Data Governance at the National Information Processing Institute in Poland

Emil Podwysocki

Maria Bylina

"The 21st century belongs to those who treat information as a strategic company resource. Information management, data resource management, and information quality and its improvement should become a natural element in the strategic planning process, both in the short and long term."

Larry English - INFORMATION IMPACT International, Inc.

OPI PIB

The Data Governance Act (DGA)

The Data Governance Act (DGA) aims to create a legal framework for data sharing for the benefit of the European single market, ensuring neutral access to data and interoperability, as well as helping to avoid lock-in effects (recital 2 DGA). The European Parliament adopted the DGA on **6 April, 2022**. As a regulation, the DGA does not have to be transposed into national law by EU member states. It entered into force 23 June, 2022. It will be applicable after a transitional period of 15 months from **24 September, 2023**.

https://cms-lawnow.com/en/ealerts/2023/02/the-data-governance-act-overview
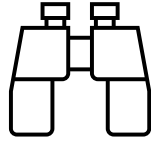
OPI PIB

The Data Governance Act (DGA)

Increase data availability

- ✓ **Reuse** of certain categories of protected data held by public sector bodies

- ✓ **Data intermediation services** as a "key role in the data economy"

- ✓ **Data altruism** – increasing the availability of data by voluntarily donating it
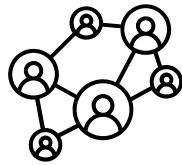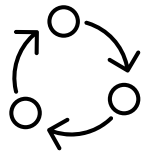
The FAIR data principles
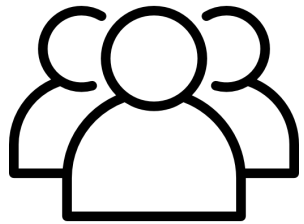
Findability

Accessibility

Interoperability

Reusability

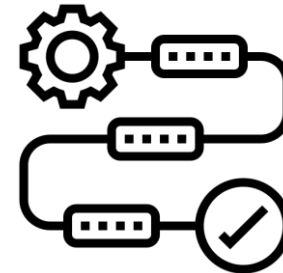# Introduction to the data governance framework

Data governance

Data governance is a framework and set of processes that ensure the proper management of an organisation's data assets.

People
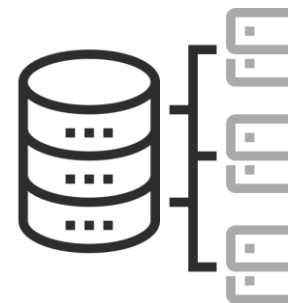(Chief Data Officer, Data Steward, Data Owner, Data Quality Manager, etc.)

Processes
(formalised procedures for data discovery, classification, quality control, and access management)

Policies
(clearly defined rules and guidelines that govern data and metadata management, usage, security, and quality practices)

Standards
(data quality and data architecture)

Data governance - DAMA

DAMA International created the Data Management Framework (the DAMA Wheel) to illustrate how different subjects are interconnected.

The DAMA Wheel includes the 11 knowledge areas and places data governance at its centre to highlight how it underpins every data management activity.

