

Data Spaces Symposium

11:00

Societal impact of data spaces – Sharing language data and media resources trustworthy and effectively

Domain session [language | media | public procurement]

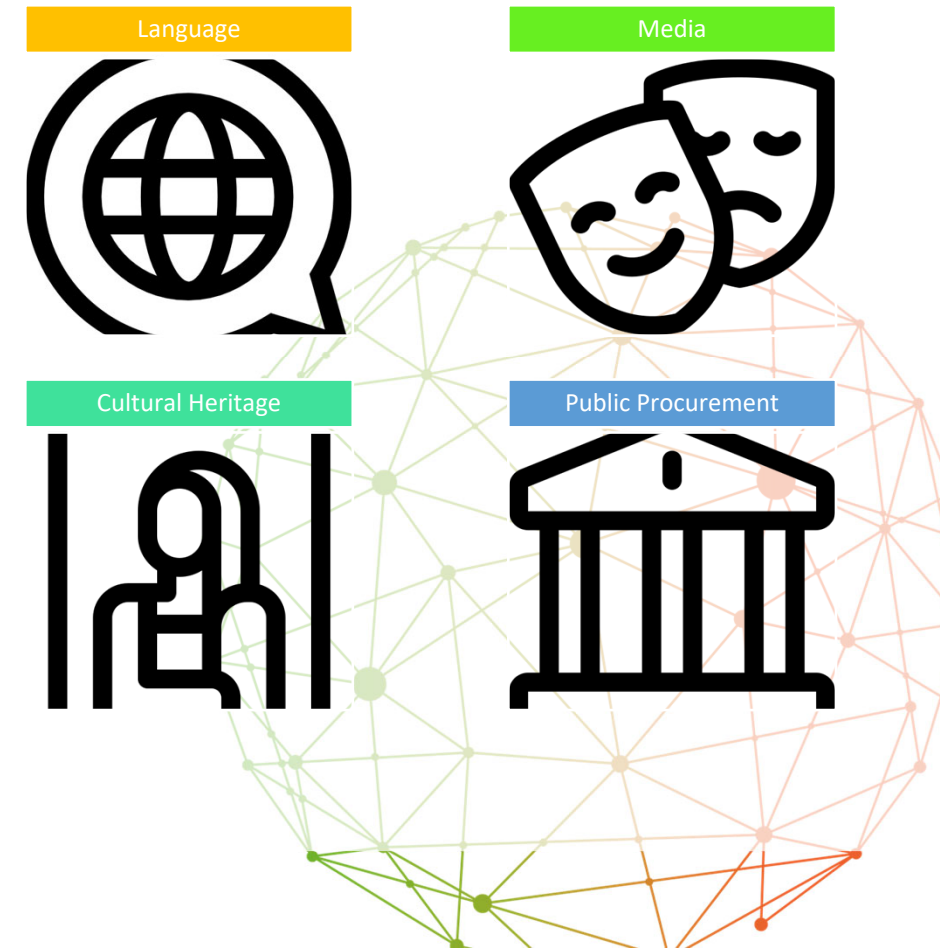
Data Spaces Symposium

Setting the scene

Savvas Rogotis

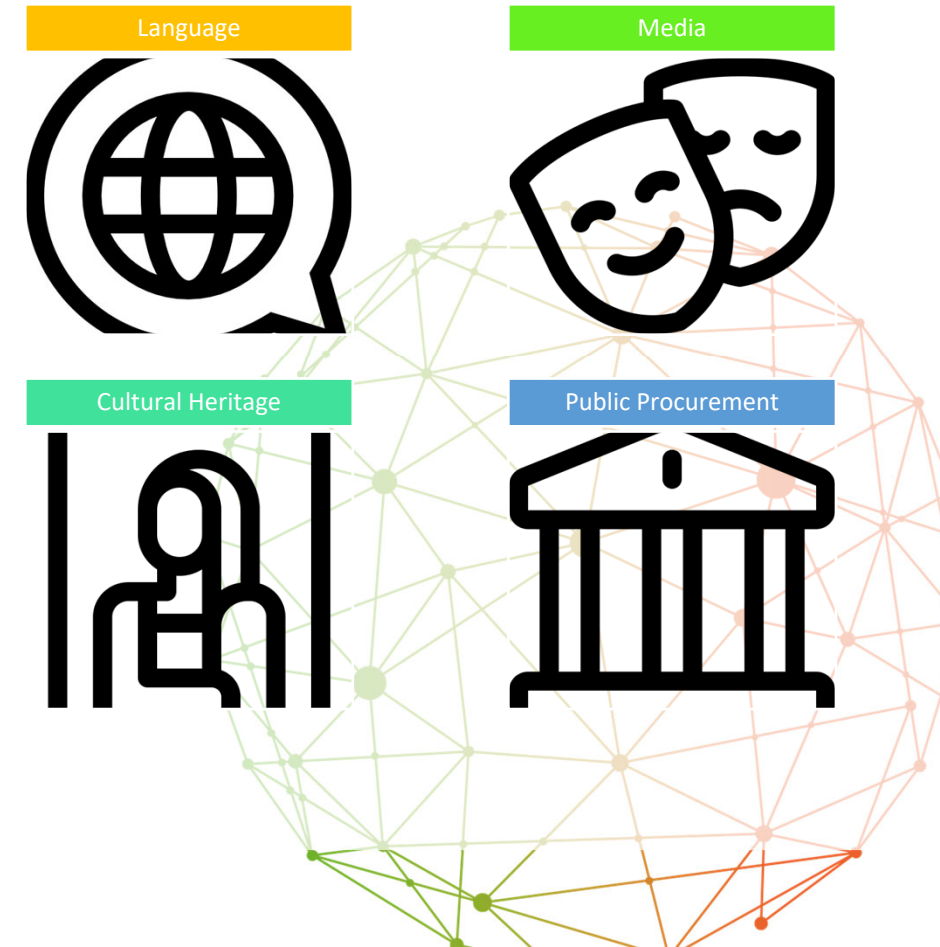
Language, Media, Cultural Heritage & Public Procurement. How these come together?

