European Commission

Directorate-General for Informatics
Directorate B – Digital Services
B2 - Interoperability

Guidelines on Access to Base Registries

v3.00

Action 'Promoting semantic interoperability amongst the European Union Member States (SEMIC)'

Specific Contract 006 BEACON

April 2023

interoperable europe

SEMIC

Table of Contents

Introduction **1**

Key concepts and definitions 2

Scope and target audience 3

Structure of the document 4

List of guidelines 1

List of abbreviations 2

1. LEGAL ASPECTS (of access to base registries) **3**

1.1 Applicable law and rules and their impact on ABR 4

1.1.1 EU and national law applicable to Base Registries 4

1.1.2 Impact of the current legal context on data sharing 5

1.2 Legal principles 8

1.2.1. Security by design 8

1.2.2. Centralised or decentralised architecture 8

1.2.3. Cloud computing agreements 8

1.3 Guidelines – Legal arrangements 10

2. GOVERNANCE ASPECTS **18**

2.1 Common governance and organisation 19

2.1.1 Data Governance 19

2.1.2 Organisational interoperability 20

2.2 Governance principles 21

2.2.1. Collaboration - Define and establish a common data governance model 21

2.2.2. Transparency and reliability - Agree on the governance aspects and maintain them throughout the data lifecycle 21

2.2.3. Accountability - Clearly define responsibilities and liabilities 22

2.3 Guidelines for base registries governance 24

2.3.1. Guidelines on Data Policies 24

2.3.2. Guidelines on governance structure 29

2.3.3. Guidelines on Change Management 31

2.3.4. Guidelines on Business Continuity 33

3. SEMANTIC INTEROPERABILITY ASPECTS **35**

3.1 Data models and master data management 36

3.2 Guidelines – Data models 37

3.2.1 Guidelines on Master Data Management 37

3.2.2. Guidelines on semantic data models and standards 43

4. TECHNICAL INTEROPERABILITY ASPECTS **53**

4.1 Data infrastructures interconnectivity 54

4.1.1 Base Registries Interconnecting Framework 54

4.1.1 Interoperability through APIs 55

4.2 Guidelines – Data architecture 56

Introduction

Information on basic data items, such as people, companies, vehicles, licences, buildings, locations and roads, is stored in authoritative databases, also known as **base registries.** Base registries 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 improving 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. From 2016-2020 the program continued under the name ISA² Programme and has been transformed in 2021 into Interoperable Europe initiative[1]. Since late 2022, the ABR Action is continuing supporting 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, under the SEMIC Initiative.

Among other aspects, the Interoperable Europe initiative aims at ensuring a common understanding of interoperability through the **European Interoperability Framework[2] (EIF)** and its implementation in MS' public administrations. The Framework can be seen as one of the main axes permeating the actions supporting interoperability policy. In particular, it aims at the following:

- Promoting and supporting the delivery of European public services by fostering cross-border and cross-sectoral interoperability;
- Guiding public administrations in their work to provide European public services to businesses and citizens;
- Complementing and tying 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 **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.

[1] Interoperable Europe Initiative: https://joinup.ec.europa.eu/interoperable-europe

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

Key concepts and definitions

This document provides practical guidance on base registries interconnection and interoperability. Thus, it is essential to define and have a mutual 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"*.[3]
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 Figure 1.

Interoperability[4] 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. 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.

The interoperability of base registries is also key for the development of the **Single Digital Gateway** (**SDG**[5]**)**, a system 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 Once-only technical system (OOTS), based on **the once-only principle (OOP)** [6]. The OOTS will make possible that citizens data, which are submitted once to at least one MS, could be reused by any public authority across the EU.

Interoperability by design

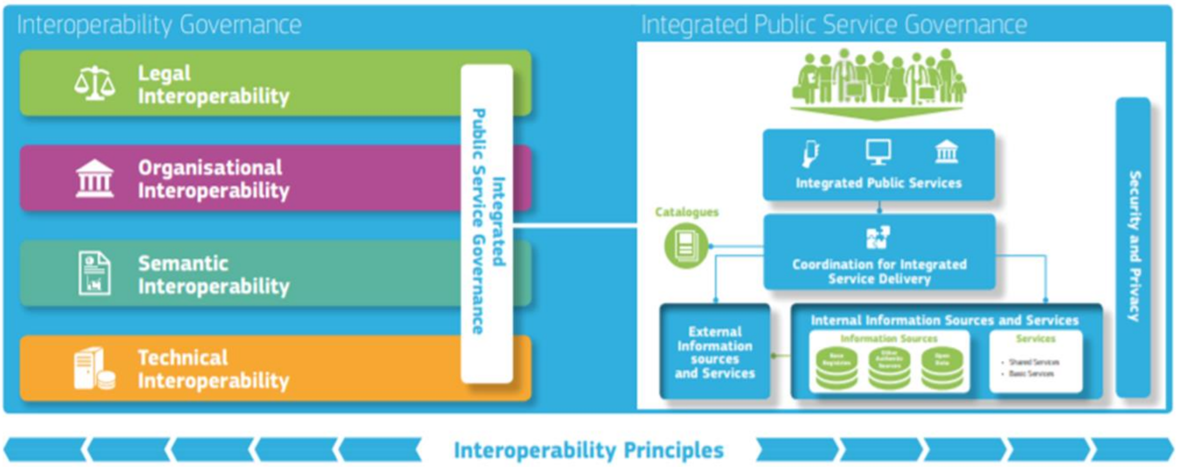
An **interoperability-by-design**[1] concept should always drive the development and evolution of public services. New public services should reuse existing information and available services from public administrations, namely, those already available in base registries.

The EIF defines an interoperability model[7] 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**[8];
- A cross-cutting component to the four layers called "**Integrated Public Service Governance**";
- A background layer called "**Interoperability Governance**".

[3] The EIF: https://ec.europa.eu/isa2/eif_en
[4] 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."
[5] Single Digital Gateway: https://ec.europa.eu/growth/single-market/single-digital-gateway_en
[6] Once OnlyTechnical system (OOTS): https://ec.europa.eu/digital-building-blocks/wikis/display/DIGITAL/Once+Only+Technical+System
[7] The EIF conceptual model: https://ec.europa.eu/isa2/sites/isa/files/eif_leaflet_final.pdf
[8] 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

EIF conceptual model


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

This document provides several examples of good practices examples of Member States' work on base registries and Registry of registries (RoR), according to the EIF layers of interoperability.

In addition, these Guidelines are complementary to the Base Registries Interconnecting Framework (BRAIF)[9] of which the goal is to support the interconnection of base registries through a suggested sequence of defined phases and steps.

Scope and target audience

This Guidelines document has the following objectives:

- To discuss main aspects and **challenges** that may hinder the interconnection and interoperability of base registries at national and cross-border level;
- To provide **recommendations** for the successful and sustainable development of base registries interconnections
- To share **good practice examples on solutions** adopted across Member States for effective cross-border and cross-sector interoperability

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 base registry data interconnection and interoperability models across public administration services.

Accordingly, these guidelines can be interesting for a variety of roles, but in particular for **non-technical roles (Business and policy)** which might be interested in **having a full picture of base registries interconnectivity** before discussing implementation plan with more technical roles. Such non-technical roles might include including staff of national and EU public administrations, such as:

- **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.

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

Structure of the document

Based on the European Interoperability Framework (EIF), the document consists of four chapters:

Chapter 1	***Legal aspects***
Chapter 2	***Governance aspects***
Chapter 3	***Semantic interoperability aspects***
Chapter 4	***Technical interoperability aspects***

Each chapter contains a set of useful information, guidelines and examples which should support different aspects of base registries interconnection. More specifically, the chapters are organised according to the sections described below:

Introduction and current context

The introductory part which will provide you an overview of the EU policies and strategy.

Principles

In some sections, the second part focusses on the principles that should be taken into consideration.

Guidelines

Third part lists the relevant the Guidelines on how to achieve the desired goal.

Best practices

Most of the Guidelines include interesting examples and best practices from the EU and its Member States which can be inspirational for achieving the same results.

List of guidelines

🟩	G1	**Establish the overall Governance**
🟩	G2	Define the hierarchy of norms and regulations
🟩	G3	Define and prerequisites
🟩	G4	Delimitate the scope
🟩	G5	Make data accessible through open source specifications
🟩	G6	Facilitate data sharing beyond national borders
🟩	G7	Define sharing principles
🟩	G8	Assess and define economic capacity and funding
🟩	G9	Set-up a legal framework favouring enforcement of data protection laws
🟪	G10	Ensure the right users access original and authentic data
🟪	G11	Ensure the security of the data access and its communication
🟪	G12	Ensure and control the quality of the data by all means
🟪	G13	Envision the global (holistic) organisational picture
🟪	G14	Establish interoperability agreements to ensure base registries and public services sustainability
🟪	G15	Draft a change management plan
🟪	G16	Implement and release change requests
🟪	G17	Ensure digital preservation and permanent access to data
🟪	G18	Agree on flexible data availability levels
🟦	G19	Define an MDM style
🟦	G20	Define data types and their management approach
🟦	G21	Identify unique and unambiguous instances of your master data
🟦	G22	Define the data domain
🟦	G23	Distinguish scope and use of metadata
🟦	G24	Define semantic assets of (master) data
🟦	G25	Reuse semantic assets: Ontologies and taxonomies
🟦	G26	Reuse semantic assets: Core Vocabularies
🟦	G27	Reuse semantic assets: standard Application Profiles
🟦	G28	Publish data as Linked Data
🟧	G29	Choose a data architecture model adapted to your organisational model
🟧	G30	Reuse data architectural approaches on data exchange platforms
🟧	G31	Use common testing tools to ensure for interoperability conformance
🟧	G32	Enable data access supported by APIs
🟧	G33	Develop specific strategies to steer APIs implementation

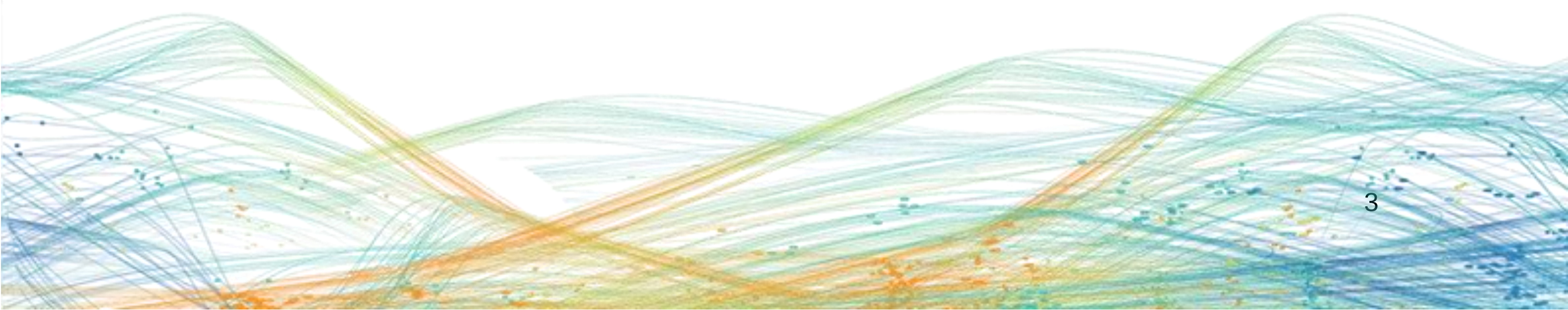
List of abbreviations

Find below a list of the main acronyms and terms used in this document.

