European Commission

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

D04.01 Framework for Base Registry Access and Interconnection

ISA² Action 2016.28: Access to Base Registries

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

Date: 30/11/2020
Doc. Version: 3.00

DOCUMENT CONTROL INFORMATION

Settings	Value
Document Title:	SC380_D04.01. Framework for Base Registry Access and Interconnection
Project Title:	Specific Contract n° 380 under Framework Contract n° DI/07624 - ABCIV Lot 3: ISA² - Action 2016.28: Access to Base Registries
Document Author(s):	Emmanouil Alivizatos Ismini Savvala Axel Haumont Ksenia Bocharova Petro Dudi
Project Owner:	Natalia Aristimuno Perez
Project Manager (PM, European Commission):	Peter Burian
Contractor's Project Manager (CPM):	Ksenia Bocharova
Doc. Version:	3.00
Sensitivity:	Public
Date:	30/11/2020

Revision History

Date	Version	Description	Author(s)	Reviewed by
29/04/20	0.01-0.02	Update of the document structure, review of Executive Summary and Data Governance section	Petro Dudi	Ksenia Bocharova
30/04/20	0.03	Submittion to PO for review	Ksenia Bocharova	Peter Burian
18/05/20	0.04	Technical review of the document	Petro Dudi	Ksenia Bocharova
19/05/20	0.05	Update of Introduction and add reference to European Data Strategy and	Ksenia Bocharova	-
26/05/20	0.06	Technical quality peer review by data management team	Reviews conducted by: Evangelos Nittis Nikolaos Kollaros	Ksenia Bocharova
27/05/20	0.07	Implementation of feedback from peer reviews	Ksenia Bocharova	-
27/05/20	0.08	Quality review	Ludovic Mayot	-
28/05/20	1.00	Submittion to PM for review and approval	Ksenia Bocharova	Peter Burian

08/06/20	1.01-1.02	Incorporation of comments	Ksenia Bocharova Emmanouil Alivizatos Ismini Savvala	-
09/06/20	1.03	Review of the comments and incorporated text by Technical Writer	Petro Dudi	Ludovic Mayot
10/06/20	1.04	Confirmation of the implemented comments with PM.	Ksenia Bocharova	Peter Burian
10/06/20	2.00	Publication of the document for collecting public feedback.	Ksenia Bocharova	-
06/08/20	2.01	Implementation of feedback from Joinup and public webinar on 29 June 2020.	Ksenia Bocharova	Ludovic Mayot
26/11/20	2.02	Mapping of use-cases from interviews with MS and NIIS.	Ksenia Bocharova	Peter Burian
29/11/20	2.03	Quality review of updates.	Petro Dudi	Ludovic Mayot
30/11/20	3.00	Submission for approval.	Ksenia Bocharova	Peter Burian

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION[1]

This document represents the deliverable under Task-04 in the framework of the specific contract n°380 under ABCIV-Lot 3, regarding the project on the continuation of an Action running under the ISA² programme (Action 2016.28), namely Access to Base Registries (ABR).

The purpose of the aforementioned task is to review the Framework document with Technical Writer and peer reviewers (specialists in data management), and discuss it with public, thus finalising it.

The project team performed the following activities concerning this deliverable:

- Review by Technical Writer, re-structuring the content in the document and improving the logical flow of the information;
- Reference to European Data Strategy and how the framework contributes to it;
- Technical quality review, namely, peer review by data management specialists within Trasys International, providing suggestions and recommendations;
- Public discussions with public (e.g. Joinup and LinkedIn), confirming the Framework answers their potential questions;
- Mapping of use-cases received during various activities (e.g. webinars), analysis and incorporation of best practice from case studies (interviews with Member States (MS));
- Presentation and discussion with public during webinar, confirming the Framework would serve the needs of the relevant professionals and roles within public administrations.

The work was also based on the following reference material:

- Relevant aspects from existing documentation on Action 2016.28;
- Existing and/or similar initiatives on the European (EU) level;
- Best practices and challenges that MS face in creation of their registry of registries, shared during various activities (e.g. interviews, webinars);
- Suggestions and other feedback from MS representatives and other ABR working group (WG) members, shared during the work on a draft of specification of registry of registries.

On the long term, the Framework for Base Registries Access and Interconnection (BRAIF) aims to fulfill one of the goals of the Action, namely, the creation of a Framework that will serve as a guidance to MS on how to grant access to data in base registries and how to interconnect such base registries.

[1] Introduction to the deliverable D04.01. of SC380

EXECUTIVE SUMMARY

In recent decades, European public administrations have been modernising their internal operations, improving the services they offer to citizens and businesses, and reducing their costs by investing in information technologies.

Administrations continue to face considerable barriers, when exchanging information and/or collaborating electronically, despite the significant existing progress and obtained benefits. The legislative barriers, diversity of technologies used, incompatible business processes and inconsistent information models, are the main obstacles encountered by public authorities[2].

Base Registries play a significant role in overcoming these barriers. The new European Interoperability Framework (EIF) defines base registries as ''[…] the cornerstone of European public service delivery.''[3] This vision emerges from the fact that many Member States are dissatisfied with their return on investment (ROI) of information technologies set in place to modernise their public administrations. They are facing increasing expectations from their citizens for the provision of digital public services, as well as the need to focus on solving the public sector's information management issues. It is acknowledged that **information as a critical asset** has to be managed and shared[4]. Such information should be trusted, reusable and authoritative, which is precisely the role of Base Registries.

The true potential of e-Government information management can be unleashed only by removing the electronic interaction barriers between Base Registries. In addition, the needs and expectations of citizens should be taken into account, facilitating the provision of high-quality public services. Therefore, the need for a harmonised initiative to define Base Registries' interoperability is central to allow the provision of integrated public services.

Given the situation described above an effective framework should:

- Show how to materialise the EIF basic principles into tangible outcomes for enabling access to Base Registries;
- Include a conceptual model that serves as the basis for interoperability-by-design;
- Provide an intuitive and methodological way for defining high-quality integrated public services;
- Build on top of existing tools and initiatives included in the EIF, like the IMAPS[5] model.

The European **Base Registries Access and Interconnection Framework** (BRAIF) encompasses and adheres to these requirements.

[2] European Commission, COM(2017) 134 final, "communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: European Interoperability Framework – Implementation Strategy"

[3] European Commission (2017), "New European Interoperability Framework: Promoting seamless services and data flows for European public administrations", p. 17.

[4] European Commission, Interoperability solutions for public administrations, businesses and citizens. [Online].

[5] European Commission, Interoperability Maturity Assessment of a Public Service. [Online].

1 OVERVIEW

The "Implementation Strategy" document of the revised European Interoperability Framework defines a Base Registry as:

> *"A trusted and authoritative source of information which can and should be digitally reused by others, where one organisation is responsible and accountable for the collection, use, updating and preservation of information. Base Registries are reliable sources of basic information on data items such as people, companies, vehicles, licences, buildings, locations and roads. This type of information constitutes the master data for public administrations and European public service delivery."*[6]

Here, the term ''authoritative'' implies that information is correct, up-to-date and of the highest possible quality and integrity. Thus, Base Registries information is considered authoritative and plays a crucial role in public administrations' modernisation.

Member States are willing to implement integrated public services by improving access and interconnection of Base Registries. With this regard, BRAIF aims to support the Member States (MS) in building and interconnecting their base registries on national level, and also facilitate the interconnection of base registries on the European level. The interoperability of base registries is key for the development of the **EU Single Digital Gateway[7],** a platform that aims to be the single point of access to public EU Member State's services, facilitating digital public services among public administrations and citizens. Its implementation relies on **the once-only principle**, ensuring that data – which are submitted at least once to an EU Member State – could be reused by any public authority across the EU. Thus, this initiative aims to produce a tool that will require the data from national base registries to be interoperable and shared on a cross-border level.

Data interoperability is already being addressed by the ISA programme[8], and targeted by its corresponding Actions. In this regard, **BRAIF** aims to guide on the creation of an ecosystem of interconnected base registries which exchane data and, therefore, facilitate the set up of integrated public services. In particular, BRAIF provides guidance to Member States' public administrations on what elements should be present in an optimal scenario in their base registries and registry of registries, and for the implementation of cross-border access and interconnection of Base Registries. This way Member States should be able to[9]:

- be more efficient and have effective access to information across borders, when establishing European-level integrated public services;
- establish quicker and easier national and European-level public services[10];

[6] European Commission, COM(2017) 134 final, http://eur-lex.europa.eu/resource.html?uri=cellar:2c2f2554-0faf-11e7-8a35-01aa75ed71a1.0017.02/DOC_3&format=PDF.

[7] https://ec.europa.eu/growth/single-market/single-digital-gateway_en

[8] European Commission, "Improving national and cross-border access to government data" online https://ec.europa.eu/isa2/actions/improving-cross-border-access-government-data_en.

[9] Ibid, https://ec.europa.eu/isa2/actions/improving-cross-border-access-government-data_en.

[10] The conceptual model, phases, roles etc. are applicable to Member States and European level. I.e., each Registry contributes to the Interconnected Registries and then through the Governance layer to the Integrated public services at EU level.

- reduce administrative burdens;
- accelerate the development cycle of Base Registries;
- promote the discovery and reusability of existing solutions;
- improve transparency.

1.1 DATA TERMINOLOGY & DEFINITIONS

Data can be used in different ways and for various purposes. For example, public administrations could utilise it as follows:

- Creation of different integrated public services, based on life-events;
- Data-informed policy-making, where data is used when policies are being drafted.

Information is defined as data, placed in a context that creates meaning. This may be as simple as combining two data sets, or as complex as placing information in a political scenario. Both data and information may be stored and manipulated digitally in slightly structured documentary systems, and/or highly structured data systems.

The words 'data' and 'information' are often used in the same context, although their meaning varies. Moreover, these terms correspond to different meanings for people working in different areas (e.g. archivists, data scientists, document management experts etc.). **Data** are most commonly defined as elementary objective facts. Individual data items are discrete and often measurable. Data may or may not have meaning in isolation. Traditionally, data captured and stored in computer systems have been almost exclusively textual or numeric, but other forms of data are being captured and stored, too (e.g. symbols or images). Data becomes **information**, when it is put into context and related with other pieces of data. It is at this moment when information acquires meaning and becomes useful for taking decisions.

Typically, a **Register** is a structured list of units, containing a number of attributes for each of those units, and has a regular updating mechanism. Under this definition, many Administrative data files can be considered to be Registers[11].

Base registries are reliable administrative sources of basic information on data items, such as people, companies, vehicles, licences, buildings, locations and roads. This type of information constitutes the '**master data**' for public administrations and European public service delivery.

Master data describes the people, places, and things that are involved in an organisation's business. Examples include people (e.g., customers, employees, vendors, suppliers), places (e.g. locations, sales territories, offices), and things (e.g., accounts, products, assets, document sets). The master data type is synthesised with different data types that provide a holistic view of the public administration information. Because this data tends to be used by multiple business processes and IT systems, standardising master data formats and synchronising values are critical for successful system integration. Master data also tends to be grouped into master records, which may include associated **reference data**.

Reference data are sets of values or classification schemas that are referred to by systems, applications, data stores, processes, and reports, as well as by transactional and master records.

[11] Register, ESS Vision 2020 ADMIN (Administrative data sources), https://ec.europa.eu/eurostat/cros/content/register-0_en.

Examples include lists of valid values, code lists, status codes, state abbreviations, demographic fields, flags, product types, gender, chart of accounts, and product hierarchy. Standardised reference data are key to data integration and interoperability and facilitate the sharing and reporting of information. Reference data may be used to differentiate one type of record from another for categorisation and analysis, or they may be a significant fact such as country, which appears within a larger information set such as an address. Organisations often create internal reference data to characterise or standardise their own information. Reference data sets are also defined by external groups, such as government or regulatory bodies, to be used by multiple organisations.

