management plan** involving awareness-raising, which would convince everyone on the benefits of its implementation. In this direction, dissemination activities and training are good occasions to explain and illustrate those benefits and make sure they are not missed. One could also use base registries that are currently offering their data for free reuse, and consider them as good practices by illustrating the change they went through and how they managed that change.

Member States and other countries show various examples in the area.

In **Great Britain**, the three main business registries, located in England, Scotland, and Wales respectively, coordinated through the **Companies House**[103], planned their adaption and alignment with the **Public Sector Information (PSI) reusing Directive**[104] and national legislation, and worked to open their data[105]. Previous experience existed in the UK in the private sector related to business registries, e.g. **OpenCorporates**[106], which offers data on companies under an Open Data Commons Open Database License (ODbL).

In **Belgium**, business registries offer practically all their data for free and without access restrictions (in business registries there are very few data subject to data protection, such as industrial property, national security, or other superior restrictive legal systems). In parallel, some Belgian organisations and administrations promoted early the change into Open Data paradigms. Some examples are the **Belgium**

[103] Companies House (UK): https://www.gov.uk/government/organisations/companies-house

[104] Directive 2013/37/EU, amending Directive 2003/98/EC on the re-use of public sector information, https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1561563110433&uri=CELEX:32019L1024

[105] Article as an example of the benefits of opening the business data (UK): https://www.journalism.co.uk/news/companies-house-opens-up-uk-business-data/s2/a557386/

[106] Opercorporates (UK): https://opencorporates.com/companies/gb

Business Register[107] (KBO BCE), the datasets open at the national level at **data.gov.be portal**[108], and the **Open Access Belgium**[109] that provides access to relevant repositories and projects related to base registries and public services interoperability, such as VDAB (employment) or OSLO (Open Standards for Administrations in Flanders), among others.

Another set of tools facilitating the change towards the structuring, digitisation and opening of data have been the work by **SEMIC**[110], an initiative within the ISA^2 Programme. The Core Vocabularies have played an essential role in this effort. It is also worth mentioning the **Core Business Vocabulary**[111], which is based on the aforementioned one, that organisations and initiatives like **OpenCorporates** and **OSLO** (or the **Greek Tax Agency**[112]) have further developed or piloted with their own data models and interoperability vocabularies (see more details in sub-section 'Reuse semantic assets for reference: standard ontologies, core vocabularies, taxonomies').

In the case of **Land Registries and Cadastres**, this role was covered by the developments around the **INSPIRE**[113] Directive. The INSPIRE Directive allowed the Commission to establish a community geo-portal through which Member States should provide access to their infrastructure, as well as through any access points they themselves decide to operate. Based on the work developed in the framework of INSPIRE, one proactive example of a national base registry that analysed and implemented an early approach to this, aimed at structuring and sharing its data under the Open Data principles, was the **Spanish Cadastre**[114].

Implement and release change requests

Implementing changes, as a result of applying a change management plan, includes taking into consideration some technical processes and their practical applications. The most common ones are described below.

During the deployment and evolution of base registries interoperability initiatives, the **semantic assets**[115], services or tools may change due to requests issued by the community of users or by the governance bodies themselves. This need can become a real challenge, especially when the number of stakeholders is high and their characteristics are heterogeneous (i.e. different countries and languages, different legislations and processes, etc.).

A common way of addressing this situation consists of implementing a **twofold management plan**: on the one hand, it should manage **the change requests** and, on the other hand, it should put in place a **release management method** through the implementation of a **release planning**. The responsible entities for both types of management should be clearly defined in the data governance model, as well as the workflows and interactions between them and the rest of the governance bodies (see section Data Governance).

[107] Belgium Business Register (KBO BCE) in Open Data: http://es.slideshare.net/FrankDeSaer/open-data-vl

[108] Data Gov (Belgium): http://data.gov.be/en

[109] Oen Access (Belgium): https://openaccess.be/open-access-in-belgium/open-data/

[110] SEMIC Collection on Joinup: https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic

[111] Core Business Voc on Joinup: https://joinup.ec.europa.eu/release/core-business-vocabulary/100

[112] Article Towards Linked Open Business Registers (Greece): http://dl.acm.org/citation.cfm?id=2839282

[113] INSPIRE: http://inspire-geoportal.ec.europa.eu/, http://inspire.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_CP_v3.0.1.pdf

[114] Electronic Office of the Cadastre (Spain): http://www.catastro.meh.es/esp/sede.asp (In Spanish)

[115] The ISA glossary defines a "semantic interoperability asset" is a collection of highly reusable metadata (e.g. xml schemata, generic data models) and reference data (e.g. code lists, taxonomies, dictionaries, vocabularies) which are used for e-Government system development: https://data.europa.eu/euodp/en/glossary

During the lifecycle of public services, the recommendation is to make sure that the governance model, maintenance policy, and all related plans are thoroughly understood by involved people and other stakeholders, and applied. Thus, it is recommended to regularly gather feedback from the stakeholders and review whether the policy is flexible enough and adapted to the real needs.

The studies on the governance and maintenance of specifications conducted by the ISA^2 Programme include examples of implementing change requests. One of these studies is **"A change management release and publication process for structural metadata specifications**[116]", and another study is on the maintenance of metadata by SEMIC, the "**Methodology and tools for Metadata Governance and Management for EU Institutions and Member States**[117]".

Other aspects to consider in change management

Big changes are not likely to be implemented fast, hence, technical solutions can lose some of their benefits due to delays. Also, when introducing changes, it is good to explore what possibilities are out there – is it really important to change the whole infrastructure or can we achieve the same results with a smaller intervention?

X-Road supports a **research-focused change management approach**. Namely, there are two collaboration projects with the University of Tartu and Tallinn University of Technology, which are focusing on expansion of X-Road communication capabilities. Currently, X-Road supports synchronous data exchange via request-response message pairs. Based on feedback received from users, the messaging capabilities should be expanded to cover asynchronous one-to-many messaging. Therefore, one of the ongoing collaboration projects is studying whether **Apache Kafka** – a well-known open source messaging solution – could be integrated into X-Road – instead of implementing everything from scratch.

Luxembourg focuses on **organisational aspects** of **change management**. In general, changes occur not only in the IT landscape, and requirements are always from businesses or users. In order to benefit users and businesses, it is essential to check if new technology would be able to comply or fulfil the new requirements. Luxembourg aims at achieving, as far as possible, a **proactive government approach** which means applying more automation in back office procedures, and eliminating those which are not essential or are redundant. Thus, instead of trying to integrate the Once Only principle in online procedures, it is better to completely **automate the procedures** by applying that principle to the back office. In other words, transmitting the required data in the back office directly to the organisations that need them, so they do not have to ask the user to do this procedure. This way, a number of administrative procedures will not be used anymore. When new data are created, the relevant back offices should be as far as possible, and in compliance with data protection and privacy legislation, informed automatically that this information is available. Achieving automation leads to reduced administrative burden.

Business Continuity

A business continuity plan is necessary to prevent the disruption of flow of operations, by implementing and applying **a disaster recovery plan** when needed. It also aims to ensure the preservation and sustainability of the data, the base registries themselves, and the digital public services over time.

This calls for common agreements between institutions, organisational planning, process alignment and well-managed workflows. However, legal support and a governance strategy ensuring the maintenance of the base registries remain the most important aspects.

[116] This document is not avaibla online, contact isa@ec.europa.eu

[117] Document: Methodology and tools for structural metadata management and governance for EU Institutions and Member States

Ensure digital preservation and permanent access to data

Ensuring the availability of data is not a minor challenge, as data, especially very old data, have to be always accessible and readable (**technical continuity**) when needed. In the case of base registries, this may be a significant challenge as they normally keep the data for extremely long periods of time (or even permanently). Furthermore, the transfer of the data to historical archives or their definitive deletion (**data disposition**) is rarely performed.

Data preservation is also about ensuring the authenticity and validity of the data. Thus, base registries should also ensure that the original values of the data are not lost as a result of an operation (business continuity). Digital preservation is one of the principles that the **EIF**[118] puts a special focus on — thus all the necessary ICT solutions that grant the correct, regular and exhaustive execution of a preservation plan should be put in place.

Preserving data that have no more legal or informative value can be costly, unnecessary and in some occasions also unlawful. Base registries authorities should analyse whether the data kept in their registries are affected by specific regulation and implement the organisational and technical means to effectively comply with the regulation.

The person responsible for the preservation of the base registry data should allow the data to be assessed by experts in records management and archivists within their organisation.

Examples of recommended consolidated standards related to the preservation of base registry data are **PREMIS**[119], which aims at supporting the preservation of digital objects and ensuring their long-term usability, and **PRONIM,** a web-based technical registry to support digital preservation services.

For an example of permanent storage of records, the existing **WORM**[120] (Write Once Read Many) solutions can be studied, that assure that data cannot be tampered with once written to a device.

Agree on flexible data availability levels

In order to ensure that the data of base registries is accessible, it is essential that every system is **up and running properly** at least during the agreed periods of time necessary to ensure the public services. Base registries systems, however, may fail or may need to stop their services at certain moments for different maintenance reasons, such as publishing new services releases or publishing updated data.

Base registries and the public services they feed data with, should agree on common time-windows for interrupting access to the data for maintenance purposes ("***Mean Time to Maintenance***"), and in case of failure ("***Time between Failures***"). In general it is a good practice to also propose a common time-window for the availability of the data, considering the fact that (except for critical services) "24x7 service availability" is not always necessary and may impose on the registries extraordinary and superfluous investments.

Other useful measures include:

- Preparing a **disaster recovery plan** in case of an accident or great disaster; the plan should also be tested regularly, monitored and improved if necessary;

- Compiling all measures and requirements in a **Service Level Agreement (SLA)** between each base registry and its stakeholders;

[118] Underlying principle 10: Preservation of information.

[119] PREMIS: http://www.loc.gov/standards/premis/index.html

[120] https://en.wikipedia.org/wiki/Write_once_read_many

- Logging and monitoring the **responsiveness** and **the level of compliance** of the base registries to their SLAs, reporting also the results to the responsible governance bodies.

For example, **TES**[121] systems provide very detailed documentation about the scope, quality and responsibilities of the service. In the case of BRIS, the levels of detail were drafted in the **System Wide Requirements**[122], describing the technical requirements expected from all the registries and the ones expected from the services offered by the central platform.

[121] TES Cartography on Joinup : https://joinup.ec.europa.eu/solution/tes-cartography-1

[122] [BR22] System Wide Requirements on ABR Collection on Joinup.

4. Common data models and master data

This chapter provides information on common data models which allow public administrations to organise data – that originate from multiple sources – into a standard structure for further processing. With the help of master data, which are mostly identifiers and attributes of an organisation's assets, these data models can be conceptualised to offer more efficient management of data in base registries.

Master Data Management

Despite the fact that several initiatives, including some Large Scale Pilots[123], have demonstrated that even in large and complex fields (e.g. public procurement, health, finance, and justice) information can be represented as structured data and shared digitally, the notion that information cannot always be represented as structured data is still anchored in some public administration authorities.

There are authorities who still keep heavily narrative information (i.e. non-structured data) in documents, most often on paper or in digitised images. This is the case, for instance, in base registries that work with legal documents like deeds, founding charts and statutes and complex financial accounts, or for registries which deal with scientific information that is difficult to structure (e.g. scientific formulae or raw data captured by sensors or produced massively by powerful algorithms).

The digitisation and structuring of the entire information of base registries is the proper way to facilitate the interoperability, world-wide visibility and accessibility to base registries' data.

For data structuring, one should start with the assets of any base registry that are its data (master data), since such data are essential for the delivery of public services. Thus any effort invested in ensuring thorough management of the data's lifecycle is worthy. This includes managing the data, as well as securing their quality, validity, authenticity, preservation and continuous availability.

Master Data Management (MDM) refers to applications designed to create a single view of a core entity for an organisation (e.g. a public administration) across all operational and analytical uses, and independent of any other repository of similar data. As a discipline, MDM is focused on consistency and quality of data that describes the core entities of an organisation.

MDM's entity definitions and reference data facilitate the accurate sharing of data. Many of the challenges addressed in these guidelines are covered by MDM practices and tools, largely used in the private sector and, increasingly, in public administrations, too.

Thus, three of the greatest challenges addressed by MDM are:

- **Data Governance** (DG) – in MDM context, it is the creation and enforcement of policies and procedures for the business use and technical management of data[124]. It is usually the responsibility of an executive-level board or committee and its scope can vary greatly, from the data of a single application to all the data of an entire organisation.
- **Data Stewardship** (DS) – is usually performed by a business manager who knows how data affects the performance of the organisation (or of a unit within the organisation). A data steward's tasks, in addition to daily management responsibilities, involve the collaboration with data management specialists and data governors to direct MDM work that supports business goals and priorities.