Abbreviation	Explanation
ABR	Access to Base Registries
BR	Base Registries
BRAIF	Base Registries Access and Interconnection Framework
CEF	Connecting Europe Facility
DPA	Data Protection Authorities
DPO	Data Protection Officer
EDPB	European Data Protection Board
EIF	European Interoperability Framework
EIRA	European Interoperability Reference Architecture
EUPL	European Union Public Licence
MS	Member State(s)
NSA	National Supervisory Authorities
SDG	Single Digital Gateway
SEMIC	Semantic Interoperability Community
RoR	Registry of registries
TES	Trans-European Systems

1 Guidelines for legal interoperability

This chapter aims to support decision-makers across European countries who are in the process of drafting or updating their legislation with impact on base registries/data management/data sharing, providing advice and key considerations for finetuning the national legal base.



1.1 Applicable law and rules and their impact on ABR

1.1.1 EU and national law applicable to Base Registries

Base registries are reliable sources of basic information on items such as persons, companies, vehicles, licences, 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, car, company, etc.

Among the EU legal texts and guidelines that have an impact on organising the access to base registries, the following are particularly relevant:

1. **General Data Protection Regulation (GDPR)**
2. **Directive on the reuse of public sector information (PSI)** - which encourages Member States to make public information available for access and reuse as open data[10]**;**
3. **INSPIRE Directive -** which is handling the matters of processed/shared spatial or geographical information, and requires sharing of spatial datasets and services between public authorities with no or as few as possible restrictions or practical obstacles for its reuse[11]**;**
4. **European Interoperability Framework (EIF)** - which recommends the use of open licences for both data and software (EIF recommendations 2 & 3)[12]**.**

Each of the above-mentioned documents is providing valuable information and principles that are tackling different aspects of the base registries' interconnection. In this section, The GDPR will be presented in detail, as it is the most comprehensive, thus representing the main legal basis. The additional three texts provide further legal context and should be consulted as needed.

1.1.1.1. The General Data Protection Regulation (GDPR)

Since 25 May 2018, the GDPR Regulation (EU) 2016/679 protects natural persons regarding 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[13] countries Norway, Iceland and Liechtenstein) 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 an EU citizen[14] or 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 countries and defines the competences of national supervisory authorities (NSA), data protection authorities DPA) and of a European Data Protection Board (EDPB)[15].

[10] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information: https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1561563110433&uri=CELEX:32019L1024

[11] INSPIRE Directive: https://inspire.ec.europa.eu/inspire-directive/2

[12] European Interoperability Framework: https://ec.europa.eu/isa2/eif_en

[13] European Economic Area.

[14] More information for UK citizens: https://www.gov.uk/data-protection and https://ico.org.uk/for-organisations/dp-at-the-end-of-the-transition-period/data-protection-and-the-eu-in-detail/the-uk-gdpr/

[15] 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.

1.1.1.2. Data protection rules beyond the GDPR

Other rules, beyond GDPR, are applicable to data protection, in the following cases:

- When the European Union Data Protection Representatives (EUDPR)[16] 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[17];
- When data are processed and exchanged in the framework of EUROPOL[18];
- When focusing on specific aspects, like the ePrivacy Directive (and upcoming regulation) i.e., complementing the GDPR regarding on-line/direct marketing or like the privacy aspects of telecommunications;
- When there are activities processing electronic data that are non-personal[19].

1.1.1.3. 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[20].

1.1.2 Impact of the current legal context on data sharing

Based on the same principles, the GDPR has reinforced (with the possibility of strong penalties) the data protection already provided under the previous Directive 95/46. It would be out of scope here to detail the GDPR principles[21] and conditions for personal data processing where at least one is enough (consent, legal obligation, public interest etc.).

Without delving in the details about the specific GDPR provisions (rights of data subjects, obligations of the controller, role of the Data Protection Officer (DPO), etc.), neither the fact that service providers or sub-contractors acting on behalf of data controllers (as data processors) are equally committed to comply with these provisions. Though, it would be useful to highlight some points that are relevant for implementing ABR sharing systems.

1.1.2.1. Data transfers to third countries

The transferring of personal data **outside the GDPR zone (EEA)** can take place only when a GDPR-equivalent level of protection is ensured. For instance, there might be the case in which an authority

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

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

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

[19] Regulation (EU) 2018/1807 of 14 November 2018 on a framework for the free flow of non-personal data in the European Union

[20] 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.

[21] namely, i) lawfulness of processing, ii) legitimate purpose, iii) minimisation, iv) accuracy, v) limitation in time, vi) integrity, vii) confidentiality, and viii) accountability

from Country X would need residence data on citizens from another EU Country Y to understand how many of its citizens moved to that country in the past year. In case Country X is an EEA country, GDPR would immediately apply, but also when such a country is an extra-EEA one, certain conditions for the data flow should be satisfied. This must be legally enabled through an ad-hoc EC 'adequacy decision' as the conditions for data transfers to third countries are quite stringent in the GDPR (GDPR Articles 44 – 50).

One of the main examples of this is the **United Kingdom**. After the country left the EU on 28 June 2021, the European Commission adopted two adequacy decisions to acknowledge the equivalence of personal data protection for transfers in of personal data to in the UK, in relation to two main EU legal sources, namely being the General Data Protection Regulation (GDPR) and the Law Enforcement Directive (LED) respectively.

Another relevant case of data transfers to an extra-EEA territory is represented by the **United States**, which hosts some of the world's most important ICT, cloud computing and social network providers. In 2000 the EU Commission passed the 'Safe Harbour' decision, as one of the first attempts to seek alignment between US principles on data protection and EU Directive 95/46. The agreement resulted in a first alignment in the customer data protection practices between EU and US business actors according to "safe harbour" privacy principles, established to prevent the accidental loss or disclosure of personal data The decision was subsequently substituted by a new framework for transatlantic data flows, through another adequacy decision, the 'EU-US Privacy Shield' which was nonetheless invalidated /2020 by the CJEU, as its predecessor. To remedy the situation, a new agreement in principle was announced between EU and US in March 2022, opening the door for a new agreement.

The current status of the EU-USA legal context in the field of data transfers implies 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).

It is also important to note that on 4 June 2021, the European Commission published its final Implementing Decision adopting new Standard Contractual Clauses ("SCCs") for the transfer of personal data outside the EEA. Given that the USA has not been deemed to have an adequate level of data protection, the SCCs will become a critically important mechanism for the transfer of personal data.

1.1.2.2. Impact on 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[22] because of the following frictions:

- The GDPR assumes 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 a specific body is assigned with the responsibility.

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

- The GDPR assumes 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. Still, it should be underscored that the notion of ‘erasure’ in Article 17 has been interpreted broadly by the CJEU, which may leave room for compliance with the GDPR in case the data are “delisted” or become inaccessible in the ledger.

- 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 to store data anywhere in the world, without any restrictions posed by a closed number of localised partners, Blockchain technologies would not be 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.

Notwithstanding these challenges, 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 way, but instead must be determined on a case-by-case basis. Despite the numerous experiments, no concrete use of Blockchain technology can be observed so far in ABR projects, like the Nordic countries’ X-Road project (see further below for more details).

1.2 Legal principles

The existing legal framework related to data, its management and transfer, impacts technologies and infrastructures that are relevant for base registries. It is therefore possible to identify several principles related to the set-up/implementation of data sharing infrastructures, inspired by key legal sources and mostly the GDPR.

In this respect, this paragraph introduces a few relevant principles focusing on the following:

- Security by design
- Centralised or decentralised architecture
- Cloud computing

1.2.1. 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).

1.2.2. 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[23], it maintains responsibilities at local level, implying that Member States and authorities are those responsible for the civil or commercial records databases and the processing of personal data within these databases.

However, it is worth highlighting that extremely decentralised architectures may become problematic in case the wide dissemination of data controllers makes it difficult to allocate responsibilities (see section 1.1.2.2. related to Blockchain technologies).

1.2.3. Cloud computing agreements

Cloud computing agreements should be concluded making sure that data is treated in compliance with existing laws and especially GDPR, even in the case of cross-border computing activities involving data of EEA citizens.

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’ personal devices. 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, for instance China, India, Europe and the US work at different hours. Consequently, the location of data, and the applicable legal protection, may vary and,

[23] 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

thus, bypass any European protective regulation. The principal potential cause of issues is the prominent role of US 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 address the following aspects:

- 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 European Economic Area (EEA), or is any data being transferred outside it, including backups?
- Does the service comply with all GDPR obligations, including the permanent data deletion when not used anymore, and are data accessible for correction?

Apart from the principles presented above, and related to specific technology aspects, several general data sharing principles should be clearly defined also when it comes to setting up and administering base registries. Such general data sharing principles have been included in a specific guideline (**Guideline #7**) in the following guidelines section.

1.3 Guidelines - Legal arrangements

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.

The following table includes the list of Guidelines on legal arrangements for the interconnection of base registries, provided in this section.

G1	Establish the overall Governance
G2	Define the hierarchy of norms and relationships
G3	Define and prerequisites
G4	Delimitate the scope
G5	Make data accessible through open source specifications
G6	Facilitate data sharing beyond national borders
G7	Define sharing principles
G8	Assess and define economic capacity and funding
G9	Set-up a legal framework favouring enforcement of data protection laws

G1 Establish the overall Governance

Each (national) instrument is to be placed under a national authority, normally defined by the generic term 'Minister'. 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.