Data governance pertains to the full lifecycle of data in an organisation (collection, storage, sharing, archiving, and deletion).
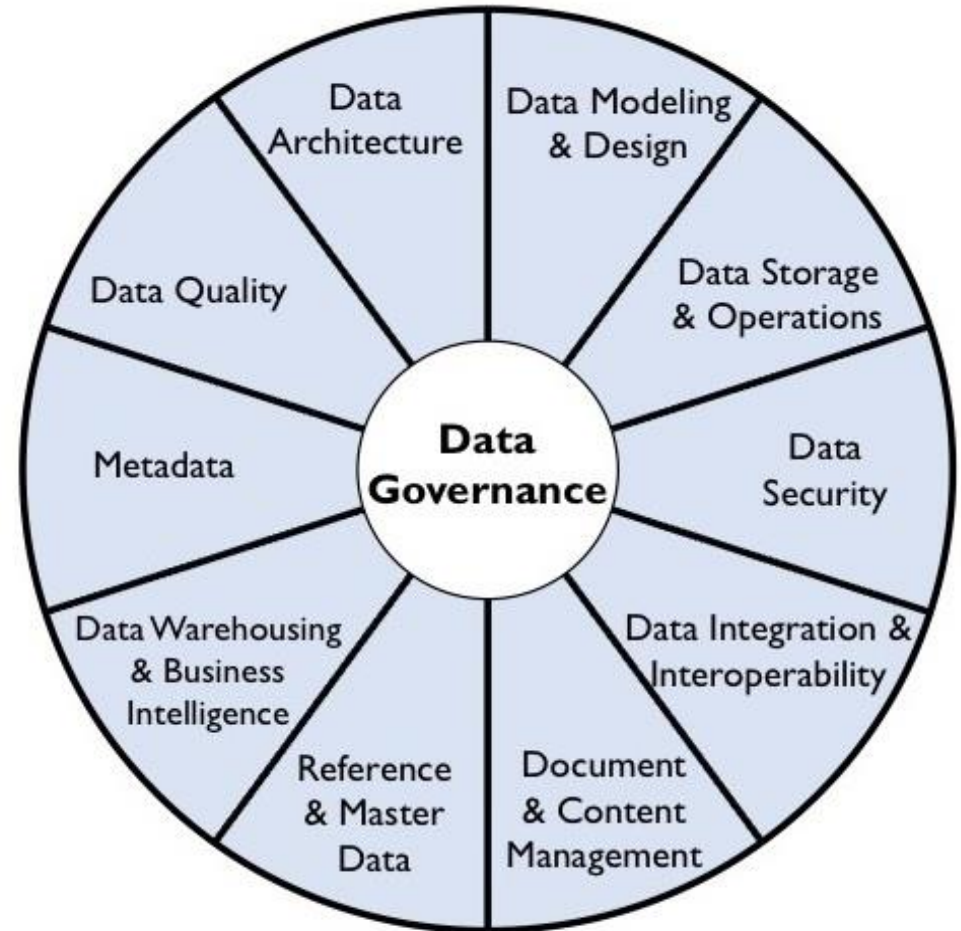
Copyright© 2017 DAMA International

Data governance benefits

What are the benefits of implementing data governance?

- ✓ Enables an organisation to manage data as an asset

- ✓ Ensures that data is used properly, and people have access to the data they need to make business decisions

