



DG Joint Research Centre

Study and guidelines on  
Geospatial Linked Data as part  
of ISA Action 1.17

Persistent Identifiers (PIDs)

5th May 2014

Alice Vasilescu – Deloitte

Christian Hauschildt Deloitte

# Agenda



- 1 Introduction
- 2 Context
- 3 Approach
- 4 GOFA objectives
- 5 TO-BE GOFA
- 6 Wrap-up, next steps

## 1. Setting the scene



## 2. Developing proposals



## 3. Refinement & recommendations



## GUIDELINES



## State-of-play webinar

We provided an outline of the study and our work so far with regards to PIDs. This webinar was also an opportunity to provide feedback and to exchange experiences: **What barriers could there be for governing PIDs?**

## Today's webinar: Guidelines on methodologies

Presentation of the first version of the guidelines for a common approach to PID governance: **opportunities for community-led improvements.**

Persistent  
Identifiers  
(PID)

- 1 Introduction
- 2 Context
- 3 Approach
- 4 GOFA objectives
- 5 TO-BE GOFA
- 6 Wrap-up, next steps



This study has been prepared in the context of the Interoperability for European Public Administrations (ISA) Programme and, in particular **A Reusable INSPIRE Reference Platform** (ARE3NA, ISA Action 1.17)

This study should provide:

1. Shared evidence about the current status in Europe of linked (geospatial) data related to INSPIRE.
2. An initial common/agreed methodology and guidelines towards **RDF encodings for INSPIRE**
3. Recommendations for how location **PIDs** could be governed for INSPIRE and other relevant activities.

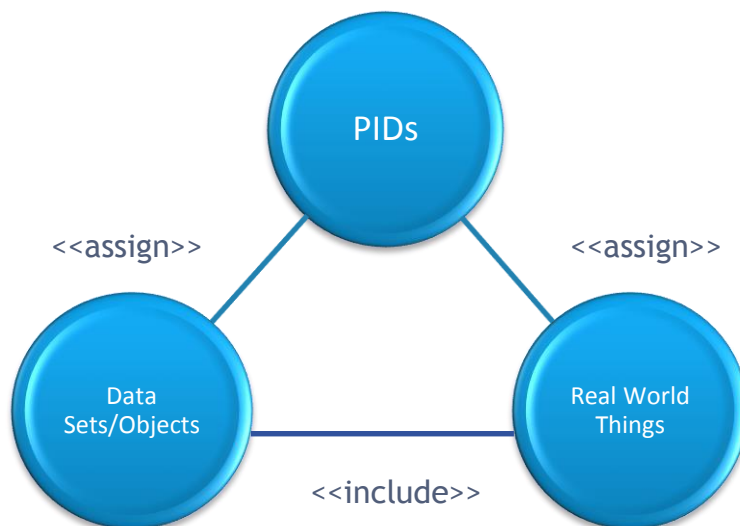




**PID** = A persistent identifier (PI) is a **long-lasting reference to a “real-world things” or a digital object**—a single file or set of files. Noted persistent identifier systems include: Archival Resource Keys (ARKs), Digital Object Identifiers (DOIs), Persistent Uniform Resource Locators (PURLs), Uniform Resource Names (URNs), and Extensible Resource Identifiers (XRIs)