Additionally, **transactional data** describes an internal or external event or "transaction" that takes place as an organisation conducts its business. Examples include sales orders, invoices, purchase orders, shipping documents, passport applications, credit card payments, and insurance claims. These data are typically grouped into transactional records, which include associated master and reference data.

Another type of data is **Metadata** which literally means "data about data." Metadata label, describe, or characterise other data and make it easier to retrieve, interpret, or use information. Some common metadata types are technical metadata that could include field names, length, type, lineage, and database table layouts, and business metadata that could involve field definitions, report names, headings in reports and on Web pages, application screen names etc. However, it could be argued that there are other types of metadata that make it easier to retrieve, interpret, or use information. The label for any metadata may not be as important as the fact that it is being deliberately used to support data goals. Any discipline or activity that uses data is likely to have associated metadata. We provide a more detailed description of Metadata and their management in section 2.3.2 Metadata Management.

Historical data contain significant facts, as of a certain point in time that should not be altered except to correct an error. They are important to security and compliance. Operational systems can also contain history tables for reporting or analysis purposes. Examples include point-in-time reports, database snapshots, and version information.

Temporary data are kept in memory to speed up processing. They are not viewed by humans and are used for technical purposes. Examples include a copy of a table that is created during a processing session to speed up lookups.[12]

There are different types of data, according to different data classification criteria, for example[13]:

- **Open data** are publicly shared with minimal restrictions, available to everyone interested to reuse or republish them, an example of this type of data is non-personal data;

- **Shared data** are shared with other government entities, and might be shared with private-sector entities as well, based on appropriate privacy protection and interoperability agreements. Examples of this type of data are confidential (non-sensitive) personal data and sensitive personal data, such as statistical data for medical research;

[12] Definitions of Data Categories.

[13] Example for inspiration available here: https://government.ae/en/about-the-uae/digital-uae/data/data-operability.

- **Closed data** cannot be shared with anyone, they represent secret / high-level protected data.

Base registries from the public sector contain **Open data** that focus on releasing machine-readable data for use by others to stimulate transparency, fair competition, innovation and a data-driven economy. To ensure a level playing field, the opening and reuse of data must be non-discriminatory, meaning that data must be interoperable so it can be found, discovered and processed. Ideally, basic metadata and the semantics of open datasets should be described in a standard format readable by machines.

It gives the opportunity to address any kind of administrative challenge and provide growth options to Member States. Moreover, publishing of open data should also allow reuse of such data in order to reach its full potential, along with legal interoperability and certainty. Hence, the right for anyone to reuse open data should be communicated clearly among Member States, and legal regimes to facilitate the reuse of data, such as licences, should be promoted and standardised as much as possible.

1.2 SCOPE AND TARGET AUDIENCE

The objective of BRAIF is to describe the agreements and infrastructure for operating Base Registries and the relationships with other entities, as already stipulated in the new EIF. However, BRAIF does not enter into the details of the Public Service Governance but it still is part of the Interoperability Governance, since it offers guidance on how the different base registries should be interoperable with each other. Both these aspects (Public Service Governance and Interoperability Governance) are covered by other works (EIF and EIRA, among other), except where it directly relates to base registries, such as Data and Operational Governance (for example, section 2.1.1 Data Governance, describes how different governance bodies can set up the rules for data interoperability).

BRAIF is meant to be a framework applicable to all public administrations in EU Member States. It is expected to act as a common denominator for relevant initiatives at national, regional and local level. It aims at inspiring European public administrations in their efforts to design or adapt Base Registries, their access and their interconnection. It is foreseen to facilitate base registries' interoperability and, thus, the delivery of integrated public services based on them.

The BRAIF conceptual model is based on the previous work that has been done in Europe. This is beneficial for all Member States, as they can be aligned with best-practice recommendations.

Additionally, the knowledge transfer and partnership is crucial for the targeted state of a data framework, where the base registries will cooperate and reuse data and services in an effective, secure and consistent way using API-led technologies.

BRAIF is addressed to the owners of Base Registries, who actually are State Bodies (Ministries and Fiscal Services), including both administrative staff and experts. It also targets policy makers, public administration officers and ICT developers who are involved in the definition, design, coordination and delivery of European public services that rely on the data kept in the Base Registries.

1.3 VISION AND STRATEGY

In the context of the European Strategy for Data[14], BRAIF supports its implementation by provision to Member States of a set of good practices in the area of base registries interconnection, allowing secure

[14] European Commission, 2020, European Data Strategy.

cross-border sharing of data, and establishing a base for the future European Registry of Base Registries.

The current state of Base Registries in Member States indicates they are in different phases of maturity, hindering, thus, their ability to implement any common data framework.

Some parts of public administrations in many Member States function independently from each other without a common frame or goals. They work in silos, collecting, storing and maintaining data in individual registries. This results into collecting data in different contexts, but with similar meaning, enlarging the issue of information duplication, reducing the quality of data, etc.

Additionally, the processes that serve citizens and businesses with information from public administrations are often too complicated and not efficient. This can result in cases where the same data information can be provided to data consumers more than once[15].

Overall, data is being treated inconsistently within a legacy environment, affecting not only public administrations, but also citizens and businesses.

The main aspects of the legacy environment scenario could be summarised, as follows:

- Different architecture is used in different base registries;
- Data are being collected, stored and maintained individually in different base registries;
- Duplication in data collection and its storage in various base registries;
- The same data are being shared with consumers more than once;
- Quality of data, shared by different sources, is low;
- Bilateral agreements on data usage and sharing between public administrations;
- Low level or lack of data reuse, facilitating silos.

The future state of affairs in Member States should target optimisation of data usage and implementation of the culture of responsible, secure, efficient and transparent data reuse for the benefit of policy makers, citizens and businesses.

The move towards optimised data management and regulated data-sharing can be achieved by:

- Implementation of effective legislation enabling concept of base registries and data reuse;
- Simplification of governance through establishment of a common governance, setting the standards, leaning the processes and creating the data domains;
- Creation of common meanings through reuse of common data models and master data management;
- Implementation of authoritative base registries;
- Development and implementation of interconnecting infrastructure;
- Continuous improvement of data quality.

The move from a legacy environment to a future state of affairs could be summarised in the figure below:

[15] ISA² Access to Base Registries: Peter Burian, DG DIGIT, European Commission.

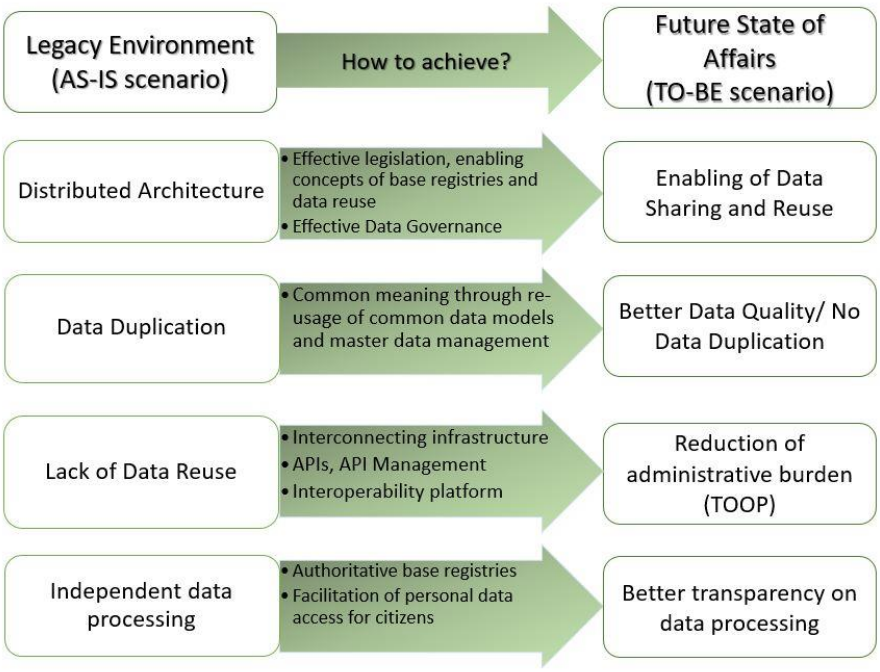

Figure 1: Legacy scenario versus Future state of affairs

Some Member States have already made the first steps to the adherence to the "Once only" and "Privacy by design" principles, focusing on effective legislation and governance, enforcing data-sharing and implementing core interoperability platforms. It is important to emphasise that with regard to the interoperability platform, data should remain in the source-base registries, thus an interoperability platform aims to discover and collect data from some base registries and enable its reuse by other base registries. Overall, implementation of an interoperability platform facilitates the technical interoperability between different base registries on different levels (in some countries, on local, regional, national – and even cross-border levels). Industry-standard data management frameworks recommend the steps for moving from a legacy environment to total master data management. The Data Management Capability Assessment Model[16] (DCAM) proposes the following steps:

- To achieve simplification, a reconciliation of complex data environments should be done (e.g. by definition of authorised data domains, elaboration of end-to-end process workflows);
- To achieve alignment to meaning, harmonisation of data should be based on contractual precision (e.g. usage of core vocabularies for conceptual glossaries, unique identifiers, cross-referencing among the systems);
- To achieve data quality management, the fit-for-purpose data should be provided without reconciliation and transformation (e.g. by establishment of data quality criteria, data quality controls, rules and tolerances);
- To move to the technical implementation, integration of data into operational and production environments should be done (e.g. by development and implementation of the platform, usage of semantic data model).

[16] https://dgpo.org/wp-content/uploads/2016/06/EDMC_DCAM_-_WORKING_DRAFT_VERSION_0.7.pdf.

In order to support Member States in their work towards such a future state of affairs, BRAIF comes with a substantial added value on the practical level to its target audience. It is further enhanced by juxtaposing the challenges faced by Base Registries and the roles performed. As it is known, the main functionalities for the majority of Base Registries are the collection, storage and processing of information for decision-making purposes.

Additionally, Base Registries need to store qualitative data and foster personal data de-duplication. It is common knowledge that low data quality undermines stakeholders' and citizens' trust and, in consequence, lowers the consumption of digital integrated public services.

A systematic (and methodological) approach has to be defined for reforming the state of Base Registries that supports the aspects developed in the following sections of the document. Unfortunately, there are quite a lot of barriers that could possibly hinder this process, namely:

- Lack of legal frameworks and relevant regulations, which negatively affect the implementation of Legal Interoperability;
- Lack of coordination between officials, which hinders Organisational Interoperability;
- Difficulties to apply the basic principles of EIF, which lowers the trust in EC recommendations;
- Poor data quality, which results in wrong decision-making, poor operations and cost increase;
- Lack of strategic orientation, which corresponds to poor (or non-existent) policy-making;
- Unclear technical and semantic specifications, which hinder proper implementation of the respective EIF levels and the adoption of standards.

The purpose of BRAIF is to help overcome the above-mentioned barriers. It is foreseen to support relevant stakeholders in the following aspects:

1. With a clear methodology and a systematic approach, BRAIF enables Base Registry access, interconnection and integration.
2. BRAIF defines a common terminology and regulated procedures on national and cross-border levels between Member States. This should be the basis for a common language for Base Registry Interoperability (BRI).
3. It will provide the means for data quality and management plans by defining the systematic approach for integrated digital public services.
4. It will improve the coordination and control between State officials, as the common language of BRI, thus, fostering Organisational Interoperability.
5. It will support the development of technical and semantic requirements that in turn will support the creation of Base Registries.
6. BRAIF will strengthen policy drafting and a national and cross-border interoperability strategy.
7. It will enable the efficient, agile and secure reuse of data and services to better serve citizens and businesses.