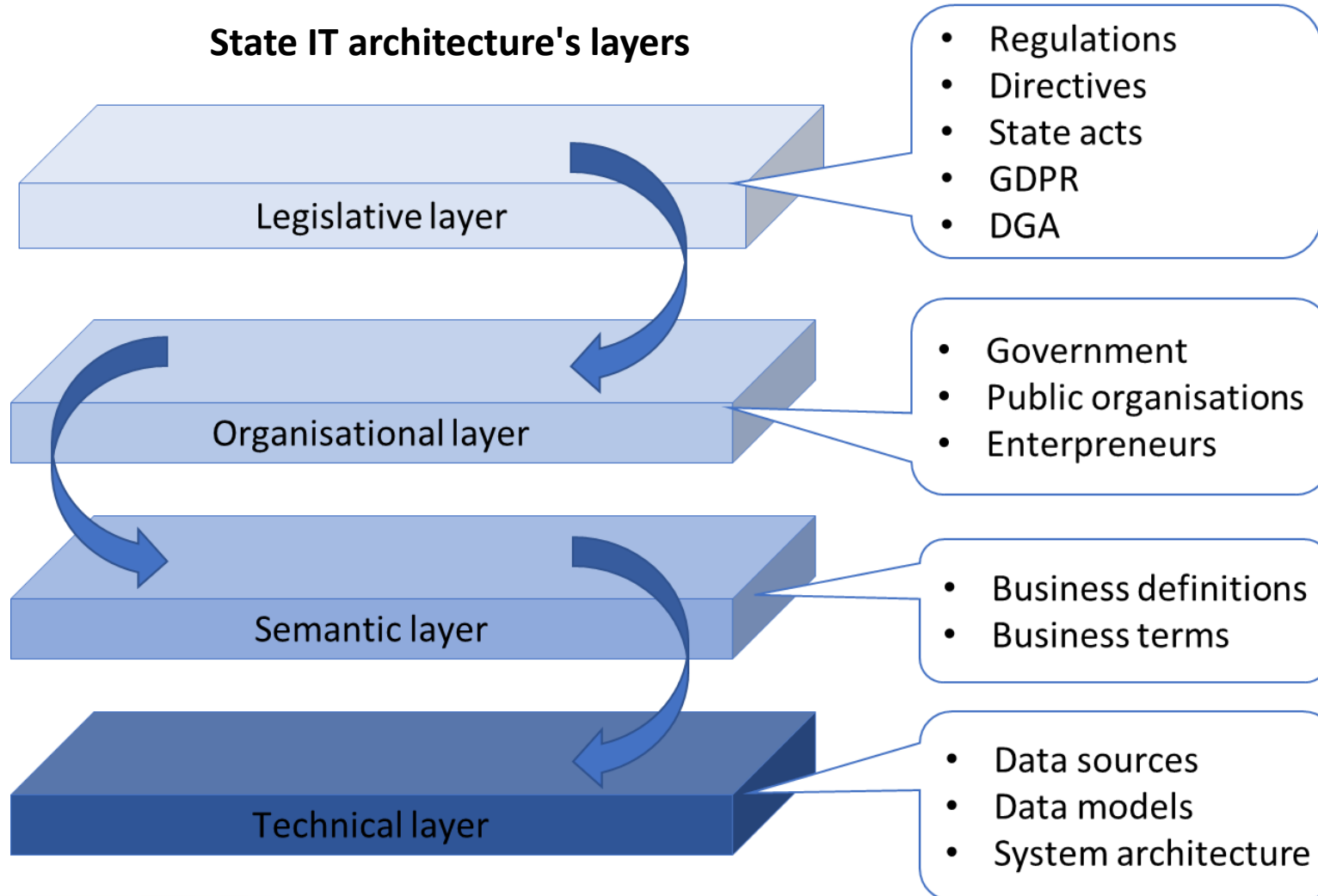- ✓ Standardises data business definitions

- ✓ Improves data quality

It's not a project, it's a process—a long term process.

OPI PIB

Data governance in higher education and other sectors

✓ A wide variety of data governance interpretations in many sectors (e.g. finance, energy, telecommunications)

✓ Some universities in the United States have created new data governance units using different labels (e.g. data governance, institutional research, or data management/analytics), while others have extended IT governance or information governance to data governance.

✓ EDUCASE – a nonprofit association (USA), whose mission is to advance higher education through the use of information technology, published *2023 EDUCAUSE Horizon Action Plan: Data Governance*.

✓ Some institutional data governance programmes (such as that of the University of Toronto in Canada) have been implemented at particular universities, but not centrally.

✓ In science and higher education in Poland, only OPI PIB can undertake such tasks.

✓ Within the European Union, the CERIF and HERM models have been implemented; however, they do not cover the entire thematic range of data collected at OPI PIB.

Data governance – part of the State Information Architecture (AIP)

State IT architecture's layers

Legislative layer

- Regulations
- Directives
- State acts
- GDPR
- DGA

Organisational layer

- Government
- Public organisations
- Enterpreneurs

Semantic layer

- Business definitions
- Business terms

Technical layer

- Data sources
- Data models
- System architecture

Data governance is a cornerstone of the state policy for IT system architecture.