## Linked Data Principles

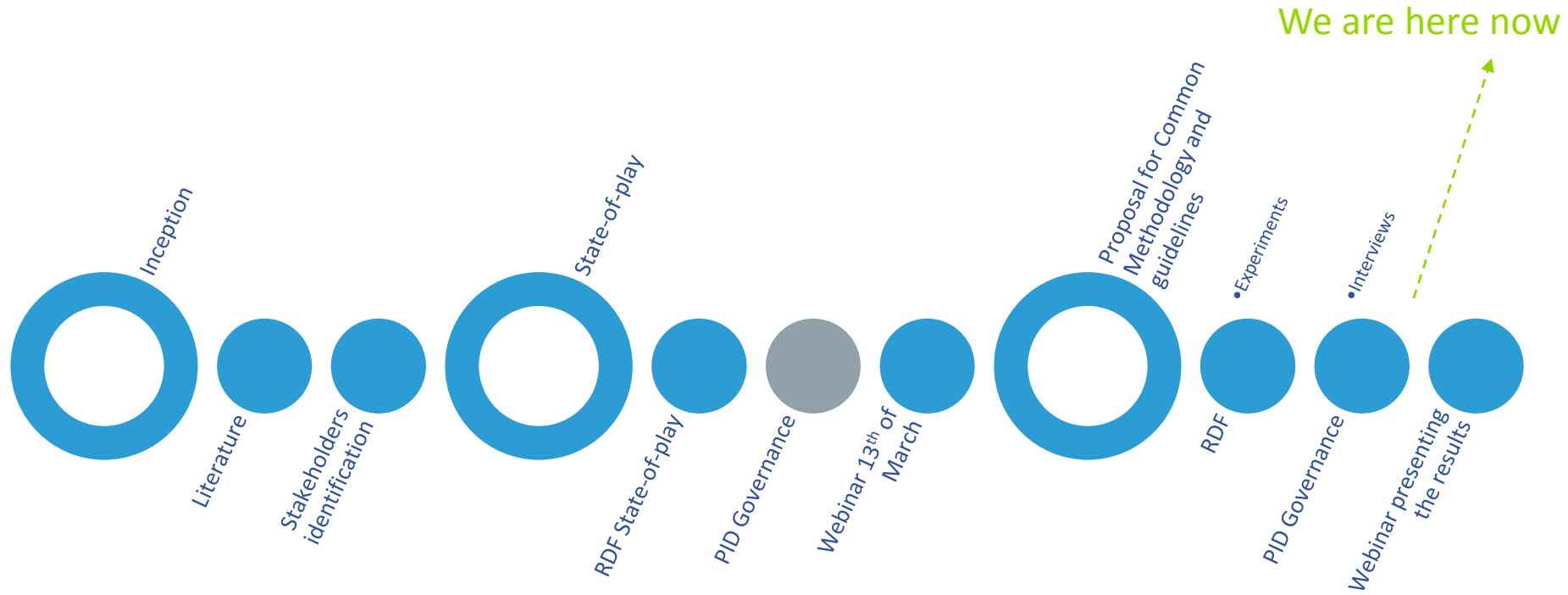
- 1) Use URIs as names for things
- 2) Use HTTP URIs, so that people can look up those names
- 3) When someone looks up a URI, useful information should be provided, using the standards (RDF, SPARQL)
- 4) Include links to other URIs, so that they can discover more things



**Resources** on the web that contain data about real-world things

**Real-world things** such as a river basin, train station, a country, etc.

# Timeline



- 1 Introduction
- 2 Context
- 3 Approach
- 4 GOFA objectives
- 5 TO-BE GOFA
- 6 Wrap-up, next steps