[123] LSPs: https://ec.europa.eu/digital-agenda/en/large-scale-pilot-projects

[124] The detailed information is covered in Section 1 'Common governance and strategy', sub-section 'Data Governance'.

- **Data Quality** (DQ) – a set of related data-management techniques and business-quality practices aimed at assuring that data are accurate, up to date and fit for the intended purpose. The most common data quality techniques are data cleansing and data standardisation; other techniques include verification, profiling, monitoring, matching, merging, geocoding, data enrichment.

Master Data Management represents a "unified" approach and solution to the challenges faced by base registries themselves and the public administrations' services that need to interoperate with them. This is why, more frequently, Member States are considering the adoption of MDM to manage their data lifecycle.

MDM tools – which are mainly proprietary commercial developments – are good for a great number of data-related activities. Therefore, before considering the possibility of developing one's own solutions, it is recommended to study the existing practices and solutions, and assess how these could be reused and applied in a public administration.

Define an MDM style

There are different MDM styles that a public administration can choose from. The choice depends on whether it needs to have **a central hub** to manage its data, or to **synchronise it with existing sources**[125]. The focal points in establishing a model should be on data governance, enhancing data quality, and ensuring that data can be easily managed and accessed.

The **four (4) more commonly used styles**[126] are the following:

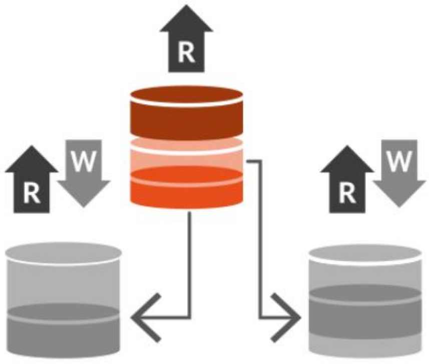


Figure 3: The registry style

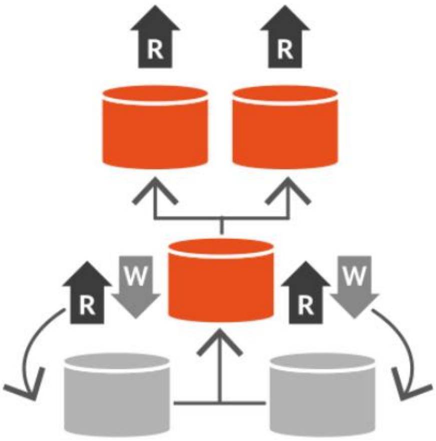


Figure 4: The Consolidation style

[125] Understanding Metadata - National Information Standards Organization, NISO, e.g. publication: https://www.niso.org/publications/understanding-metadata-2017

[126] Understanding Various MDM Implementation Styles: https://towardsdatascience.com/understanding-various-mdm-implementation-styles-5b4c8fcbbecf

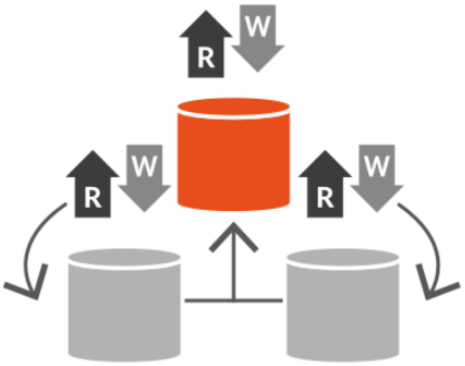


Figure 5: The co-existence style

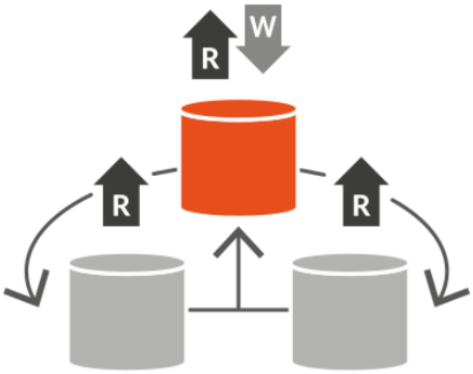


Figure 6: The Transaction/Centralised style

The information on the advantages and disadvantages of adopting one of the styles should help a public administration identify which style – or combination of styles – would be most suitable for their legal and organisational situation, as presented in the table below:

Table 2. Advantages and disadvantages of MDM Styles

MDM Style	Advantages	Disadvantages
Registry	• a large number of source systems, each with its own rules and complexities that are difficult to be modified • screen data and run cleansing and matching algorithm, assign a unique global identifier to duplicate records and finally establish the single version of the truth • the most low-cost way of implementing MDM	• nothing will ever change in the source systems • it can be difficult to establish an authoritative source • latency
Consolidation	Besides the ones from the Registry style: • the stewardship capability is available in the MDM hub	The same as in the Registry style
Co-existence	Besides the ones from the Consolidation style: • real-time synchronisations between MDM hub and sources by sending back the golden record to each respective source systems • a significant improvement in master data quality by updating master data in source systems and MDM hub	Besides the ones in Registry style: • more intrusive • the sources have data cleansing capabilities in order to maintain consistency with the hub • more expensive to deploy than the Consolidation style

MDM Style	Advantages	Disadvantages
Transaction / Centralised	• the hub becomes the single provider of master data • any systems outside the hub can no longer be allowed to create or amend the master, instead, they are obliged to subscribe to the hub for any update	• demanding on resources • time required for the style to be implemented

Member States adopt different styles or combination of styles, based on the local needs and legal requirements. Taking an example from the Nordic countries, namely Sweden, Norway and Denmark, different styles are adopted within the same region.

For example, in **Denmark**[127], registries are a **combination of coexistence and consolidation styles**, as they are autonomous, but can communicate with each other. This interconnection is based on law, as well as on the authoritative source and ownership of data. Denmark has a long tradition and history of having authoritative registries and data sets. One of the first digital registries in the world was established in this country – and it was a Civil Registration number registry (hereafter CPR registry), dated from 1968. There is also a centralised platform for the base registries, the already mentioned Data Distributor, which collects data from all registries and is the unique point of centralisation of data in Denmark.

In **Sweden**[128], two different data management styles are represented – the **registry style** and **consolidation style**. The rationale behind the selection of these two styles lies in the fact that registries are regulated by different laws, given there is no common law defining a base registry in Sweden.

In **Norway**[129], too, a **combination of styles** is used depending on the registry. Thus, there are three main national registries in Norway, namely, land registry, business registry and civil registry. Land registry is based on a **consolidation style**, while the remaining ones follow the **transaction/centralised style**.

On the **EU level**, the **hybrid search**[130] can serve as an example, where data must be searched in a central repository and in distributed base registries based on the **coexistence style**. This is the architecture developed for **BRIS** by the EC. This approach was dictated by the requirements imposed by at least three Member States that could not afford to centralise any of their data, and included a reduced set of data for indexation and performance enhancement purposes, such as legal entity names and registered address of the business.

[127] Information from interview with Danish public authorities, available here: https://joinup.ec.europa.eu/collection/access-base-registries/news/interview-highlights-denmark

[128] Information from interview with Swedish public authorities, available here: https://joinup.ec.europa.eu/collection/access-base-registries/news/interview-highlights-sweden

[129] Information from interview with Norwegian public authorities, available here: https://joinup.ec.europa.eu/collection/access-base-registries/news/interview-highlights-norway

[130] [BR02] Architecture approach - Hybrid, [BR03] Motivation of the architectural approach on ABR Catalogue of Solutions on Joinup.

Define data types and their management approach

Once the MDM style is defined, the next recommended step is to define which data types in the base registries should be managed, and which approach is the best for this purpose.

Critical data kept in the registry concern one specific type of **"core" entity** (e.g. person, vehicle, business, land, etc.), making base registries the **primary source** of "master" data. Core entities include **parties** (e.g. citizens, businesses, employees, vendors, suppliers and trading partners), **places** (including locations, offices, regional alignments and geographies), and **things** (such as vehicles, real estate, accounts, assets, policies, products and services).

As such, the data kept in the registries acquire legal value, making the registry a legally recognised source of "authentic" data and, in addition to **master data**, MDM also takes into account the management of **"reference data"** (constants that define permissible values for data). However, even though it is not strictly master data, reference data may be managed in a similar way. One public administration should coordinate and, when possible, harmonise which reference data should be used throughout the administration and, namely, for the base registries.

The **e-Certis2 - of the Directive 2014/24/EU**[131] and **ESPD (European Single Procurement Document[132] Service)** are perfect candidates to illustrate how the "Once-Only" Principle can be implemented based on master data and reference data that are exchanged among cross-border and cross-sector base registries (i.e. social security and tax agencies, business registries and BRIS, service providers, etc.).

Regarding the **data management approach**, one can learn from the way **X-Road** is dealing with **data management** as it provides a lot of **flexibility** to those implementing it. Although the Once-Only (TOOP) principle is a guiding principle, X-Road does not impose it and, moreover, supports different approaches to data management. To illustrate this example, we can mention Estonia and Finland and the way **master data management of personal data** is handled by each of them. Estonia implemented the once only principle, which means that data are fetched directly from the responsible authority. In Finland, the situation is different – although it has TOOP implemented when needed – another common approach used is to replicate (a relevant subset of) the master data registry in organisations' database and download updates regularly. However, exchange of data between the two countries is achieved through the master base registries.

Another interesting **data management approach** has been adopted by **Luxembourg**. Firstly, one should note the context of the access to base registries that concerns online procedures, which is performed in Luxembourg mainly through the personal space proposed to each user at MyGuichet.lu (part of the **one-stop-shop guichet.lu** managing online procedures). Guichet.lu is, simultaneously, a repository of 1500 descriptions of administrative procedures for citizens and businesses, as well as of more than 200 interactive online procedures (via MyGuichet). This online platform operates, to a very large extent, on the **Once Only (TOOP) principle** on the level of **procedures**, and allows users to see which data are kept on them from the most important authentic sources, providing also the reuse of such data via prefilling of the forms in the context of online procedures. The situation in Luxembourg can be rightly described as **a user-centric** or **a user-driven approach**. Data is not directly connected or exchanged between authentic sources, but rather driven by users who request specific information. This means that citizens actually submit the information only once. Moreover, organisations acquire the data through procedures which are defined by law, and they need to ensure the accuracy of the data before storing in their own database.

[131] EUR-LEX : http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.L_.2014.094.01.0065.01.ENG

[132] ESPD : https://ec.europa.eu/isa2/solutions/european-single-procurement-document-espd_en

Another popular type of data management is the **records management**, also known as **records and information management**. In the context of base registries, its main focus is on management of data throughout their lifecycle, from the moment of data input to their disposition.

One example on how to approach the lifecycle management of all types of data records, including e-documents with structured data, is the DLM-Forum's specification **MoReq**[133] (**modular requirements for records systems**). The latest version of MoReq (MoReq2010®) emphasises on the interoperability between systems responsible for the management of the different phases of the life of a record, or as expressed in the official website, defines an "approach to implementing a records management solution by establishing a definition of a common set of core services that are shared by many different types of records systems, but which are also modular and flexible, allowing them to be incorporated into highly specialised and dedicated applications that might not previously have been acknowledged as records systems".

Two pioneering initiatives in Europe covering the **records structuring and digitisation**, records life-cycle identification and management, and records management systems interoperability, were the **DOMEA Concept**[134] with its **XDOMEA specification (Germany)** and the **SIGeDA policy and model**[135] (**Catalonia, Spain**). These two projects have inspired different national and regional public administrations in the EU to develop their own records management policies and systems.

One example of management of data implemented within the governance policy is from **Spain**. Its National Interoperability Framework was included in the Law through Royal Decrees and deployed in the (mandatory) Technical Interoperability Norms (NTI). This policy covers almost all aspects of interoperability and public services governance. Among them is the **management of data** which is mainly reflected in a norm for the interoperability of data ("protocolos para la mediación de los datos", literally "**data mediation protocols**[136]").

Identify the instances of your master data uniquely and unambiguously

An **identifier** does not define a concept, but represents a particular instance of an object and facilitates the access to all data about it. It is important to differentiate between unique and context-specific identification (or core data):

- **Unique identification** is universal and concerns multi-contextual data; it can be used in every country regardless of context. Examples are the list of vehicle plate numbers in a Vehicle Base Registry, or the list of ID card numbers in a Civil Base Registry.