2 BRAIF CONCEPTUAL MODEL AND PHASES

The BRAIF conceptual model defines a series of steps to take in order to establish structure-based Base Registries that enable the delivery of cross-border integrated public services. Integrated public services are services, which are delivered under different legal frameworks, policies and strategies. These services should be one of the cornerstones of the Digital Single Market[17]. They will appear as a direct consequence of the cross-border sharing of information, which is a *sine qua non* condition for the accomplishment of the European strategic vision in terms of a digital landscape. Additionally, the BRAIF conceptual model also supports the principles and recommendations defined in the European Interoperability Framework (EIF).

The conceptual model proposed by BRAIF is the following:

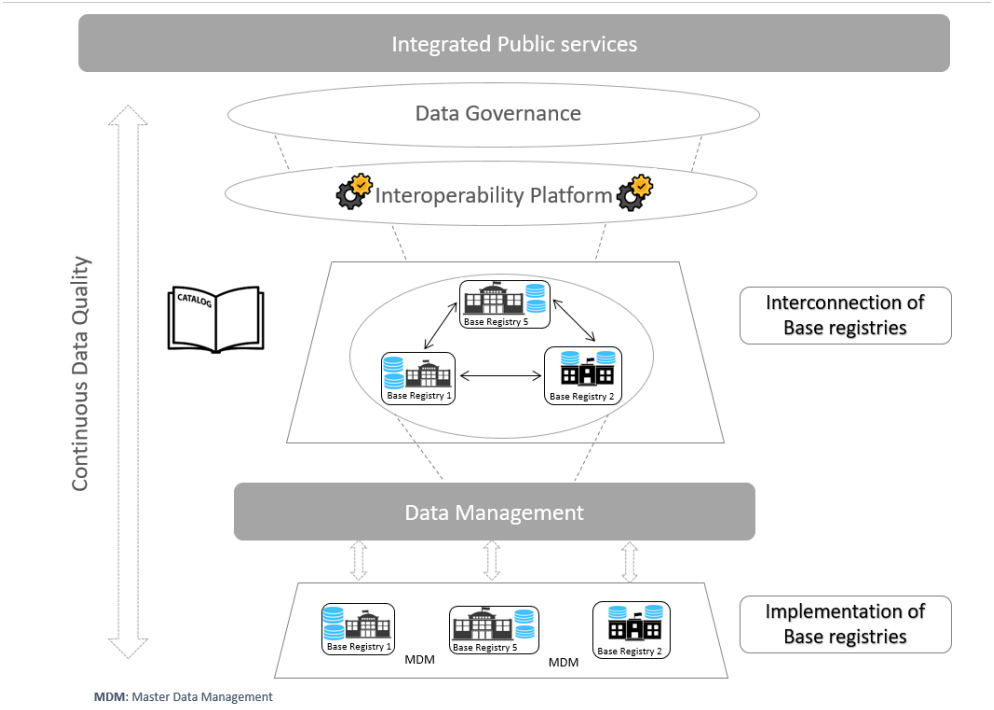

Figure 2: BRAIF conceptual model[18]

The BRAIF conceptual model presents the necessary phases needed to establish cross-based Base registries within the State, and interoperability governance based on such Base Registries. The process for the implementation of these services is not linear, i.e. it has no predefined start/end point. Therefore, it is a cycle that starts based on the level of maturity each organisation has obtained or achieved.

The cycle of activities, as depicted in Figure 2, help the deployment of integrated public services. As discussed previously, the starting point of the cycle differs among services, depending on their respective maturity levels. There are different phases of the cycle, which can be grouped in these categories[19]:

[17] A policy belonging to the European Single Market that covers digital marketing, E-commerce and telecommunications. Available here: https://ec.europa.eu/digital-single-market/en

[18] The Conceptual model encompasses the master data management elements, as identified by DAMA International and defined in DMBOK.

[19] ISA² Access to Base Registries: Peter Burian, DG DIGIT, European Commission.

- Passing from the legacy environment into a Common Governance with a specific Strategy under certain interoperability agreements;
- Setting the standards and creating more lean processes;
- Using common data models and mastering the data under profound Data Security and Quality.

The justification of this approach is illustrated below by the two most common maturity use cases:

Use case 1: No Base Registry exists before the definition of the service
This case corresponds to the scenario where all the information required to convey the integrated public service is not collected by existing base registries. This use case would take place, when the information is not completely or partially stored in an already existing Base Registry. The process will start in the ''Strategy'' phase of the cycle, as the conception of the service itself needs to be developed (including the sources of information necessary for its delivery). From that point onwards, the organisation would advance in the cycle developing the full structure needed for the delivery of the service, including the establishment/upgrade of a Base Registry.

Use case 2: The Base Registry already exists before the definition of the service
This case corresponds to the scenario where an integrated public service is built from a group of existing base registries, which are currently collecting and storing the information required to convey the service. After performing all relevant actions and assessments (based on the CMMI – DMM model), maturity levels have to be applied – depending on the results – with recommendations and proposals. At this stage, it is important to determine the characteristics of data already stored in the Base Registry(ies) and ascertain that the data are compliant with the (interoperability) requirements established for the delivery of the service. In case the data are not compliant, a further treatment would be needed, and the organisation should follow the necessary recommendations and phases of the cycle from that point onwards. In this case, the process can start in the level of the assessment result, which can be the *"Base Registry quality/audit and data curation"* phase of the cycle, to ensure that high-quality data will be shared in the future integrated public service, the "*MDM*", or at any step in the interconnection of Base Registries.

Finally, the proper design and implementation of master data management policies will be a critical factor for the correct functioning of services based on Base Registries. The nature of the data held by Base Registries, and its high-level quality, underpin the adoption of Master Data Management (MDM) best-practices during the development of the integrated public service.

2.1 COMMON GOVERNANCE AND STRATEGY

This group of phases includes activities required to create the common ground on governing the data in base registries and to unify the strategy of these actions. The detailed description of each phase is provided in the table below.

Table 1 Phases concerning the interconnection of base registries

Phase	Description/Scope
Model Data Governance	This phase aims to establish a cross-domain data governance organisation, which then sets the rules and policies on how to manage the data. Thus, the governance model should be defined that will steer the development and, later on, the delivery of the integrated public service. The intrinsic organisation of the base registries for the delivery of the integrated

	public services is essential during this initial phase, and, therefore, should be reflected in the governance model. The definition of the governance model involves the following components: - Organisational structures, roles and responsibilities, including their interactions and reporting and escalation processes; - Standards, rules and policies, and the corresponding processes required to support their adoption by all involved stakeholders in the provision of the service; - Processes to monitor the deployment, measure the impact, audit the application and evaluate the effectiveness of the above-mentioned organisational structures, policies and standards. As a result, the defined governance model should be inclusive to facilitate that the 'voice' of the stakeholders is heard but, at the same time, be efficient to ensure that the objectives and the planning are met.
Reuse the data in base registries via integrated public services	This phase aims to set up the integrated public services that reuse data from base registries, and establish the interoperability agreements.
	<u>Setup of the integrated public services:</u> The delivery model of the integrated public service should be designed accordingly, in order to comply with the requirements emerged from its legal basis. This entails the definition of the mission, vision and the general principles, which will lead the construction and evolution of the service. Furthermore, high-level requirements and needs should be elicited to propose reliable service indicators and service level agreements. The necessary financial needs for the construction and future evolution of the service should be secured in order to ensure the long-term sustainability of the process. The main focus of this phase, in terms of access and interconnection of base registries, is the identification of the users, who will access the service and the interconnection of the service with other services[20]. The specific needs for data and information exchanges should be duly analysed to identify the existing and/or new sources of data (Base Registries) that will be involved in the delivery of the service, and the formats by which the data will be provided to its users. This process involves also the identification of all infrastructure components that will be needed by the service. Finally, the agreements and stakeholders of the service shall be identified and included to the service strategy.

[20] Note: the interconnection requirements can be "internal", where interconnection among the base registries is part of the integrated public service, or "external", where, for example, the interconnection with other services and base registries is not directly involved in the provision of the service.

	Establishment of the interoperability agreements: Interoperability is considered as formal agreements between different organisations to ensure interoperability, and can be formalised at legal, organisational, semantic and technical levels. These agreements establish the basis of the collaboration of all stakeholders involved in the delivery of the public service. In the case of Base Registries, these interoperability agreements shall be put in place to formalise the roles and the exchange of data between the different public administrations or authorised entities involved in the delivery of the service and the Base Registries themselves. The implementation of interoperability agreements should be complemented by operational agreements (related to operational matters) and change management procedures in order to adapt to the changes in the service and its data[21].

2.1.1 Data governance

The Data Governance Institute defines data governance 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."*[22] Data Governance is the creation of rules, the execution of those rules, and the resolution of any violation of the rules[23]. The implementation of those rules requires the establishment of a data governance model, which focuses on the management of data throughout its whole lifecycle. In the context of access and interconnection of base registries, the implementation of the data governance requires primarily the following aspects:

- organisational structures, roles and responsibilities **for the management of data, its access and interconnection**;
- standards, rules and policies **to formalise data management** across integrated public administrations;
- processes to monitor and evaluate the **adoption effectiveness of the data management practices** by organisations.

All above-mentioned elements should be formalised and detailed in a data management plan representing the reference document for the management of data across public administrations. It should be noted that the management of all data-related activities is not limited to the data management plan only. Instead, activities related to change management or data recovery should be included, respectively, in the change management, business continuity and recovery plans.

[21] The need to complement interoperability agreements with operational agreements and change management procedures follows the EIF Recommendation 26.

[22] The Data Governance Institute, http://www.datagovernance.com/adg_data_governance_definition/.

[23] Data Governance vs. Master Data Management.

2.1.1.1 Structures, roles and responsibilities

The first step towards the implementation of data governance is the definition of the organisational structure that handles data management activities. During this process, the following elements could be defined:

- the **empowerment and the role** of the data governance organisation with horizontal power to set standards across the entire public administration;
- **decision-making process** and coordination mechanisms;
- **organisational bodies** – i.e. Committees, Working Groups, etc. – with their corresponding members, working rules, periodicity of meetings, etc.

The definition of the data governance model requires a compromise between providing stakeholders with adequate means to channel their requirements, needs and/or complains, and a flexible decision-making process, allowing to cope with changes in a timely fashion. During the description phase, different organisational roles are identified such as steering committee member, service owner, etc. – including specific data-related roles such as data owner and data steward. Specific responsibilities regarding the management of data shall be clearly determined for every data role. The final step to complete the governance model is to appoint actors to the different roles.

As presented in the table below, and inspired by DAMA DMBOK (version 2), typical IT- and business-related Data Governance committees need to be established for a fully operating governance framework.

Table 2. Typical Data Governance Committees/Bodies

Data Governance Body	Description
Data Governance steering committee	Consists of senior and executive members of the organisation, responsible for overseeing, supporting and funding DG activities.
Data Governance Council (DGC)	Handles data governance initiatives, issues and risks
Data Governance Office (DGO)	Keeps continuous interest in enterprise-level data definitions and data management standards.
Data Stewardship	Teams of data analysts, business and technical data stewards concerned with specific projects and areas. They collaborate and provide consultation on data definitions and data management standards.
Local Data Governance committee	Being part of large organisations that work under the DGC enterprise. Small organisations are not advised to be divided into more local DG committees.

In practice, any established governance model requires collaboration between different stakeholders and public administration entities (for example, different entities might need to convene in Working Group meetings, to discuss and agree on certain topics like a common data model, etc.).