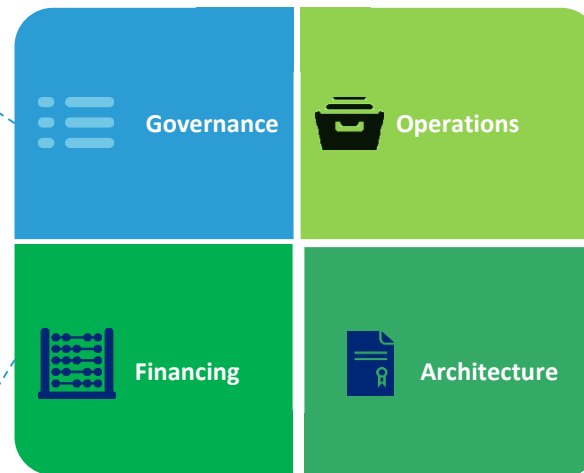


# Approach - GOFA



**G**overnance refers to the control and the steering of the decisions on PIDs

**O**perations is about the processes and tools needed to run PIDs

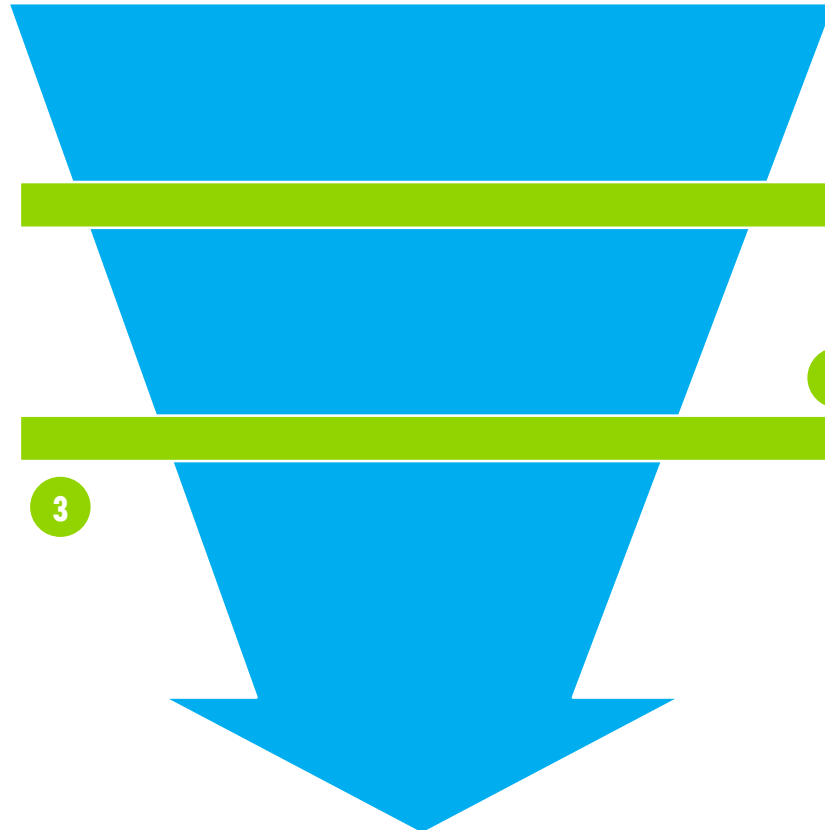


**F**inancing is about the resources needed for the operations and the architectural updates

**A**rchitecture corresponds to a formal specifications around PIDs



1



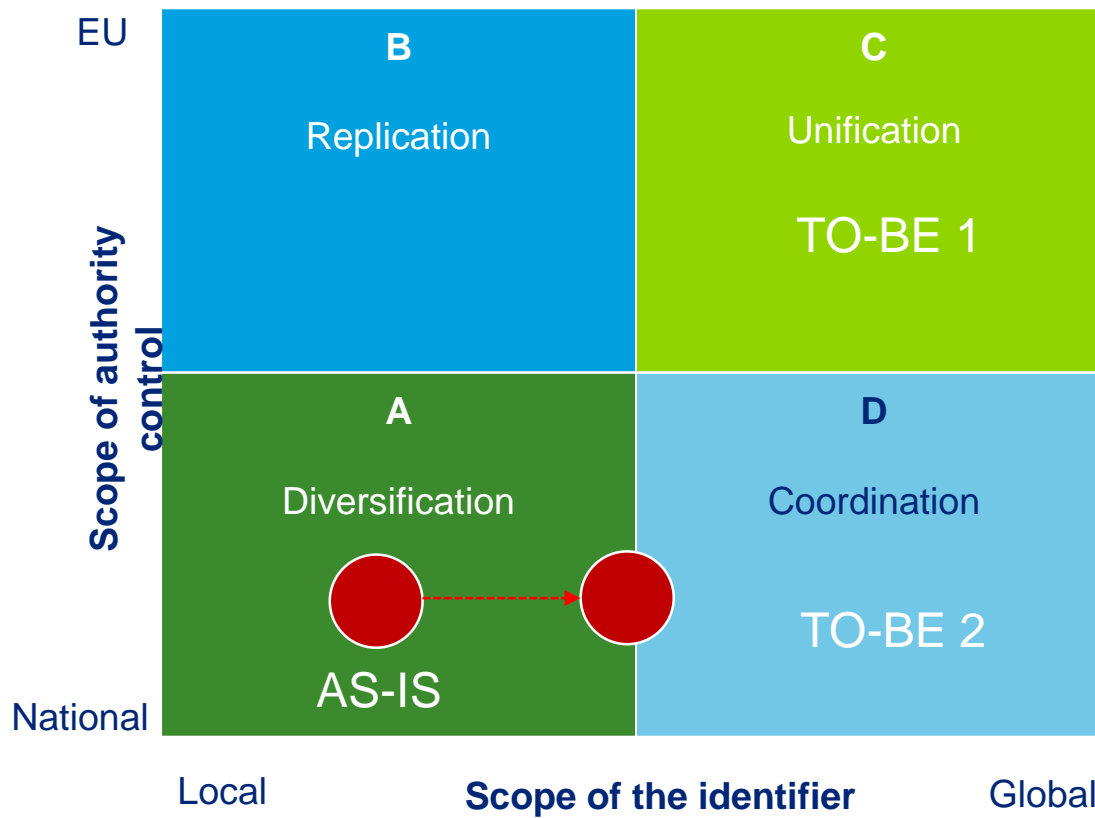
2

Evaluate using GOFA

3

1. AS-IS GOFA
2. Challenges for the TO-BE GOFA
3. Requirements of the TO-BE GOFA

# Step 1 – Data collection



# Step 1 – Data collection



Collecting Best practices, that meet the Requirements and tackle the Issues



## Governance



URIs Policy		
URIs Organisational structure		



## Operations

Domain Name Registrar		
Namespace Register		
Redirection Register		
Long Term Preservation Register		