- **Specific identification** regards data that are adapted to the context of one country or service, but not necessarily for other ones. This is the case, for example, of middle names (or second surnames) in the Civil Base Registry in Spain.

[133] MoReq: http://www.moreq.info/

[134] The Federal Government Commissioner for Information Technology (www.cio.bund.de/), publication (Germany): https://www.verwaltung-innovativ.de/SharedDocs/Publikationen/Presse__Archiv/domea_konzept_organisationskonzept_2_1.pdf?__blob=publicationFile&v=1 (in German)

[135] Publication (Spain): http://cultura.gencat.cat/web/.content/dgpc/arxius_i_gestio_documental/09_publicacions/altres_publicacions/Guia-implantacio-documentalV506DEF.pdf (in Catalan). Information in English is available through this contact: https://ovt.gencat.cat/gsitfc/AppJava/generic/conqxsGeneric.do?webFormId=251&set-locale=ca_ES

[136] Data mediation protocols interoperability standard NIF Spain: http://administracionelectronica.gob.es/pae_Home/pae_Estrategias/pae_Interoperabilidad_Inicio/pae_Normas_tecnicas_de_interoperabilidad.html (English version available online).

Therefore, public administrations should consider generating or keeping **Universal and Unique Identifiers (UUID).** This is especially relevant for cross-border and cross-sector initiatives as **master data** – though improbable – could have **identical identifiers** in **different base registries**.

The establishment of a new UUID system may have an impact on base registries, that is why it is recommended – before inventing a new identification system – to search for existing ones and try to reuse them. In any case, before adopting an existing solution or developing a new one, public administrations should assess the impact it could have on stakeholders' systems and, if any, try to minimise it.

There are a few other recommendations concerning the **use of identifiers**:

- In particular, when designing exchange data models, **consider using multiple identifiers** for one instance of an object. This is convenient for cross-sector initiatives, where different authorities may identify the same instance in different ways, and it also makes possible the automated production of identifier mappings.

- **An identifier should never be modified over time** once it has been assigned to one particular entity. One way of ensuring the long-term existence of an identifier is to assign a URL to it, which also facilitates the description of the entity being identified.

The challenge of uniquely (and universally) identifying entities has haunted many different business domains, among them, base registries. At the EU level, the BRIS project came up with a solution for this: the **EUID** (a unique European ID for companies). This solution was inspired by a research project[137] funded by the EC and is quite similar to the IBAN solution adopted to uniquely identify bank accounts.

For an example of the usefulness of providing multiple identifiers, one can study how the **OASIS Universal Business Language (UBL) TC[138]** specification is used in documents for electronic procurement, like tenders or invoices. This specification is currently used for different applications in various Member States and EU Institutions (namely the European Commission).

One way of ensuring the long term survival of identifiers is to assign them to **Permanent Uniform Resource Locators (PURLs[139])**. The ISA^2 Programme recommends the use of PURLs and defines principles and practices[140] for their use.

One interesting example is the **EULF**[141] initiative where PURLs are used to identify locations, verify that users are the ones they claim to be, and are entitled to use the requested information or functionality.

In addition to the aforementioned, other important aspects directly related to the data lifecycle management are ensuring data security and authenticity, e.g. integrity, provenance, privacy and personal data (see section 'Data Policies'), and establishing an efficient policy and methodology for semantic assets and reference data maintenance.

[137] The BRITE REID Identifier: http://www.ict-21.ch/com-ict/IMG/pdf/REID-Unique-Company-Identification-12-March-2008.pdf

[138] OASIS UBL TC: https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=ubl

[139] E.g. from Wikipedia: https://en.wikipedia.org/wiki/Persistent_uniform_resource_locator

[140] Publication on SEMIC Collection on Joinup: https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/document/10-rules-persistent-uris

[141] EULF on Joinup: https://joinup.ec.europa.eu/collection/european-union-location-framework-eulf/about

Semantic data models and standards

Semantics has the goal of developing a common meaning across public administrations through adaptation and implementation of common data models. It states that it is important to use metadata as much as possible to document the meaning of each concept and define and distinguish between the types of metadata.

This way it is feasible to eliminate ambiguity by providing common terminology and a glossary for each concept. Overall, it is recommended to reuse the existing semantic standards, creating application profiles on them. Thus, a good approach is to reuse “core vocabularies” to model information by using them as a starting point to customise a data model. Other good practices are the use of English as the reference language to accommodate reuse among Member States, translating the terms in the native languages where applicable, and include key metadata in normative documents (such as regulatory frameworks). When native languages are used, equivalent English terms should be provided to allow for concept terms to be included in an international context, and make mapping against other international models possible.

Define the data domain

Managing master data and handling data models (among others) have in common the need to agree on a definition of what master data is, and on which data models they will rely on. This is important since, when organisations grow, their business processes need additional support and the data models of their core entities increase. This imposes a data consistency breach with new quality rules emerging, among other things.

A consistent way to tackle this involves two major actions:

- Define and agree on the data domain(s);
- Once in agreement, choose the most appropriate standard(s) to express the (master) data models.

The first action involves a contextualisation of what are the master data (and the consequent models) in an organisation. Analysis techniques could involve the following:

- Differentiate between master data, reference data, application-specific data and content;
- Identify metadata that do not change often;
- Use root cause analysis and information classification techniques to determine which data and models need to be governed.

Identify and distinguish the concepts of the domain

After defining the data domain, and before designing the data model – or ideally reusing an existing one –, it is necessary to identify and distinguish the core concepts of the domain.

In an effort to reduce semantic conflicts due to the heterogeneity of the actors (i.e. information and services of different Member States), the ISA^2 Programme introduced the **Core Vocabularies**[142] as a way to model the core concepts that are widely used among Member States and, thus, enable their reuse and facilitate semantic interoperability. Core Vocabularies are simplified, reusable, and extensible data models that capture the fundamental characteristics of a data entity in a context-neutral fashion. They are by definition syntax-neutral, indicating that they focus on fundamental characteristics of data entities rather than on the specific representation.

[142] https://ec.europa.eu/isa2/solutions/core-vocabularies_en

The Core Vocabularies consist of:

- **Core Person**: captures the fundamental characteristics of a person, e.g. name, gender, date of birth, location;
- **Core Business:** captures the fundamental characteristics of a legal entity (e.g. its identifier, activities) which is created through a formal registration process, typically in a national or regional registry;
- **Core Location:** captures the fundamental characteristics of a location, represented as an address, a geographic name or geometry;
- **Core Criterion and Core Evidence:** describe the principles and the means that a private entity must fulfil to become eligible or qualified to perform public services. A Criterion is a rule or a principle that is used to judge, evaluate or test something. An Evidence is a means to prove a Criterion.
- **Core Public Organisation:** describes public organisations in the European Union.

The Core Vocabularies are context-neutral semantic building blocks that can be extended into context-specific data models.

A good example of reuse of Core Vocabularies with successful modelling of the data domain and definition of core concepts is the **OSLO**[143] project **(Open Standards for Linked Administrations in Flanders),** which started in February 2012 and facilitated a working group with ICT experts from local, regional and federal public administrations and ICT service providers. The project aimed to develop a semantic agreement and build a consensus on standards for information exchange. The project's outcome, the **OSLO vocabulary** is a simplified, reusable and extensible data model that captures the fundamental characteristics of information exchanged by public administrations in the domains of: contact information, localisation and public services. The standards of the Flemish OSLO project are local extensions of the core Person, Business, Location, and Public Service vocabularies of the ISA^2 Programme's Core vocabularies. e-Government OSLO Vocabularies are simplified, reusable, and extensible specifications and serve as the starting point for developing interoperable e-Government systems as they allow mappings with existing data models. This helps public administrations attain cross-border and cross-sector interoperability.

OSLO initially aimed to have one governance model with 30 **domain models**; more than 300 people contributed by implementing more than 3000 **definitions**. Currently, OSLO contains over 18 domain models consisting of more than one thousand definitions made by more than 250 contributors[144]. One of the most interesting aspects of the OSLO domain models is the modelling of persons, organisations and roles.

The importance of identifying and clearly defining the core concepts of a domain is evident in cases where different terms are used by different registries to signify the same concept. A successful example of dealing with such issues – through the establishment of common concepts – is demonstrated by the **Danish Basic Data Programme**[145]**,** which created a government shared registry for data distribution, called Common Public-Sector Data Distributor, introducing the once only principal.

In Denmark, there was a complex connection between property data in the primary property registries. Whereas, at state-level, registration of property was handled by the Cadastral, Land Registry and Tax authorities, the municipalities also registered information about property. Since the supporting IT systems and procedures were created at a time where dynamic data exchange between the property registries was not a given, there were instances of the same information being registered and maintained in several places. As a consequence, each basic registry used its own property concept and different

[143] https://joinup.ec.europa.eu/collection/oslo-open-standards-local-administrations-flanders/about

[144] https://data.vlaanderen.be/

[145] https://www.academia.edu/29858925/The_Basic_Data_Programme_A_Danish_Infrastructure_Model_for_Public_Data

keys for identification of property. For instance, instead of using one common concept, three different property concepts were used in the Cadastral Register, the Land Register, the BDR and the Property Register respectively. In an effort to tidy up core property concepts, the sub-programme 1 of the Danish Basic Data Programme incorporated the term 'particular property' as one common term in all foundational registries to be used in related base registries in a 'uniform and safe' way. Thus, a common concept was defined ('particular property'), and the roles of the following base registries were clarified and distinguished:

- **Cadastral Register**: this registry aims to contain information about all properties in Denmark, becoming the authoritative basic registry for all types of property;

- **Building Dwelling Register (BDR):** this registry aims to continue to contain information about all buildings and dwellings, remaining an authoritative basic registry for this type of data, but should also include reference to the Cadastral Register's registrations for the properties to which the building or dwelling belongs;

- **Land Register:** this registry aims to continue to contain information about the rights that are registered against the property, but should also be based on information from the Cadastral Register when the property is registered there.

In many cases, different physical pieces of the infrastructure that hold information on data models and datasets such as data catalogues, metadata registries, registry of registries are in different locations in public administrations and, hence, it is important to distinguish them as different concepts and define what concept should be used in each case, and implement them rationally.

A **Data Catalogue** "serves as an inventory of available data and provides information to evaluate fitness data for intended uses"[146]**.** It has a collection of metadata and, combined with data management and search tools, a data catalogue informs its users on the available data sets and metadata, enabling users to locate them quickly.

A **Metadata Registry** represents "an information system for registering metadata"[147]. Within the ISO/IEC International standard 11179, a metadata registry is a database of metadata that supports the functionality of registration. Thus, it serves as a central location in an organisation where metadata definitions are stored and maintained in a controlled method[148].

A **Registry of Registries** (RoR) represents a catalogue of base registries, enabling data discovery in base registries. It aims at the following:

- identifying and describing the data and the data source;
- mapping and preventing duplication;
- developing public services using the catalogue and the data directly from the base registries.

The existence of a RoR introduces many **benefits**, the first being the increment of data quality in terms of non-redundancy and consistency, since duplicated data are eliminated as only one authoritative source is referred to in the catalogue. It also helps to facilitate the access and interoperability between registries, and it makes possible the automatic (and massive) reuse of the data.

Only a few countries in Europe have a RoR, such as **Denmark (Data Distributor)**, **Finland and Estonia (X-Road)**, **Belgium** (**MAGDA**), **Netherlands (Stelselcatalogus)**, benefiting, thus, from a reduced number of base registries and from gained data quality that comes with it. In other countries, interoperability and intermediation platforms – or reliable data aggregators and service providers (such

[146] E.g. on data catalogs: Gartner report: augmented data catalogs, https://www.dataversity.net/what-is-a-data-catalog/#

[147] Glossary of statistical terms: https://stats.oecd.org/glossary/detail.asp?ID=5137

[148] E.g. on https://en.wikipedia.org/wiki/Metadata_registry

as the National Statistics Agencies or Linked Open Data Portals) – are in place, but they tend to be domain-based and therefore incomplete RoRs.

In **the Netherlands**, the "**Stelselcatalogus**"[149] is available as an online catalogue, as well as a Linked Open Data Store and, therefore, is accessible to everyone. It represents a catalogue of base registries that allows relevant stakeholders (e.g. lawyers, civil servants, citizens and companies) to identify available (authentic) data and concepts. It also provides the option to search base registries that keep data and decide on whether the data are relevant for their own work processes. It is also accessible through the Dutch e-Government portal "Mijn Overheid[150]" from where the abovementioned stakeholders can view data directly from the base registries.

Distinguish types of metadata

In cross-border initiatives, and especially in cross-sector ones, stakeholders do not always interpret master data the same way. Master data attributes and identifiers may differ largely in number and nature, and this is where metadata comes into play.