- Long standing communities & notable investments backing current developments
- Public value orientation
- Data is often scattered across many stakeholders (and different types), countries, formats, etc..
- Strong value creation elements
- All prioritized in EU Data Strategy
- Several common themes and use cases



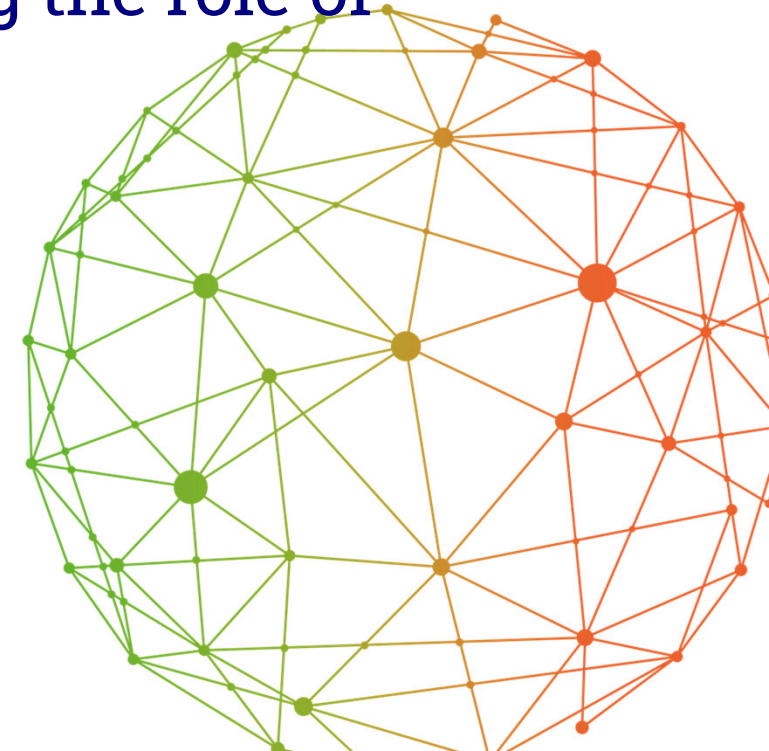
But also unique features

- Data level:
 - Multilingual corpora (language),
 - Rich in rights-protected content (media),
 - highly curated, long-archived (cultural heritage),
 - transactional, highly sensitive (public procurement)
- Key Stakeholders:
 - Private vs public sector stakeholders
- Data spaces level:
 - in different maturity levels
 - different in nature (procurements vs grants)
 - different priorities and opportunities ahead
- ...



Objectives of the session

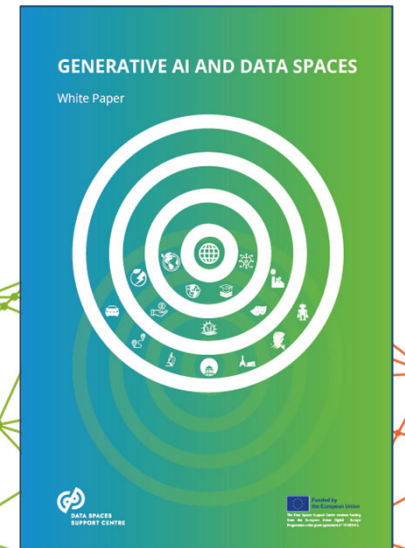
- Understand the status and potential of data sharing / spaces in these sectors, along with the challenges ahead as the ecosystem matures.
- Explore how data can generate value and drive positive impact—societal, business, etc.- while examining the role of GenAI in this transformation.
- Identify synergies, opportunities and next steps for collaboration.



Whitepaper on Gen AI and Data Spaces

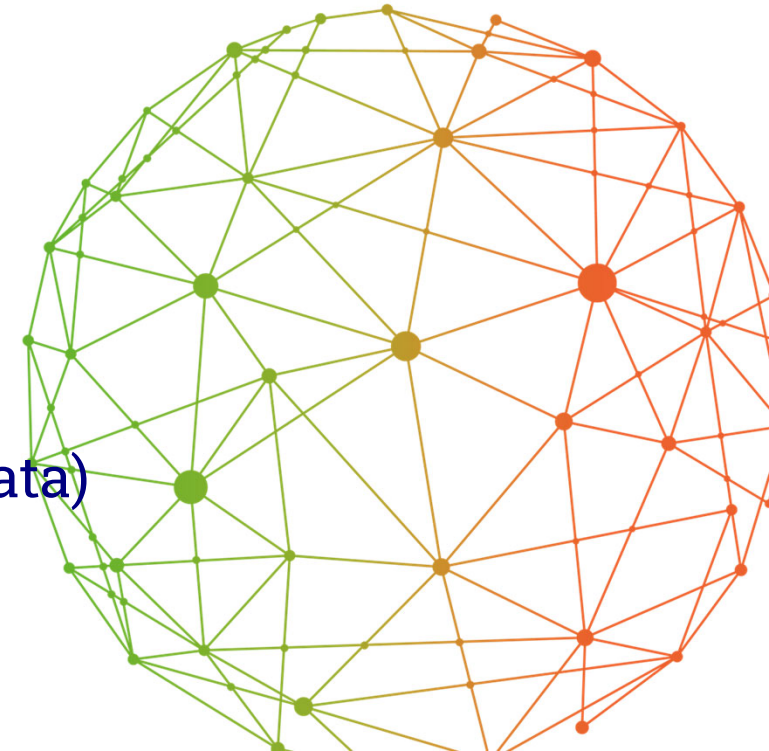
Driven by DSSC Strategic Stakeholder Forum

- The rise of Generative AI, underpinned by new foundation models, offers numerous opportunities for businesses and society in Europe
- Synergies between Generative AI and Data Spaces hold significant transformative potential
- Exploring this interplay can be a key for unlocking the value of industrial, commercial and personal data
- Potential for value creation identified in Language, Media, Public Procurement



Agenda

- Introduction & scene setting – Savvas Rogotis (BDVA)
- Data space revolution: Unleashing media's potential, navigating challenges, and shaping the future – Veronique Demilly (France Télévisions)
- Facing the challenges of data spaces: the LEIA case- Asunción Gómez (UPM)
- Panel Discussion: How to generate value and impact in cultural heritage, media & language data spaces
 - Moderator: Clara Pezuela (FIWARE)
 - Georg Rehm (DFKI)
 - Óscar Lázaro (Innovalia)
 - Valentine Charles (Europeana)
- Public Procurement Data Space – Daniel Arosa Otero (NTT Data)
- Q&A and Closure