The data governance framework implemented at OPI PIB should be consistent with its assumptions.
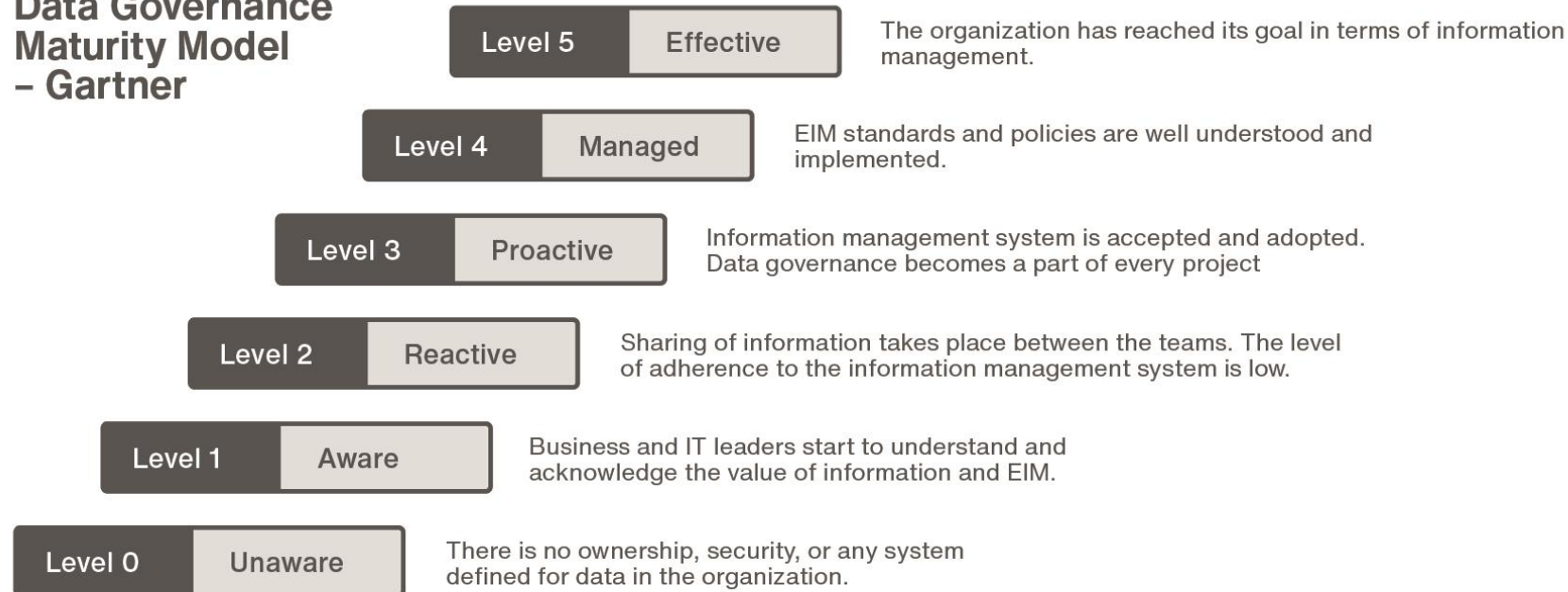
OPI PIB

Implementing data governance at OPI PIB

The Data Governance Maturity model and business case

Data Governance Maturity Model – Gartner

Level	Stage	Description
Level 5	Effective	The organization has reached its goal in terms of information management.
Level 4	Managed	EIM standards and policies are well understood and implemented.
Level 3	Proactive	Information management system is accepted and adopted. Data governance becomes a part of every project
Level 2	Reactive	Sharing of information takes place between the teams. The level of adherence to the information management system is low.
Level 1	Aware	Business and IT leaders start to understand and acknowledge the value of information and EIM.
Level 0	Unaware	There is no ownership, security, or any system defined for data in the organization.

**source:** https://www.hitechnectar.com/blogs/data-governance-maturity-models-explained/

The first exercise involved assessing OPI PIB's awareness and maturity in data management. According to the Gartner Data Governance Maturity model, we evaluated the institute's awareness between levels 1 and 2. This indicated that OPI PIB needed to work on implementing a data governance framework.