## Financing

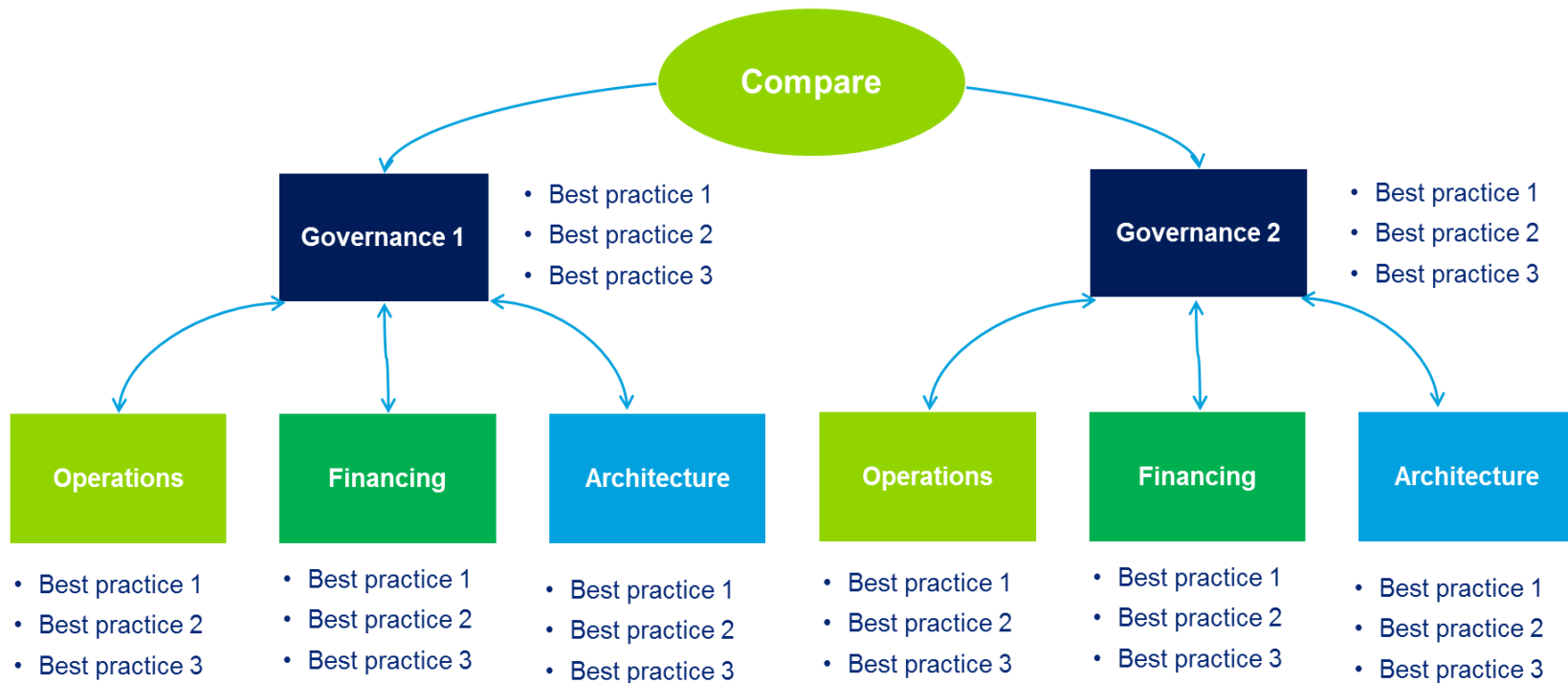
Business Case		
Cost Model		



## Architecture

URIs Naming and Design rules		
Digital services		
RDF		

# Step 2 – Compare TO-BE GOFAs



# Step 3 – Recommendations



## Governance 1

- Best practice 1
- Best practice 2
- Best practice 3

## Operations

- Best practice 1
- Best practice 2
- Best practice 3

## Financing

- Best practice 1
- Best practice 2
- Best practice 3

## Architecture

- Best practice 1
- Best practice 2
- Best practice 3




Recommendations  
Check list  
Conclusions

- 1 Introduction
- 2 Context
- 3 Approach
- 4 GOFA objectives
- 5 TO-BE GOFA
- 6 Wrap-up, next steps