Case example: Definition of governance and roles in Ireland

To facilitate, accelerate and provide a comprehensive legal framework for sharing data between base registries, Member States may implement a general-purpose NIS. Such a framework has been implemented by Ireland through the ***Data Sharing and Governance Act 2019***[24].

The implementation of data sharing must be documented in a standard, written agreement that is concluded voluntarily between relevant bodies. Such agreement must specify the stakeholders, the shared information, the purpose and functional link, the legal basis, and who will or may use/disclose

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

data. Accordingly, the Act defines the roles and the categories of specific public service information, including:

- **Lead Agency:** which normally acts as data controller (or a joint controller).
- **Minister**: it designates the base registries for use by public bodies, so that they can access BR personal data without having to collect it directly from service users.
- **Registries owners**: under the governance of the minister, when the information collected is qualified as a "base registry", registry owners must ensure that the personal data are accurate and up to date.
- **Data subjects**: which are individual citizens. Their right to assess their own data is given through a personal data access portal (including information on any data breaches or data sharing agreements affecting them).

The 2019 Irish Act is one of the most recent and complete national legislations governing data sharing across public bodies. It follows previous efforts undertook by other countries, such as Spain, which in 2015 obliged public bodies to share data without requesting paper certificates from citizens. In a country characterised by multiple authorities at national, regional and local levels and different local languages, this represented another relevant example on the feasibility of legal tools aimed at fostering base registries data sharing and interoperability.

G2 Define the hierarchy of norms and regulations

By defining the hierarchy of norms and regulations, it will become clear which other acts, or higher (EU) regulations are impacting the specific ABR regulation. An ex-ante inventory is needed as it may be that other specific acts contain ABR restrictions or conditions.

Case example: Relationship between European and national legislation in the context of X-Road

The X-Road[25]® data exchange layer is continuously developed and managed by the Nordic Institute for Interoperability Solutions (NIIS). Two different X-Road ecosystems can be connected, which enables easy and secure cross-border data exchange between countries using X-Road. Among EU Member States, X-Road is used nationwide in the Estonia (*X-tee* data exchange layer) and in Finland (*Suomi.fi Data Exchange Layer* service). Outside of the EU, X-Road is also used in Iceland (national X-Road environment *Straumurinn)* and in the Faroe Islands (*Heldin* environment).

NIIS members have their own X-Road ecosystems, and they are responsible for operating them. Any legal constraint related to X-Road is 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 referring to **X-Road as an open-source software there are specific no laws or regulations** directly related to it, but when referring to it as an 'ecosystem' in Estonia or Finland, the use of **the national X-Road ecosystem is regulated at the national level**, and each country has its own laws and regulations.

When it comes to non-EU countries, such as Iceland the same overarching EU rules on data protection apply, since it is part of the GDPR zone as an EEA state (the Icelandic Parliament passed its 2018 Data Protection Act - Act 90/2018- to implement the GDPR). The Faroe Islands – despite not part of the EU – have agreed to adopt similar level and modalities of data protection to the GDPR. This was recognised by the EU Commission through an adequacy decision which certified the Faroe Islands as a country providing GDPR-comparable levels of data protection[26].

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

[26]https://commission.europa.eu/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en

G3 Define the concepts applied to the specific legislation

All concepts already defined in other general instruments (i.e., the GDPR data controller, processor, data subject, personal data, special categories of data, etc.) should be mirrored and contextualised into specific legislation aiming at regulating data sharing, including rules on Base Registries.

This is relevant since, key concepts, such as ‘base registry’, have often not yet been legally defined in national frameworks. Find a list of examples below:

- **Base Registry (BR)** should be defined as a database designated 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 Registry of Registries (RoR)** can be considered as a cornerstone and a prerequisite for any global sharing service. The RoR highlights 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.
- **Base registry owner** describes the entity/organisation that is the appointed controller of the data in the base registry.
- **Electronic record** relates to structured information in electronic form produced by a software application or as a result of digitisation, e.g., paper scanning.

G4 Delimitate the scope

Legislation aiming at regulating the exchange of public data in compliance with European and national legislation should clearly define the scope of its action by:

- **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 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).

Case example: Scope delimitation in the Irish Data Sharing and Governance Act

The **Irish Data Sharing and Governance Act** is also a relevant example of scope delimitation relevant for base registries. The Irish Act regulates the sharing of information between Irish public bodies, including personal data – but also specifies special categories of data which are excluded.

These comprise all kind of crime-related data used for crimes prevention or prosecution as well as data related to state security, espionage, and defence. In accordance with GDPR, 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 relevant for the provision of a public service (performance, verification, avoidance of burden, assessment, …).

The Act – which must be reviewed on a regular basis (< 5 years) – also states that all relevant public bodies that must comply with the Act should be defined and listed by name (i.e. the Attorney General, the Police/Guardia Siochána), or by category (i.e. recognised school, bodies delivering public services under an agreement with a public body).

G5 Make data accessible through open source specifications

In order to foster the distribution and reuse of public sector assets, open source publishing of specifications should be encouraged, possibly under a recognised open-source licence. This should preferably be done for free and in a non-commercial way.

Main examples of open-source licences normally fall under one of the following two types:

- Reciprocal licences - such as e.g. the European Union Public Licence (EUPL)[27] - which has been specifically designed to reuse, validate, edit and distribute open source software needed to keep track of further distributed improvements. This is because its 'reciprocity' implies that any derivative developed by a re-user/third party will not become property of such third party, but any source code should be made accessible an reusable as well.
- Permissive software licences - such as e.g. an Apache or MIT licence – which, differently from reciprocal licences, are more 'permissive' in the sense that allow the creation of ownership of derivatives generated by open-source licences.

In both cases, publishing open-source specifications under open-source licence is beneficial since it favours harmonisation on the approaches given the absence of access costs and the possibility to reuse solutions that are already in use by similar public services.

Case example: X-Road as open-source framework of cross-border interoperability

X-Road[28]® is one of the main cross-border interoperability frameworks currently in use in Europe, mostly by northern European countries (including e.g. Finland, Estonia and Iceland). Despite operated centrally, X-Road is based on a distributed architecture where each member organisation of the ecosystem manages its own data and controls who is allowed to access it.

X-Road software is open source and provided for free under the MIT licence[29]. 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, permissive (recipients can implement it in the way most suitable for their needs) 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 between MIT licence and the European Union Public Licence (EUPL) that is reciprocal: in the case the software is re-distributed (and providing access online via a network as 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).

[27] Introduction to the EUPL licence, Joinup: https://joinup.ec.europa.eu/collection/eupl/introduction-eupl-licence

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

[29] https://github.com/nordic-institute/X-Road/blob/develop/src/LICENSE.txt

G6 Facilitate data sharing beyond national borders

When sharing data beyond national borders, the following aspects should be taken into consideration:

- **Implementation and reciprocity**: this aspect is normally missing (not forbidden according to the GDPR, but not specifically regulated) in national instruments, but 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).
- **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.
- **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[30]. 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.

A central legal question when implementing such systems is about 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.

It is possible to identify three high-level scenarios:

- a common platform is interconnecting decentralised databases, where all the data displayed from the central platform are "transient data" (not stored in any central component) and where no tracing or logs from user-run queries through the service are centrally recorded or kept. In such cases 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
- a system stores core data in a central component, and queries are run against this central database. In this scenario the owner of this central system looks as the most convenient data controller.[31]
- The stakeholders involved in the set-up of the cross-border data sharing agreement, jointly decide to assign the responsibility of data controller to a specific body

[30] 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.

[31] This is the case for instance of European portals hosting national base registries that are kept updated at regular intervals). where the legal notice privacy statements 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 Portal is the European Commission".

Case example: Trans-European Systems (TES)

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. Trans-European Systems (or solutions) are operational interoperable solutions for cross-border exchange of data between public administrations’ base registries and, in some cases, between citizens or businesses.

TES are normally owned by the European Commission or by private initiatives and foundations often co-funded by Member States in the framework of EU-polices on cross-border cooperation. The table below lists some examples of running (or planned) TES projects:

TES projects	Details
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)

Given their cross-border dimension, TES implementations often require the negotiated adoption of specific legal instruments (legal basis) and are normally built on voluntary participation where members can opt-in (e.g. for Land Register Interconnection). This process normally requires multiple years of consensus-building activities and national legal frameworks reconciliation.

The projects above include examples of different ways of dealing with data control:

- Multiple data controllers in a platform interconnecting decentralised databases (e.g. EPRIS, EUCARIS or for the systems managed by the eu-LISA Agency.)
- Owner of a central component storing core data acting as single data controller (e.g. EU Commission for BRIS/ e-Justice portal)

Specific body is assigned the role of data controller through agreement among different stakeholders (e.g. e-CODEX or EVIDENCE2, where a foundation is assigned with the role of data controller for all personal data processed through the system).

G7 Define sharing principles

As mentioned above in the section on legal principles, when interconnecting the base registries, the following sharing principles should be considered:

- **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 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, and 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 (e.g., 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.

G8 Assess and define economic capacity and funding

A key aspect to clarify at central management level is related to funding and fee collection: beyond the business plan of the infrastructure provider, some BRs provide access free of charge in one country, while in other countries equivalent BRs may require payment for accessing specific data. This aspect, along with the covering of the operational costs of the infrastructure, is normally one of the key topics when it comes to cooperation on base registries interconnectivity.

When it comes to funding, the infrastructure provider should plan some efforts in actively promoting the economic advantages related to the set-up of interconnecting infrastructures of base registries (i.e., costs cutting for requests submissions 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 related to the multiannual implementation of cross-border initiatives justifies an additional set of skills in the field of accounting and financial audit/transparency. This might be needed to manage funding and fees in cross-border exchanges, which are typically issues that are problematic since based on national regulations. Specific support should be foreseen to the end of harmonising Base Registries fee practices across the various Member States according to FRAND[32] (Fair, Reasonable And Non-Discriminatory) conditions and to make sure that eligible projects find their way to receive funding from European funds in initial phases[33].

[32] Fair, reasonable and non discriminatory

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

G9 Set-up a legal framework favouring enforcement of data protection laws[34]

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 the GDPR. In order to implement this regulation, competent EU authorities (i.e., EDPS), MS and other countries defined clear guidelines on how to ensure citizens' rights in relation to their personal data protection.