Building the team – challenges

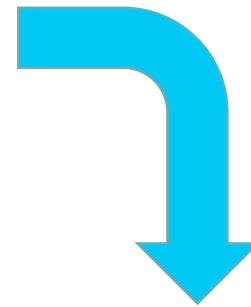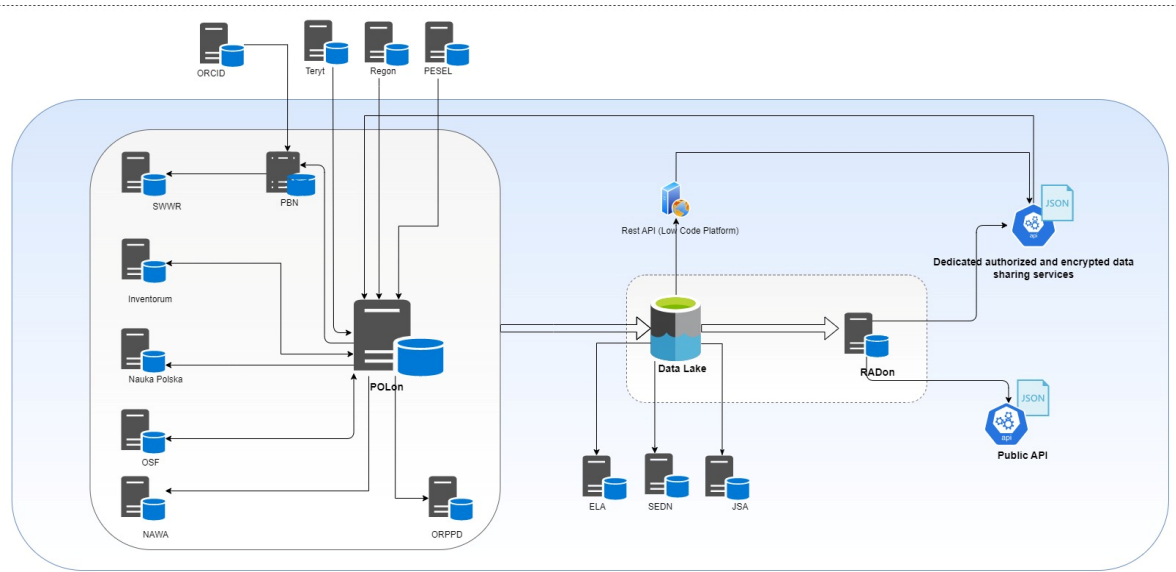Data governance is created by people for people, and people are the most valuable part of the process.
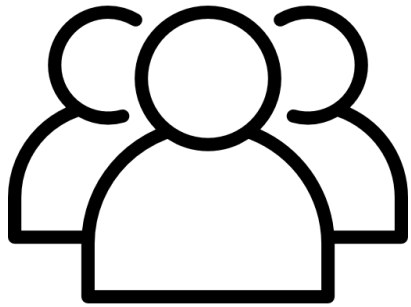
People in the process
- ✓ Chief Data Officer
- ✓ Data Steward
- ✓ Data Owner
- ✓ Data Custodian Officer
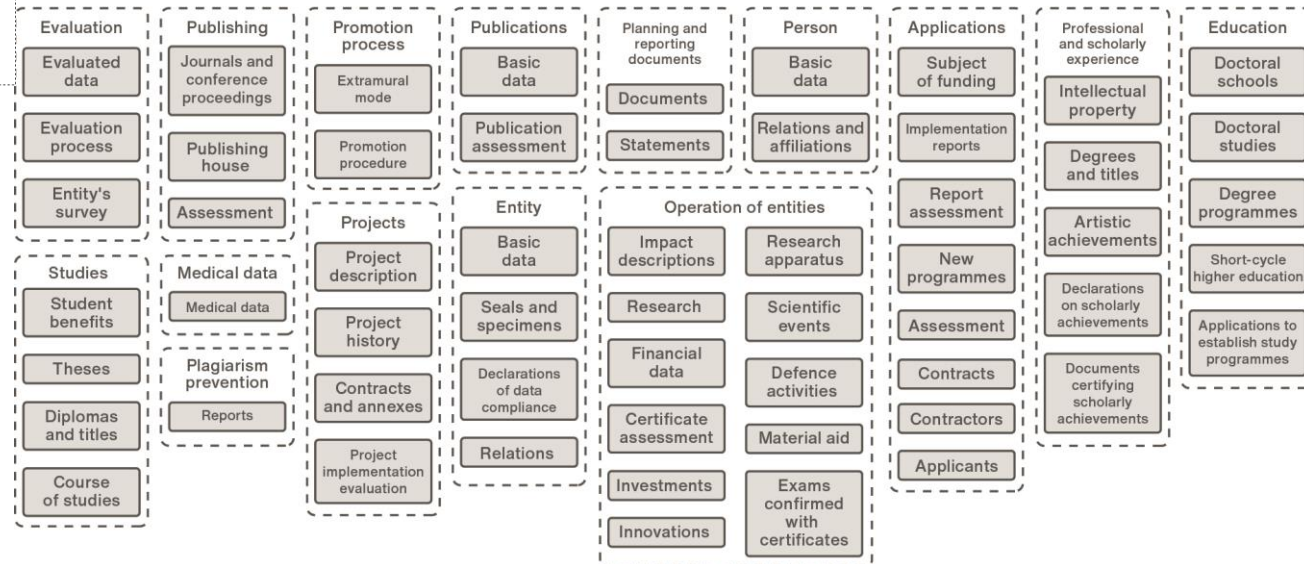- ✓ Data Quality Manager