# Coverage



 Input reviewed

 Unknown yet



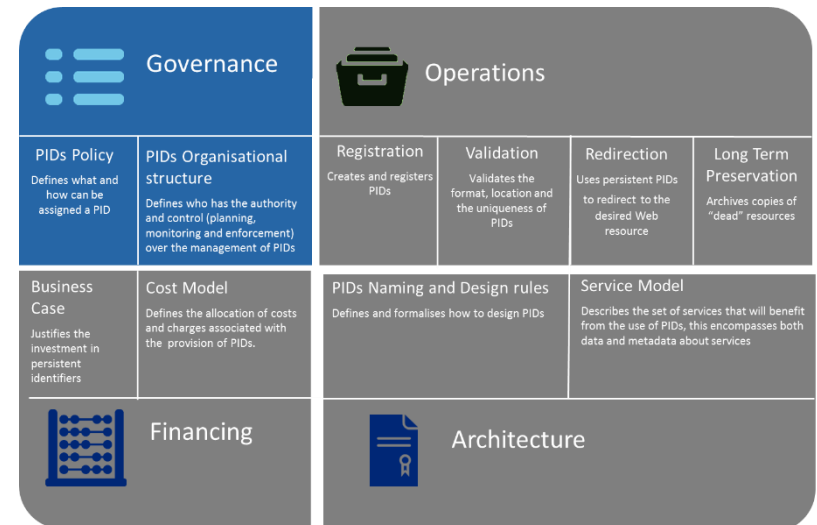


It is important to note that the choice of governance scenarios affects:

- What needs to be governed (In unification much more needs to be governed, than in coordination)
- Who should be involved at what level
- The governance structure
- The role of the European Commission
- The role of national and sectorial communities

The governance scenario also affects

- Operations
- Financing
- Architecture



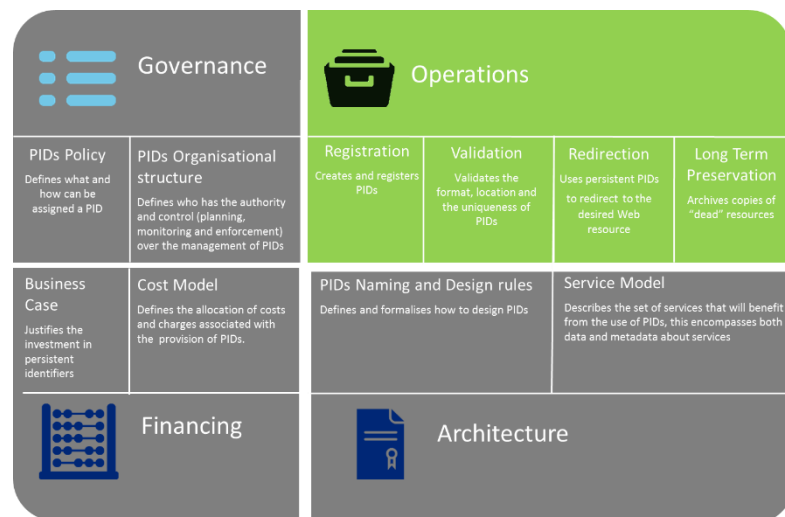


As a result of the chosen scenario, the Operations will also be affected:

- The scope, objectives and responsibilities of operations
- The needed processes and procedures
- The type of organisation that is needed
- The number and type of staff needed

The operations affects:

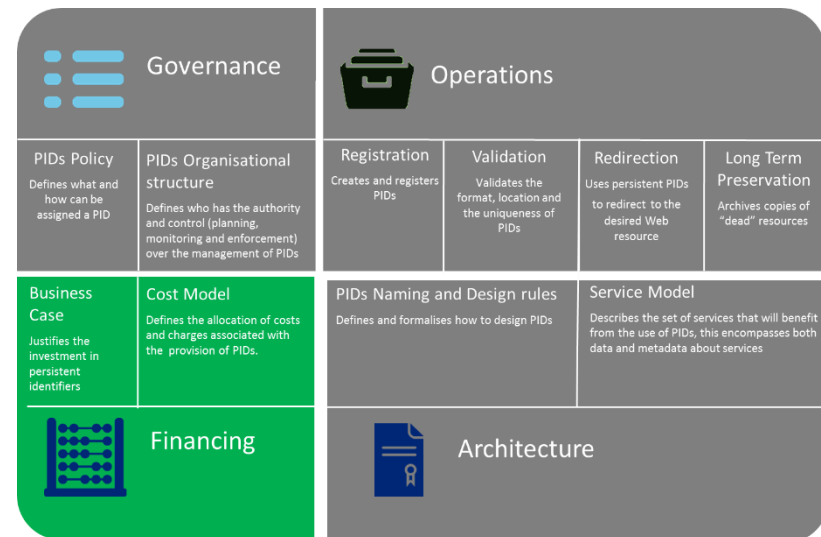
- The needed level of financing
- To some extent, the operational (and organisational) structure (taking into account the structures within the Member States) can affect architecture to some extent – e.g. in terms of naming and design rules