Data Spaces Symposium

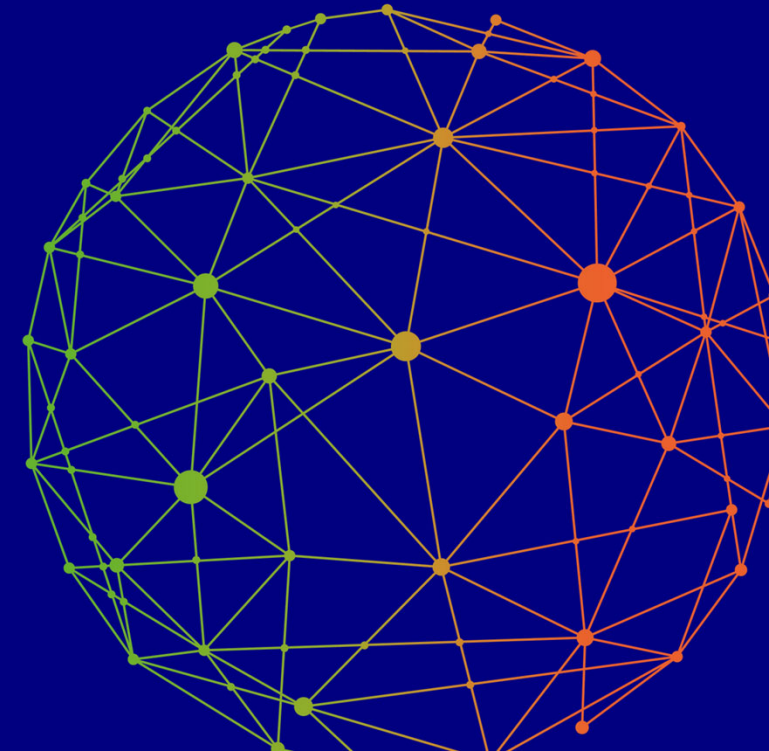
Data space revolution: Unleashing media's potential, navigating challenges, and shaping the future

Véronique Demilly

Media : what are we talking about?



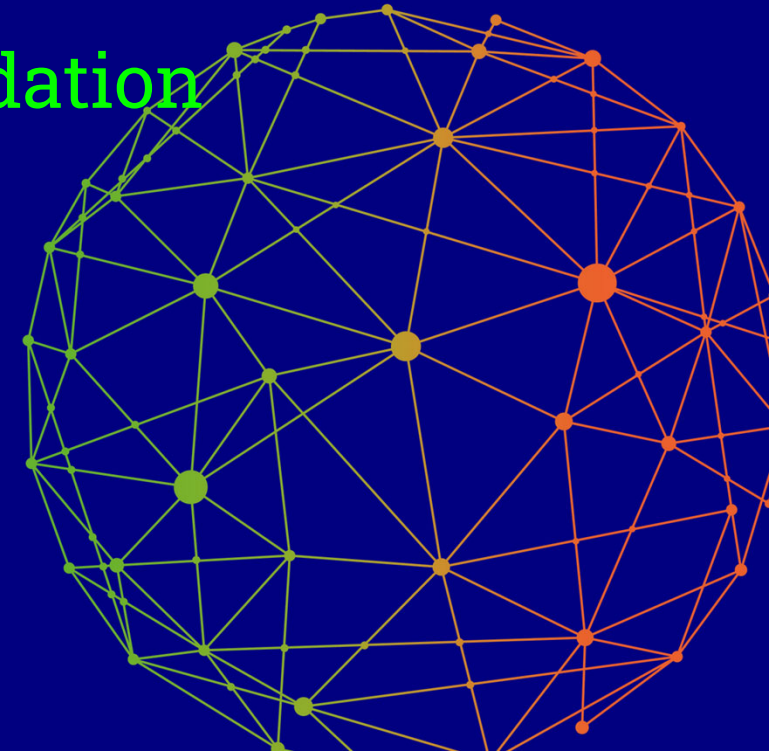
- In Europe :
 - 200k companies – 99,8% of SMEs
 - 1 million jobs
 - Revenues expected in 2025 : 323 bn€