Challenges
- ✓ Flat organisational structure
- ✓ No additional resources (FTE)
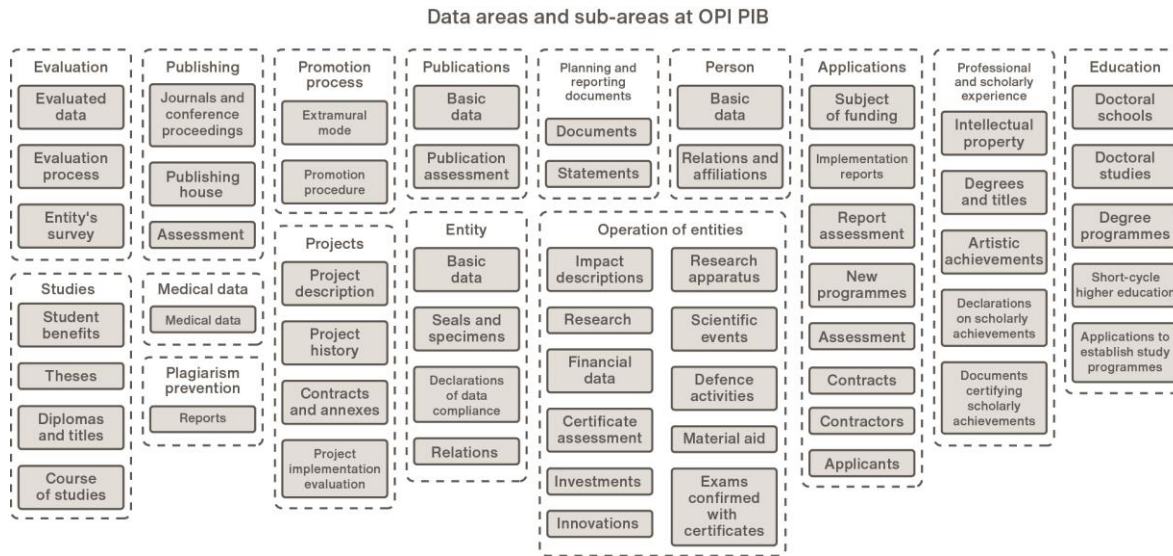- ✓ Change management (fear of change)

Data inventory

OPI PIB holds approximately 12 TB of data.
The institute provides answers to questions from journalists, agencies, and other stakeholders, as well as clarifying issues related to the definition of data, and the possibility of reprocessing it or making it available to the public.


Data areas and sub-areas at OPI PIB

Evaluation	Publishing	Promotion process	Publications	Planning and reporting documents	Person	Applications	Professional and scholarly experience	Education
Evaluated data	Journals and conference proceedings	Extramural mode	Basic data	Documents	Basic data	Subject of funding	Intellectual property	Doctoral schools
Evaluation process	Publishing house	Promotion procedure	Publication assessment	Statements	Relations and affiliations	Implementation reports	Degrees and titles	Doctoral studies
Entity's survey	Assessment					Report assessment	Artistic achievements	Degree programmes