## Financing is also affected by the choice of scenario:

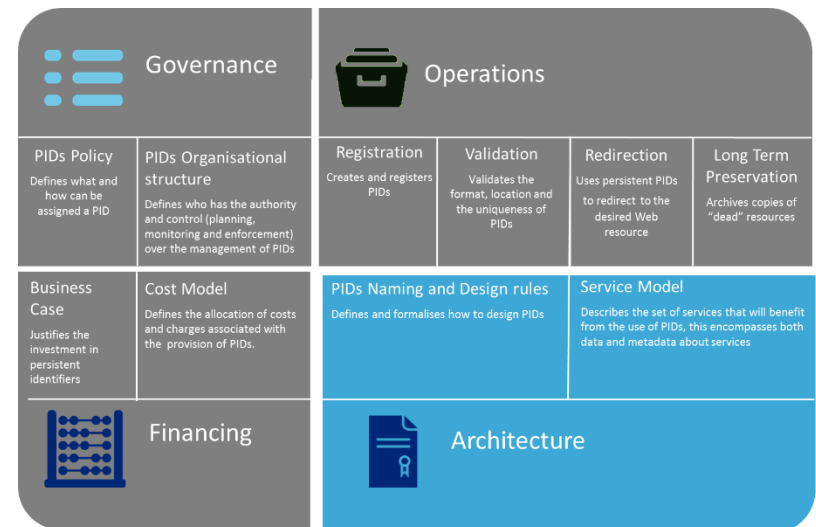
- The level of financing needed, will depend on the governance scenario and the operational setup
- Whether it is mainly EU or Member State financing will depend on the scenario
- The source of financing will depend on the scenario





## Architecture is also affected by the choice of scenario:

- In unification, the architecture is centralized, in coordination, the architecture is decentralized/federated
- The design patterns – for unification there could be going towards only one design pattern; for federation, the choice of design pattern remains with the data owner
- The root domain used for the PID registration (for unification there would be one, for coordination there would be as many as issuing authorities)





## What needs to be governed:

- **Roles and Responsibilities**
- **Delegation of responsibilities**
- **Liabilities (who is liable for what)**
- **Standards to be used**
- **Life-cycle approach**
- **Policies**
- **Validation of PIDs**
- **National communities vs. sectorial (cross-border) communities**
- **How to drive the maturity of the communities – moving from less mature national communities to sectorial (cross-border) communities**
- **Monitoring the external environment and developments within it**
- **Development and use of common models and reference data**
- **Considering the “LOST” perspectives of interoperability, e.g. are there any legal constraints in the way things should be dealt with at National level (obviously things that are not covered by INSPIRE, PSI Directive and any other European policies)**



- 1 Introduction
- 2 Context
- 3 Approach
- 4 GOFA objectives
- 5 TO-BE GOFA
- 6 Next steps



## European Commission

- 1) Take **bottom-up** approach, keep it **simple** and **transparent**
- 2) **Coordinate** the national and sectorial communities
- 3) **Develop** and put forward **guidelines** and best practices for all the areas of the GOFA
- 4) **Identify** and **analyze** national and sectorial **communities**. Every community has its specific needs and Use Cases. Define those Use Cases.
- 5) Recommend generic **roles** and **responsibilities** to be further developed within a community
- 6) Set up a **network of MOU's** at least with respect to important reference sets
- 7) Encourage and support the **communication** and **consultation** within each community
- 8) Enhance **interoperability** by promoting the use of common standards and best practices

## Organizations

- 1) **Coordinate** within the relevant community
- 2) Define **URI Policy** using the EC guidelines:
  - a) **What** you are identifying
  - b) **How long** is it persistent
  - c) **What** are the processes, including life cycle
  - d) **Who** is responsible for what
- 3) Define the **roles** and **responsibilities** with regards to PID management following EC recommendation
- 4) The **PID management** is the responsibility of the data owner and should be part of the written procedures and processes of an organization
- 5) Define the **life cycle** within the relevant community, following best practices
- 6) Agree on how the **shared resources** should be managed and who is responsible
- 7) Define how the resources that **change frequently** should be managed



## DOs

- 1) Do define a **URI policy**
- 2) Define and publish how long is the **persistence** and who is **responsible**
- 3) Do define and publish **life-cycle**
- 4) The **PID management** should be part of the written procedures and processes of an organization