The attention economy

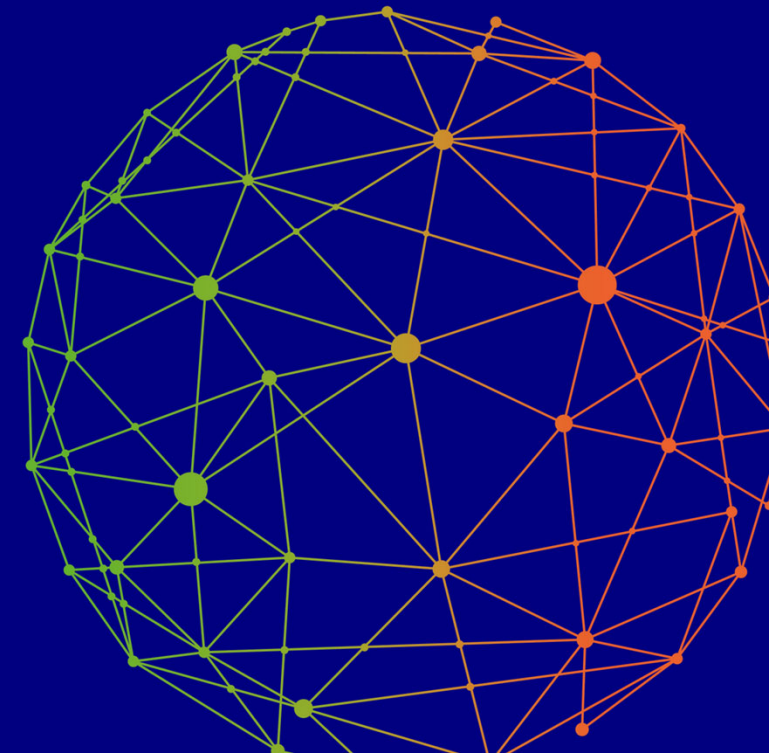


- Personalization
- Recommendation
- Adaptation

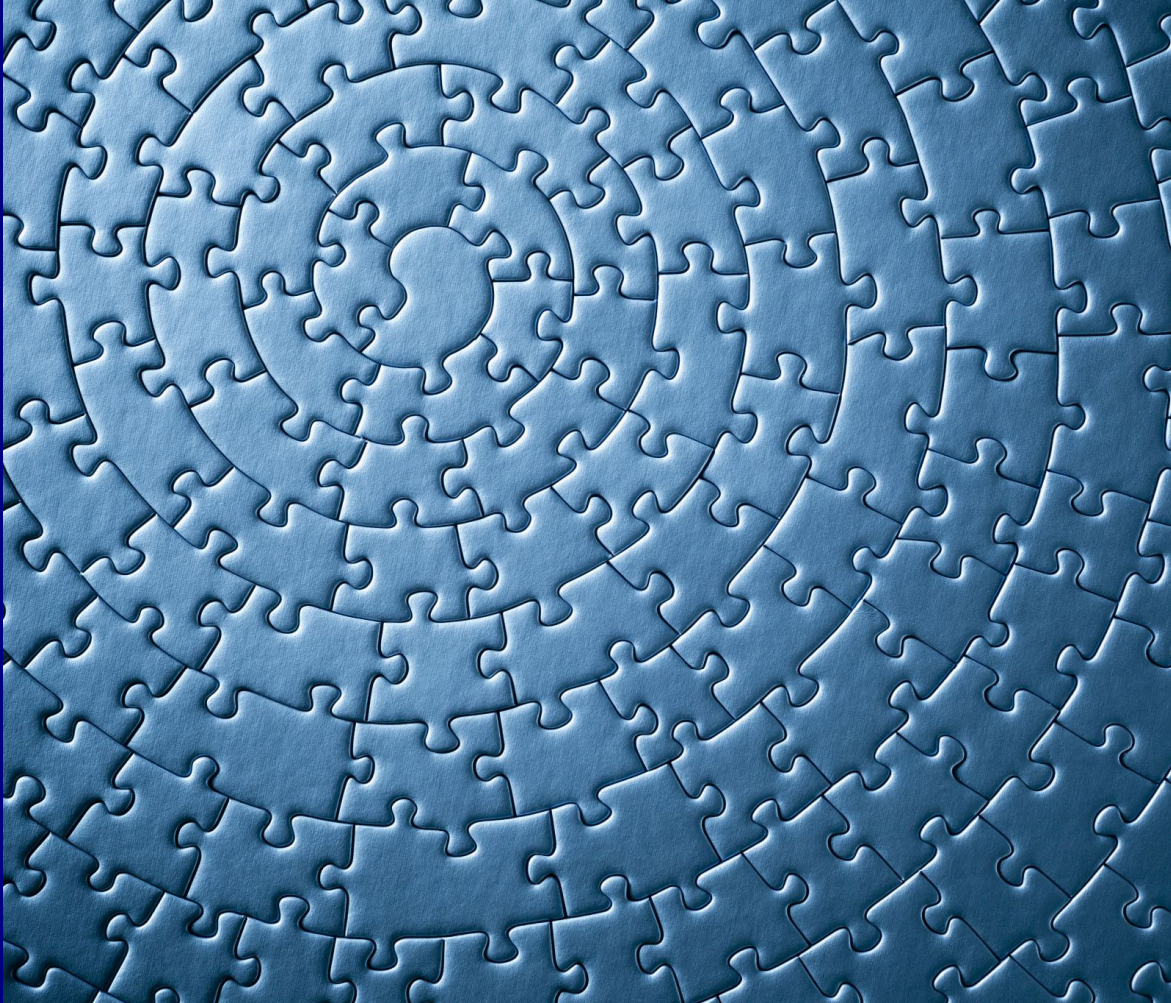


And Gen AI?

- Opt-out?
- Right holders compensation?
 - Loss in revenues in musical and audiovisual industries :
 - 2028 : over 20% loss*

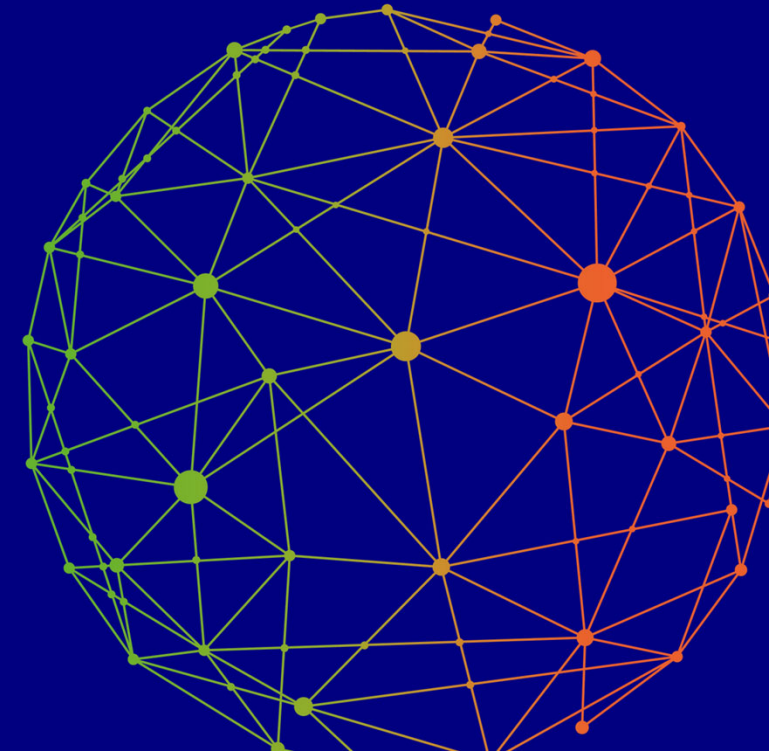


Challenges : become data-driven?