Metadata (often called 'data about data') are structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource[151]. In order to be useful, metadata needs to be formalised. This includes agreeing on language, spelling, date format, etc.

Metadata are made up of a number of elements which can be categorised into the different functions they support. According to the NISO[152] definition, there are three main types of metadata:

- The **descriptive metadata** which describe a resource for purposes, such as discovery and identification, and can include elements, such as title, abstract, author, and keywords;

- The **structural metadata** that indicate how compound objects are put together, for example, how pages are ordered to form chapters;

- The **administrative metadata** that provide information to help manage a resource, such as when and how it was created, file type and other technical information, and who can access it.

Metadata need to be structured. Therefore, a key component of metadata is the schema, which defines the overall structure, describes how the metadata elements are arranged, and usually addresses standards for common components of metadata like dates, names, and places (see next sub-section).

Metadata also need to be published over the web, or internally in an organisation. Metadata registries are good candidates for the latter, while a set of URIs is a typical implementation of having metadata published over the web as Linked Data (described further below in this chapter).

Metadata registries make use of **code lists** (equivalent to controlled vocabularies; see next sub-section) and **identifiers** (more generic than URIs). The first ones are usually maintained by standardisation organisations (e.g. ISO), while identifiers are usually maintained by public authorities (e.g. base registries). Additionally, officers and developers have a general tendency to define their own code lists and to generate internal identifiers.

[149] System Catalog (Netherlands): https://www.logius.nl/diensten/stelselcatalogus

[150] My Government (Netherlands): https://mijn.overheid.nl/

[151] Understanding Metadata - National Information Standards Organization, NISO, e.g. publication: https://www.niso.org/publications/understanding-metadata-2017

[152] https://www.niso.org/

When possible, the recommendation is to **reuse code lists** that are maintained by international or **European Standardisation Development Organisations (SDO**[153]**)**, but always after having assessed the reusability of the initiative. If no reusable code lists are available, the alternative would be to use the legal texts (if any) which could be used as a basis for defining the desired reference data. If possible, those should be provided in English and their reusability promoted in other initiatives at the national and EU levels.

One example of an EU initiative that defines unambiguously complex legal concepts and cross-sector and cross-borders reusable code lists is **ECRIS**. ECRIS defined an exhaustive list of terminology and concepts used by stakeholders for the exchange of criminal records. It also created a code list defining criminal offences that are recognised by all Member States, allowing the possibility for it to be used in other sectors (e.g. in e-Tendering, for the identification of certain exclusion criteria). Both are referred to in the regulatory framework supporting ECRIS.

Examples of European SDOs that maintain codes are CEN, CENELEC and ETSI, enforced through the European Union (EU) Regulation (1025/2012) which settles the legal framework for standardisation. For examples of Code Lists, one could use the **CEN search engine**[154] and type "Code List" in the field "Title / Scope".

Define semantic assets of (master) data

A **semantic interoperability asset** is defined[155] as highly reusable metadata (e.g. xml schemas, generic data models) and reference data (e.g. code lists, taxonomies, dictionaries, vocabularies) which are used for e-Government system development. They are considered key resources for achieving semantic interoperability and can be in the form of data models, taxonomies, ontologies, code lists and semantic data exchange formats (e.g XML and RDF schemas). Semantic assets, and the agreements associated with them, are essential elements for organisations to understand the meaning of the information they exchange – without which information would be of little use. These are elements that specify the format and the content according to the concept of the represented information, i.e. they specify names of elements and their semantics, content and representation rules and allowable content values.

In a general perspective, these assets are typically used in the following situations:

- Integrating structured knowledge into knowledge bases, in order to solve complex problems;
- Extracting knowledge from information sources, maintaining this knowledge, and making it available to users;
- Applying knowledge representation and maintenance techniques (rules, frames, semantic nets, ontologies) and using knowledge extraction techniques and tools;
- Optimising and enhancing semantic search.

A **summary** of semantic assets and **useful recommendations** is presented in the table below.

Table 3. Semantic Assets

[153] ESOs: https://www.cenelec.eu/aboutcenelec/whoweare/europeanstandardsorganizations/index.html

[154] European Committee for Standardization: https://standards.cen.eu/dyn/www/f?p=204:105:0

[155]European Commission Joinup e-Library. Towards Open Government Metadata. https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/doc ument/towards-open-government-metadata

Semantic Asset	Short description	Recommendation
Controlled vocabulary	In library and information science, controlled vocabulary is a carefully selected list of words and phrases, which are used to tag units of information (documents or work) so that they may be more easily retrieved by a search[156].	Use controlled vocabularies when you work with taxonomies, thesauri, indexing schemes and subject headings. Always try to see if controlled vocabularies exist for a text list that you want to produce (before you actually produce it).
Taxonomies and Thesauri	Taxonomies and thesauri are closely related species of controlled vocabularies that describe relations between concepts and their labels, including synonyms, most often in various languages[157].	Use taxonomies and thesauri when the relations between the concepts are hierarchical "broader" and/or vice versa "narrower". Use thesauri when you see that non-hierarchical relations exist, like the symmetric property "related" and also when you see poly-hierarchy (where a concept can be the child-node of more than one node). Finally, use these structures as a basis for domain-specific entity extraction or text classification.
Ontology	An ontology, like a thesaurus, is a kind of taxonomy with structure and specific types of relationships between terms. In an ontology, the types of relationship are greater in number and more specific in their function[158].	Use ontologies when the knowledge domain is more contextually rich. In ontologies, the relations between the concepts go beyond broader or narrower and their semantics are richer. Use ontologies when you want to include and relate more than one taxonomy/thesaurus.
Metadata schema	A schema is a logical plan showing the relationships between metadata elements (normally through establishing rules for the use and management of metadata, specifically as regards the semantics), the syntax and the optionality (obligation level) of values[159].	Use a schema, as opposed to an application profile when you need to come up with new metadata elements, their logical relations and their organisational structure.
Application Profile	An application profile delineates the use of metadata elements declared in an element set. While an element set establishes concepts, as expressed via metadata elements, and focuses on the semantics or meanings of those elements, an application profile goes further and adds business rules and guidelines on the use of the elements. It identifies element obligations and constraints and provides comments and	Use an application profile when your focus is more on applying business logic (rules, constraints and guidelines) rather than defining metadata elements (concepts, terms). Use application profiles when metadata schemas exist that capture the knowledge of the domain you are describing.

[156] Controlled Vocabulry: https://en.wikipedia.org/wiki/Controlled_vocabulary

[157] Semantic Web Company: https://semantic-web.com/2014/07/15/from-taxonomies-over-ontologies-to-knowledge-graphs/

[158] Taxonomies & Controlled Vocabularies SIG: http://www.taxonomies-sig.org/about.htm#ontology

[159] ISO 23081.1 s3 Terms and Definitions: https://www.iso.org/obp/ui/fr/#iso:std:iso:23081:-1:ed-2:v1:en

Semantic Asset	Short description	Recommendation
	examples to assist in the understanding of the elements. Application profiles may include elements integrated from one or more element sets thus allowing a given application to meet its functional requirements.	In short: use application profiles when you want to apply a metadata schema in your organisation.
Folksonomy	A system in which users apply public tags to online items, typically to aid them in finding again those items.	Use folksonomies when you use social tagging for knowledge acquisition. This means that you can use these structures when you want users to apply tags online, in social media, in order for them to be able to find again the items they tagged.

Reuse semantic assets for reference: standard ontologies, core vocabularies, taxonomies

In order to achieve and facilitate interoperability in the e-Government field, the reuse of existing semantic assets is highly encouraged to structure metadata and exchange master data. Therefore, a variety of tools can be employed, including controlled vocabularies and common classification taxonomies. The metadata schemes will depend on strict ontologies and common vocabularies that model the domain of base registries and the master data. These schemes must be **standard** and **internationally recognised** proposals, such as the **Application Profiles for describing Public Services** (**Core Public Service Vocabulary**[160]), **Public Organisations** (**Core Public Organisation Vocabulary**[161]), **DCAT Application Profile for data portals in Europe** (**DCAT-AP**[162]), the **European Legislation Identifier** (**ELI**[163]) ontology and, especially, the **Specification of Registry of Registries** (**BRegDCAT-AP**[164]) that offers solutions to represent master data in base registries.

Along with the metadata schemes, **classification taxonomies** are recommended to be used. Knowledge organisation systems, such as taxonomies, glossaries and complex thesauri offer collections of concepts and associated multilingual terms (that provide common representations of data in different languages), enabling the semantic interoperability among systems.

For example, the **EU Publications Office** (**OP**) maintains a **Metadata Registry with Named Authority Lists** (**NALs**[165]), that are sets of controlled vocabularies or value lists for inter-institutional data exchange. Examples relevant to work performed on base registries are as follows:

- country codes;
- organisation type;
- time periods;
- language tags, etc.

[160] CPSV Collection on Joinup: https://joinup.ec.europa.eu/solution/core-public-service-vocabulary

[161] Core Public Organisation Vocabulary : https://joinup.ec.europa.eu/solution/core-public-organisation-vocabulary

[162] DCAT on SEMIC Collection on Joinup: https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/dcat-application-profile-data-portals-europe

[163] ELI: https://op.europa.eu/en/web/eu-vocabularies/eli

[164] BRegDCAT-AP: https://joinup.ec.europa.eu/collection/access-base-registries/solution/abr-bregdcat-ap

[165] NALs in Metadata registry of OP: https://op.europa.eu/en/web/eu-vocabularies/authority-tables

Also, the EU Publications Office maintains the **EuroVoc**[166], a multilingual and multidisciplinary **thesaurus** with domains and sub-domains that describes topics of legal documents.

An example to follow on developing, customising or extending a core vocabulary is provided by the "**Handbook for using the core vocabularies**"[167] on Joinup. This handbook describes how the Core Vocabularies can be used by public administrations to attain a minimum level of semantic interoperability for e-Government systems. It aims to form a generic approach for designing and mapping data models based on the Core Vocabularies. The proposed approach is syntax-neutral (i.e., independent of any technical representation), and can be used together with other methodologies for creating information system data models, information exchange data models or linked data models. The handbook provides guidance on:

- the design of new data models that extend the Core Vocabularies by using the latter as building blocks;
- the mapping of existing data models to the Core Vocabularies, thus allowing to bridge different data models by using the Core Vocabularies as a common foundational data model.

One interesting core vocabulary to consider is the **Core Criterion & Evidence Vocabulary[168]**. This vocabulary addresses the representation of criteria, independently of the area or domain where the criteria are defined or used, and links one criterion to one or multiple pieces of evidence attesting the fulfilment of the criterion. A criterion in this context can be understood as "a condition defined or imposed by an authority that must be fulfilled to reach a specific objective".

The vocabulary was developed initially for EU systems, such as the **e-Certis-2[169]** which is a type of evidence repository, or the **ESPD Service[170] (European Single Procurement Document)** to help model the criteria that the Economic Operators tendering in a public procurement procedure must fulfil to (i) not be legally excluded, and (ii) be selected as good candidates before a final award decision. The main idea underlying both these services is to use the Core Criterion & Evidence vocabulary to refer to data and documents kept by the base registries, thus eliminating the need for resending the same information to public authorities and enforcing, this way, the once-only registration principle. This is a promising vocabulary because it aims at being largely reused in cross-sector interoperability. But it should be of interest particularly to any base registry providing electronic pieces of evidence to public services through interoperability services.

Some Large-Scale Pilots and Trans-European Systems data models, like **e-Codex**[171], **BRIS** and **EULF**[172] are also inspired by the ISA Programme's core vocabularies. There are other examples from variours public sectors in Member States, too.

In **Finland**, a major research initiative, namely, the **National Semantic Web Ontology Project** (**FinnONTO)**[173], was carried out during 2003–2012 with the goal of providing a national-level semantic web ontology infrastructure based on centralised ontology services. Since 2008, a prototype of such a system, the ONKI Ontology Service, has been used in a living laboratory experiment with more than 400 daily human visitors and over 400 registered domains using its web services, including the ONKI mash-

[166] EuroVoc: https://eur-lex.europa.eu/browse/eurovoc.html

[167] Handbook: https://joinup.ec.europa.eu/site/core_vocabularies/Core_Vocabularies_user_handbook/Handbook-for-using-the-Core-Vocabularies_v0.50.pdf

[168] https://joinup.ec.europa.eu/solution/core-criterion-and-core-evidence-vocabulary#:~:text=Abstract%20The%20Core%20Criterion%20and,criteria%20by%20means%20of%20evidences

[169] e-Certis: https://ec.europa.eu/isa2/solutions/e-certis_en

[170] ESPD: https://ec.europa.eu/isa2/solutions/european-single-procurement-document-espd_en

[171] Information on e-Codex: https://www.e-codex.eu/

[172] EULF: https://joinup.ec.europa.eu/collection/european-union-location-framework-eulf/about

[173] FinnONTO: https://seco.cs.aalto.fi/projects/finnonto/

up widget for annotating content in legacy systems and semantic query expansion. The FinnONTO infrastructure also includes the notion of creating and maintaining a holistic Linked Open Ontology Cloud (KOKO) that covers different domains, is maintained in a distributed fashion by expert groups in different domains, and is provided as a national centralised service.