As a general legal guideline for the alignment of data sharing national legislation with the main legal framework, policy makers and administrators of public systems and services should focus on rethinking public services and the interoperability with base registries to give them a more user-centric dimension. This implies the definition of clear roles and responsibilities regarding the management of the data, providing citizens with '**proportionate' legal instruments for self-control of personal data**.

In this process it is advisable to involve a wide range of professional figures in assisting the public authorities dealing with the set-up and provision of digital government services in the configuration of a legal framework. Such figures include policy makers, legal drafters, service/business managers and IT experts who can also validly assist with the definition and implementation of appropriate processes for public services.

Case example: Approaches to data protection across European countries

In the **Netherlands** the approach to data protection has entailed the adoption of a 'standard' data protection schema to provide individuals control of their data. The schema - named **QiY**[35] and provided by the homonymous foundation that includes private and public organisations (e.g. Ministry of Economy and Ministry of Interior) - defines a trust framework for individual users, companies, and government organisations. Such a schema allows users to obtain full, secure and private control of their personal data and the possibility to choose which data to share with third parties. The schema is provided in open format[36] with specific guidelines for its implementation.

Related to the 'proportionate' instruments, one example could be the objective of unburdening citizens, reflected in the legislation in **Spain** through Laws 39 and 40 of 2015. These amend the previous legal framework on public administrations and common administrative procedures. The current context is more flexible and, similarly to other EU countries, Article 28 of the new Law 39 allows the sharing of data, unless citizens express 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 MyData[37] on **Finland** and Datove Schranky[38] in **Czech Republic**, that have strong roots in their national legal frameworks and developed user-centric approaches to manage the relationship between legal entities and public administrations.

[34] See Legal section for detailed information.

[35] QiY factsheet (Joinup): https://joinup.ec.europa.eu/sites/default/files/inline-files/NL03.pdf

[36] QiY schema (GitHub): https://github.com/qiyfoundation/Qiy-Scheme

[37] 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

[38] 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

2 Guidelines for governance and organisation

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 into an effective strategy.

2.1 Common governance and organisation

2.1.1 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.

The **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"*[39]

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.

There are many challenges[40] that Member States must tackle which are relevant for base registries data governance. For instance, some MS have different authorities responsible for operating and maintaining base registries, which can generate challenges in the definition of data flows and data across them and cross-border. 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.

As introduced previously, the EIF defines a framework to support interoperability based on four 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, agreements and other aspects of ensuring and monitoring interoperability at national and EU levels"*.[41] 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".*[42]

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

[40] See Annex 1.

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

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

EIF recommendations on governance and organisation

To support public administrations in their work on interoperability governance, the EIF offers public administrations a set of **47 concrete recommendations.** Such recommendations suggest 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[43].

Data sharing at national and European levels can be challenging since the exchange of data occurs in an ever-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 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).

2.1.2 Organisational interoperability

Organisational interoperability, being part of the EIF interoperability model 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.

> *"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"[44].*

Organisational interoperability in the case of base registries presents challenges both when data must be exchanged at national and at EU level:

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

- **EU level**: a national registry must interoperate with other registries, but this process is hindered by country specificities that see each MS having its own administrative rules, business processes, data collection and maintenance roles.

Effective organisational interoperability therefore requires not only putting in place flexible architectures capable of reconciling national differences, but also implies relevant organisational efforts. Such efforts should be addressed at **aligning all the processes** executed in each of the different stakeholders' systems.

Normally the achievement of organisational interoperability requires the relevant stakeholders to be actively involved in standardisation efforts and ultimately agree on the use of **common standards**[45] which would enable data flows, while, at the same time**,** allow for the smooth prosecution of usual processes. However, in some cases, an agreement might be unfeasible or extremely complex to reach

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

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

[45]EIF:https://joinup.ec.europa.eu/collection/nifo-national-interoperability-framework-observatory/solution/eif-toolbox/recommendation-24

technically. In such cases the stakeholders should jointly identify or design **common processes** for selecting relevant standards and monitoring their implementation[46] and compliance.

2.2 Governance principles

There are several principles that should be considered when establishing governance models aiming to support data interoperability.

2.2.1. Collaboration - Define and establish a common data governance model

A **common data governance model** allows to clarify roles and processes, increasing timeliness and reliability of data management and sharing over time.

The principle of collaboration should be guaranteed by establishing organisational bodies that will facilitate decisions taken at two key different levels of governance:

- **upper-level governance** (e.g., national level authorities),
- **lower-level governance** which relate to the individual base registries level

The organisational bodies should facilitate the reconciliation of the decisions taken at these two governance levels (e.g., the compliance of an access policy to a base registry taken by a base registry owner against a data security policy decided by a national authority.

In other words, the common data governance model should aim at governing the interoperability of base registries and that of public services to which they provide their data. To this end and to foster collaboration, the common governance model should take into account that data itself is the main asset of any base registry, meaning that the object of its governance should pertain to each aspect related to data (data security, data quality, etc.).

2.2.2. Transparency and reliability - Agree on the governance aspects and maintain them throughout the data lifecycle

The **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[47]:

- **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

[46]EIF:https://joinup.ec.europa.eu/collection/nifo-national-interoperability-framework-observatory/solution/eif-toolbox/recommendation-21
[47] These topics are being discussed more in detail in next sections of the document.

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.

2.2.3. Accountability - 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 complaints, and a flexible decision-making process, allowing them 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.

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.

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

Example: Initiatives interconnecting business registries

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.

More examples from Member States are available in the section on Establishment of data governance models in different Member States.

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

2.3 Guidelines for base registries governance

This section provides a number of guidelines on base registries governance and organisational aspects, divided in five main blocks:

- Guidelines on Data Policies
- Guidelines on Governance Structure
- Guidelines on Change management
- Guidelines on Business Continuity

The following table list of guidelines provided under each of the above-mentioned blocks:

G10	Ensure the right users access original and authentic data
G11	Ensure the security of the data access and its communication
G12	Ensure and control the quality of the data by all means
G13	Envision the global (holistic) organisational picture
G14	Establish interoperability agreements to ensure base registries and public services sustainability
G15	Draft a change management plan
G16	Implement and release change requests
G17	Ensure digital preservation and permanent access to data
G18	Agree on flexible data availability levels

2.3.1. Guidelines on 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 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[50], 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.

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

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

G10 Ensure the right users access original and authentic data (Data Policy on Authorisation and Accessibility)

The data kept in a base registry is, by definition, public data. However, this does not imply that any individual or any system should be granted access at any time or at all. Only **legitimate** users can 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 (and their access rights and responsibilities), should be defined and specified in each national legal framework. Moreover, the process of identifying, tracking and logging 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**[51] that concerns the electronic identification and trust services for electronic transactions in the internal market[52], to which all new solutions should comply with. In a nutshell, the regulation defines how people and businesses should be enabled to use their own national electronic identification schemes to access public services in other EU countries where electronic IDs are available.

To understand how Regulation 910/2014 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 former CEF[53] programme. In particular, a **catalogue of reusable building blocks** was developed by the EC and MS to support service providers and national eID infrastructure owners involved in the study **eID**[54] and in the implementation of eIDAS nodes according to EU guidelines.

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)[55]. 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**).

An option which ensures a strong certainty of non-repudiation consists of signing electronically the shared data by using **qualified signatures**[56]. From the legal perspective, the eIDAS Regulation

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

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

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

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

[55] Such as the European Single Procurement Document (ESPD) Service, which exports electronic documents with references to service providers and base registries (examples in France and Spain)

[56] 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

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**[57] (**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)[58]**, in which the monitoring process is carried out 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).

Case example: Danish Basic Data Programme

The **Danish Basic Data Programme[59]** is a good example of data governance model implemented at national level. Its intended purpose was to overcome typical challenges related to base registries data, such as duplications, differences in definitions, data silos not connected to each other (resulting in shadow registers)[60], and flaws in data management.

Focusing on 5 categories of **basic data**, selected by the Programme (Spatial Data, Address Data, Property Data, Company Data, Personal Data) and covering more than 10 authoritative registries, the programme identified **five processes** for data access to be implemented[61], as a “way to open easy-to-access high-quality basic data”:

1. 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.
2. 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.
3. To make it possible to link data, efforts will be made to ensure that all data conforms to the same technical requirements.
4. 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.
5. To ensure efficient, effective, and coordinated development and use of basic data, a cross-institutional basic data committee is to be established.

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.

[57] 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.

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

[59] 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

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

[61] 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

G11 Ensure the security of the data access and its communication (Data Security Policy)

The aim of a data security policy is to secure the communication channels and the means used to transmit the data from one base registry to others. It should define a strong security environment as the basis to ensure secure and trusted data exchanges.

To this end, useful input comes from **EU Regulation 910/2014** on electronic identification and trust services. The Regulation also entails the possibility of having **certificates for website authentication**, as means of users - and in particular, citizens and SMEs - to verify the identity of a legal person behind the website as a source of trustworthy information.

The Regulation sets clear requirements for website authentication certificates which can be taken as cornerstone principles to follow when setting up a data security policy for base registries. It also defines some obligations for providers of such certificates, regarding the security of their operations, their liability and their (light-touch) supervision regime which can be directly implemented by public administrations.

Case example: Approaches to Data security policy in Denmark and Italy

The **Danish e-Boks platform**[62] is an interesting case of data security arrangements relevant for base registries. The idea behind e-Boks is 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.

The solution has been first made accessible to the entire population, with the idea to extend its enforcement to cover the use of data-boxes by companies. In this sense Denmark is one of the few countries that has been able to develop a comprehensive legal and policy framework to effectively support the implementation of public data-related security policy on the relation between citizens and public administrations.

Another example at national level on the governance of base registries can be found in **Italy,** where a national code for digital administrations establishes which databases are of national interest and which administration oversees 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

G12 Ensure and control the quality of the data (Data Quality Policy)

Data quality is essential especially when it comes to base registries data, which are at the basis of digital public services and procedures. The primary goal of an effective data quality policy should be therefore the provision of accurate, understandable, complete and consistent data, ensuring technical support to this end.

To implement a data quality policy, base registries owners are recommended to adopt a **Data Quality Assurance Plan.** Its purpose is to correctly identify procedures, roles, responsibilities, liabilities and workflows.

Such plan should take into consideration both legal and implementation implications:

- **Legal implications**: The importance of data quality is reflected in the legislations and strategies and translated into practical solutions implemented at national European level, often on a multiannual timeframe. Therefore, it is crucial to correctly define 'data quality' in national

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

legislations and define the legal scope and mandate of the national stakeholders dealing with public data management. Such understanding is an essential part of a national strategy related for data management.