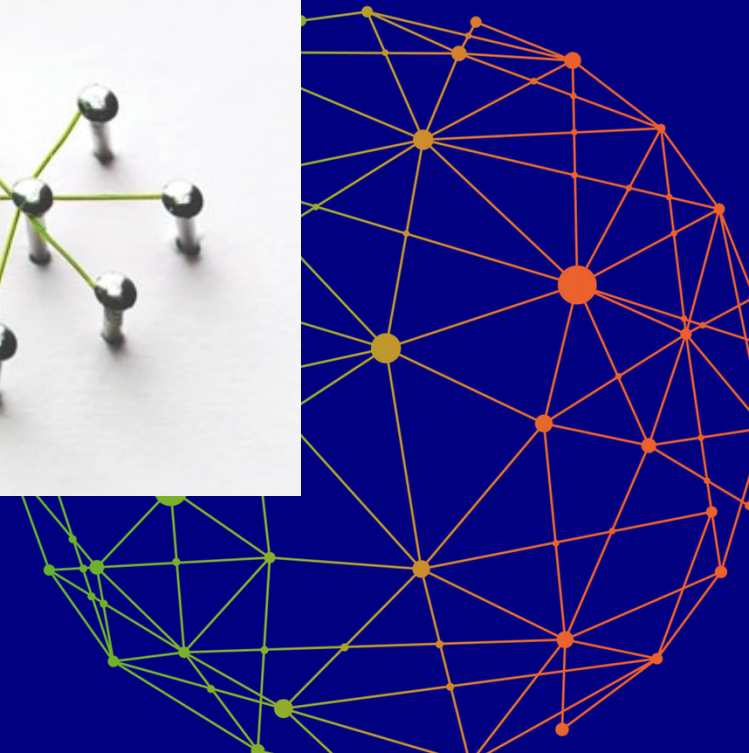
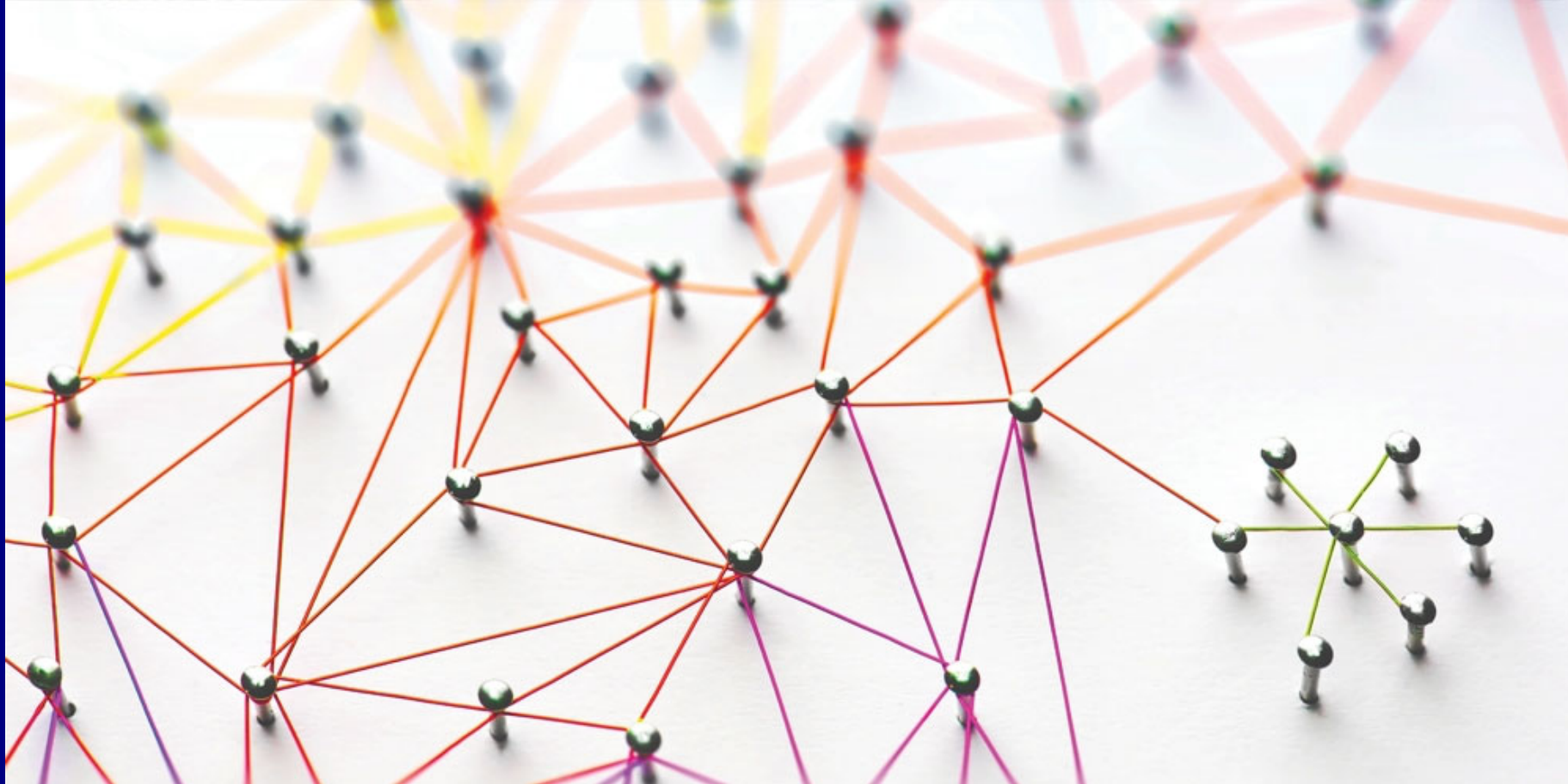


- Skills

- Costs



Role of Data Space

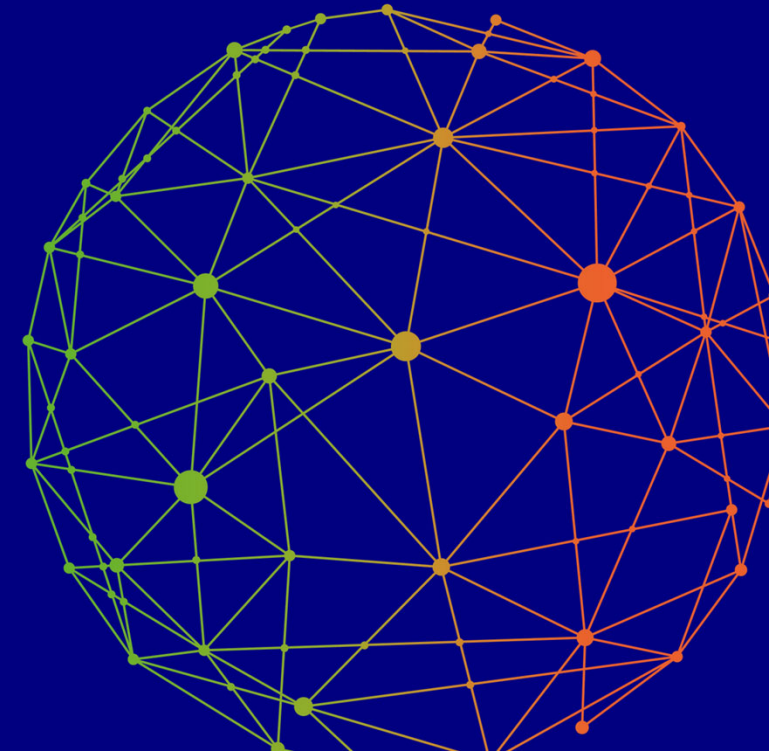


Expectations

Break down data silos



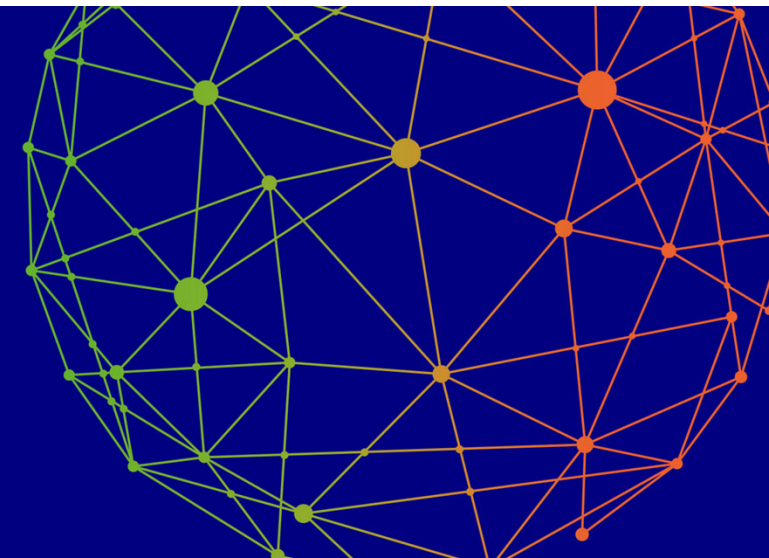
- Make it easy !
- Sustainable
- Future proof