Studies	Medical data	Projects	Entity	Operation of entities		Applications	Professional and scholarly experience	Education
Student benefits	Medical data	Project description	Basic data	Impact descriptions	Research apparatus	New programmes	Declarations on scholarly achievements	Short-cycle higher education
Theses	Plagiarism prevention	Project history	Seals and specimens	Research	Scientific events	Assessment	Documents certifying scholarly achievements	Applications to establish study programmes
Diplomas and titles	Reports	Contracts and annexes	Declarations of data compliance	Financial data	Defence activities	Contracts		
Course of studies		Project implementation evaluation	Relations	Certificate assessment	Material aid	Contractors		
				Investments	Exams confirmed with certificates	Applicants		
				Innovations				

Enterprise data model

Data areas and sub-areas at OPI PIB

An enterprise data model is a type of integration model that covers all (in practice, most) of the data of an enterprise.

An enterprise architecture may include enterprise-wide data models that are also conceptual, logical, or physical (West M. 2011).
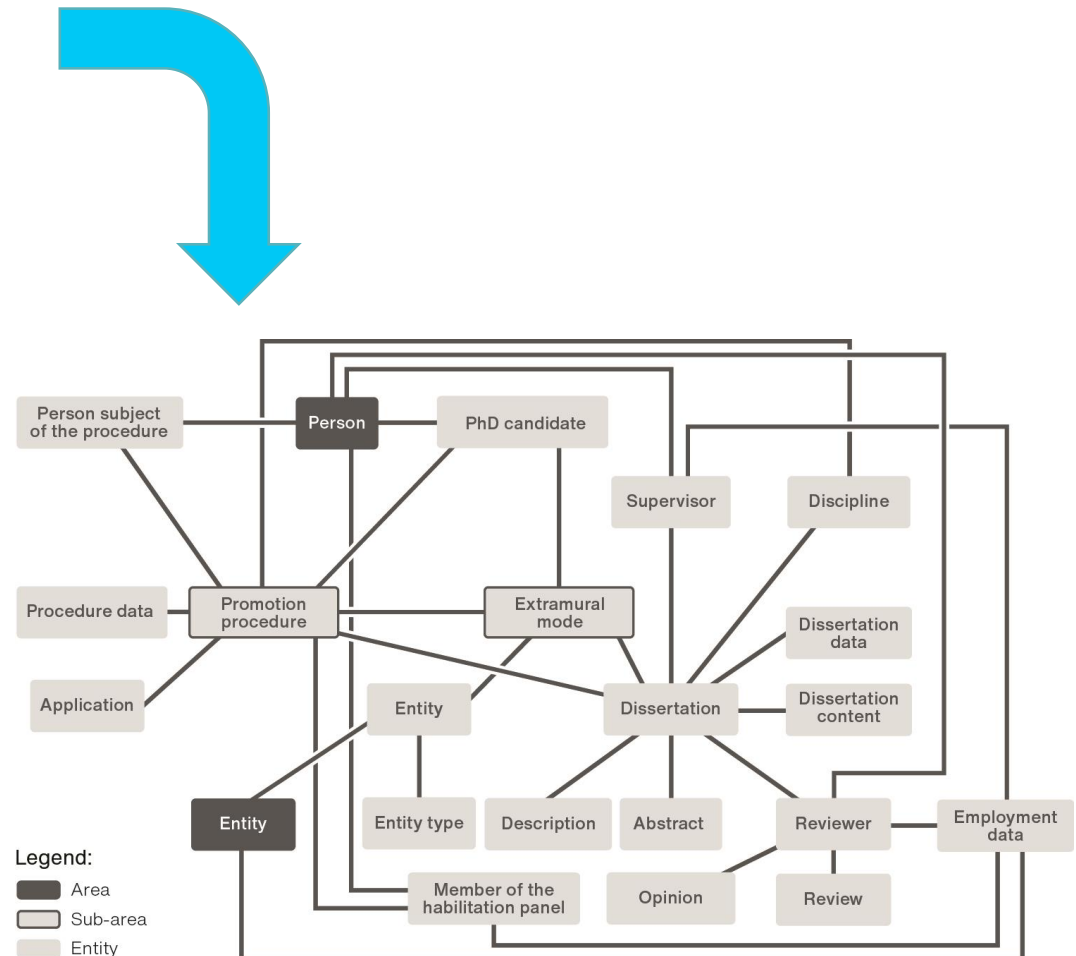
Data metrics