It is clear that the development of goals, principles and policies as a result of a defined Data Governance Strategy will guide the organisations into the desired future state. Policies may materialise in different stages as follows:

- The Data Governance Office will certify data for use by the organisation;
- Business owners will be approved by the Data Governance Office;
- The Data Stewards will have daily responsibility for coordinating data governance activities;
- Certified users will be granted access to Certified data for ad hoc/non-standard reporting.

These policies must be effectively communicated, monitored, enforced and re-evaluated periodically. The Data Governance Council can delegate this authority to the Data Stewardship Steering Committee.

A strong governance model enables the entire ecosystem of data to operate by securing a well-functioning internal organisation, the definition of clear policies and its systematic practising and supervision – this also goes for data conventions and nomenclature. There could be a centralised governance model, a decentralised one, as well as tailored versions (such as a federated model).

Governing organisations set high-level technical and administrative standards for how to model a public sector common data model. Therefore, the initial focus has been on governance aspects regarding logical modelling of metadata and high-level data concepts. More detailed designs of data concepts related to specific administrative systems are owned and managed in a decentralised fashion but the decentralised actors all have to comply with the guidelines from the agency/body. Creation of a full-blown public sector enterprise data model will thus happen progressively, which will likely be the chosen model in other Member States with large legacy data registers[24].

In order to implement a centralised governance model, one organisation should be made responsible for establishing, implementing and managing a common IT centre where data are consumed from established authentic sources. DCAM defines a need to establish data domains thus there could be bodies and authentic sources (base registries) responsible for a specific sector: central government, social security, municipalities, health, education etc. Therefore, these authorities follow the standards and rules established by a common IT centre which is often the de facto one responsible for setting the common governance structure and rules on how to standardise data.

Besides the centralised form, there is also the **federated governance** model. In most cases, it consists of a registry, where the responsibility is distributed among different entities. A federated organisation requires the necessary collaboration of different "leaders" during the design and the implementation processes of data governance across the integrated public service. Ad hoc structures are required to ensure coordination across entities and alignment with the defined management plan. The federated governance model can be implemented by establishing one governance body that positions itself as an organisation which other institutions and citizens can reach out to when interested in either exposing their data or expressing their needs for establishing certain standards or data source interconnections. Adoption of data management practices triggers the need for new resources to cover data-related roles, identified in the data management plan. The Data Steward is essential in this context since his/her role is to focus on the definition and surveillance of policies and rules for the use

[24] UNLOCKING THE POTENTIAL OF eGOVERNMENT.

of information. The Data Stewards may be differentiated by their place within an organisation, the focus of their work or both. Thus, there are different variations of this role, such as:

- Chief Data Stewards;
- Executive Data Stewards;
- Enterprise Data Stewards;
- Business Data Stewards;
- Data Owners;
- Technical Data Stewards;
- Coordinating Data Stewards.

The Data Steward role, in general, concentrates on supporting the implementation of the data management plan within the organisation's own systems. The Information Steward, specifically, also ensures the correct access to data, information collection, usage, update, maintenance and deletion, as established by the EIF[25].

2.1.1.2 Data policies, standards and requirements

The deployment of data management practices relies on the definition of data policies, standards and requirements. Data policies refer to the practices required to ensure the correct management of data throughout all the processes. Meanwhile, standards, requirements and data definitions are used to define processes and procedures necessary to implement these data policies.

The development of goals, principles, standards and requirements derive from the Data Governance Strategy that will guide the organisation into the desired future state. These are initially created by data management professionals, business policy staff or a combination of thereof. But, these are always done under the supervision of the data governance body and strategy. Subsequently, the Data Stewards review and refine them. In the end, the Data Governance Council, or any similar body, conducts the final review, revision, and adoption.

There are traditional data policies to study and implement, as listed below:

- **Data policy on authorisation and accessibility** is based on the national legal frameworks, and it defines legitimate users that can be authorised to access the data in base registries, define types of access rights, define consent management[26], etc.;
- **Data protection policy** is based on data protection-related legal frameworks[27], 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 registry of registries) are protected, which security protocols should be used, etc.;

[25] European Commission (2017), European Interoperability Framework, p. 37.

26 Art. 6 GDPR Lawfulness of processing

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

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

The scope of data policies, standards and requirements in this realm goes far beyond traditional data management practices. Aspects like data linking and enrichment, management of master data, publication in various formats (including open data), preservation, anonymisation and expiration are relevant for (semantic) access and the interconnection of base registries.

2.1.1.3 Monitoring, Auditing and Evaluation

The implementation of data governance implies structural changes in the organisation. This requires periodic monitoring and evaluation processes in collaboration with the established change management process. In addition, timely audits are required to assess the alignment of the practices and processes with the data management plan. Therefore, monitoring, auditing and evaluation are key activities in simplifying and streamlining the governance structure as much as possible, as well as eliminating unnecessary complexities in order to improve their performance.

2.1.2 Metrics

Metrics are standards for measurement or evaluation of performance, progress, quality, efficiency or any other effect. These are measurable entities of work that are applied in each area mentioned in this Framework. A framework must be able to measure progress and success through metrics that demonstrate whether the policies, strategies, and governance procedures have added business value and have attained the objectives.

Identifying specific Key Performance Indicators (KPIs) and building up scorecards is a method to measure, in a timely manner, the progress of rollout of the different Framework areas. Examples of such KPIs follow below:

- Data Governance:
    - Reduction of risk;
    - Achievement of goals and objectives;
    - Speed of change adoption;
    - Performance of policies and processes.
- Data Architecture:
    - Proportion of new architecture facts versus reused;
    - Costs of delays;
    - Time to correct mistakes.
- Master Data Management:
    - Confidence of entity for use across the organisation;
    - Rate of change of data values;
    - Software licences costs;
    - Data sharing volume.
- Metadata Management:

    - Ideal versus actual coverage of organisation's Metadata;
    - Metadata usage, i.e. repository logins;
    - Metadata documentation quality.
- Data Security:
    - Contingency planning and business continuity plan status;
    - Performance metrics of security systems;
    - Alerting incidents and investigations.
- Data Quality:
    - Number and percentage of errors within the data set, for example, completeness of entity (i.e. null values), uniqueness of the entity record, etc.;
    - Time to resolve issues;
    - Issues by priority and severity.

Another way to measure effectively the state of implementation of the data framework is by using **maturity models**. This can be done in different states of the Framework.

The CMMI Institute's Data Management Maturity (DMM) model is a formal framework for identifying the maturity of data management for any kind of industry. This Institute has many years of experience in implementing models for process improvement of any type of business and/or organisation that has many collaborators and clients.

The CMMI – DMM contains best practices for establishing, building, sustaining, and optimising effective data management across the data lifecycle, from creation through delivery, maintenance, and archiving[28]. The DMM is structured in such a way that it can be used by organisations to not only assess their current state of capabilities but also to build a customised roadmap for data management implementation.

It is a data management framework providing best practices in 5 + 1 categories. These categories are:

- Data Management Strategy;
- Data Quality;
- Data Operations;
- Platform and Architecture;
- Data Governance;
- Supporting Processes.

For each category, there are Process areas[29] (25 in total). These help the organisation to define the point of reference for its capabilities and build upon these. They also assess strengths and weaknesses and acknowledge any gaps. In the end, the framework's areas assist organisations in taking most of

[28] What is the Data Management Maturity (DMM) model?

[29] Data Management Maturity (DMM) Model At a glance, by CMMI Institute.

their data assets and applying proper data management maturity leading to business success. All this occurs in conjunction with the infrastructure support practices, too.

The main phases of the CMMI – DMM for assessing the elements of the Process areas – which enable identification of the level of maturity for each category – are the following[30]:

- Phase 1 – Kick-off: Consists mainly of general and assessment preparation and its duration can be up to 4 weeks;
- Phase 2 – Assessment: Consists mainly of surveys, interviews with key-point data management people for capabilities evaluation, workshops, reviews of work products, and its duration can be up to 1 week;
- Phase 3 – Report: Includes the Planning Guidance which mainly consists of the assessment report and executive briefing containing scoring, findings, gaps, strengths, weaknesses, observations, proposals for effective use of data assets and the data management framework implementation, and its duration can be up to 2 weeks.

In Phase 3, the results of performing all relevant actions and the assessment lead to the maturity levels that could be applied; these, in turn, produce the resulting recommendations and proposals[31].

An example of a CMMI - DMM data management maturity assessment deliverable, along with how an organisation can be rated in the 25 different Process areas, is shown in the figure below:[32]

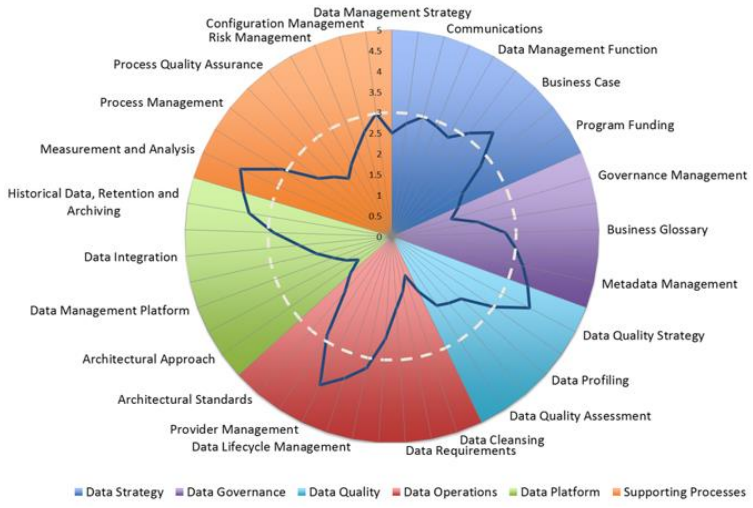

Figure 3: Example of a CMMI - DMM data management maturity assessment

2.2 STANDARDS AND ADDITIONAL LEAN PROCESSES

This group of phases focuses on the actual implementation based on the global principles defined for the delivery of the integrated public service. The phases covered in this group are detailed in the table below:

Table 3. Standards and Processes

Phase	Description/Scope

[30] What is a DMM Assessment and how to get more information about it?

[31] Data Management Maturity (DMM) Model At a glance, by CMMI Institute.

[32] What is a DMM Assessment and how to get more information about it?

Development of a trusted registry	This phase takes place when important pieces of authoritative information are not managed by Base Registries. Hence, two main scenarios can occur to trigger the need for this phase: either a new base registry should be built from scratch, or the mandate or scope of existing base registries has to be significantly changed. The goal of this phase is to implement/upgrade the platform(s) that primarily will be in charge of the creation and acquisition of data from end users. Base Registries will be developed according to the service requirements, the defined data policies and the interoperability arrangements defined during the establishment of integrated public services. The construction of this base registry will be done using the guidelines defined during the strategy phase, aiming at reusing existing practices and available building blocks. At the conclusion of this phase, an up-to-date base registry will be up and running. It will be ready to be involved in the construction of the integrated public service, too.
Information Exchange	Information exchange is the first step towards the construction of an integrated public service, where multiple base registries are really collaborating and contributing technically to build a future integrated public service. The data stored in the base registry will be exchanged with other stakeholders following the defined data policies and the interoperability agreements. The exchange of information requires not only the technical means to publish data, like APIs or datasets, but also has to comply with the standards and definitions agreed in terms of formats, content, metadata, constraints, business rules, etc. At the conclusion of this phase, the master data (raw data) will finally be available to the service.
Once-Only Principle	The Once-Only Principle is the EIF Underlying Principle 9: Administrative simplification. It prohibits the use of different databases for the collection of the same data. The Once-Only Principle entails that citizens and businesses provide diverse data only once in contact with public administrations, while public administration bodies take actions to internally share and reuse these data – even across borders – always in respect of data protection regulations and other constraints. The Once-Only Principle outlines that public bodies should collect data only once from citizens and businesses, and reuse that data (as opposed to recollecting it). The benefits of adopting this approach are numerous in data quality aspects, systemic performance issues, organisational simplification, decision making etc.
Aggregation / Enrichment	Aggregation and enrichment are transformations that can be made to raw data in order to adjust the information to the needs of the end users. Aggregation refers to the processes required to compile and consolidate raw data into datasets, so raw data will be transformed from their original format to the

	representation used by these datasets. Also, data can be grouped for statistical analysis, or filtered to reduce the number of data fields, or the number of records. Enrichment refers to enhancing and/or refining raw data with the use of specific techniques. Enrichment involves the use of reference data in the form of external vocabularies, taxonomies and/or thesauri. Enrichment is used, among other, to create new data fields, update/replace the actual value of data fields, or link data with external entities. Metadata, for example DCAT, will be used to tag and label the information stored in the dataset with the aim to facilitate its future reusability.
Data Evaluation	The service, once provided, should be monitored and evaluated. It should be assessed whether it meets the objectives established during the definition of the service. The result of this assessment is the driver for possible changes to the service, its objectives and/or its organisation. The identified changes should be managed through the change management process. During this process, the plan for implementation (or not) of those changes – according to their impact and priority – should be set.