Reuse existing data models for catalogues of base registries

Solutions that were designed for other purposes, such as the **Asset Description Metadata Schema (ADMS)**[174] and the **DCAT-AP specification**, are now being approached as basic models and applied successfully in some registries to make possible the automatic discovery of semantic assets, the federation of base registries, and the sharing of datasets between portals and services connected to registries.

In **Norway**, the **National Data Catalogue[175],** based on the DCAT-AP model, is already implemented. Since DCAT-AP does not cover all the aspects, Norway created extensions and incorporated them in **DCAT-AP-NO**[176], that is currently under review to be aligned with the newest version of DCAT-AP.

Here is some interesting information about related aspects of this catalogue:

- Two major open data catalogue sets are automatically harvested by the National Data Catalogue: https://geonorge.no/en and https://data.norge.no;

- In addition to the automatic harvesting, an application was developed aiming to register data sets with the National Data Catalogue;

- The National Data Catalogue contains descriptions of all major base registries in Norway (i.e. the central registry of population, registry of legal entities, Land Registry and Cadastre, Norwegian Digital Contact Information Register);

- The National Data Catalogue is on the data sets-level, and the work is ongoing with information models, concepts, etc., with data owners to describe the elements of data sets;

- Due to the connection with the European Data Portal, the EU vocabulary is being used to cover the themes of data sets, in addition to the national vocabulary.

Other extensions (profiles) designed to facilitate specific needs not covered by DCAT-AP include the Italian profile **DCAT-AP_IT**[177] , the Belgian profile **DCAT-AP-BE**[178] and the Swedish and Norwegian profiles **DCAT-AP-SE**[179] and **DCAT-AP-NO[180]**. The last is aligned with DCAT-AP 2.0.0, DCAT 2.0, and BRegDCAT-AP v2.00.

[174] ADMS: https://joinup.ec.europa.eu/asset/adms/home

[175] National Data Catalog (Norway) : https://data.norge.no/

[176] Information on DCAT-AP-NO (Norway): https://data.norge.no/specification/dcat-ap-no/

[177] Information on DCAT-AP-IT (Italy): https://www.dati.gov.it/content/dcat-ap-it-v10-profilo-italiano-dcat-ap-0; presentation of AGID in ABR webinar: https://joinup.ec.europa.eu/sites/default/files/event/attachment/2020-11/WebinarABR_Lodi_IT.pdf

[178] Information on DCAT-AP-BE (Belgium): http://dcat.be/

[179] Information on DCAT-AP-SE (Sweden): https://docs.dataportal.se/dcat/en/#intro

[180] Information on DCAT-AP-NO (Norway): https://data.norge.no/specification/dcat-ap-no/

The reuse of Core Vocabularies, ADMS and DCAT-AP, can be seen also in **Belgium,** with **MAGDA**[181]. It shows how the Flemish government uses core vocabularies for interoperability between complex multilingual and multi-administrative levels.

Another example of the **reusability of existing international standards** (**W3C[182], EU ISA Core Vocabularies, DCAT-AP, INSPIRE**, etc.), is the already-mentioned OSLO project **(Open Standards for Linked Administrations)** that focuses on increasing semantic and technical interoperability. The OSLO semantic agreement concentrates on three domains: Contact Information, Localisation, and Public Services. The OSLO standards are local extension of the ISA^2 Core Vocabularies on Person, Business, Location, and Public Service. Additionally, OSLO also extended DCAT-AP. The current data standards of OSLO are Open Standards, listed in vocabularies, application profiles and code lists. Moreover, the project offers open source tools that can be reused for the implementation of data models.

And for common standards, protocols, vocabularies and interfaces, one could refer to those Member States with advanced implementations of their National Interoperable Frameworks, such as the Estonian **X-Road**. For more Member State developments one can check the **National Interoperability Framework Observatory** (**NIFO[183]**) Collection on Joinup.

The fact that a RoR facilitates interoperability is demonstrated by the majority of the Trans-European Systems. **BRIS**[184] represents an example of a registry of registries at the EU level. BRIS compiles a list of all the existing Business Registers in Europe (at any administrative level). This catalogue was initially based on the **Legal Entity Identifier**[185] (LEI) catalogue of world-wide legal entities. The catalogue contains a unique identifier per each business registry existing throughout Europe regardless of whether it is central, regional or local. The business registry identifier is key for the implementation of a true European Unique Identifier (EUID) of companies, since the EUID results from the concatenation of the country code, the business registry identifier, and the identifier that was assigned to the company by that country's registry.

Other examples are the new ideas for the federation of data catalogues or the use of **GeoDCAT-AP** in the implementation of the **INSPIRE[186] Registry** of metadata.

Apart from the technical implementation, putting in place a registry of registries requires an important organisational effort, as well as a strong will to reach agreements. The upper level of political authorities (policymakers, lawyers, business experts and technicians) need to be aware of the opportunity, meet and draft, then govern and execute the strategy on RoR. Ultimately, it should be placed at the core of data governance plans in Member States that aim towards digital maturity of national public administrations.

For the roles responsible for defining new interoperability strategies at the national or EU levels, the recommendation is to assess the existence of business cases that would justify the development of a European Registry of Base Registries (ERBR). For example, use cases from the **Plan for Registry of Registries**[187], namely, aggregation of base registries, and analysis, reports and decision making. This would especially benefit cross-border and cross-sector initiatives.

[181] MAGDA's interconnecting infrastructure of base registries and its use of the ISA's Core Vocabularies: https://joinup.ec.europa.eu/community/epractice/case/magda-20-platform

[182] W3C: https://www.w3.org/

[183] NIFO: https://joinup.ec.europa.eu/collection/nifo-national-interoperability-framework-observatory

[184] The Business Registers Interconnection System: https://e-justice.europa.eu/content_business_registers_at_european_level-105-en.do

[185] The Legal Entity Identifier initiative: http://www.leiroc.org/

[186] Inspire: https://inspire.ec.europa.eu/metadata-codelist

[187] Plan for Registry of Registries on ABR Collection on Joinup: https://joinup.ec.europa.eu/solution/abr-specification-registry-registries/document/plan-registry-registries-released

The **IMOLA**[188] **project** (**Interoperability Model for Land Registers**), launched by the **European Land Registry Association (ELRA)**[189] is a good example. It interconnects land registries on the EU level in order to accommodate the need for a standard means of accessing basic land registry information reusing ISA^2 Core Vocabularies, which help create common data models for standrardised land registry outputs. This European Land Registry Document (ELRD) standard provides an interoperability solution to the variations found in individual land registries and the different formats they use. It formulates reference information within a structure of common fields (a template) and is developed by means of an XML Schema that allows the semi-automated processing of information through shared rules with thesaurus-derived metadata.

The project's vision is an incremental approach emerging from the increased demand for Land Registry information that pertains to the registration of foreign documents and judicial decisions, establishing local equivalents for foreign legal rights in order to get an efficient implementation of EU Regulations on civil and commercial matters. Among the results of the project are:

- A Knowledge Repository integrated with the e-Justice portal as a controlled vocabulary (Thesaurus);
- The use-controlled vocabularies as part of the descriptive metadata to characterise the content of the information objects of the Land Registries;
- A common and shared semantic model to facilitate the implementation of EU Regulations and to consolidate the European Single Market within the frame of a digital administration, linked directly to the semantic domain represented by IMOLA-controlled vocabularies, glossaries and thesauri (Knowledge Repository);
- Standardised and customisable web services adapted to the ELRD schema to facilitate the harmonisation of LRI.

The ELRD validation[190] is supported by the ISA^2 Test Bed[191] project, available both as a standalone online validator and as a community in the ISA² Test Bed. Validation of ELRD content in XML format is possible for versions 3.0 and 3.1, both in an anonymous and stateless manner (via the validator), and also with persistent results and reporting (via the Test Bed).

Reuse DCAT-AP for Base Registries in Europe (BRegDCAT-AP)

The development of a European Registry of Base Registries (a pan-European registry of base registries), should improve the interoperability of individual base registries and harmonise the existing registries of base registries, enabling a one-stop platform for citizens, businesses and public bodies to access and manage base registries across the European Union, and across different domains.[192]

European Union bodies initiated the activities[193] to elaborate, test and implement the **specification of a registry of registries,** namely, **BRegDCAT-AP**, that should provide Member States with a common data model for the creation of their registry of registries, as well as a base for ERBR.

The data model aims to support the description of base registries and registries of registries in Member States and drive the creation of a vocabulary to represent base registries, and registries of registries.

[188] IMOLA - Interoperability Model for Land Registers: https://www.elra.eu/imola-iii/

[189] ELRA - European Land Registry Association: https://www.elra.eu/

[190] ELRD Test Bed support: https://joinup.ec.europa.eu/collection/interoperability-test-bed-repository/solution/interoperability-test-bed/news/itb-supports-elrd-v31

[191] ISA^2 Test Bed: https://joinup.ec.europa.eu/collection/interoperability-test-bed-repository/solution/interoperability-test-bed

[192] Plan for Registry of Registries on ABR Collection on Joinup: https://joinup.ec.europa.eu/solution/abr-specification-registry-registries/document/plan-registry-registries-released

[193] Information and status of BRegDCAT-AP is on ABR – Specification of a Registry of Registries on Joinup: https://joinup.ec.europa.eu/collection/access-base-registries/solution/abr-bregdcat-ap

Together with a proper set of taxonomies and value schemas, it becomes a key component for sharing information between national registries and ERBR.

Both the data model and the vocabulary are based on recognised schemes and ontologies, such as the **ISA^2 Core Vocabularies**[194], **DCAT**[195] (W3C Data Catalogue Vocabulary), **EUROVOC**[196], **NUTS**[197](Nomenclature of Territorial Units for Statistics), and **ELI**[198] (European Legislation Identifier).

Since the ERBR will manage registries (i.e., catalogues of data, and catalogues of catalogues), the development of the vocabulary is based on the W3C DCAT specification, a standard for describing data catalogues and their content. More specifically, the ERBR will extend **DCAT-AP** (DCAT Application Profile for Data Portals in Europe)[199], a technical specification that the ISA^2 Programme developed for describing public sector datasets in order to achieve a successful exchange of metadata among data portals in Europe.

Thus, a new **specification of base registries in Europe** (**BRegDCAT-AP**) was created, as a DCAT-AP extension for describing base registries, their contents, and the services they provide.

Currently, **version 2.00**[200] of the **BRegDCAT-AP** is available online in the ABR Collection on Joinup, with all serialisation.

This specification has been produced by the Access to Base Registries (ABR) Working group (WG), that included leading semantic experts from W3C, nominated members from the ISA^2 Committee, various semantic and other experts from Member States, representing more than 50 organisations and individuals from 19 Member States, that helped shape a stable model and vocabulary.

During the evolution of the specification, from March 2019 to December 2020 [201], the group collected use cases, requirements, provided feedback and discussed the issues and challenges of Member States, along with possible solutions. With the support of reusable tools, the specification was tested by some Member States during a pilot period, by implementing proof of concepts[202].

The BRegDCAT-AP is aligned with the **European Legislation Identifier (ELI)**[203], and takes into consideration feedback from the **Core Public Service Vocabulary Application Profile (CPSV-AP)**[204], **SDG** (**TOOP**[205]), **SEMIC**[206] (more specifically, DCAT-AP), and other EU initiatives.