Thank You!

info@tems-dataspace.eu

<https://tems-dataspace.eu/>



Funded by the Digital Europe Programme of
the European Union Under Grant Agreement
No 101123423

Data Spaces Symposium

Facing the challenges
of data spaces: the LEIA case

Asunción Gómez



REAL ACADEMIA ESPAÑOLA

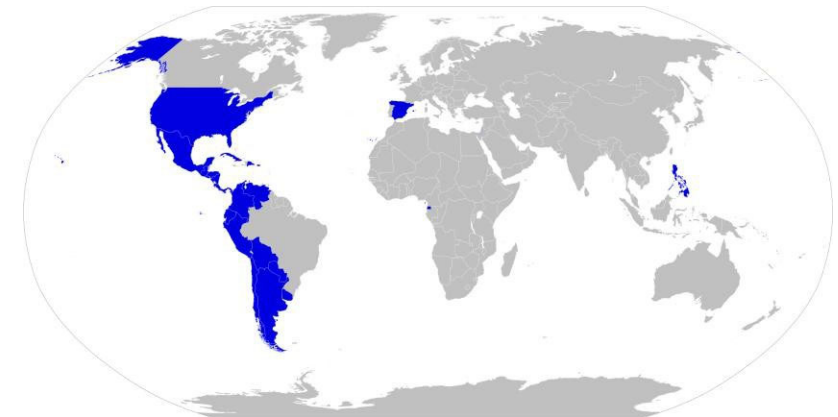
Facing the challenges of data spaces: the LEIA case

Asunción Gómez-Pérez

Real Academia Española
Universidad Politécnica de Madrid

Real Academia Española (RAE) is the authority on the Spanish language

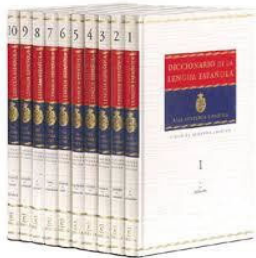
1. Spanish is spoken by almost 600 million people around the world.
2. The **Royal Academy of Spain** (RAE) founded in 1713, is the ultimate
 - Authority on the Spanish language conferred by the Spanish-speakers for over 300 years.
 - To ensure the changes the Spanish language undergoes as it constantly adapts to its speakers' needs do not break the fundamental unity of the language across the Hispanic world.
3. **Association of Spanish Language Academies** (ASALE)
 - to “work to further the unity, integrity and growth of the Spanish language, which constitutes the richest element of shared heritage among the Spanish-speaking community”.
4. **Pan-Hispanic Plain Language Network**
 - This is not limited to legal language but covers all forms of communication addressed to citizens.



Conclusion: Change of paradigm



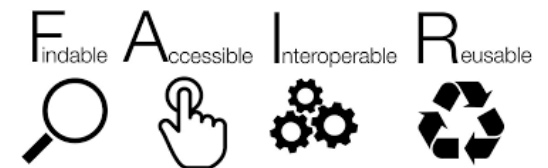
The first dictionary (1726-1739)



The dictionary in a CD-ROM (2007)



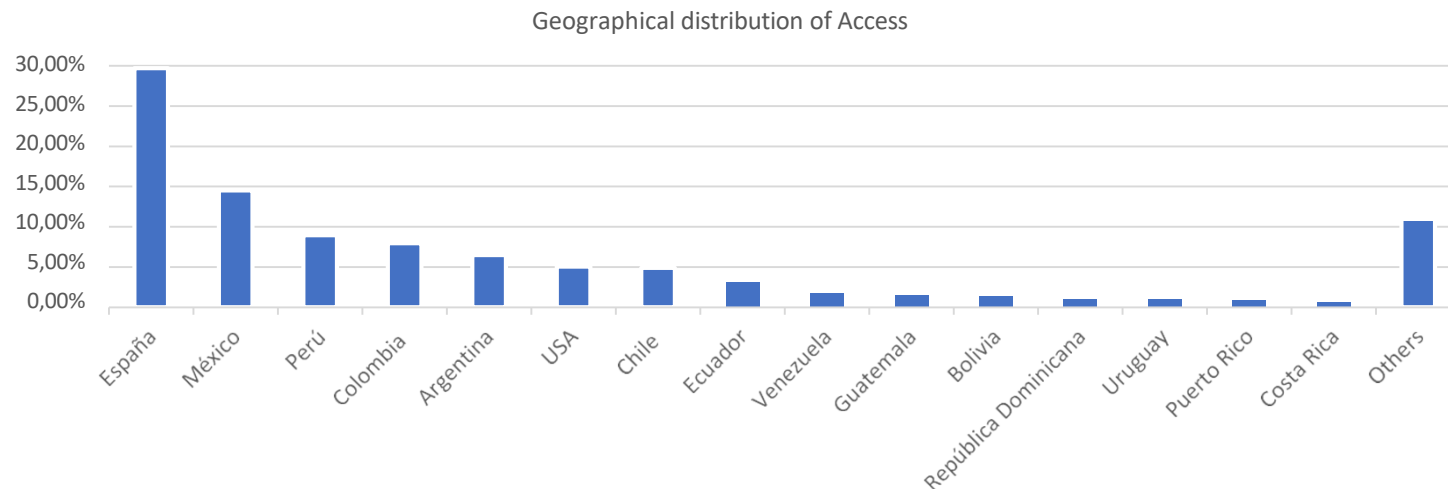
November 1998



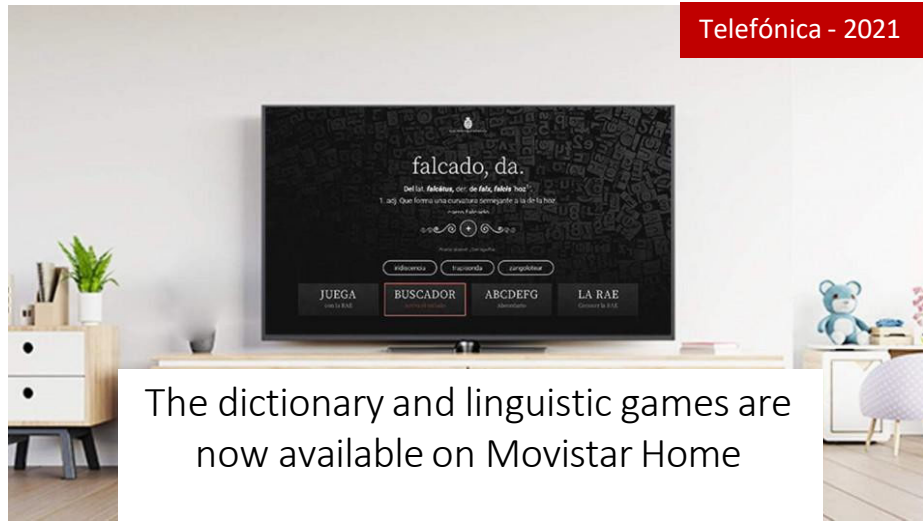
Users get access to RAE materials through web applications