2.2.1 Data architecture

Data architecture can be defined as *''a set of rules, policies, standards and models that govern and define the type of data collected and how it is used, stored, managed and integrated within an organization and its database systems. It provides a formal approach to creating and managing the flow of data and how it is processed across an organization's IT systems and applications.''*[33] Data architecture is the main process required to manage data and their relationships - i.e. actors, processes and sources of data and information, which facilitate future interoperability and reusability of data. It helps to transform data requirements into storage and persistence with the use of data standards, vocabularies, taxonomies, thesauri, etc. It is a requirement for establishing an Enterprise Data Architecture, by evaluating the current status of data in the organisation and expressing strategic data requirements. Inline with the above, it is helpful to develop a solid data architectural roadmap, and outline high-level integrated designs to meet and incorporate with the Overall Enterprise Architecture roadmap.

The main goals of data architecture in the realm of access and interconnection of base registries can be summarised as follows:

- The identification of entities and attributes required to manage the information within the organisation, i.e. the common data model;
- The definition of processes related to acquisition, guidance to data integration, transformation, enrichment, publication, control data assets and storage of data;
- The provision of data exchanges with both data suppliers and consumers.

Traditionally, data architects using enterprise architecture methodologies have performed this activity. This section will present the main practices that lead into an Enterprise Data Architecture, the Common

[33] https://www.techopedia.com/definition/6730/data-architecture.

Data Model – that contains the business vocabulary, the Data Flow Design, and how the influx data are organised throughout the different elements.

The Enterprise Data model *"[…] represents a single integrated definition of data, unbiased of any system or application. It is independent of 'how' the data is physically sourced, stored, processed or accessed. The model unites, formalizes and represents the things important to an organization, as well as the rules governing them."*[34]

It is based on the Enterprise Data Model (EDM), which is a data architectural framework used for integration. *"It enables the identification of shareable and/or redundant data across functional and organizational boundaries. Integrated data provides a 'single version of the truth' for the benefit of all. It minimizes data redundancy, disparity, and errors; core to data quality, consistency, and accuracy."*[35] EDM is a holistic implementation-independent conceptual or logical data model providing a common consistent view of data across the organisation.

It includes key organisation data entities (i.e. business concepts), their relationships, critical guiding business rules, and critical attributes. *"It is independent of 'how' the data is physically sourced, stored, processed or accessed. The model unites, formalizes and represents the things important to an organization, as well as the rules governing them."*[36] EDM is required to be reviewed and accepted by stakeholders, depicting the real representation of the organisation.

EDM can serve as a means to create the business vocabulary of the organisation and prevent semantic interoperability conflicts. These are caused by discrepancies in the interpretation of administrative procedures and legislation, the lack of commonly agreed data models, the absence of universal reference data, and others.[37] In light of this, the EDM can be enriched with Core Vocabularies, which are defined as simplified, reusable and extensible data models that capture the fundamental characteristics of a data entity in a context-neutral and syntax-neutral fashion[38]. These Vocabularies assist in creating new data models with building blocks, and guarantee a minimum of cross-sector interoperability. Moreover, the existing data models that have mappings to the Core Vocabularies allow using them as a common foundational data model permitting the bridging of different data models.

In order to have a functional data model, it is important to identify a common meaning and understanding of data, by defining a common data model at the data-governance bodies' level. Thus, common terms for the same objects should be defined for each of the data domains operating related base registries to ensure a common interpretation of data across different entities.

The semantics become an essential topic while defining common meaning, and there are two main options on how to create them:

- A governance body could reuse the Core Vocabularies as authentic models for different databases;

[34] The Enterprise Data Model.

[35] Enterprise Data Modelling, Wikipedia.

[36] The Enterprise Data Model.

[37] e-Government Core Vocabularies handbook.

[38] Core Vocabularies leaflet v11, 2017.

- A governance body can use the Core Vocabularies as a translation layer to map the different data models into a common one.

Many Member States are already using the Core Vocabularies, such as EU ISA Core Vocabularies[39], and are extending them with local extensions on different levels, such as person, business, location, and public service.

Thus, semantic interoperability is enabled by creating a simplified, reusable and extensible data model that captures basic fundamental characteristics of information that is being exchanged by public administrations in different data domains.

2.2.1.1 Data Flow design

Data Flow defines the requirements and master blueprint for storage and processing across databases, applications, platforms, and networks. Such data flows map the movement of data to business processes, locations, business roles, and technical components. Data flows indicate where the data originated from, where it is stored and used, and how it is transformed as it traverses among consumers, systems, and processes. They can map and document relationships between data and:

- applications within a business process;
- data stores or databases in an environment;
- network segments;
- business roles for creating, updating, using and deleting data;
- locations where local differences occur.

2.3 COMMON DATA MODELS AND MASTER DATA

This group of phases focuses on the conceptualisation of data models through management of Master data, the main asset of a Base registry, with a continuous cycle of Data Quality, having the data protected and secured. The phases are described in further detail in the table below.

Table 4. Common data models with Master Data under a Good Data Quality

Phase	Description/Scope
Data models	Common data models shall be defined to foster data exchange, therefore, interoperability, in all such cases where Base Registries need to interoperate, i.e. "talk" to each other. This is a method for reducing semantic interoperability conflicts and create a common ground to interchange data across registries and Member States.
Master Data Management (MDM)	In this phase, managing the master data of the base registry is an essential step to meet organisational goals and reduce any risks associated with data redundancy. Moreover, it helps to ensure data quality from the previous step of the development stage and reduce the cost of data integration.

[39] https://joinup.ec.europa.eu/page/core-vocabularies; EUROVOC.

Metadata Management	In this phase, certain activities related to planning, implemention and control are required to achieve high quality, integrated metadata, as described in detail in section 2.3.2 of this document.
Data Protection Policy and Security	Data policies define the fundamental rules to manage data during the whole data lifecycle in a secure environment. These data policies should be taken into account afterwards to govern all the processes that deal with data, such as data creation and/or acquisition, transformation, enrichment and linking, publication, etc. The scope of the protection policies defined in this phase are not only applicable to the integrated public service, but also to the base registries and stakeholders involved in the service. Therefore, they can be seen as a reference policy to ensure data interoperability and will be the basis for the (data) interoperability arrangements. The implementation of data policies relies on data standards and procedures that detail how the data is handled by an organisation in a trusted and secured data ecosystem. Their uptake ensures a thorough implementation and the adoption of data policies by the whole organisation.
Data Quality audit and curation	In this phase the assessment of data quality, stored in the Base Registry, will be performed in order to determine if actions are needed to assure certain data quality that enables service delivery. The quality aspects to be considered in this phase are further elaborated in section 2.3.4 of this document.

2.3.1 Master data management

Master Data Management (MDM) defines the core business processes and applications of an organisation. As defined by Gartner, MDM is *"[…] a technology-enabled discipline in which business and IT work together to ensure the uniformity, accuracy, stewardship, semantic consistency, and accountability of the enterprise's official shared Master Data assets. Master Data is the consistent and uniform set of identifiers and extended attributes that describes the core entities of the enterprise including customers, prospects, citizens, suppliers, sites, hierarchies and chart of accounts."*[40]

Therefore, high-quality master data specifically governed to foster consistency and reuse, will have a great impact on the performance of the organisation's processes. On the other hand, the lack of properly designed and implemented data management policies can have devastating effects on the organisation's performance.

Moreover, *"[..] Focus is placed on improving data quality, establishing guidelines for data governance and ensuring data can be easily managed and accessed across the business."*[41]

The establishment of a common set of criteria and requirements are the biggest challenges regarding master data management. A common set of criteria describes the business reality of the organisation in a sufficient degree of detail and accuracy. In this sense, MDM is a holistic programme that

[40] Master Data Management (MDM), IT Glossary.

[41] 4 Common Master Data Management Implementation Styles.

documents the information's governance that is oriented towards the optimisation and the reuse of data.

Requirements in terms of assessing an organisation's MDM include the identification of:

- which roles, organisations, places and things are referenced repeatedly;
- what data are used to describe people, organisations, places and things;
- how the data is defined and structured, including the granularity of it;
- where the data are created, stored, made available and accessed;
- how the data changes as it moves through systems within the organisation;
- who uses the data and for what purposes;
- what criteria are used to understand the quality and reliability of the data and its sources.

It is important to emphasise that the data policies are set up at the data governance level, and the data management cycle's aim is to implement them.

In the specific case of Base Registries, master data management is especially important due to the distribution of the information that services use as the basis for their delivery. Within the context of data exchange, master data management requires to set one authoritative source for a particular piece of data (for example, it is recommended to set up an address registry to serve as an unique source of data on addresses, avoid this way the duplication of address data in different public administrations entities).

Master data management is explored throughout the phases of the data lifecycle proposed in this Framework, as the data lifecycle describes the processes performed to manage the data assets of an organisation. In Base Registries, the data assets should be considered the master data.

Table 5. Master Data Management phases

Phase	Description/Scope
Plan	This phase consists of a sequence of actions, intended for the creation of a data management plan. In a data management plan, the sources from which data will be retrieved are identified and secured, and the processes for data gathering, management and use are defined. The data management plan also covers the provision of funding and the identification of technical and staff resources to carry out the full process of data lifecycle management. The system is designed based on the requirements established by this plan.
Specify	The aim of this phase is to ensure a comprehensive and consistent data definition. This is done through the standardisation of data in the data model, which establishes the relationship between data itself, and the properties of the entities that they represent. The data models need to be aligned with each other, thus it is recommended to reuse existing data models, or define the common data model collectively, gathering different stakeholders for discussions and decision taking (e.g. in cross-organisational committees). It also defines the manipulation and integrity aspects of the data stored in Base Registries. This process should foster the compatibility of the data between itself and its consumption by integrated public services.