[194] Core Vocabularies on Joinup: https://joinup.ec.europa.eu/page/core-vocabularies

[195] W3C, DCAT: https://www.w3.org/TR/vocab-dcat/

[196] EuroVoc: http://eurovoc.europa.eu

[197] NUTS: http://ec.europa.eu/eurostat/web/nuts/background

[198] EUR-LEX: https://eur-lex.europa.eu/eli-register/about.html

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

[200] Release v2.00 is downloadable here: https://joinup.ec.europa.eu/collection/access-base-registries/solution/abr-bregdcat-ap/release/200

[201] Discover all details about evolution of BRegDCAT-AP on ABR Collection on Joinup: https://joinup.ec.europa.eu/collection/access-base-registries/solution/abr-bregdcat-ap

[202] Check an article on ABR Collection on Joinup: https://joinup.ec.europa.eu/collection/access-base-registries

[203] ELI: https://publications.europa.eu/en/web/eu-vocabularies/eli

[204] CPSV-AP: https://ec.europa.eu/isa2/solutions/core-public-service-vocabulary-application-profile-cpsv-ap_en and on Joinup: https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/cpsv-ap-tools

[205] TOOP: https://www.toop.eu/

[206] SEMIC: https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic

Once the major releases of DCAT and DCAT-AP are confirmed as standards, the model will need to be reviewed, but until then it represents a stable version to be reused by public administrations and interested EU initiatives to describe the data contained in base registries.

Validation support for BRegDCAT-AP is available through the Interoperability Test Bed[207]. This service provides an easy and configuration-driven approach to set up validation for RDF-based specifications in XML, RDF, and JSON formats, benefiting from the Test Bed's automation processes and hosting resources. To complement this service, the Test Bed has launched a new validator for SHACL shapes[208], allowing specification experts to validate their content before exposing it to their user communities (by checking their RDF serialisation against the specification's expectations, expressed as SHACL shapes). The validator – a service based on the Test Bed's generic RDF validation capabilities – is public and can be used anonymously with no recording of data or validation reports. It is available as a web interface (for users), and as a REST or SOAP API (for machine-to-machine integration), implementing the GITB validation service API that allows potential usage in conformance test cases on the Test Bed platform.

Publish data as Linked Data

Public organisations and administrations, having completed the steps described by the guidelines in the previous sub-sections, are able to publish their data as Linked Data to further facilitate semantic interoperability. Additionally, linked data offers data integration with a low impact on legacy systems; and enables creativity and innovation through context and knowledge creation.

Linked Data, as an enabler of semantic interoperability, is a set of design principles for sharing machine-readable data on the Web for use by public administrations, businesses and citizens[209].

The four design principles of Linked Data (by Tim Berners Lee):

1. Use Uniform Resource Identifiers (URIs) as names for things;
2. Use HTTP URIs so that people can look up those names;
3. When someone looks up a URI, provide useful information, using the standards (Resource Description Framework – RDF, SPARQL Query Language for RDF);
4. Include links to other URIs so that they can discover more things.

As implied above, URIs, RDF and SPARQL form the foundational layers for Linked data. URIs are used for naming things, RDF for describing data, and SPARQL for querying them.

Linked data are different from open data since they can be linked to URIs from other data sources using open standards such as RDF, but without being publicly available under an open licence. Whereas **Open Data** can be published and be publicly available under an open licence but without linking to other data sources. In the case of linked data with an open licence we refer to **Linked Open Data**.

Public organisations and administrations can refer to the following **Good Practices** [210] for the publication of linked data:

- Model the data;
- Reuse vocabularies whenever possible;

[207] Interoperability Test Bed: https://joinup.ec.europa.eu/collection/interoperability-test-bed-repository/solution/interoperability-test-bed

[208] Test Bed SHACL validator: https://joinup.ec.europa.eu/collection/interoperability-test-bed-repository/solution/interoperability-test-bed/news/validator-shacl-shapes

[209] EC ISA Case Study: How Linked Data is transforming eGovernment: https://ec.europa.eu/isa2/sites/isa/files/publications/how-linked-data-20140711_en.pdf

[210] W3C Cookbook for Open Government Linked Data https://www.w3.org/2011/gld/wiki/Linked_Data_Cookbook

- Name things with persistent URIs;
- Publish human- and machine-readable descriptions;
- Convert data to RDF;
- Specify an appropriate licence;
- Host the Linked Dataset publicly and announce it.

The **ISA^2 Programme** provides good practices and practical examples to help public administrations apply Linked Data technologies to e-Government. Specifically, the **SEMIC** Action of the ISA² Programme aims to improve the semantic interoperability of e-Government systems, facilitate information exchange and promote the provision of cross-border and cross-sector EU digital public services. A number of pilots were executed by the SEMIC Action in close collaboration with public administrations in several EU Member States, as well as European Commission services and other EU bodies and agencies as a proof-of-concept to demonstrate the applicability of Linked Data[211]. Among them are the **Registered organisation data pilot** and the **Core Location Pilot: interconnecting Belgian National and Regional Address Registers.**

One of the key Linked Government Data initiatives in Europe is the **European Union Open Data Portal** (EUOPD)[212], which provides access to an expanding range of data from EU institutions and bodies, that can be reused for commercial or non-commercial purposes. It provides, among others, a standardised catalogue, giving easier access to EU open data, a SPARQL endpoint query editor, and REST API access. For the metadata, it has in place a vocabulary which was created using the Data Catalogue Vocabulary (DCAT) and the Dublin Core Terms (DCT) vocabulary. The vocabulary is provided as a worksheet specification and as an ontology. It has been aligned in general terms to be compatible with the Asset Description Metadata Schema (ADMS).

In **Belgium,** the **Flemish government** with **OSLO²** (Open Standards for Linked Organisations)[213] is committed to an unambiguous standard for the exchange of information, ensuring greater consistency and better discoverability of data, so everyone can use easily aggregated information from different national, regional and local e-Government information systems. OSLO² is the logical succession of the OSLO (Open Standards for Linked Administrations) initiative which laid the basis for an open semantic information standard.

In this context, the **Linked Base Registry** for addresses is the effort of the Flemish Government administration to align the base registry for Addresses with the design principles of Linked Data, by unfolding the process followed for raising semantic interoperability based on Linked Data principles.

In the **Netherlands**, the **Dutch Addresses and Buildings key register (BAG)**[214] is published as linked data[215]. BAG is an automated system in which Dutch municipalities keep their information about local addresses and buildings up to date. Municipalities store this information in the National Facility for Addresses and Buildings (BAGLV). The Land Registry Office (Kadaster) manages the National Facility and makes the data available to governments, companies, institutions and citizens.

[211] SEMIC linked data pilots: https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/our-pilots#PilotLOD

[212] European Union Open Data Portal EUODP: https://data.europa.eu/euodp/en/data/

[213] OSLO (Open Standards for Linked Organisations): https://joinup.ec.europa.eu/collection/oslo-open-standards-linked-organisations-0/about

[214] Addresses and Buildings key register (BAG): https://business.gov.nl/regulation/addresses-and-buildings-key-geo-register/

[215] https://lov.linkeddata.es/dataset/lov/vocabs/bag

5. Interconnecting infrastructure

This chapter explores the relevant infrastructure technologies that can allow interconnection and interoperability between base registries. It delves into the data architecture and various approaches that help define facilities and systems for setting up successful platforms.

Searching through distributed base registries is usually one of the use cases that poses organisational and technical challenges. Those challenges can be derived from different facts, such as base registries' owners not being legally allowed to give control of their data to third parties, or some data not being free of charge (by law), or not being open data.

A solution to overcome these challenges is to create and implement an interconnecting infrastructure[216] that allows secure exchange of data between different base registries and enables the reuse of data for public services.

Choose a data architecture model adapted to your organisational model

An organisational model and its challenges always have an impact on data management and, consequently, on the **use of a data architecture model**. This is usually the case when base registries and public services are not centralised but distributed over the Member State's territory. When competences are distributed, each administration tends to strictly control how these competencies are performed within their layer of administration – by choosing an architecture that varies compared to other administrations – thus further impacting the potential of interoperability.

For example, in **Spain[217],** there is currently no registry of registries, but there is **a platform for interconnection of the base registries.** The platform was built based on **the four dimensions of interoperability governance**, namely, legal, organisational, semantic, and technical. The following were defined and implemented[218]:

- First, the legal obligation of public administrations to share information, in the context of the 'Once-Only' Principle (hereafter TOOP[219]), was established, but it is focused on citizens only;
- The next step was the creation of a semantic specification, and, technically, the creation of the platform on a national level. It should be mentioned that there are still autonomous communities that have their own intermediation platforms, but are all interconnected nonetheless;
- From the organisational point of view, the authority (here, the owner of the platform) has the final responsibility for the security of citizens' personal information. This authority established bilateral agreements between public bodies that either consume or provide data. The agreements are handled by a platform manager, both in terms of technical and organisational intermediation.

Currently, there is an increase in the number of consumers and providers that exchange data in Spain, and the authorities are working at this stage on the refinement of the semantic model.

On the other hand, in **Belgium[220],** the competences depend on the political landscape, which includes **federal, regional and local levels**, thus the situation is complex. Belgium is still working

[216] A variety of solutions on interconnecting platforms from MS can be found on ABR Catalogue of Solutions on Joinup: https://joinup.ec.europa.eu/collection/access-base-registries/do-you-need-create-interconnection-platform

[217] Spain' Factsheet on Joinup: https://joinup.ec.europa.eu/sites/default/files/inline-files/Spain%20Factsheet%20Final.pdf

[218] The use case shared by Spain during the collaboration in an ABR working group on definition of a data model, and in a public ABR webinar, https://joinup.ec.europa.eu/collection/access-base-registries/news/abr-webinar-highlights-0910

[219] More information on TOOP: http://toop.eu/info, https://cordis.europa.eu/project/rcn/207635/factsheet/en

[220] Presentation of Belgian authorities in ABR webinar: ABR Webinar 20191009 - Presentation Belgium

on the collection of information in order to obtain an overview on what type of data is stored where. One key issue here is that many data users are not aware of a) what data exists, and b) where they could find sources of certain data.

To address this challenge, Belgian authorities are trying to promote the use of master data sources.

Luxembourg has a **highly centralised IT landscape** and usually there is **only one IT centre responsible** for a **specific sector**: central government, social security, municipalities, health, education, etc. Therefore, this authority is often the de facto one responsible for setting the rules on how to standardise data in a specific sector.

As is evident, **there is no "one size fits all" type of solution** for all Member States. Apart from the organisational model, solutions depend on business and IT requirements (which tend to change over time), and national legal limitations. In the end, most scenarios normally lead to the adoption of hybrid solutions, such as coexistence of two partially implemented approaches; for example, the organisational and master data management implementation approaches (see ‘Define an MDM style’).

When choosing a data architecture model, what is also important is to question the following:

- **General topologies**, e.g. should a central platform be developed or would a distributed model fit better the initiative’s purposes and requirements? Will the chosen model fit all the use cases?
- **Data-sharing model** and how this model affects architectural decisions, e.g. should the data be “delivered”, “consumed” or otherwise conveyed and treated?

Many different projects have already faced these challenges and solved them. Therefore, before designing and implementing new architectural approaches, the recommendation is to study those initiatives that have defined common business processes and services, tackling the use cases where the base registries have to interoperate between themselves and with other public administrations’ services.

One example of such an initiative is represented in **Estonia and Finland,** which initiated and succeeded in their cross-border data exchange by utilising the Estonian **X-Road**[221].

First, the cooperation was signed on the high governmental level by the prime ministers of Estonia and Finland in 2013 (the Memorandum of Understanding about the cooperation in the field of ICT), and in 2014 Estonia provided X-Road to Finland under the EUPL licence, thus the project of the Finnish X-Road implementation started.

Soon afterwards, both countries found out that they needed to share the same X-Road core system and maintain the interoperability between X-tee and the Suomi.fi Data Exchange Layer to enable cross-border data exchange between Estonia and Finland, thus the collaboration initiative was extended in 2015-2016 for the joint development of X-Road. Finland's Population Register Centre and the Republic of Estonia's Information System Authority were assigned as responsible entities for the coordination of the X-Road core development, and a set of practices and guidelines have been also agreed to for managing the cooperation.

Lastly, a shared organisation was established, namely, the Nordic Institute for Interoperability Solutions (**NIIS**)[222], which took over the development of the **X-Road open-source technology**. It is interesting to learn that for the X-tee and Suomi.fi-palveluväylä member organisations nothing changed, namely, Finland's Population Register Centre and the Republic of Estonia's Information

[221] X-Road (Estonia and Finland): https://x-road.global/

[222] NIIS (Finland): https://www.niis.org/blog/2018/5/27/changes-in-the-x-road-development

System Authority remain responsible for their national systems and provide the same support services to their members.

Considering that X-Road is an **open-source software** and ecosystem solution that provides unified and secure data exchange between organisations, free of charge, any interested country or organisation can implement it.

What is usually underlined as an important aspect of X-Road is its **growing ecosystem**. Establishing connections between different data sources is good and, of course, needed. However, what is described as a key matter is enlarging the number of services which can be interconnected. Benefits from the technical perspective are also numerous, namely, with X-Road, cross-border and national data exchanges are implemented through the same channel. This means that there is no need of adding new integrations when exchanging data with a different partner organisation/country

There are some interesting examples of national exchange platforms that have based their solutions on data-sharing models.