- **Implementation implications**: the adoption of increasingly user-centred or business-centred solutions (such as Data Boxes and e-Government personal workspaces) should be regarded as an indicator of effective implementation of an interconnecting solution. Solutions involving citizens, businesses and service providers in the process result advantageous as they facilitate data checks and validation directly by users through secured electronic channels. This ideally contributes to structurally increase the overall data quality.

The drafting of a data quality assurance plan should also be regarded as an occasion and means to check the actual quality of the data before this is shared, targeting both the moment of the registration of new master data as well as the moment in which these are modified.

Case example: Approaches to data quality in multiple European countries

Denmark: When it comes to data quality, in Denmark[63], 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 case the data is incorrect, an update has to follow based on a pre-determined set of rules. For cases of simple mistakes, if stated in the law, citizens can correct them directly by providing their data to entity responsible for the first contacted registry (e.g. at local level the municipality). Subsequently other registries reusing these data are responsible to approve the data inputted by the first registry.

Luxembourg: citizens can contribute to raise the accuracy of data in Luxembourg, thanks to officially authenticated means. In particular, through a **personal space** on the **MyGuichet.lu,** citizens are enabled to directly view their data and submit online requests for corrections when inaccuracies are identified. 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.

Netherlands: the Netherlands is one of the countries where is also **responsibility of the 'data subject'** to ensure that the provided data is correct. This implies an obligation to report a move to another address to the 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).

Belgium: the **OSLO Framework**[64] represents a group of data standards created in Flanders (Belgium), which allows notifications and feedback to be provided for a topic. A user who has a comment on a topic 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: as part of its efforts to control data quality throughout the entire data lifecycle[65], Norway launched the **Altinn**[66] data sharing platform. Data from the public sector can be shared through the portal, providing a **data pre-filling function**, when reporting to the government. About 95% of reporting from businesses to the government in Altinn is represented by structured (machine-readable) data. Thanks to the fact that such data ids frequently checked/validated or updated by the platform users themselves (citizens and businesses), data quality in Altinn remains high over time.

[63] 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

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

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

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

2.3.2. Guidelines on governance structure

A good governance plan should take organisational challenges 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. The following guidelines provide recommendations in this sense.

G13 Envision the global (holistic) organisational picture

In many 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. 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.

Setting the foundations for collaboration between organisations requires the alignment of cross-organisation business processes and smart service orchestration, aiming for seamless interaction and data exchange between distinct systems using standards and common interoperability interfaces.

At methodological level, in the past years the EU developed a range of tools to support the implementation of the programmes in public administrations. For example, **PM²** Project Management Methodology[67] is designed to support the organisational aspects of the implementation, while **Agile@EC[68]** would provide details on how to realise technical implementations. Many base registries owners and policy makers are also familiar with the **ABR Catalogue of Solutions**[69] which has been available for some years and collects examples of base registries interconnecting solutions, structured around these methodologies.

Case example: Organisational structures across Europe

Stelselcatalogus in the Netherlands

The System catalogue (*Stelselcatalogus*[70]), 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.

X-Road approach to roles and responsibilities

In the X-Road ecosystems the different roles, responsibilities and liabilities in the data management process are clearly defined. Each ecosystem consists of an X-Road Operator, Member

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

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

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

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

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. Trust Service Providers are organisations in charge of providing two key services for the functioning of a trusted X-Road ecosystem, namely: 1) a time-stamping authority (TSA) and 2) a certification authority (CA).

G14 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.

The following represent typical categories of agreements that can be put in place to formalise the involvement of multiple parties into interoperability initiatives:

- **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 encryptions 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 diverse 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.

Case example: interoperability agreements at national and European level

National bilateral agreements between - Estonia and Finland in the X-Road example

At national level, the existing cross-border **bilateral agreement between Estonia and Finland** illustrates the development of a **joint data exchange platform**[71] in order to make digital services mutually accessible for inhabitants, by reusing existing national infrastructure (i.e. the Estonian X-Road[72]). In practical terms, this means that the data kept in base registries (tax boards and social

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

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

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.

Interoperability agreements in the context of EU initiatives

Interoperability agreements are usually modelled on **templates**. One project that is providing simple, but practical templates, is the **Centre of Excellence for Information Sharing**[73] **(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, a relevant example is the **European Business Registers (EBR)** organisation (and platform) which is based on the cooperation between the participating registries on an **Information Sharing Agreement**[74]**.** In this one in 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 in the design of the **Business Registers Interconnection System (BRIS)**[75].

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)**[76] developed by the European E-invoicing Service Providers Association (EESPA) for the transmission and processing of electronic invoices and other business documents.

2.3.3. Guidelines on Change Management

One of the main challenges to building successful interconnections of base registries is to ensure that the solutions in place are flexible and capable of adapting to changes, be them planned and organised or deriving by exogenous factors connected to changes in the legal or political framework or in the available budget. To this end, a comprehensive change management plan is useful tool to define the procedures and processes needed to steer, control, implement and monitor changes.

The overall purpose is to ensure that the information services delivered to other public administrations, businesses, and citizens through base registries are constantly evolving to address changing needs, while keeping highest accuracy, and reliability of data and systems. In this context it is also crucial to maximise the awareness on governance structures and processes across all levels of public administrations.

G15 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.

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

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

[75] BRIS: https://e-justice.europa.eu/content_business_registers_at_european_level-105-en.do

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

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 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 of 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.

Case example: National change management plans – Base registers in Belgium

In **Belgium**, business registries offer most of 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 Business Register**[77] (KBO BCE), the datasets open at the national level at **data.gov.be portal**[78], and the **Open Access Belgium**[79] 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.

G16 Implement and release change requests

During the deployment and evolution of base registries interoperability initiatives, the **semantic assets**[80], services or tools may change due to requests issued by the community of users or by the governance bodies themselves. This aspect 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 **two-fold management plan** with the following main objectives:

1. manage **the change requests**, identifying priorities, magnitude of the intervention and assessing impacts
2. put in place a **release management methodology,** to be implemented through a **release planning** which should ensure continuity but also timely introduction of the required changes

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 (*consult section 2.1.1 on 'Data Governance' for more information*).

Studies on the governance and maintenance of specifications, conducted under the former ISA² Programme, include examples of implementing change requests. One of these studies is **"A change management release and publication process for structural metadata specifications**[81]", and

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

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

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

[80] W3C deinfition of semantic assets (as object of the ADMS profile) https://www.w3.org/TR/vocab-adms/#:~:text=ADMS%20is%20a%20profile%20of,used%20for%20eGovernment%20system%20development.

[81] Link to document 'Description of a change management release and publication process for structural metadata specifications'

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**[82]".

In the regular operativity and maintenance of public services, the recommendation is to make sure that the governance model, maintenance policy, and all related plans are thoroughly understood across all levels of the public administration. Thus, it is recommended to regularly conduct feedback and review cycles involving all actors participating in the management of base registries, to timely identify whether a policy is flexible enough and fit for the real needs.

Case example: X-Road approach to change and innovation management

X-Road supports a **research-focused change management approach**. that in Finland and Estonia relies on collaboration projects with two universities (University of Tartu and Tallinn University of Technology).

These conduct research on the expansion of X-Road communication capabilities. For instance, an ad-hoc research project was launched in order to study the possibility of extending X-Road messaging capabilities – based to on synchronous data exchange (via request-response message pairs), to cover asynchronous one-to-many messaging.

This need emerged from feedback received from users, so the integration between X-Road and existing open-source messaging solutions was analysed in order to rapidly develop a solution that could meet the newly emerged needs.

2.3.4. Guidelines on 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.

G17 Ensure digital preservation and permanent access to data

Ensuring data availability over long periods of times is not a minor challenge. This includes a number of activities that can be summarised by three key concepts:

- **Technical continuity** – correspond to the concept of keeping data always accessible and readable over time, which is particularly important for public data. In the case of base registries, it may be a significant challenge since base registry data is normally kept for extremely long periods of time (or even permanently).
- **Data disposition** – it corresponds to the process of deleting or archiving electronic data that is no longer needed. Normally public administrations archive data in digital archives which also require to be maintained and kept accessible if needed.
- **Data preservation**[83] – relates to ensuring the authenticity and validity of the data. Base registries should also ensure that the original values of the data are not lost as a result of regular

[82] Link to document 'Methodology and tools for structural metadata management and governance for EU Institutions and Member States'

[83] Data preservation is also one of the key principles identified by the EIF in its list of Underlying principles of European public services (Underlying principle 10: Preservation of information)

operations (business continuity) and their administrators should put in place all the necessary ICT arrangements to this end.

As a general rule, mostly related to GDPR - it is relevant to mention that preserving data which ceased to have any legal or informative value can be costly, unnecessary and generate legal non-compliance. Normally this is not the case of base registries, nonetheless the competent authorities should monitor whether the data kept in their registries could be affected by any regulation and make sure to be compliant by establishing the necessary organisational and technical means.

Case example: Digital preservation approaches

Examples of recommended consolidated standards related to the preservation of base registry data are **PREMIS**[84], 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**[85] (Write Once Read Many) solutions can be studied, that assure that data cannot be tampered with once written to a device.

G18 Agree on flexible data availability levels

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, since (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;
- 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**[86] 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**[87], describing the technical requirements expected from all the registries and the ones expected from the services offered by the central platform.

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

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

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

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

3 Guidelines for semantic interoperability

This chapter focuses on the design and simplification of business processes. It 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.

.

3.1 Data models and master data management

Even though several initiatives, including some large-scale pilots[88], have demonstrated that even in broad and complex fields (e.g., public procurement, health, finance, and justice) information can be represented as structured data and shared digitally, it is not uncommon to still encounter scepticism across public administrations and national authorities over the idea that information can potentially always be represented as structured data.

In fact, in several domains the competent authorities and policy owners authorities still heavily rely on non-structured data (i.e., digitised documents such as pdfs or digitised images), if not physical documentation. 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).

In this context, the digitisation and structuring of the base registries' information is key to facilitate the interoperability and maximise public data accessibility and reusability. The structured data produced will then need to be properly managed, to secure its quality, validity, authenticity and ultimately its availability.