Enable	This phase has an objective of ensuring all necessary infrastructure for the adoption of the predefined data models in place. This should include several aspects, starting from the construction of the repositories and interfaces of the solution. Once built, they should be tested in order to ensure their full operability. If the tests are satisfactory, the components of the solution are deployed. Finally, this phase will also include the maintenance of all solution components while the service is running.
Create	Once all the necessary infrastructure is in place and in normal operation, the data shall be designed. This data has to comply with the data model and other characteristics, previously defined in the service's data policies, set by the data governance body. It also has to be relevant from the point of view of the service in terms of business value.
Acquire	This phase consists of the acquisition of data from other repositories to store it in a Base Registry. The acquired data has to be analysed in order to determine their degree of compliance with the different sets of characteristics for data, defined by the Base Registry itself, like the format and the terminology in which data are expressed.
Maintain	This phase consists of performing the necessary activities to ensure the integrity and comprehensiveness of data during the service's lifecycle. One of the main activities performed in the context of this task is data cleansing. Data cleansing covers detecting and correcting inaccurate or corrupt data stored in the Base Registry, identifying the affected parts of the data, and replacing them with other data.
Archive	In this phase, the data is archived in the Base Registry. The actual repository in which the data is archived depends on the layout of the base registries for the delivery of the integrated public service. A centralised Base Registry will archive its data primarily in a main repository, while a federated Base Registry will rely on a set of repositories distributed across its Base Registries.
Retrieve	In this phase, the archived data is being retrieved, when needed. It is essential to study the related regulatory requirements for data archiving and retrieving in order to adopt the best solution. For instance, less restrictive regulations might require that the data is retrievable, and more restrictive regulations might require the data to be archived together with the source – application code and the hardware on which the application runs. It is recommended to adopt a comprehensive archiving solution that should support any data retrieving, even if the application data structures have changed or if the source application / system is no longer available[42].
Purge	This is the last phase of the data lifecycle and it implies the elimination of data that have been evaluated as not relevant or necessary anymore, or which an entrusted actor has requested their elimination. One example is a request for the elimination of personal data. This phase should also include the anonymisation of data. It is recommended to consider this activity as standard due to recently conceived procedures following

[42] Managing Data in Motion. Data Integration, Best practice techniques and technologies, by April Reeve, 2013

	the release of the General Data Protection Regulation (GDPR)[43] in the European Union.

2.3.2 Metadata management

Metadata is, simply put, "data about data". Even if this definition sounds quite simple, metadata itself is not a simplistic entity. In everyday life, whatever we use is or contains metadata: content information, names, features, history, etc. Moreover, *"Metadata is key to the functionality of the systems holding the content, enabling users to find items of interest, record essential information about them, and share that information with others."*[44]

Without reliable metadata, an organisation does not know what data it has, what the data represents, where it originates from, how it moves through systems, who has access to it, or what it means for the data to be of high quality. Furthermore, without metadata, an organisation cannot manage its data as an asset, if at all. To be data-driven, an organisation must be metadata-driven.

As defined in DAMA – DMBOK v2, metadata requires to be managed like every other data in an organisation. To be able to have access to high quality, integrated metadata, certain activities related to planning, implemention and control are essential.

The main goals to achieve by properly handling metadata are:

1. Provide organisational understanding of business terms and usage, to ensure that people in the organisation understand the data content and can use the data consistently;
2. Collect and integrate metadata from different sources, in order to make sure that people understand the similarities and differences of data from diverse parts of the business;
3. Provide a standard way to access metadata, and make it accessible to all consumers in the organisation (people, systems, and processes);
4. Ensure metadata quality, consistency, currency and security.

Specific business processes and activities need to be followed, which help deliver the desired deliverables based on certain inputs. Certain key roles orchestrate the continuous flow of correct information, such as data stewards, business analysts, data architects etc. In addition to these, technical methods, techniques, tools and metrics encompass the whole metadata management area of an organisation.

[43] https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en.

[44] Understanding Metadata - National Information Standards Organization, NISO Press, 2004.

2.3.2.1 Types of Metadata

Table 6. Types of Metadata

Type	Definition	Example(s)
Descriptive metadata	For finding or understanding a resource	- Title - Author - Subject
Structural metadata	Relationships of parts of resources between each other	
Markup languages	Integrates metadata and flags for other structural or semantic features within content	
Administrative metadata		
Technical metadata	For decoding and rendering files	- Physical database table and column names - Column properties - Data lineage documentation
Business metadata	Include non-technical information, attribute properties, calculations, algorithms, business rules and valid domain values and their definitions	- Data models - Data standards - Data quality rules - Definitions and descriptions of data sets, tables and columns
Operational metadata	For describing details regarding the processing and accessing of data	- Error logs - History of extracts and results - Technical roles and responsibilities, contacts

2.3.2.2 Standards supporting Metadata harmonisation

The delivery of integrated public services will trigger the need for Base Registries interoperability. It will also facilitate the identification and promotion of the standard reference data that may be (re)used commonly by all Base Registries. It can be delivered as part of the information provided by the integrated public services, too. Reference data, which are based either on open standards or have been

harmonised at the European/international levels, should be reused as much as possible (see, for example, the Publications Office Metadata Registry – MDR[45]).

Common data models shall be defined to foster data exchange, therefore, interoperability, in all these cases that Base Registries need to interoperate, i.e. "talk to each other". Such data models depict, implement, and document very well each concept, relationship, restriction, business rule etc., making them the best candidates for "schemata" or "Application Profiles"[46]. A quite efficient way of addressing such a common exchange data model (supported by the ISA^2 Programme, too) is to identify, assess and select the existing Core Vocabularies and other ontology-supported standards. These foster semantic interoperability and help harmonise the various existing (meta) data models.

Such specifications, which describe metadata and aim at sharing data as linked open data, can reveal themselves as powerful tools to share data between registries and services, and between services and final users' systems.

Table 7 in Annex 4.3 introduces a list of well-known standards that are commonly used to standardise metadata for the exchange of information/datasets in interoperability scenarios.

Following a specific way to harmonise data among registries, and selecting a metadata architecture to ensure the accessibility and its lifecycle, is the proper way to move closer towards the interconnection of base registries. The creation, maintenance and public cataloguing of metadata can help harmonise the free exchange of data among Base Registries and support a semantic harmonisation, as well as creation of unique elements of the single point of truth.

In the base registries' ecosystem, there are typically two main types of users: a) data providers, and b) data consumers. Data providers are mainly base registries owners, the competent authorities managing base registries (i.e., registering, updating information, and maintaining registries), and data consumers are typically third parties, including individuals, businesses, public authorities, public administrations, and any other stakeholders, benefiting from the data providers (collecting, processing, reusing data, etc.).

Moreover, data consumers can benefit from a registry of registries, which connects different base registries by collecting data from them and, in turn, distributes the data to these users.

A registry of registries acts as a one-stop platform for the information contained in base registries, enabling public authorities to have access to official data of specific domains at the national level and, in the future, at the cross-border and European levels[47]. Thus, through secured, interconnected infrastructure, the registry of registries provides information on existing national base registries (with future support for additional European ones) and facilitates the continuous improvement of data quality.

The development and implementation of a registry of registries improves the interoperability of individual base registries and provide a one-stop data platform for citizens, businesses and public bodies, allowing them to access national data across different domains.

[45] MDR.

[46] For instance, like in the case of DCAT. An application profile, in a broad sense, is a set of guidelines, elements and policies defined for a particular application.

[47] Plan for the Registry of Registries.

2.3.3 Data security

According to DAMA, *"[…] data security management is the planning, development and execution of security policies and procedures to provide proper authentication, authorization, access and auditing of data and information assets."*[48] The main objective of these actions is to create a trust environment for the organisation by ensuring that the right actors manage and update data at every moment. In order to determine the right actors at any given time, a set of criteria should be established.

In the case of Base Registries, the construction and maintenance of this trust environment are essential to ensure the ''authoritative'' status of the information stored and processed by them. This is because the ''authoritative'' status is binded to the compliance with a more restrictive set of security requirements than an average organisation. Mainly, data security concerns the data protection from the unauthorised use. In order to protect the information assets of the organisation, data security management has to be aligned with a series of requirements, as identified in the sections below.

2.3.3.1 Stakeholder concerns

Organisations have to design their information security and privacy policies taking into account the needs and characteristics of their stakeholders. Here, Base Registries stakeholders are not the only users of the integrated public service since citizens, too, have their data stored in them.

2.3.3.2 Government regulations

The requirements, stemming from the stakeholders' needs and characteristics, also affect the organisation's stakeholders and shall be taken into account when designing security and privacy policies. In the case of Base Registries, these requirements are particularly important, as personal information is often stored in them. This means that the restrictions in terms of security and privacy, with which they have to comply, are more restrictive than the average ones. In particular, the General Data Protection Regulation (GDPR), which came into force in May 2018, is the main document of reference. The restrictions have to be further analysed in order to determine the viability of the construction of certain Base Registries, which handle EU citizens' data.

2.3.3.3 Proprietary business concerns

Proprietary business requirements concern the organisation's data, which represents its source of competitive advantage. In the case of Base Registries, these requirements are not applicable, as Base Registries are designed to share information through integrated public services. Thus, there is no need to share any information to maintain a competitive advantage.

2.3.3.4 Legitimate access needs

Finally, the last category of requirements refers to the individuals or entities in the organisation that shall have access to data and in which context. In the case of Base Registries, these requirements are particularly bound to the role of Information Steward that was previously described in chapter 2.1.1 Data governance.

The creation of dashboards that allow the visualisation of the use of data, with a particular focus on individuals (citizens), is a measure that could increase the control over the access to data. These dashboards could allow individuals to check who used their data, when, and for which purpose. In that regard, another level of audit for the access and use of data could be implemented, such as conducting

[48] DAMA, 2009, p. 20.

regular enforcement checks against security controls and ensuring that baseline standards (i.e. system security baseline settings and configuration rules) are carried out to monitor compliance.

The following are some recommended practices that ensure a trusted environment for the data to be kept in and shared through different systems[49]:

- Establishing, documenting, communicating and implementing policies and procedures around data, information and system management;
- Developing and implementing redundancy plans for single points of failure that can bring down the entire system or network;
- Implementing and testing the robustness of defensive mechanisms such as defence-in-depth techniques (e.g., deep packet analysis, traffic throttling and black-holing) for detection and timely response to network-based attacks;
- Protecting data endpoints by blueprinting all network infrastructure before implementing and/or maintaining secure standardised network protocols (e.g., TCP/IP, FTP, HTTPS) that involve communication between two or more devices;
- Protecting confidential information stored in all types of endpoint devices with strong encryption;
- Isolating or segregating unencrypted data in online or offline backups for replication and storage concerns data safety (data protection against the loss).

2.3.4 Data quality

EIF establishes that *"[…] citizens and businesses should be able to check the accuracy, correctness and completeness of any of their data contained in base registries."*[50] Hence, the integrated public administration should ensure that the integration of data coming from various base registries maintains the same high-level standards. To that end, the integrated public administration should establish an organisation and processes to achieve high data quality.

Data quality, which can be defined as *"[…] an essential characteristic that determines the reliability of data for making decisions,"*[51] is a necessary enabler for the attainment of different interoperability principles such as openness, transparency, reusability, data portability, etc. The implementation of data quality relies on a data quality organisation, and on the execution of specific activities to ensure the quality of the data, managed by the service.

2.3.4.1 Data Quality Primary Dimensions

Below are the primary dimensions of data quality to be taken into account in data management:

- **Accuracy** – The degree to which data correctly describe "real world" objects or events;
- **Completeness** – The proportion of data stored against the potential for 100%;
- **Timeliness** – The degree to which data represent reality in a specific point in time;
- **Uniqueness/Deduplication** – No data will be recorded more than once;

[49] Trusted data sharing framework.

[50] European Commission, (2017), "European Interoperability Framework", p. 37.

[51] https://www.ibm.com/analytics/data-quality.

- **Validity** – Data are valid if they conform to a syntax (i.e., format, type, range) of their definition;
- **Consistency** – The absence of variations, when comparing two or more representations of data against a definition.

2.3.4.2 Data Quality Organisation

The implementation of data quality requires two main elements: the establishment of an organisation in charge of data quality (**data quality organisation**) and the creation of the **data quality assurance plan,** which helps the organisation define to which extent its data matches high-quality standards. Nevertheless, the data quality organisation and the data quality assurance plan are usually a part of the overall quality organisation and global quality assurance plans, respectively.