1300M Access/year to public RAE resources

- 1000M Access/year to the dictionary (dle.rae.es)
- 100M Access/year to www.rae.es
- 1.2M Access/year to the corpus only by linguistic researchers



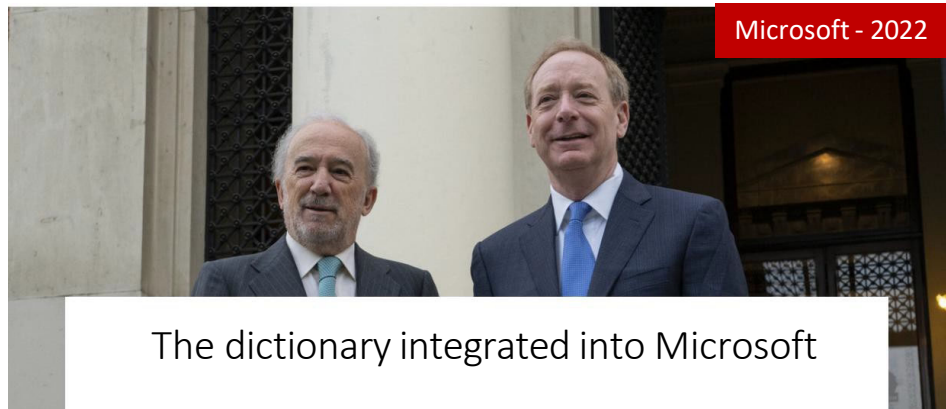
Linguistic materials being used by big tech companies



The dictionary and linguistic games are now available on Movistar Home



The Academy and AWS have developed an IA tool to investigate the status of the Spanish language on the Internet



The dictionary integrated into Microsoft



The dictionary integrated into Google search facility and the lexicón used by Gboard (Android keyboard)

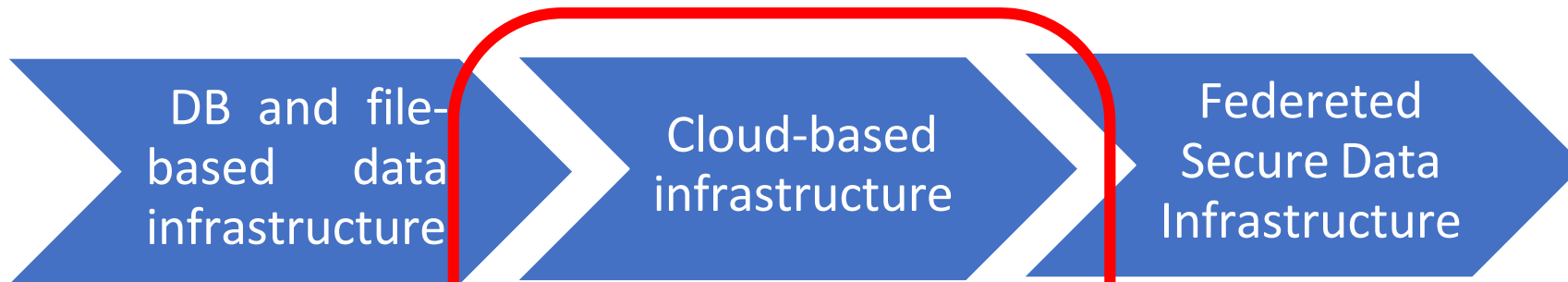
Paving the way towards data spaces

- But we can not move into the data spaces without a secure data infrastructure in the cloud



- RAE resources are served though:
 - Legacy applications with some issues:
 - Obsolete technology in some cases
 - Reduced interoperability
 - Decentralized data sources
 - On-premises servers
 - Poor scalability
 - Security issues
 - High maintenance costs
 - Limited connectivity