AREA	HIGHER EDUCATION
AREA DESCRIPTION	Data related to universities, research institutes, and educational and higher education institutions
INFORMATION SCOPE	The area covers identification, registration and basic data, addresses of registered offices and correspondence addresses, contact details, and details of the head of the unit.
DOMAIN	POL-on, OSF, NAWA
GROUPING LEVEL	Category
STAKEHOLDERS	Data Owner, Head of …

Proof of concept with the data catalog

After several months working with data governance, we began to search for an IT tool that could help OPI PIB in data management processes. We conducted research and compared five of the most prominent ones:

- ✓ Oracle Data Catalog (ODC) - https://docs.oracle.com/en-us/iaas/data-catalog

- ✓ AtaccamaOne (ATA) - https://www.ataccama.com

- ✓ Atlan (ATL) - https://docs.atlan.com

- ✓ Collibra (COL) - https://marketplace.collibra.com

- ✓ Talend (TAL) - https://help.talend.com

Comparison of data catalog tools

Available (selected) data connectors:

	ODC	ATA	ATL	COL	TAL
ORACLE DB	✓	✓	✓	✓	✓
MYSQL	✓	✓	✓	✓	✓
POSTGRESQL	✓	✓	✓	✓	✓
MICROSOFT SQL SERVER	✓	✓		✓	✓
MONGO DB		✓		✓	✓
CSV	✓	✓		✓	✓
XLSX	✓	✓		✓	✓
XML	✓	✓		✓	✓
JSON	✓	✓		✓	
APACHE KAFKA	✓	✓		✓	✓
HIVE	✓	✓	✓	✓	✓
ELASTIC		✓			✓
JIRA			✓		
AZURE	✓	✓	✓	✓	✓

Comparison of data catalog tools

Available (selected) functionalities:

	ODC	ATA	ATL	COL	TAL
DATA MODELING CONCEPTUALISATION	✓	✓	✓	✓	✓
DICTIONARIES	✓	✓	✓	✓	✓
CLASSIFICATIONS	✓	✓	✓	✓	✓
DATA QUALITY		✓		✓	✓
DATA LINEAGE		✓	✓	✓	✓
EMBEDDED AI (CLASSIFICATIONS, PROFILING, RELATIONS...)	✓	✓		✓	✓

OPI PIB

Data catalog at OPI PIB

Although the comparison above contains empty cells, all tools available on the market are similar. When functionalities are unavailable 'out of the box', we can source add-ons from the same vendor (although they require additional payment). For instance, Oracle and Atlan have separate tools related to data quality and data lineage. No data catalog tools have data modeling functionalities included; separate tools are necessary.
Ultimately, we decided to use Oracle Data Catalog, because it is included in Oracle Cloud Infrastracture and does not require any additional fee.

Conclusions

Legal aspects

- ✓ As an organisation, we are ready for the Data Governance Act

- ✓ We are fully GDPR compliant

- ✓ We respect the FAIR data principles (radon.nauka.gov.pl)

- ✓ We have adapted to State's Information Architecture (AIP)

OPI PIB

Data governance experience

✓ The work involved in implementing elements of the data governance programme at OPI PIB has led to deeper understanding of how crucial proper data management is for organisations.

✓ From the perspective of OPI PIB, the implementation of data governance has contributed to increasing the quality of data, but above all to increasing the awareness of those responsible for the analysis and design of IT systems.

✓ The implementation of data governance has already delivered measurable benefits in shortening the time necessary for analysis and obtainment of information for external stakeholders.

✓ Nevertheless, challenges remain.

Thank you

National Information Processing Institute
al. Niepodległości 188 B
00-608 Warszawa

tel.: +48 22 570 14 00
fax: +48 22 825 33 19
email: opi@opi.org.pl
www.opi.org.pl