The **data quality organisation** falls under the responsibility of a data quality manager, who is in charge of the whole data quality organisation and provides guidelines to define the appropriate data quality strategy. Within the data quality organisation and its strategy, it is needed to decide on the scope of an initial assessment. Identifying critical data and existing rules and patterns is a part of it, using comparisons with high-quality data and performing several assessments. As an example, the data quality manager regularly monitors the quality of the data through a set of scoreboards. The scoreboards cover applicable key performance indicators (KPIs), based on the information gathered from operational tools. With these KPIs, it is easy to recognize issues and to perform root cause analysis. In centralised base registries organisations, the data quality organisation falls under the responsibility of the *"lead"* organisation whereas, in federated base registries organisations, it is distributed among different organisations led by a shared quality management committee or working group.

The **data quality assurance plan** is the reference document for the governance of data quality of the entire integrated public service. It ensures that the (interoperability) data arrangements are respected and describes the following information:

- **data quality activities,** included in the different data governance processes by identification and implementation of improvements with prioritising actions based on business impact, developing preventative and corrective actions;
- **processes, quality indicators** and their corresponding **thresholds** to be met, in order to ensure and maintain a high level of quality in the process;
- **data quality controls**, which involve operational technical tools and checklists to ensure that quality standards are met, and to measure and monitor data quality;
- **data quality audits** that determine whether the data quality assurance plan is properly adopted by an organisation, correct data quality defects, and report on data quality levels and findings.

2.3.4.3 Data Quality Activities

As explained before, the data quality assurance plan defines the activities to be performed for ensuring data quality in the different processes dealing with data.

However, from a technical perspective — and in order to support the access and interconnection of base registries — the data quality plan should be primarily focused on the following processes:

- **Data acquisition** is the first step in the data processing cycle. It consists of the *"ingestion"* of data from various sources via different channels. It is followed by a pre-validation of their format, making sure that only data, which fulfils data interoperability arrangements, are accepted.
- **Source trustworthiness** aim at rating the sources on the basis of the quality of data they provide with other sources in an open or peer-to-peer context.
- **Transformation and validation** aim at transforming the incoming data into the internal format(s) used by the integrated public service. During this process, business rules' validations are executed to verify the received data and to reject invalid information. These validations can be either automatic or manual.
- **Data curation and cleansing** deal with data, which are already validated (and/or not rejected). This process is usually implemented as a processing chain, which involves activities like data cleaning, i.e. dismissal of unnecessary information, enrichment and completion with the use of taxonomies and/or vocabularies, knowledge extraction to identify relevant entities, linking with external ontologies and/or open data repositories, and other.
- **Data publication** involves the publication of validated and curated data for end-users' access (through user interfaces) and/or computers (via machine-to-machine interfaces). Data publication encompasses extracting the subset of the fields, which are included in a dataset, and filtering the records according to a specific criterion, like in the case of open data.

3 INTERCONNECTING INFRASTRUCTURE

By surveying the Member States' landscape, we see that the development of base registries has been, in most cases, an isolated process. Situations in which a registry has been created independently from the development of other base registries – corresponding to a specific need or policy that requested its creation – are quite frequent. Consequently, this has led to a current state where the technical stacks used in base registries are quite heterogeneous. Thus, technical heterogeneity is a constraint to interconnecting base registries within – and across – Member States. This disarray, though, has started to diminish by the need to create a unified data framework across Europe. In many Member States, base registries follow specific processes, while in some cases they have also implemented existing frameworks that adhere to accepted best practices. As a result, technology architecture relevant to base registries' interconnection is being explored. In light of the above, this section will describe how architecture can allow interoperability between interconnected base registries.

The attainment of technical interoperability can be achieved by using modular, loosely coupled service components which are interconnected through an infrastructure layer. **Service Oriented Architecture (SOA)** is an implementation of this concept and it is emerging as the architectural style of choice for interconnecting base registries.

SOA refers to a paradigm for organising and using distributed capabilities that may be under the control of different ownership domains. As defined, "*[…] it provides a uniform means to offer, discover, interact with and use capabilities to produce desired effects consistent with measurable preconditions and expectations.*"[52]

A service-oriented architecture is composed of services that are self-contained and independent from the context or state of the other services they communicate with. While services communicate with each other within an infrastructure, Web services shall be used to support interoperable machine-to-machine interactions over the network. Moreover, Web services can enable a formal separation between the provider and consumer, as long as the interface between them complies with a set of defined technological conventions.

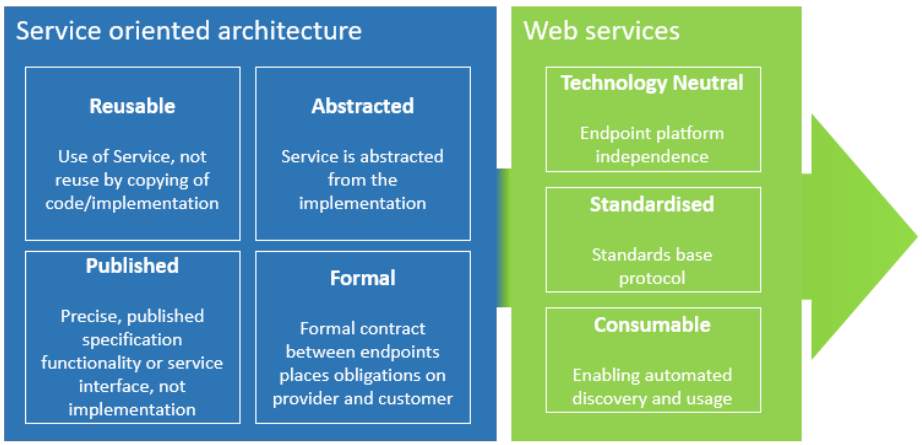

Figure 4: Service Oriented Architecture (SOA) and Web services benefits

[52] OASIS Reference Model for Service Oriented Architecture 1.0 Committee Specification 1, 2 August 2006, p. 29, available online https://www.oasis-open.org/committees/download.php/19679/soa-rm-cs.pdf.

Although Service Oriented Architectures do not necessarily use Web services, their compatibility shall be used for interconnected base registries.

In order to make the base registries' interconnection possible, there is a need to ensure an interconnection infrastructure which, in practice, is already being established in Member States with the creation and implementation of **interconnection platforms** that enable data access supported by **APIs**[53].

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. As mentioned, some Member States have already created and implemented interconnecting platforms[54] that vary from data management and technical points of view, enabling:

- A service-oriented data exchange infrastructure for accessing base registries with citizens' and enterprises' data, at regional, local and federal levels (where applicable);
- Independent data exchange layer, allowing secure internet-based data exchange, based on interoperability agreements between data providers and data consumers;
- Data and its 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.

Regarding harmonised interfaces, many Member States are setting up strategies and policies around an API approach for data access and reuse, enabling open API-driven services.

APIs are well-known technological solutions that exist many years now. However, in the last two decades, they have been used mostly by applications for exchanging data over the web. A lot of Member States have invested in the adoption of APIs – with their associated changes for more digitally-enabled services and internal workflows – because they enable interoperability and can be used in different environments (including information systems and networks).

There are several important aspects regarding the relevance of APIs for interconnecting infrastructures:

- Public bodies should provide the data using APIs;
- Data access on the interoperability platform level is driven by APIs;
- An APIs Catalogue on governmental level is recommended.

In practice, this could be achieved by reusing existing standards for building Web services, for example, the REST[55] architectural style. This implies the following:

- 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.

EU bodies are studying and creating overviews of the feasibility and practicality of API adoption by the public sector, by identifying the concepts, terms, technical specifications and relevant API ICT standards that could help Member States choose the best API approach[56] for their requirements.

[53] Application Programming Interfaces.

[54] A variety of solutions on interconnecting platforms from MSs can be found in the ABR Catalogue of Solutions on Joinup at this link.

[55] Representational State Transfer.

[56] One of the studies on API approach: JRC Technical Report.

In conclusion, through harmonised interfaces the data are standardised, managed and controlled. This enables data exchange in the following ways:

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

4 ANNEXES

4.1 ADDITIONAL DEFINITIONS

Openness is directly related to the notion of Open Data, as stated in the EIF principles, and when applied specifically to data. It is important to stress that not all data managed by Base Registries can be published as open data. Though Open Government initiatives foster free availability of public data for use and reuse purposes, there are certain restrictions applicable to it (e.g. for protection of personal data, confidentiality, intellectual property rights, and other). Nonetheless, Openness is also taken into account for the development and interconnection of Base Registries.

According to IMAPS[57], a **digital public service** is the digital delivery of a public service via channels such as interactive digital collaboration (chat, cognitive agent), mobile app, web portal/website, e-mail and a machine-to-machine interface.

Interoperability is the capacity of one system to interact with another through related interfaces. According to EIF, *"[…] interoperability is the ability of organisations to interact towards mutually beneficial goals, involving the sharing of information and knowledge between these organisations, through the business processes they support, by means of the exchange of data between their ICT systems."*[58]

The interoperability aspects in the context of EIF are legal, organisational, semantic and technical.

The underlying 12 EIF principles **(Subsidiarity and proportionality, Openness, Transparency, Reusability, Technological neutrality and data portability, User-centricity, Inclusion and accessibility, Security and privacy, Multilingualism, Administrative simplification, Preservation of information, Assessment of Effectiveness and Efficiency)** are grouped into four categories[59]:

1. Principle setting the context for EU actions on interoperability (No 1);
2. Core interoperability principles (Nos 2 to 5);
3. Principles related to generic user needs and expectations (Nos 6 to 9);
4. Foundation principles for cooperation among public administrations (Nos 10 to 12).

Data Governance defines the rules of engagement, necessary for a Data Management Programme's implementation. It includes the definition of policies, procedures and standards as the mechanisms for the alignment among stakeholders.

Data Quality establishes the concept of fit-for-purpose data, defines the processes associated with establishing data control, and addresses the implementation of governance mechanisms for the management of the data manufacturing supply chain.

Master Data Management is a comprehensive method of enabling an organisation to link all of its critical data to one file, called a master file that provides a common point of reference. Critical data are the Master data, which indicates business entities like employees, customers, citizens, assets, locations etc. When properly done, master data management streamlines data sharing among

[57] European Commission, Directorate-General for Informatics, (2018), "Interoperability Maturity Model: Report on Benchmark Results. 2017 Edition". Luxembourg: Publications Office of the European Union.

[58] European Commission (2017), "New European Interoperability Framework", p. 5.

[59] New European Interoperability Framework, Promoting seamless services and data flows for European public administrations.

personnel and departments. In addition, master data management can facilitate computing in multiple system architectures, platforms and applications[60].

Metadata describes what data an organisation has, what they represent, how they are classified, where they came from, how they move within the organisation, how they evolve through use, who can and cannot use them, and whether they are of high quality[61].

Maturity Assessment is the process to evaluate the level and state of the implemented data management framework using models. Preformatted questionnaires are used that define the state, and retrieve strengths and weaknesses of the implementation of the framework found in its various phases.

Data architecture is described by an integrated collection of master design documents at different levels of abstraction, including standards that govern how data are collected, stored, arranged, used and removed. It is also classified by descriptions of all the containers and paths that data takes through an organisation's systems.

Technology architecture is a description of the structure and interaction of the platform services, and logical and physical technology components[62].

4.2 SOURCES OF INPUT

This Framework is grounded on both previous works developed by the European Commission (EC) and on relevant initiatives carried out in the private sector (such as Governance Frameworks, Information and Data Management Guidelines). These initiatives are referenced directly or indirectly throughout the whole document in the sense that they were used as a source of inspiration.