To this end, the following concepts are crucial to be introduced:

- **Master Data** - key assets of any base registry and for the delivery of public services. These represent the key objects and single point of truth for any public service transaction or procedure and are at the foundation of base registries (it can be lists of natural or legal persons, lists of vehicles, lists of businesses and associations, lands/parcels/administrative divisions, etc.).
- **Master Data Management (MDM)** - a "unified" approach to deal with base registries, their data lifecycle and the public administrations' services that need to interoperate with them. MDM is focused on consistency and quality of master data and includes all the processes and tools that are functional to that.
- **Master Data Tools** - proprietary commercial instruments, useful to manage Master Data. Normally they help 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.

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[89]. 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.
- **Data Quality*** (DQ) – a set of related data-management techniques and business-quality practices aimed at ensuring 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 (*additional guidance and examples provided in Chapter 2 on Governance*).

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

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

3.2 Guidelines – Data models

This section provides a number of guidelines on base registries data models and semantic aspects, divided in two main blocks:

- Guidelines on Master Data Management (MDM)
- Guidelines on semantic data models and standards

G19	Define an MDM style
G20	Define data types and their management approach
G21	Identify unique and unambiguous instances of your master data
G22	Define the data domain
G23	Distinguish scope and use of metadata
G24	Define semantic assets of (master) data
G25	Reuse semantic assets: Ontologies and taxonomies
G26	Reuse semantic assets: Core Vocabularies
G27	Reuse semantic assets: Standard Application Profiles
G28	Publish data as Linked Data

3.2.1 Guidelines on Master Data Management

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. Therefore, more frequently, Member States are considering the adoption of MDM to manage their data lifecycle.

The guidelines on Master Data Management relate not only to semantic aspects, but actually touch upon technical arrangements. However, the two dimensions are closely related and the technical arrangements affect the semantic aspects, hence the reason for discussing both in this section.

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.

G19 **Define an MDM style**

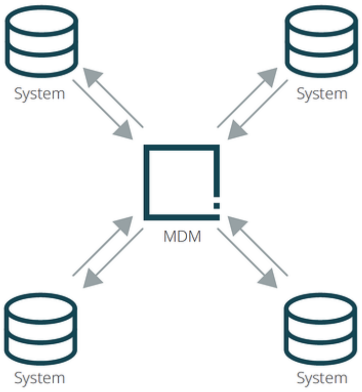
There are different MDM styles that a public administration can choose from. The choice depends on whether it needs to have **a central hub** to manage its data, or to **synchronise it with existing**

sources[90]. 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 more commonly used styles**[91] are the following:

- Registry Style
- Consolidation Style
- Co-existence Style
- Transaction / Centralised Style

Public administrations in different countries may adopt different styles or combinations of styles, based on the local needs and legal requirements[92]. 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.

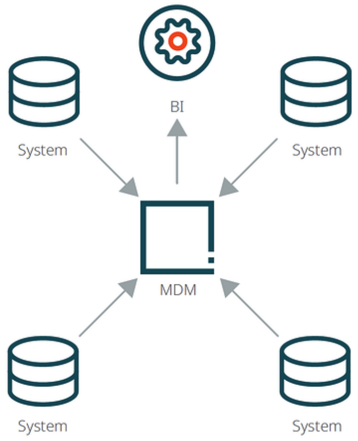
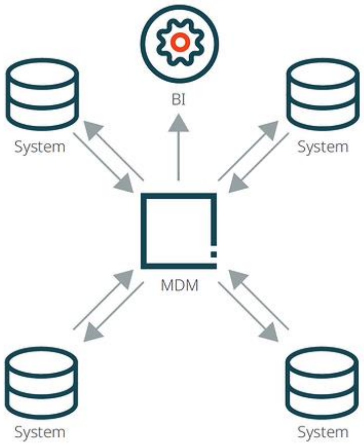
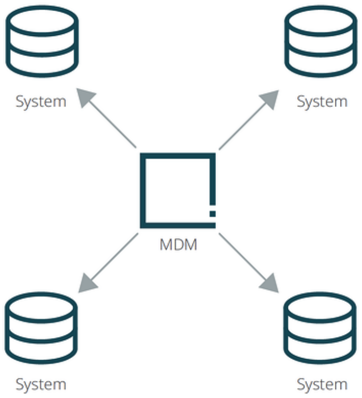
MDM Style	Advantages	Disadvantages
Registry style  *Source: StiboSystems*[93]	● Solves the complexities that are related to having a large number of source systems, each with its own rules (difficult to be modified) ● Selects the 'best' version: the central registry screens data and runs a cleansing and matching algorithm which assigns a unique global identifier and finally establish the single soure of truth ● Low-cost / easier way of implementing MDM	● Data from source systems does not change (not updated) ● Can be difficult to establish an authoritative source ● Data latency *(same as in the Consolidation style)*

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

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

[92] 4 Main Master Data Management Implementation Styles: https://www.reltio.com/blogs/4-main-master-data-management-implementation-styles/

[93] The 4 most common master data management implementation styles: https://www.stibosystems.com/blog/4-common-master-data-management-implementation-styles

MDM Style	Advantages	Disadvantages
Consolidation style  *Source: StiboSystems*	Besides the ones from the Registry style: ● the stewardship capability is available in the MDM hub	● Data from source systems does not change (not updated) ● Can be difficult to establish an authoritative source ● Data latency *(same as in the Registry style)*
Co-existence style  *Source: StiboSystems*	Besides the ones from the Consolidation style: ● Real-time synchronisation between MDM hub and sources by sending back the golden record to each respective source systems ● Significant improvement in master data quality by updating master data in source systems and MDM hub ● Possibility for incremental expansion (progressively add more domains)	Besides the ones in Registry style: ● Synchrnonisation issues might arise ● The sources have data cleansing capabilities to maintain consistency with the hub ● More expensive to deploy than the other MDM styles
Transaction / Centralised style  *Source: StiboSystems*	● The central component/system becomes the single provider of master data (works as sort of repository) ● The updates are done directly in the central system and then distributed to the others) - any systems outside the central system can no longer be allowed to create or amend the master data	● Demanding on resources and potentially costly ● Time required for the style to be implemented

Case example: MDM styles definition in Scandinavia and at EU level

For example, in **Denmark**[94] 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**[95], 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**[96], 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**.

At **EU level**, the **hybrid search**[97] 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.

G20 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 considers 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**[98] and **ESPD (European Single Procurement Document**[99] **Service)** are perfect candidates to illustrate how the "Once-Only" Principle can be implemented based on master data and reference data that are

[94] Information from interview with Danish public authorities, available here: https://joinup.ec.europa.eu/collection/access-base-registries/news/interview-highlights-denmark
[95] Information from interview with Swedish public authorities, available here: https://joinup.ec.europa.eu/collection/access-base-registries/news/interview-highlights-sweden
[96] Information from interview with Norwegian public authorities, available here: https://joinup.ec.europa.eu/collection/access-base-registries/news/interview-highlights-norway
[97] [BR02] Architecture approach - Hybrid, [BR03] Motivation of the architectural approach on ABR Catalogue of Solutions on Joinup.
[98] EUR-LEX : http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.L_.2014.094.01.0065.01.ENG
[99] ESPD : https://ec.europa.eu/isa2/solutions/european-single-procurement-document-espd_en

exchanged among cross-border and cross-sector base registries (i.e., social security and tax agencies, business registries and BRIS, service providers, etc.).

Case example / best practice

Data management approach from X-Road in Estonia and Finland

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. Despite having Once-Only as the guiding principle, X-Road does not impose it and, moreover, it 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 OOTS 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.

Data management approach from MyGuichet in Luxembourg

Another interesting frontend **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 principle (OOP)** 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.

Technical Interoperability Norms (NTI) in Spain

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**[100]").

G21 Identify unique and unambiguous instances of your master data

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

[100] 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).

- **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 -** concerns 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.

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 following represent basic recommendations concerning the **use of identifiers**:

- When designing exchange data models, **consider using multiple identifiers**[101] 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 URI to it, which also facilitates the description of the entity being identified.

The challenge of uniquely (and universally) identifying entities is common to 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[102] funded by the EC and is quite similar to the IBAN solution adopted to uniquely identify bank accounts.

One way of ensuring the long-term survival of identifiers is to assign them to **Persistent Uniform Resource Identifiers (PURIs**). SEMIC recommends the use of PURIs[103] and defines principles and practices for their use. One interesting example is the **EULF**[104] 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.

Case example / best practice

Multiple Identifiers - OASIS Universal Business Language TC

For an example of the usefulness of providing multiple identifiers, one can study how the **OASIS Universal Business Language (UBL) TC**[105] 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).

[101] Study on 10 rules for persistent URIs, Joinup: https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/document/10-rules-persistent-uris

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

[103] SEMIC PURIs: https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/document/10-rules-persistent-uris

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

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

3.2.2. Guidelines on semantic data models and standards

The goal of semantics is developing a common meaning across public administrations through adaptation and implementation of common data models. 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** and refer to them when proceeding to create application profiles.

SEMIC definition of semantic assets reuse

Specification reuse is the act of sourcing an established specification into a new one (referred to as the "current" or "own" specification).

Technically, reuse may mean either or both of the following:

- Importing (using owl:imports) the contents of another ontology or data shape specification into a current one

  OR
- relying on URI dereferencing to get the formal definition of the ontology (this, however, does not work for the data shapes) In the SEMIC context, for the purpose of interoperability, we need to carefully constrain the meaning of the term "reuse" with a fine-grained description on a construct-by-construct basis, and this is the aim of the remainder of this section.

The following guidelines focus on semantic best practices and the reusre of existing semantic assets a starting point to model information or customise data models.

G22 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

G23 Distinguish scope and use 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[106]. 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 National Information Standards Organisation (NISO)[107] 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 the next subsection).

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.

When possible, the recommendation is to **reuse code lists** that are maintained by international or **European Standardisation Organisations (ESOs**[108]**)** officially recognised by the EU and EFTA. 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.

Case example: European Criminal Records Information System (ECRIS)

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.

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

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

[108] DG GROW List of key ESOs: https://single-market-economy.ec.europa.eu/single-market/european-standards/key-players-european-standardisation_enhttps://www.cenelec.eu/aboutcenelec/whoweare/europeanstandardsorganizations/index.html

The ESOs that maintain code lists are **CEN**, **CENELEC** and **ETSI**. They do so, based on the legal mandate provided by Regulation (EU) 1025/2012 on European standardisation also known as 'Standardisation Regulation'. , which settles the legal framework for standardisation. The three ESOs provide their assets and code lists online in open format[109].