In a scenario where there are no legal and business constraints, the suggestion is to consider the **Data Distributor**[223] platform of **Denmark**. This solution is similar to the centralised model, as it gathers and centralises the data from all base registries and consolidates the distribution of the data defined by the Basic Data Program, thus enabling a 'one-stop shop' for data within cross-domain services.

Reuse data architectural approaches on data exchange platforms

On the **EU level**, the following initiatives should be consulted for details and adoption of the data architectural approaches:

- **European Interoperability Reference Architecture (EIRA)[224];**
- **European Interoperability Framework (EIF)[225];**
- **Connecting Europe Facility (CEF)[226].**

One of the key architectural approaches acting as a supportive measure is the **European Interoperability Reference Architecture** (**EIRA**)[227], which offers a global and exhaustive analysis of all interoperability[228] aspects. EIRA is a **four-view reference architecture** for delivering **interoperable digital public services** across borders and sectors. It defines the required capabilities for promoting interoperability as a set of **architecture building blocks** (ABBs), which aim to support public administrations to model and design their business processes and capabilities.

EIRA's main characteristics are the following:

[223] Danish Data Distributor: http://datafordeler.dk/ (available only in Danish)

[224] EIRA: https://joinup.ec.europa.eu/solution/eira

[225] EIF: https://ec.europa.eu/isa2/eif

[226] INEA: https://ec.europa.eu/inea/en/connecting-europe-facility

[227] All information on EIRA can be found on EIRA Collection on Joinup: https://joinup.ec.europa.eu/collection/european-interoperability-reference-architecture-eira. Other non-less relevant are: ensuring the alignment between interoperability business goals and solutions, providing the EIF, providing and maintaining a repository for sharing and reusing generic interoperability artefacts, providing and maintaining a catalogue of services, monitoring interoperability maturity and achievement of goals, creating awareness and strengthening widespread interoperability diffusion, etc. As said, all of them present in many other interoperability layers but mainly in interoperability governance and, within it, change management.

[228] EIRA: https://joinup.ec.europa.eu/collection/european-interoperability-reference-architecture-eira/solution/eira

- **Common terminology** to achieve a minimum level of coordination: ABBs provide a minimal common understanding of the most important building blocks needed to build interoperable public services;
- **Reference architecture** for delivering digital public services: a framework to categorise (re)usable solution building blocks of an e-Government solution. It allows portfolio managers to rationalise, manage and document their portfolio of solutions;
- **Technology and product neutrality** adopts a Service Oriented Architecture (SOA) style and promotes Archimate as a modelling notation (EIRA can be seen as an extension of the Archimate's concept model);
- **Alignment** with the EIF and **TOGAF**[229]: complies with the European Interoperability Strategy (EIS) context. The views of EIRA correspond to the interoperability levels in the EIF and it reuses terminology and paradigms from TOGAF, such as architecture patterns, building blocks and views.

The general recommendation for both policy makers and system developers of base registries/public services is to study EIRA. Additionally, it is to assess whether the chosen strategy or development plan takes into account all the aspects, approached and documented here.

As these features cover a large range of aspects — some of them complex — the recommendation is to assess and also reuse existing **e-Delivery** solutions. The Large Scale Pilots and the Trans-European Systems (mentioned in this document) developed generic, complete and reusable **e-Delivery architectures** that solve most of those complex aspects. An interesting use case concerns the request of certifications (and documents in general), in those countries where e-Delivery solutions are not in place. Instead, these countries develop and publish their own services for direct consumption. This situation leads to a *peer-to-peer* network of base registries, where the interconnected nodes ("peers") expose data to third systems without the use of an intermediation system. The problem with this approach is that each base registry implements its own web service interfaces and exchange data models, which fosters the non-reusability of data and of common semantic assets (e.g. data models, vocabularies and protocols).

Interconnecting platforms in Member States vary in data management and technical points of view and – among good practice examples – it is recommended to study the following ones:

- **X-Road[230] (Estonia and Finland):** independent data exchange layer for information systems, allowing secure internet-based data exchange, based on interoperability agreements between data providers and data consumers;
- **Data Distributor[231] (Denmark**): intermediation platform enabling data distribution, serving as a common authoritative data distribution point, to make it easier for public administrations to publish and use the authoritative type of data;
- **MAGDA[232] (Belgium):** A service-oriented data exchange infrastructure for accessing base registries of citizen and enterprise data, at regional, local and federal levels (where applicable).

When it comes to interconnection platforms, another aspect to be considered is related to the notifications amongst base registries and third party authorities. In the case of notifications, specific workflows, protocols and technical solutions are necessary to ensure the following:

- Whether the notification was received or not and if it was received by the intended addressees;
- What to do in case the recipient's system is down or it is not responding adequately;

[229] The Open Group Architecture Framework (TOGAF) is a framework for enterprise architecture that provides an approach for designing, planning, implementing, and governing an enterprise information technology architecture.

[230] X-Road Data Exchange Layer (Estonia and Finland): https://x-road.global/

[231] Data Distributor (Denmark): www.datafordeler.dk

[232] MAGDA (Belgium): [BE01] Magda in the ABR Collection on Joinup.

- Which steps to follow in case the notification is incomplete or not in conformance with the expected business and semantic rules;
- Undertake the necessary measures to prevent the retraction of the notification act by the sender or the recipients and guarantee that the notification content was not altered or interfered with.

A solution to overcome some of these challenges is to reuse the **Interoperability Test Bed (ITB)** [233], developed by the EC. This solution provides "generic testing facilities to initiatives and public administrations that create interoperability solutions in a cross-border context or linked to European Legislation". The use cases that ITB would support are:

- Simulation of a web service for clients to test against;
- Validation of content sent through various channels;
- Conformance testing against a message exchange protocol;
- Testing of an entire message exchange choreography.

Thus, ITB allows users and systems to connect for the execution of test cases against simulators or reference implementations of specifications that are transparently hosted on its infrastructure, and it offers a test registry and repository (TRR) to store test artefacts (assertions, test cases, validation schemas etc.), and federate test services (validation services, simulator services etc.).

Enable data access supported by APIs

Another data architectural approach concerning data exchange platforms is represented by the implementation of interconnection infrastructure, as such is the case already in some Member States, by creating and implementing interconnection platforms, and enabling data access supported by harmonised interfaces, in particular, by APIs (Application Programming Interfaces).

Regarding harmonised interfaces, the existing standards for building web services could be consulted and reused, for example, the REST (Representational State Transfer) architectural style, by:

- The use of the data structure standards: XML, JSON or their derivatives (e.g. JSON-LD);
- The use of SAML and/or OAuth 2.0 for exchanging authentication and authorisation data.

In terms of harmonised interfaces, many Member States are setting up API strategies and policies, adopting the API approach for data access and reuse, enabling open API-driven services.

The concept of an API strategy refers less to technology itself and more to the development of a community of service or data providers – often referred to as "ecosystems" by the strategists in question – that can be accessed through a central platform. An API strategy is helpful in defining API design parameters and developer-focused criteria for creating an API. One implementation case of an API strategy is that of the **Netherlands**[234]. This API Strategy consists of a core, a generic set of rules that apply to all governmental APIs, and various extensions that are specific to a sector or that are not yet mature enough for the core set. The Netherlands had to deal with two major problems with the use of public services accessed through an API. Firstly, the description of these services was made with the help of a heterogeneous set of techniques used for this purpose and, secondly, the various services were scattered in the infrastructure of each

[233] Interoperability Test Bed: https://joinup.ec.europa.eu/solution/interoperability-test-bed/about

[234] API's Strategy (Netherlands): https://www.youtube.com/watch?v=4vzAk3bdJe8

public organisation that hosted it. The Netherlands wanted to have one single way to describe all APIs and a central point for sharing their descriptions. To achieve these goals the Netherlands adopted the **Open API Specification**[235] (**OAS**).

OAS is a standardised format that makes it easy to generate documentation that always matches the architecture of an API. In addition, OAS implementations provide the possibility of importing and hosting OAS API definitions in one central platform.

Another initiative by the Netherlands, to support their API strategy, was the creation of a knowledge platform, supported by public and private participation aimed at making APIs more responsive to demand, exchanging knowledge on API implementation, and coordinating approach across organisations[236]**.**

One of the biggest challenges that Member States face in developing an API strategy is listing the services they offer, the relevant metadata that accompanies them, as well as using a central point of management for their APIs.

Two popular concepts that are widely used for dealing with the above issues are API catalogues and base registries.

API catalogues are used by organisations to organise their private internal or public APIs. The main characteristics of every API catalogue include documentation, search functionality and accessibility. An API catalogue makes it easy to format and maintain documentation about APIs, supports the ability to search and sort through the various API listings, as well as access the catalogue and understand the listings.

On the other hand, base registries, according to the **European Interoperability Framework**, refer to a trusted and authentic source of information controlled by a public administration or organisation appointed by a government. The EC attaches great importance to the use of public services hosted by base registries and has therefore compiled and provided good practice guidance for setting them up and interacting with them[237].

The **reusable and automated nature** of the services offered by **API frameworks** allows to achieve **reduced administrative burdens**, while at the same time their ability to **connect heterogeneous systems** makes the exchange of information within and across borders easier for European public organisations, as well as companies. In addition, the use of APIs for the combined use of information hosted in various base registries could lead to an **increased degree of automation** for the processing of a number of requests submitted by citizens, legal entities or organisations, by allowing them to provide as little data as possible when filing their requests to public services.

The abovementioned possibilities arising from the use of API catalogues and base registries have already led many European countries to proceed with the creation of such structures at national level, which comprise most of the public electronic services provided by the central government. A typical example is the **French government**[238] which has created a portal that lists and provides access to all state-related APIs. The services of this portal combined with a digital identity platform

[235] OpenAPI Specification: http://spec.openapis.org/oas/v3.0.3

[236] Knowledge platform APIs: https://www.geonovum.nl/themas/kennisplatform-apis

[237] Base registries good practises: https://ec.europa.eu/isa2/sites/isa/files/publications/access-to-base-registries-good-practices-on-building-successful-interconnections-of-base-registries.pdf

[238] French basic registry: https://api.gouv.fr/

(dubbed France Connect[239]) offer citizens composite digital services, regardless of which agency offers the actual service.

Another good example adapted to the logic of using API catalogues and base registries is the **Data Architecture strategy**[240] in **Ireland**, that describes the decision for public bodies to provide data access supported by APIs, data discovery to be facilitated by the Government API Catalogue, and dictates the mandatory adoption of base registries to which access is supported via the usage of appropriate APIs.

The API-first development is another API strategy in which the main goal is to develop an API that puts the target developers' interests first and then builds the product on top of it (be it a website or application). By building on top of APIs with developers in mind, the organisation that adopts this strategy – and the developers who participate in it – are saving a lot of work while laying down the foundations for others to build upon. **UK** has been very successful in implementing this strategy for its legislation service[241].

While an API strategy dictates the overall objectives for achieving seamless and secure connectivity to the right APIs, an API design would use those objectives as an API is planned. API design guidelines could help both API service providers and consumers of these services to adopt best practices when designing development details that relate to API performance, versioning, language and errors handling. An application of such an approach exists in Belgium where the Belgian government maintains a guide of best practices for building Restful web services. This guide aims to improve compatibility between services provided by the government agencies and is a living document, updated when new interoperability issues arise or when REST-related standards evolve[242].

Another important issue that an API strategy addresses has to do with the choice of technologies that will be used for the implementation of API services. An **API strategy** should meet the **expected business and functional needs** by incorporating the **most appropriate API standards** for the implementation of the respective API service.

Nevertheless, there are times when an initial technology choice needs to be modified or adapted to current technological developments or customer requirements. This is the case with **X-Road ecosystem**, a community of organisations using the same instance of the X-Road software for producing and consuming services[243]. X-Road was first built by Estonia in 2001, and initially offered all of its services on top of **SOAP** (Simple Object Access Protocol) which at that time was the de facto standard for web service communication protocols. **SOAP based APIs** has built-in support for features such as security, authorisation but the protocol itself defines too many standards and it takes a considerable amount of time for a developer to grasp its services. Along the way, another web service communication mechanism came to the fore, **REST (Representational State Transfer)** which in general is faster, more lightweight and easier to use than SOAP. For these reasons, new clients of X-Road and even the old ones wanted to use the REST mechanism for interacting with X-Road services. The managing authorities behind X-ROAD responded to this request by making the platform services also available via the REST mechanism[244].