These initiatives relate, accordingly, to those that stem from the Interoperability Solutions for public Administrations[63] (ISA^2) programme, and from other EC initiatives, such as the SEMIC's core vocabularies and guidelines[64], the Sharing and Reuse (S&R) Action[65] or the Interoperability Maturity Assessment of a Public Service (IMAPS), among other. Additional sources of input and/or inspiration were:

- The European Interoperability Framework (EIF);
- The European Interoperability Reference Architecture (EIRA);
- The Corporate Information Management Framework (CIMF);
- The Enterprise Information Architecture (EIA), prepared by DIGIT's Methodology and Programme Office.
- Common assessment method for standards and specifications (CAMSS)[66];

[60] Master data management, Wikipedia.

[61] DAMA – DMBOK Data Management Body of Knowledge, second edition.

[62] ADM Phase D: Technology Architecture.

[63] Interoperability Solutions for Public Administrations, https://ec.europa.eu/isa2/home_en.

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

[65] https://ec.europa.eu/isa2/actions/promoting-sharing-and-reuse-interoperability-solutions_en.

[66] https://ec.europa.eu/isa2/solutions/camss_en.

- National Interoperability Framework Observatory (NIFO)[67];
- Commission Enterprise Architecture Framework (CEAF);
- Different Interconnected Base Registries frameworks already implemented in different Member States, such as Denmark[68], Estonia[69], Sweden[70], Ireland[71];
- Information provided by Member States during dedicated interviews[72];
- Singapore's and OECD's presentations;
- State of the art analysis on industry standard data management frameworks and maturity models.

Moreover, various frameworks for IT and Data development and governance were also analysed and used as sources of inspiration. For example, the DAMA Body of Knowledge Framework,[73] and the Data Management Capability Assessment Method[74] were taken into consideration, when defining the BRAIF conceptual model (cf. section 2) and its parts. Moreover, several maturity assessment models were considered, and we concluded in using the DMM – CMMI model whose findings of analysis have been integrated in this document.

The COBIT and ISO/IEC 38500 were also analysed — though more from their data governance perspective — and initial findings of analysis have been integrated in various parts of the document, especially in section 2.3.1 Master data management.

4.3 STANDARDS SUPPORTING METADATA

Table 7. Standards supporting Metadata

Standard	Scope	Description
ISA² Core Vocabularies[75]	Entities	The ISA² programme defines core vocabularies as *"[…] simplified, re-usable and extensible data models that capture the fundamental characteristics of an entity in a context-neutral fashion."*[76] Currently there are 6 elements in the Core Vocabularies list as defined below. - *Core Business Vocabulary* – Used to capture the main characteristics of legal entities;

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

[68] Factsheet: Access to Base Registries in Denmark.

[69] Factsheet: Access to Base Registries in Estonia.

[70] Factsheet: Access to Base Registries in Sweden.

[71] PUBLIC SERVICE DATA STRATEGY 2019 – 2023.

[72] All interviews are available on BR Collection on Joinup: https://joinup.ec.europa.eu/collection/access-base-registries.

[73] Mark Mosley and Michael Brackett, eds, (2009), "The DAMA Guide to The Data Management Body of Knowledge (DAMA-DMBOK Guide)", first edition. Bradley Beach: Technics Publications, LLC; and Patricia Cupoli, ed, (2014), "DAMA-DMBOK2 Framework", pp. 1-27, available online https://dama.org/sites/default/files/download/DAMA-DMBOK2-Framework-V2-20140317-FINAL.pdf.

[74] EDM Council, (2014), "Data Management Capability Assessment Model (DCAM)", working draft version 0.7, https://dgpo.org/wp-content/uploads/2016/06/EDMC_DCAM_-_WORKING_DRAFT_VERSION_0.7.pdf.

[75] https://joinup.ec.europa.eu/page/core-vocabularies.

[76] https://joinup.ec.europa.eu/page/core-vocabularies.

		- *Core Location Vocabulary* – Used to capture the main characteristics of a location. - *Core Person Vocabulary* – Used to capture the main characteristics of a person; - *Core Public Service Vocabulary* – Used to capture the fundamental characteristics of a service offered by public administration; An application profile of this core vocabulary (CPSV-AP) has been developed to describe public services and to group them in business events. - *Core Evidence and Criterion Vocabulary* – A set of data for private entity description, used by public entities to define the criteria for private entities for public procurement eligibility and qualification, allowing a private entity to carry out public services or take part in a public procurement; - *Core Public Organisation Vocabulary* – A data model used to describe public organisations in the European Union.
RDF[77]	Semantics	RDF, which stands for Resource Description Framework, is a standard model used to exchange data mainly on the Web. According to the W3C, RDF *"[…] extends the linking structure of the Web to use URIs to name the relationship between things as well as the two ends of the link (this is usually referred to as a 'triple')."*[78] RDF facilitates data merging and supports the evolution of schemas without requiring changes in the data consumers.
OWL[79]	Semantics	OWL, which stands for The Web Ontology Language, is a family of languages used to represent knowledge. The W3C defines OWL as *"[…] a Semantic Web language designed to represent rich and complex knowledge about things, groups of things, and relations between things,"* which is also known as an ontology.
DCAT[80]	Datasets	DCAT, which stands for Data Catalogue Vocabulary, is an RDF vocabulary defined to model catalogues published on the Web. According to the W3C, *"[…] by using DCAT to describe datasets in data catalogues, publishers increase discoverability and enable applications easily to consume metadata from multiple catalogues."*[81] This characteristic makes DCAT a key enabler of interoperability, when datasets are involved.

[77] https://www.w3.org/RDF/.

[78] https://www.w3.org/RDF/.

[79] https://www.w3.org/OWL/.

[80] https://www.w3.org/TR/vocab-dcat/.

[81] https://dvcs.w3.org/hg/gld/raw-file/default/dcat/index.html.

ADMS[82]	Assets	ADMS, which stands for Asset Description Metadata Schema, is a profile of DCAT defined by the W3C to describe assets. In the context of ADMS, an asset should be understood as "*[…] highly reusable metadata (e.g. xml schemata, generic data models) and reference data (e.g. code lists, taxonomies, dictionaries, vocabularies) that are used for eGovernment system development."*[83]
Provenance Standards	Entities	Provenance standards refer to vocabularies used to model the entities, producing specific data. The provenance standards are defined to bring quality, reliability and/or trustworthiness to the exchange of datasets. The W3C has defined for this purpose two main artefacts: - Provenance Data Model (PROV-DM)[84], used to define "*[…] the core structures forming the essence of provenance information."*[85] - Provenance Ontology (PROV-O) [86] that is a representation of PROV-DM in the form of an OWL ontology.
SKOS[87]	Taxonomies, thesauri and classification schemes	SKOS, which stands for Single Knowledge Organisation System, is defined by the W3C as "*[…] an area of work developing specifications and standards to support the use of knowledge organization systems (KOS) such as thesauri, classification schemes, subject heading lists, taxonomies (…) within the framework of the Semantic Web."*[88]
SPARQL[89]	Semantic queries	SPARQL is a query language that is used to query and manipulate data from RDF graphs. It provides capabilities for *"[…] querying required and optional graph patterns along with their conjunctions and disjunctions […]"* and *"[…] supports aggregation, sub-queries, negation, creating values by expressions, extensible value testing, and constraining queries by source RDF graph."*[90]

4.4 SUPPORT TO THE EIF PRINCIPLES

4.4.1 Defining and implementing security measures

Security is a major concern in the provision of integrated public services. It is very important to create a trust environment, where citizens and businesses are confident about interacting with public

[82] https://www.w3.org/TR/vocab-adms/.

[83] https://www.w3.org/standards/techs/rdfvocabs#w3c_all.

[84] https://www.w3.org/TR/2013/REC-prov-dm-20130430/.

[85] http://www.w3.org/TR/prov-dm/.

[86] https://www.w3.org/TR/2013/REC-prov-o-20130430/.

[87] https://www.w3.org/2004/02/skos/.

[88] http://www.w3.org/2004/02/skos/overviewtemplate.html. Note: the authentification is required to access this website.

[89] https://www.w3.org/TR/sparql11-query/.

[90] http://www.w3.org/TR/sparql11-query/.

authorities in a secure way. Access to Base Registries has to be regulated to ensure safety and, at the same time, its compliance with different legal requirements regarding data protection and privacy has to be guaranteed. These are "privacy-by-design" and "security-by-design" approaches, as defined in EIF. Furthermore, access and control mechanisms should be governed by the principles of information stewardship. The steward is accountable for collecting, using, updating, maintaining and deleting information, as well as for all the processes carried out as part of data security management.

4.4.2 Fostering semantic and technical interoperability

Systems have to be able to communicate effectively between each other in order to allow successful delivery of Base Registries' data to integrated public services. The interfaces between them should be a core component, enabling this communication.

Regarding semantic and technical means, both of them are integral elements of the interoperability model foreseen by EIF. Semantic interoperability ensures that the exchanged data are understood throughout the whole delivery process, while technical interoperability covers all infrastructures, linking the systems and services.

When delivering interoperable public services, the principle of openness refers not only to data, but also to specifications and software. This implies that all documentation needed for the reuse of solutions has to be publicly available and should ideally contain no restrictions for its use. Reusability is a key interoperability enabler, and therefore, a major contributor to the development of the Digital Single Market.

4.4.3 Improving reusability and data sharing

The registration of machine-readable, well-defined and effective metadata is an extremely valuable practice when sharing data between registries, services and their final users' systems. It plays a key role in improving data reusability and sharing, two pillars of EIF. To fully exploit the benefits of having a good metadata structure in place, the aforementioned metadata have to be produced prior to the registration of the master data. A common vocabulary shall be established, too. Registering metadata fosters the implementation of the once-only principle by enabling the retrieval of information from its original source, and, therefore, prevents different actors from re-registering information that is already held by other Base Registries or authoritative systems.

4.4.4 Ensuring quality preservation

As stated in the EIF, the accuracy, correctness and completeness of the data contained in Base Registries can only be ensured by designing and following a quality assurance plan. In order to do that, a precise strategy needs to be followed, including an assessment of the current state of affairs, the definition of clear objectives and the requirements based upon them.

4.5 INTEROPERABILITY AREAS

Interoperability areas are *"[…] the object of measurement in the IMAPS, specifying where interoperability plays a role from a service management, service delivery and service consumption viewpoint."*[91] Therefore, IMAPS distinguishes three main areas, as presented in the table below.

[91] ISA Action 2016.37: Interoperability Maturity Model, D03.03 – IMPAS Guideline, 15 January 2018, p. 7. https://joinup.ec.europa.eu/sites/default/files/solution/documentation/2018-01/SC506%20-%20ISA2%20Action%202016.37%20-%20IMAPS%20Guideline_0.pdf.

Table 8. Interoperability areas

Area	Description
Service Delivery	Related to the way public administrations deliver public services to their end users – i.e. citizens, businesses or other administrations. In the context of this Framework, it refers to the means which enable access and interconnection of base registries.
Service Consumption	Related to the consumption of reusable machine-to-machine services from other public administrations, businesses or Base Registries.
Service Management	Focuses mainly on aspects in the area of sharing and reuse and the design of public services. This area includes Service Management aspects, spanning from enterprise architecture and service level agreement, to service documentation and application source code.

The interoperability of Base Registries is triggered by the existence (or the need) of an integrated digital public service, and it is modelled based on service consumption, service delivery and service management. A Base Registry (**service consumption**) might consume a service that is delivered by another Base Registry. When both Base Registries can collaborate to deliver the digital integrated public service, then we have **service delivery**, and where the management and coordination of the whole service fall under the responsibility of a third party – like the European Commission – we have **service management**.