## DON'Ts

- 1) Do not duplicate namespaces





## Key considerations when designing the governance structure – making it sustainable:

- **Stakeholder representation** – All relevant stakeholders have to be represented either directly in the decision-making or indirectly through stakeholder consultation
- **Flexibility** – it is important that the governance structure be flexible so that it can be adapted to changes without having to be redesigned completely.  
E.g. if new Member States join, it should be possible to incorporate these easily also, if some the maturity of some sectors/communities increases from being a part of national communities to being a separate sectorial (cross-border) community, it should be possible to adapt this in the governance easily
- **Mandate** – The representatives in the governance structure have to have the mandate to be able to make decisions and ensure their subsequent implementation
- **Expertise** – In parallel with the Mandate, it is also necessary that the decision-makers have the expertise needed or at least expert input from relevant experts (e.g. by including them in the working groups or by consulting them)
- **Taking into account the external environment** – It is important to take into account that there is a PID environment outside the scope of INSPIRE and that there may be a need to link to the external environment and take developments in this environment into account



## Stakeholder Groups:

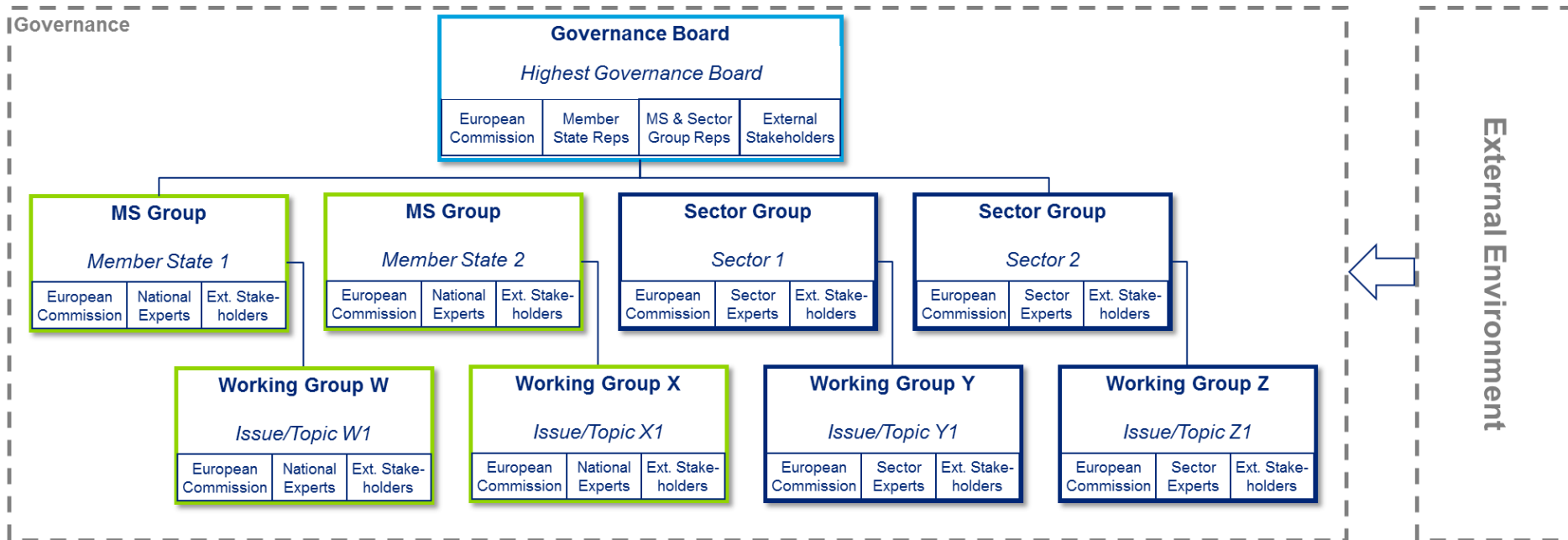
The different stakeholders that are included in the organisational structure are:

- **European Commission** – Representatives from relevant DGs, generally with the role of coordinating and driving the work (in the MS Groups, they are mainly observers, but should also provide input on how to drive the different sectors towards more maturity to create sectorial (cross-border) communities)
- **Member State Representatives** – These are representatives from the Member States with the mandate to make decisions and ensure the implementation in the Member States
- **National Experts** – These are experts from different parts of the national communities
- **Sector Experts** – These are sector experts from the different Member States (and sectorial communities)
- **External Stakeholders** – These are representatives from stakeholder groups outside the scope of INSPIRE, these can e.g. be representatives from other PID initiatives or representatives from relevant standardisation bodies or industry representatives. These are included or consulted on an as-needed basis.