European Union institutions and bodies are also studying, analysing and creating overviews of the **feasibility and practicality of APIs' adoption by the public sector**, by identifying the

[239] France Connect, a citizens SSO service: https://franceconnect.gouv.fr/nos-services

[240] Data Architecture Strategy (Ireland): https://ec.europa.eu/jrc/sites/jrcsh/files/17_06_mark-warren-ireland.pdf

[241] UK legislation API: https://www.legislation.gov.uk/index

[242] Rest Guidelines for building services in Belgium: https://www.gcloud.belgium.be/rest/

[243] X-Road technology overview: https://x-road.global/x-road-technology-overview

[244] X-Road Rest support: https://www.x-tee.ee/docs/live/xroad/pr-rest_x-road_message_protocol_for_rest.html

concepts, terms, technical specifications and relevant API ICT standards that could facilitate Member States' choice and adoption of an **API approach.** One of these initiatives is represented by the **APIs4DGov study - Assessing Government API strategies across the EU** that gathered participants from many Member States in a workshop in which various strategies on APIs were explored; the resulting presentations are publicly available to be consulted and inspired from[245].

There are several other initiatives at the **European level** that could benefit from the use of APIs. They include the publication of high value data sets in compliance with the **Open Data Directive**[246]**,** as well as access of public administrations to **artificial intelligence** (AI) and **high-speed computing**. A representative example for the development and use of AI services is that of Estonia which created the legal and strategic framework for accelerating AI development by establishing its National AI strategy[247]. The strategy describes the sum of actions that the Estonian government will take to accelerate the use of AI in both the private and public sector.

In addition, the compliance with the Open Data Directive is also important since governments have already placed a strong emphasis on open data, as a means for innovative services, and as a way of addressing societal challenges and fostering transparency. By opening up public information, the policies of various governments could aim at creating digital information markets, where new products and services are developed and citizens' participation in political and social life is fostered[248].

In particular, the exposure of government public data using APIs can offer increasing financial opportunities for private companies that make use of them and enable the creation and facilitation of Government to Government (G2G) and Government to Business (G2B) interactions with digital ecosystems. APIs can support the creation of **new useful, innovative products** making it **easier** for public and private organisations to **distribute services and information** to new audiences and in specific contexts that can be customised to provide tailored user experiences.

Member States implement various approaches to ensure interoperability on national level. For instance, **Luxembourg**[249] is working on defining standardised APIs, so that systems – having their own way of structuring and defining data – are able to communicate in a standardised way, which creates a more realistic and productive approach. Standardised APIs are less invasive and give more liberty to authorities responsible for certain domains. At the same time, this is a way to achieve interoperability on a national level. On the other hand, some Member States have already implemented cross-border collaborations with the usage of APIs, and there are good practice examples that can be reused, as mentioned above, Estonia and Finland on secure data exchange via X-Road, also **Nordic-Baltic e-ID cooperation**[250], etc.

The use of APIs is an important factor for the success of cross-border data exchange, but it alone cannot guarantee the successful outcome of these endeavours. Several other factors are involved in such a process, such as **secure data exchange mechanisms**, **compatible data models**, and **semantic interoperability**, along with the legal and administrative concerns. In many cases, this can lead to agreements and contracts being drawn up between the countries whose authorities are exchanging data, and between the parties implementing the data exchange. An example of such a data exchange mechanism that has implemented actions related to the above is the

[245] One of the studies on APIs approach: JRC Technical Report

[246] https://ec.europa.eu/digital-single-market/en/european-legislation-reuse-public-sector-information

[247] Estonia's National AI strategy: https://e-estonia.com/nationa-ai-strategy/

[248] EU open data, The basics for EU data providers: https://op.europa.eu/en/publication-detail/-/publication/c631a6de-ecd5-11e5-8a81-01aa75ed71a1

[249] Interview, October 2020: https://joinup.ec.europa.eu/collection/access-base-registries/news/highlights-interview-luxembourg

[250] https://www.difi.no/fagomrader-og-tjenester/digitalisering-og-samordning/nordisk-samarbeid

already mentioned X-Road project in Estonia and Finland, which is regulated by law and public sector organisations who are willing to access or share data among them. In addition, at **the European Union level**, an initiative has already been developed for facilitating the creation and exchange of public service descriptions called **Core Public Service Vocabulary Application Profile (CPSV-AP[251])**. CSPV-AP aims to facilitate the procedures of exchanging information between public administrations and to have an integrated view of public services offered in a specific country.

Although the use of APIs can be a key factor for the public sector to implement its digital transformation, this task carries budgetary and organisational costs, and important technical challenges including security issues, missing API governance structures, as well as difficulty in adopting proper legal instruments to adhere to current regulation. The EU aims to create a set of good practices both in **aligning the use of API technologies with policy objectives** and in **adopting best practices for the design of APIs**, from which Member States can benefit in their work with APIs.

In this manner, the EU is also moving towards implementing more standards and digital solutions to achieve interoperability. In particular, through a common **European Data Strategy[252]** that also implements the Once-Only principle – which most Member States have in the works – the road towards developing an actual European Registry of Base Registries could eventually become visible on the horizon.

[251] CPSV-AP: https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/core-public-service-vocabulary-application-profile/about

[252] https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en

6. The road ahead

Public administrations have already placed a strong emphasis on open data, as a means for innovative services and a way of addressing societal challenges and fostering transparency. By opening up public information, Member States can create policies that aim at forming digital information markets – where new products and services can be developed –, and fostering citizens' participation in the political and social life[253].

While some Member States are still implementing national-level approaches to open up their data and ensure interoperability, others are already implementing cross-border collaborations: Estonia and Finland (with a potential future implementation in Iceland) are using a secure data exchange via **X-Road**, **Nordic-Baltic e-ID Project (NOBID)[254],** etc. It seems there is a positive trend of such collaborations occurring more frequently in the near future.

The EU is also moving towards implementating more standards and digital solutions that will ensure a smoother cross-border data exchange, and help achieve interoperability via a **European Data Strategy[255]** that will eventually shape Europe's digital future and establish the **Single Digital Market[256]**. Some of the EU initiatives are already being implemented in Member States, e.g. the **SDG / TOOP** projects based on the Once-Only principle, while others are just starting off with the participation of a handful of Member States, e.g. **DE4A[257] Large Scale Pilot**.

On a larger scale, the impact of technological trends, such as **blockchain**, **eID**, even **Artificial Intelligence (AI)** is huge, but such innovative technologies evolve very fast thus Member States and the EU should keep a close eye on them, while fighting the current state of affairs of burdensome administrative processes and non-permisive legislations.

This **Guidelines** document, together with the **ABR Catalogue of Solutions,** constitute rich materials that aim to support public authorities in their journey towards base registries. They provide recommendations and good practice examples on how to start building a new base registry or registry of registries, how to securely connect an existing base registry to an existing interoperability environment (or how to build a new one), and more.

But, most importantly, the guidelines provided in this document complement the **Framework on Base Registry Access and Interconnection (BRAIF),** which constitutes an authoritative reading on how to govern and develop an interoperable base registry and registry of registries.

[253] EU open data, The basics for EU data providers: https://op.europa.eu/en/publication-detail/-/publication/c631a6de-ecd5-11e5-8a81-01aa75ed71a1

[254] NOBID project: https://www.digdir.no/om-oss/nordic-baltic-eid-project-nobid/1342

[255] EDS: https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en

[256] SDG: https://ec.europa.eu/digital-single-market/en

[257] DE4A project : https://www.de4a.eu/

Annex 1. Use cases from Member States

Challenges on the legal landscape

Like any other data exchange solution, **X-Road (used in Estonia and Finland)** also deals with certain challenges, which are common to many similar projects across Europe. The key is to share lessons learnt for the future and discuss possible solutions to overcome challenges. What often appear as obstacles are legal and budgetary requirements, which cause delays in projects. From everyday experience, we see that many organisations have the need – and are requested – to exchange data. However, they face the following issues:

a. there are not enough (or planned) resources for this type of project;
b. even when it is legally allowed – or required – to exchange data, each organisation/department still has to sign agreements regarding the data exchange.

To **overcome** these challenges, X-Road proposes the following **approach and suggestions:**

- when it comes to legal requirements, the European Commission could take this opportunity to provide one legal framework and templates to streamline the process in different countries and organisations;
- to be less aggressive towards budget, X-Road recommends an incremental model – adding new services/elements when they materialise, and replacing obsolete ones rather than replacing/building a whole new infrastructure.

In **Belgium**, in terms of legal differences for the data exchange itself, there is a legal framework, and often the regions and the Federal State have a specific cooperation agreement to work on a specific authoritative (mandatory) base registry. There are discussions on what is mandatory to use and at what level, e.g. if the region wants to recognise a base registry as "authoritative", and the Federal State does not agree to do so. Thus, the organisational model impacts further the architectural approach.

Challenges on data governance and related aspects

Belgium experiences a challenge of having many units of command and trying to establish the flows of data going across them, i.e. shared data governance even if legally each actor has its own smaller scope (e.g. a city has its scope, as does a Member State etc.). In such cases a common model that clarifies how all actors collaborate with each other would need to be established.

In the **Netherlands**[258]**,** there are 10 base registries appointed by law, which are the owners of the data and responsible for their own processes, data quality, etc. The registries provide the data to a central system, the Stelselcatalogus, which plays the role of a metadata collector. There are no common principles or guidelines on data architecture and, among current challenges, some data are defined differently in different base registries or the structure of base registries differ from each other. Currently, the focus is on collecting as much data as possible, thus everyone can view the differences between base registries. The plan is to gradually start working on common definitions. Standardisation would occur when more and more data are combined among all registries. Currently, the translations are offered from the registries as general definitions on the main objects, which will be connected to EU standard definitions.

[258] During ABR webinar, the MS representative offered information on their experiences, more details from webinar available here: https://joinup.ec.europa.eu/collection/access-base-registries/news/abr-webinar-highlights-0910

In **Norway**, one of the main challenges was keeping data up to date between different bodies. This issue was resolved by the establishment of the coordination board, with the participation of the directors of relevant bodies. Cooperation is on a good level, and this helps to achieve an agreement on prioritisation of tasks across different bodies.

In **Sweden,** legal interoperability is the biggest challenge at this point, as well as the lack of a legal base or law which would oblige different authorities on municipal and local levels to share information. Although, in theory, municipalities and government agencies are free to use any system they prefer (they can even use paper forms to collect data). Several authorities in Sweden are using a National portal for open data[259], and the processes there are digitalised. This portal is free of charge, but it is not mandatory to submit data through it.

Seven government authorities – during a government assignment – examined the current situation, analysed different models of base registries in other countries, and consulted the European Commission's guidelines, to propose a new infrastructure which would support implementation of the Once-Only Principle and the integration of base registries. The government is currently analysing this proposal. In the private sector, the systems are sometimes interoperable. This is a result of the needs of the private sector to be more interconnected.

In **Germany,** there is a challenge regarding project governance versus operational constraints. A suitable solution to this problem could be the European Commission's persistent URI Identifiers, which overcome such challenges for projects that have ended but require operational maintenance.

Challenges on standards and processes

Malta represents an interesting example. From a legal perspective, there is a "PSI Directive Transposition and Implementation" that previews the implementation of new Legal Notices on the Registry Authority, Implementing Entity, Person Register, Business Register, etc. Based on these legal requirements, a National Data Strategy was drafted, and it is currently under implementation.

In Malta, a national data infrastructure is represented by:

- An authorisation and representation platform;
- A foundation data layer;
- A Metadata Portal (Registry of Registries);
- A National Data Portal.

For example, in the Foundation data layer, apart from the registry of registries, there are Administrative registries and base registries (person, location/address and organisation).

The national data portal is a one-stop-shop for data discovery and consumption, and the main channel for requests for data. As background information, Maltese authorities have scanned around 4000 legal instruments and identified that they have around 1000 registered names. These are currently being verified, aiming to remove duplicates and ensuring the names are correct, since in Malta the names by law are in Maltese and English.

[259] The national data portal for open data and PSI (Sweden): https://oppnadata.se/#noscroll

The next steps planned by the Maltese Authorities are[260]:

- To establish a link between a registry and the government sector or function, aligning with DCAT-AP;
- To establish owners of base registries, ensuring acceptance of ownership alignment with a business role, as identified in the legal basis;
- To classify data in each base registry, allowing owners to decide whether the data in the registry will be open (for mandatory publishing to the public) or not.

[260] Presentation of Maltese authorities in ABR webinar: ABR_2019-04-08_Malta_Register_of_Registers.pptx