G24 Define semantic assets of (master) data

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 following table.

Semantic Asset	Short description	Recommendation
Glossary *Source: W3C*	A glossary is an alphabetical list of words, phrases, and abbreviations with their definitions. Glossaries are most appropriate when the words, phrases, and abbreviations used within the content relate to a specific discipline or technology area.	Use glossaries as a mechanism for locating content within a set of Web pages and for harmonising the understanding of key concepts.
Folksonomy *Source: Gartner*	Folksonomies also known as 'social tagging', are user-defined metadata collections. Users do not deliberately create folksonomies and there is rarely a prescribed purpose, but a folksonomy evolves when many users create or store content at particular sites and identify what they think the content is about.	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.
Controlled vocabulary *Source: EU Publication Office*	In library and information science, are standardised and organised arrangements of words and phrases presented as alphabetical lists of terms or as thesauri and taxonomies with a hierarchical structure of broader and narrower terms.	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).

[109] For example, in the CEN search engine or in the ETSI search engine

Semantic Asset	Short description	Recommendation
Taxonomies and Thesauri *Source: datos.gob.es and W3C*	A Taxonomy is a controlled vocabulary, where the terms are organized in a hierarchical way (with a tree structure), from the most general to the most specific terms, including those that are related. A thesaurus consists of concepts, labels for concepts, and semantic relationships between concepts. Most thesauri use a similar set of semantic relationships, which they label 'broader' 'narrower' and 'related'. Differently from Taxonomies, Thesauri focusing on terms and their immediate relationships with other terms more than following a hierarchical structure.	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.
Metadata schema *Source: ISO 23081*	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.	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.
(Core) Vocabulary *Source: W3C.*	Vocabularies define the concepts and relationships (also referred to as “terms”) used to describe and represent an area of concern. Vocabularies are used to classify the terms that can be used in a particular application, characterize possible relationships, and define possible constraints on using those terms	Use common core vocabularies to ensure that the core elements of your datasets are expressed in a standardised and unique way, aligning understanding of the key metadata across different administrations.
Ontology *Source: SEMIC Style Guide*	An ontology, also referred to as ontology specification, is a a formal specification describing the concepts and relationships that can formally exist for an agent or a community of agents (e.g. domain experts). It encompasses a representation, formal naming, and definition of the categories, properties, and relations between the concepts, data, and entities that substantiate one, many, or all domains of discourse.	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.
Application Profile *Source: SEMIC Style Guide*	An Application Profile is a data specification to facilitate the data exchange in a well-defined application context. It re-uses concepts from one or more semantic data specifications, while adding more specificity, by identifying mandatory, recommended, and optional elements, addressing particular application needs, and providing recommendations for controlled vocabularies to be used	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. In short: use application profiles when you want to apply a metadata schema in your organisation.

G25 Reuse semantic assets: Ontologies and 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. To this end, standard **classification taxonomies** are recommended to be used along with metadata schemas. (*For a definition of 'Ontologies' and 'Taxonomies' please refer to G24 'Define semantic assets of (master) data'*)

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**[110]), 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.

In addition, the EU Publications Office maintains the **EuroVoc**[111] **-** a multilingual and multidisciplinary **thesaurus** with domains and sub-domains that describes topics of legal documents – or domain-specific frameworks such as the **European Legislation Identifier** (**ELI**[112]) ontology.

Ontologies constitute the formal (machine-readable) definition of concepts and should be kept as simple as possible. In accomplishing their function of declaring the classes, properties, datatypes and controlled lists, each element (construct) should be assigned a URI and complemented with human-readable labels and descriptions. This allows to establish common references for humans and machines[113].

Case example: Web ontology infrastructure with FinnONTO in Finland

In **Finland**, a major research initiative, namely, the **National Semantic Web Ontology Project** (**FinnONTO)**[114], 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-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,it is maintained in a distributed fashion by expert groups in different domains and it is provided as a national centralised service.

G26 Reuse semantic assets: Core Vocabularies

[110] NALs in Metadata registry of OP: https://op.europa.eu/en/web/eu-vocabularies/authority-tables
[111] EuroVoc: https://eur-lex.europa.eu/browse/eurovoc.html
[112] ELI: https://op.europa.eu/en/web/eu-vocabularies/eli
[113] SEMIC Style Guide - https://semiceu.github.io/style-guide/public-review/index.html
[114] FinnONTO: https://seco.cs.aalto.fi/projects/finnonto/

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), SEMIC provides the **eGovernment Core Vocabularies**[115]**.** (*For a definition of 'Core Vocabulary' please refer to G24 'Define semantic assets of (master) data'*)

The core vocabularies represent 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.

- **Core Person Vocabulary** captures the fundamental characteristics of a person, e.g. name, gender, date of birth, location.
- **Core Business Vocabulary** 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 Vocabulary** captures the fundamental characteristics of a location, represented as an address, a geographic name or geometry.
- **Core Criterion and Core Evidence Vocabulary** 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 Event Vocabulary** captures the fundamental characteristics of a public event, its time, audience, location, etc.
- **Core Public Service Vocabulary** captures the fundamental characteristics of a service offered by public administration. Such characteristics include e.g. the title, description, inputs, outputs, providers, locations of the public service.
- **Core Public Organisation** describes public organisations with links to descriptions of public services, members of staff or other resources such as relevant legislation, policies and jurisdictional coverage in the European Union.
- **Core Assessment Vocabulary** represents and defines what an "Assessment" of "assets" is and how to perform the assessment based on "Criteria". It is a domain-agnostic vocabulary, meaning that it can be used to assess any type of assets.
- **Core Standards and Specifications Vocabulary** is used for the information exchange related to standards and specifications amongst software solutions.

Some Large-Scale Pilots and Trans-European Systems data models, like **e-Codex**[116], **BRIS** and **EULF**[117] are also inspired by the Interoperable Europe Initiative (former ISA Programme's), Core Vocabularies. An example to follow on developing, customising or extending a core vocabulary is provided by the EU "**Handbook for using the core vocabularies**"[118].
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:

[115] SEMIC eGovernment Core Vocabularies: https://joinup.ec.europa.eu/collection/semic-support-centre/specifications

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

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

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

- Design of new data models that extend the Core Vocabularies by using the latter as building blocks
- 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

Case example: Reuse of core vocabularies in Belgium – OSLO

A good example of reuse of Core Vocabularies with successful modelling of the data domain and definition of core concepts is the **OSLO**[119] project **(Open Standards for Linked Administrations in Flanders).** OSLO started in 2012 and is facilitated by a working group of ICT experts from local, regional and federal public administrations and ICT service providers. The project aim is 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 SEMIC eGovernment Core vocabularies. OSLO Vocabularies are simplified, reusable, and extensible specifications and serve as the starting point for developing interoperability across government systems as they allow mappings with existing data models. This helps public administrations attain cross-border and cross-sector interoperability.

Additionally, OSLO also extended DCAT-AP (see Guideline 27). 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. Currently, OSLO contains over 190 configurated standards in which more than 3300 terms are defined by more than 470 contributors[120]. One of the most interesting aspects of the OSLO domain models is the modelling of persons, organisations, and roles.

G27 Reuse semantic assets: standard Application Profiles

To make individual base registries interoperable and harmonise existing registries of registries[121], a variety of tools can be employed, including **standard** and **internationally recognised Application Profiles** for describing public services. (*For a definition of 'Application Profile' please refer to G24 'Define semantic assets of (master) data'*)

Currently, the two main application profiles available and supported by SEMIC are the **DCAT Application Profile for data portals in Europe** (**DCAT-AP**[122]), and the Core Public Service Vocabulary Application Profile (**CPSV-AP**)[123]. Both the data models and the vocabulary are based on recognised schemas and ontologies, such as the **eGovernment Core Vocabularies**, **EUROVOC**[124], and **NUTS**[125](Nomenclature of Territorial Units for Statistics). In addition, the ABR action had also developed a specific application profile for base registries, as a DCAT-AP extension for describing base registries, their contents, and the services they provide. The purpose was to offer a solution to represent master data in base registries.

[119] OSLO description, Joinup: https://joinup.ec.europa.eu/collection/oslo-open-standards-local-administrations-flanders/about

[120] OSLO standaardenregister: https://data.vlaanderen.be/

[121] 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

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

[123] 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

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

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

Validation support for DCAT-AP is available through the Interoperability Test Bed[126]. 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 TestBed has launched a new validator for SHACL shapes[127], 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.

Case example: National Data Catalogue – Norway

In **Norway**, the **National Data Catalogue[128],** 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**[129], 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**[130] , the Belgian profile **DCAT-AP-BE**[131] and the Swedish and Norwegian profiles **DCAT-AP-SE**[132] and **DCAT-AP-NO[133]**. The last is aligned with DCAT-AP 2.0.0, DCAT 2.0, and BRegDCAT-AP v2.00.

G28 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.

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

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

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

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

[130] 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

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

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

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

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[134].

The following four design principles represent the foundations of Linked Data:

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 rules of thumb from the **W3C Linked Data Cookbook**[135]:

- Model the data
- Reuse vocabularies whenever possible
- 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 Interoperable Europe initiative provides good practices and practical examples to help public administrations apply Linked Data technologies to e-Government. Specifically, the **Semantic Interoperability Community (SEMIC)** 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[136]. 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)[137], 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).

Case example: Implementation of Linked Data across Europe

Linked Base Registry in Belgium

[134] 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

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

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

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

In **Belgium,** the **Flemish government** with **OSLO²** (Open Standards for Linked Organisations)[138] 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.

Dutch Addresses and Buildings key register in the Netherlands

In the **Netherlands**, the **Dutch Addresses and Buildings key register (BAG)[139]** is published as linked data[140]. 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.

Felles Datakatalog with Linked Data paradigm in Norway

The Felles Datakatalog **in Norway** adopts the **Linked Data paradigm** as a basic building block. It implements a comprehensive specification known as DCAT-AP-NO[141] designed to serve as the foundation for the exchange of datasets and data services. Moreover, the choice and implementation of well-known standards (i.e., SKOS, DCAT) results in a platform that is already aligned in a significant way with the BRegDCAT-AP specification.
There is also a **SPARQL endpoint** that allows querying the RDF datasets on the platform. It is important to highlight that, although query results are presented in a user-friendly visual component, results can also be downloaded in both JSON and CSV serialisation formats for further processing on the user's side.
Although not explicitly mentioned or linked in the user-friendly section of the platform, all entities (e.g., datasets, concepts) are modelled as **RDF documents serialised in Turtle and publicly exposed on the Internet**. It is also useful to note that, thanks to the strict Linked Data implementation, machine-readable dataset distributions can be easily identified and downloaded.