## An example of what the organisational structure could look like:

- The illustration below shows a modular approach to the organisational structure, where different groups (with working groups if needed) can be added, changed and removed.
- The illustration also shows the importance of taking into account the external environment, which can be a challenge



- **Member State and Sector Representatives for the Governance Board** – These are representatives from the MS and Sector Groups so that the different groups are represented on the board



## Roles and Responsibilities:

Different roles and responsibilities should be defined once the governance scenario has been decided upon. An example of the roles and responsibilities based on the presented organisational structure could be:

- **Governance Board:**
  - **Responsibilities:** Making the overall decision that fall within the different topics/issues/objectives of the governance
  - **Composition:** European Commission, Member State Representatives, Member State and Sector Group Representatives as well as (relevant) external stakeholders
  
- **Member State Groups:**
  - **Responsibilities:** These are groups are formed at Member State level and are responsible for ensuring the lower-level governance of the sectors that are not yet mature enough to have sectorial (cross-border) communities. It is recommended that although the groups are at Member State level, that these be carried out in close cooperation with the overall governance, but also with the European Commission, in order to be able to drive the maturity of the different sectors towards sectorial (cross-border) communities
  - **Composition:** European Commission, National Experts and External Stakeholders



## Roles and Responsibilities (2):

Different roles and responsibilities should be defined once the governance scenario has been decided upon. An example of the roles and responsibilities based on the presented organisational structure could be:

- **Sector Groups:**
  - **Responsibilities:** These are groups are formed at EU level (based on existing cross-border sectorial groups) and are responsible for ensuring the lower-level governance of the sectors and discussing sector specific issues
  - **Composition:** European Commission, Sector Experts and External Stakeholders
- **Working Groups:**
  - **Responsibilities:** These working groups are responsible for analysing and providing different options and recommendations on specific topics/issues/objectives of the governance. These can be linked to Member State Groups, Sector Groups or even directly to the Governance Board if needed. Their role is to provide expert input.
  - **Composition:** European Commission, Member State/Sector Experts and External Stakeholders



## European Commission

- 1) Provide guidelines on how to document the processes that need to be in place (registration, validation, mapping, resolving, long term preservation)
- 2) Provide a register or registers for registration
- 3) Give guidelines for the **validation process** (e.g. metadata validation)
- 4) Guidelines for **SLA definition**

## Organizations

- 1) **Apply guidelines** and **best practices** coming from the EC
- 2) Within the relevant community the organizations should agree upon the **PID management processes** and document them
- 3) If needed , create an SLA for your service

## DOs

- 1) Do document your **processes and publish** them when needed
- 2) Do define the **events** that can trigger **changes** in the PID scheme and what are the next steps
- 3) Do document the **measures** that need to be taken when **change** occurs
- 4) Document how **shared resources** are managed

## DON'Ts

- 1) Don't **reassign** identifiers unless there was a mistake ( you have linked the wrong resource to the identifier)



## European Commission

- 1) Give guidelines for PID management cost model definition

## Organizations

- 1) When selecting a PID solution take into account the costs that come with the PID management, your requirements
- 2) Define who has to pay and how

## DOs

- 1) Formalize your business case for PID
- 2) Define who has to pay and how

## DON'Ts



## European Commission

- 1) Provide **evaluation framework** for tool selection
- 2) Recommend **standards** and **best practices** for PID design and management
- 3) Potentially provide a **tool** or a **service** for **validation**
- 4) Support both **URIs** and **DOIs** approach
- 5) Provide a **register of registers** to ensure uniqueness. A good starting point can be a federation of existing registries

## Organizations

- 1) Define within the community the **design pattern** that best suit the needs of the community
- 2) Select the most **suitable tool** for PID management using EC recommendation
- 3) Chose your **domain name** following best practices
- 4) Identify and document the specific **needs** for your PI System and opt for the **flexible and reliable** solution that support all types of resources

## DOs

- 1) If you do not have design principles, use some of the most referenced work **10 Rules for Persistent URI** and **Designing URI sets for UK Public Sector**
- 2) Chose your domain name that is stable and is under your control (e.g. data.gov.uk)

## DON'Ts



- 1 Introduction
- 2 Context
- 3 Approach
- 4 GOFA objectives
- 5 TO-BE GOFA
- 6 Next steps

# Next steps



- **Finalise the initial recommendations based on today's input**
- **Discuss and prioritise the next steps/actions**
- **Gather lessons learnt and continuously improve the guidelines**

## Thank you for your participation!



Join the collaborative platforms in  
other areas of Open Source &  
Semantic Interoperability on  
<https://joinup.ec.europa.eu/>

joinup