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

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

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

[141] https://data.norge.no/specification/dcat-ap-no/

4 Guidelines for technical interoperability

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.

4.1 Data infrastructures interconnectivity

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.

4.1.1 Base Registries Interconnecting Framework

A solution to overcome these challenges is to create and implement an interconnecting infrastructure[142] that allows secure exchange of data between different base registries and enables the reuse of data for public services, as also described in the Base Registries Interconnecting Framework document (BRAIF)[143].

The implementation of the interconnection platform is essential, as it acts as the intermediary, allowing the exchange of data between different base registries and enabling the reuse of data for public services. In practice, such interconnectivity is often already established in Member States with the creation and implementation of platforms that enable data access supported by APIs.

Example - Malta's national interconnecting infrastructure

Malta represents an interesting example of concrete steps towards the creation of a base registry interconnecting infrastructure. In Malta, a national data infrastructure is made up of:

An authorisation and representation platform;
A foundation data layer;
A metadata portal (Registry of Registries);
A national data portal.

Based on the "PSI Directive Transposition and Implementation" new Legal Notices were implemented in the country (on the Registry Authority, Implementing Entity, Person Register, Business Register, etc.). Based on these new requirements, a National Data Strategy was launched. Currently, Malta's national data portal is a one-stop-shop for data discovery and consumption, and the main channel for requests for data. Maltese authorities have scanned around 4000 legal instruments and identified that they had around 1000 registered names, which have been subsequently verified, to remove duplicates and ensure accuracy, considering language aspects (names are in Maltese and English). Subsequent steps included the establishment of a link between a registry and the government sector or function, aligning with DCAT-AP; the establishment of base registries ownership and the classification of data in each base registry, allowing owners to decide whether the degree of data openness in the registry.

In practice, as also mentioned in the BRAIF, interconnecting infrastructures (e.g. based on APIs) work as harmonised interfaces allowing for data standardisation, management and control. The result is facilitated data exchange making it[144]:

- **Verified and certified**, where both the sender and receiver have been identified and authenticated through agreed mechanisms
- **Encrypted**, where the confidentiality of the exchanged data is secured

[142] 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

[143] BRAIF,v.2.00:https://joinup.ec.europa.eu/sites/default/files/document/2020-06/SC380_D04.01_Framework%20for%20Base%20Registry%20Access%20and%20Interconnection_v2.00.pdf

[144] ibid

- **Logged**, where the electronic records are logged and archived to ensure a legal audit trail.

4.1.1 Interoperability through APIs

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. 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.

Member States implement various approaches to ensure interoperability on national level and some have already implemented cross-border collaborations with the usage of APIs, with several virtuous implementation examples that can be reused, as mentioned above, Estonia and Finland on secure data exchange via X-Road, also **Nordic-Baltic e-ID cooperation**[145], etc.

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 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[146].

In addition to the above, the **OOTS[147] (Once-Only Technical System)** is a key initiative aiming at making it possible for citizens, organisations and companies to provide information – such as an address – to the authorities **only once**. The design of the project dictates the storage of provided information in a single repository (base registry). Through this base registry all hosted information can be shared between authorities that have access to it. Usage of APIs fully supports implementation of this principle via its features. This is part of the broader EU activity of implementing more standards and digital solutions to achieve interoperability. In particular, through a common **European Data Strategy[148]** 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.

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**[149]**,** 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[150]. 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.

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[151].

[145] https://www.difi.no/fagomrader-og-tjenester/digitalisering-og-samordning/nordisk-samarbeid

[146] One of the studies on APIs approach: JRC Technical Report

[147] OOTS: https://ec.europa.eu/digital-building-blocks/wikis/display/OOTS/About+OOTS

[148] https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en

[149] https://ec.europa.eu/digital-single-market/en/european-legislation-reuse-public-sector-information

[150] Estonia's National AI strategy: https://e-estonia.com/nationa-ai-strategy/

[151] An example of such a data exchange mechanism that has implemented actions related to the above is the 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.

4.2 Guidelines – Data architecture

This section includes a number of guidelines on technical aspects and data architecture, to facilitate and enhance base registries interoperability.

G29	Choose a data architecture model adapted to your organisational model
G30	Reuse data architectural approaches on data exchange platforms
G31	Use common testing tools to ensure for interoperability conformance
G32	Enable data access supported by APIs
G33	Develop specific strategies to steer APIs implementation

G29 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[152]**.**

Clearly, 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 *(see Section 3.2.1 on Guidelines on Master Data Management)*. When choosing a data architecture model, what is also important is to answer the following questions:

- on **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?
- on **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.

Case example: X-Road

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**[153].

[152] See e.g. the case of Spain, which established a platform for base registries interconnection, based on the four dimensions of interoperability (governance, legal, semantic and technical) - Spain Factsheet on Joinup: https://joinup.ec.europa.eu/sites/default/files/inline-files/Spain%20Factsheet%20Final.pdf

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

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 (Estonian data exchange layer) and the Suomi.fi (Finnish 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**)[154], 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 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.

G30 Reuse data architectural approaches on data exchange platforms

The European Union has led a number of initiatives aimed at harmonising the approach to the adoption of data architectural approaches supporting interoperability. Although the set-up of data exchange platforms is a key enabler for public sector digital transformation, this task carries budgetary and organisational costs, and important technical challenges (as well as security, legal and governance issues). The EU aims to create a set of good practices both in aligning the technological set-up with policy objectives and in adopting best practices for the design of platforms (such as APIs), from which Member States can benefit. In particular the following should be taken into account for guidance:

- **European Interoperability Reference Architecture (EIRA)**[155]**;**
- **European Interoperability Framework (EIF)**[156]**;**

One of the key architectural approaches acting as a supportive measure is the **European Interoperability Reference Architecture** (**EIRA**), which offers a global and exhaustive analysis of all interoperability 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:

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

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

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

- **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**[157]: 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 (TES) 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).

Case example: base registries interconnection projects

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[158] (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[159] (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. It provides access to over five hundred different services that are exposed in the form of web services, file extractions and events on behalf of over twenty different registries.
- **MAGDA[160] (Belgium):** A service-oriented data exchange infrastructure for accessing base registries of citizen and enterprise data, at regional, local and federal levels (where applicable). The platform provides access to base registries of citizen and enterprise data, harnessing reusable technologies that can be easily adapted to the needs of different government administrations, from the regional to the local level, and increasingly to the federal level.

[157] 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.

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

[159] Data Distributor (Denmark): https://datafordeler.dk/vejledning/

[160]MAGDA (Belgium): Magda factsheet in Joinup: https://joinup.ec.europa.eu/sites/default/files/inline-files/BE01%20Magda_0.pdf

G31 Use common testing tools to ensure for interoperability conformance

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
- 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)**[161], developed by the EC. This solution provides "general 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 channel
- 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.).

G32 Enable data access supported by APIs

Another data architectural approach concerning data exchange is represented by the implementation of interconnection platforms and systems 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.

One of the biggest challenges that Member States face in developing an API 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.

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

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[162.]

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.

Case example: Implementation of APIs across Europe

The 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.

France: the French government[163] 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 (dubbed **France Connect**[164]) offer citizens composite digital services, regardless of which agency offers the service.

Ireland: another good example adapted to the logic of using API catalogues and base registries is the **Data Architecture strategy**[165] in Ireland. It 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 which can be accessed through appropriate APIs.

Luxembourg: Luxembourg[166] is 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.

Belgium: for the effective set-up of APIs, design guidelines can be very helpful for both API service providers and consumers of these services, also to support the development of aspects that relate to API performance, versioning, language and errors handling. To this end, 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[167].

G33 Develop specific strategies to steer APIs implementation

The concept of 'API strategy' refers less to technology itself and more to the development of a community of service or data providers – often referred to as "ecosystem" by the strategists in question – that can be accessed through a central platform.

[162] 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
[163] French basic registry: https://api.gouv.fr/
[164] France Connect, a citizens SSO service: https://franceconnect.gouv.fr/nos-services
[165] Data Architecture Strategy (Ireland): https://ec.europa.eu/jrc/sites/jrcsh/files/17_06_mark-warren-ireland.pdf
[166] Interview, October 2020: https://joinup.ec.europa.eu/collection/access-base-registries/news/highlights-interview-luxembourg
[167] Rest Guidelines for building services in Belgium: https://www.gcloud.belgium.be/rest/

The main function of an API strategy is therefore the definition of design parameters and developer-focused criteria for creating an API. 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. 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.

'API-first' approach

The API-first development is another API strategy design approach 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. While an API strategy Iindicates the overall approach to achieve seamless and secure connectivity, an API design would use those objectives as an API is planned.

The preparation of API strategies objectives and target focus must also closely consider key legal texts with reference to specific policies. In particular, one of these is the Open Data Directive which acknowledges the strong emphasis that some MS have placed 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[168].

In addition, strategies should be forward-oriented and take into account upgrades and changes in the initially intended design related to possible developments on both the technology and users sides, which could have an impact on the technical set-up.

Case example: Dutch OAS and X-Road's communication protocols

Dutch API strategy implementation

In relation to API strategies, it is worth mentioning the Dutch implementation[169]. The Netherlands had to deal with two major problems with the use of public services accessed through an API, namely i) the description of services was made through heterogeneous techniques and ii) the various services were scattered in the infrastructure of each host public organisation. To have a single way to describe all APIs and a central point for sharing their descriptions the **Open API Specification**[170] (**OAS**) was adopted: 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. In addition, the Netherlands, to support their API strategy, created 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[171].

Developments in the X-Road communication protocols

The **X-Road ecosystem**, Initially chose to offer its services based on top of **SOAP** (Simple Object Access Protocol) which at that time is the de facto standard for web service communication protocols. SOAP based APIs have 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**

[168] EU open data, The basics for EU data providers: https://op.europa.eu/en/publication-detail/-/publication/c631a6de-ecd5-11e5-8a81-01aa75ed71a1

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

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

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

(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[172].

[172] X-Road Rest support: https://www.x-tee.ee/docs/live/xroad/pr-rest_x-road_message_protocol_for_rest.html

European
Commission

Directorate-General for Informatics

Directorate B – Digital Services

B2 – Interoperability

interoperable
europe

SEMIC