Paving the way towards data spaces



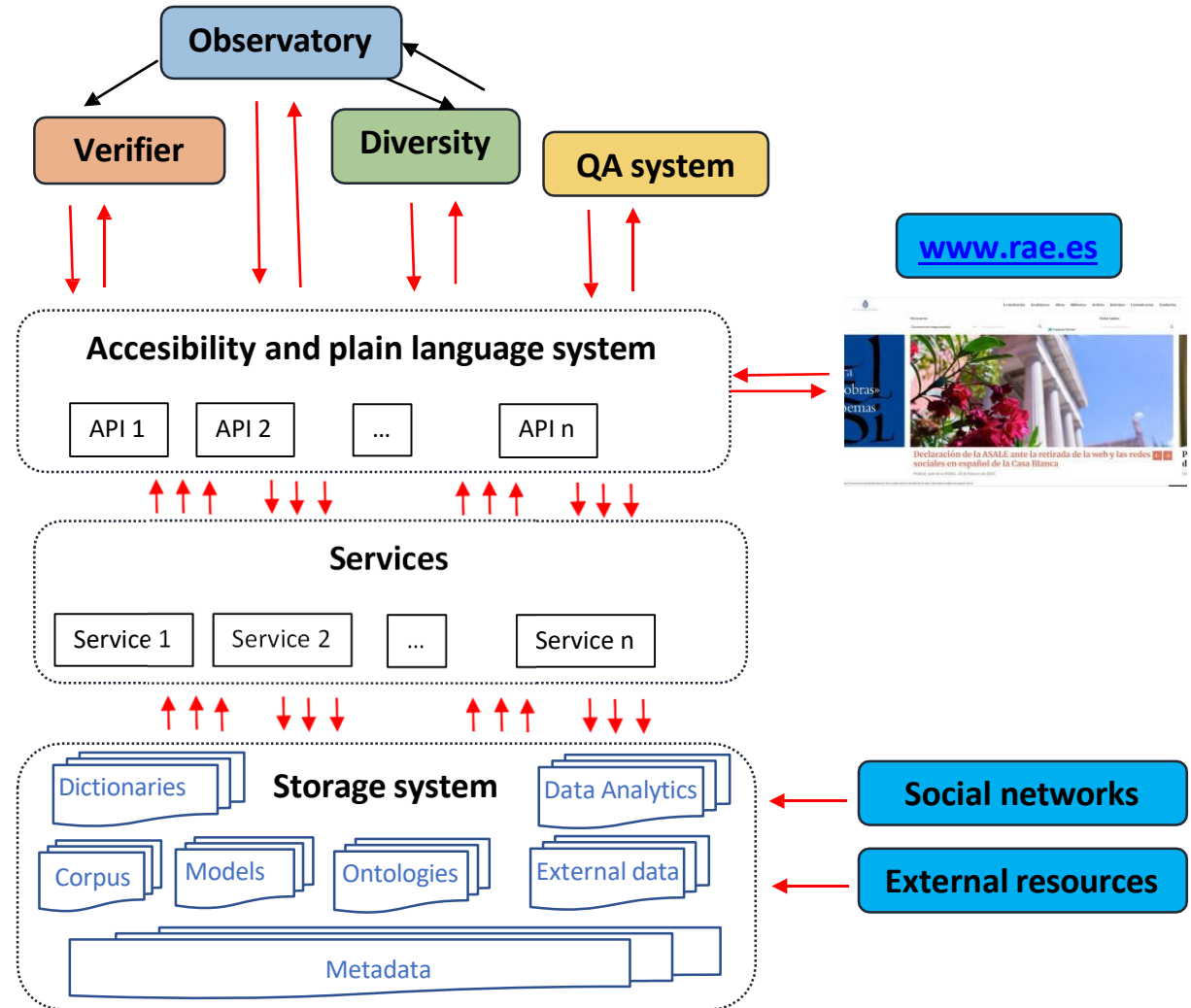
- Modern approach:
 - Native Cloud applications:
 - Up-to-date open-source technologies
 - High interoperability
 - Centralized data sources
 - Infrastructure as code
 - High scalability
 - High security
 - Medium maintenance costs
 - High connectivity



- Advantages
- Sovereignty
 - Trustworthy environment
 - Interoperability standards (data, protocols, ...)
 - EU and Spanish regulation
 - FAIR data
 - Open source software

LEIA platform

- Generate **linguistic materials** and make them **openly available** to contribute to the development of language technologies, among others.
- To observe the current state of the **use of the Spanish language on the Internet**
- To continue guaranteeing the **unity of Spanish in all its varieties**
- **Use AI** to improve the use of Spanish and its automatic processing to help lexicographers



LEIA Project: Use case

1. Plataforma

2. Four use cases

- **Observatory to discover** neologisms, derivatives, tecnicisms, regionalism and foreignness
- Linguistic verifier
- QA system to be used by linguistics to speed up the response time
- System for collecting material based on diversity

3. Accesibility

4. Ecosystem

Goal:

- Detect new words and expressions
- Record new uses of existing words

Digital sources:

- Press, specialized magazines, social networks

Source selection parameters

- Types of digital media
- Geography
- Theme
- Typology

Initial parameter configuration:

- That of CORPES

What it stores:

- The word and link to the source. It is not a corpus

LEIA Project: Use case

1. Plataforma

2. Four use cases

- Observatory to discover neologisms, derivatives, tecnicisms, regionalism and foreignness
- **Linguistic verifier**
- QA system to be used by linguistics to speed up the response time
- System for collecting material based on diversity

3. Accesibility

4. Ecosystem



Goal:

- Checking if a text in Spanish is correct

Digital sources:

- Spelling, vocabulary, syntax, grammar and style

Three stages

- Automatically detect errors
- Suggest corrections to the user
- Explain – references to RAE doctrine

AI Techniques:

- LLM
- RAG
- Synthetic error catalogue
- ...

What it stores:

- Nothing

LEIA Project: Use case

1. Plataforma

2. Four use cases

- **Observatory to discover** neologisms, derivatives, tecnicisms, regionalism and foreignness
- Linguistic verifier
- **QA system to be used by linguistics to speed up the response time**
- System for collecting material based on diversity

3. Accesibility

4. Ecosystem

Goal:

- Answer questions of a linguistic nature from the public

The system

- The system Learns from thousands of question-answer pairs
- Search engine

Scope:

- Supervised system
- Helps linguists select the answers that best respond to the question

AI Techniques:

- Semantic similarity
- Sentence transformer

LEIA Project: Use case

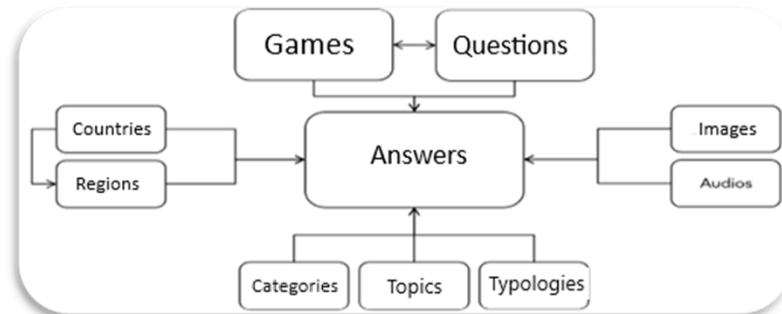
1. Plataforma

2. Four use cases

- Observatory to discover neologisms, derivatives, tecnicisms, regionalism and foreignness
- Linguistic verifier
- QA system to be used by linguistics to speed up the response time
- **System for collecting material based on diversity**

3. Accesibility

4. Ecosystem



Goal:

- Collect lexical and oral material from different geographical areas using games and challenges

The system provides multimedia material:

- Word
- Text
- Image
- Audio

Scope:

- Users respond with text or audio

AI techniques

- Generation of multimodal training dataset for AI

Conclusions

1. RAE –Royal Academy of Spain

- Is the authority in the Spanish Language
- Has quality resources
- Has millions of users.

2. The LEIA platform paves the way towards a Spanish linguistic data space with a secure data infrastructure in the cloud

3. AI helps to keep RAE data updated

- The observatory system helps to discover new Spanish words and to keep the dictionary updated
- The QA system helps lexicographers on their daily work
- The diversity system helps to gather oral and text data about the different varieties of Spanish

4. RAE is open to participate on European projects to move some data into the data spaces





Data Spaces Symposium

Panel discussion | How to generate value and impact in cultural heritage, media & language data spaces

Clara Pezuela, Valentine Charles, Oscar Lázaro & Georg Rehm

Data Spaces Symposium

Public procurement data space

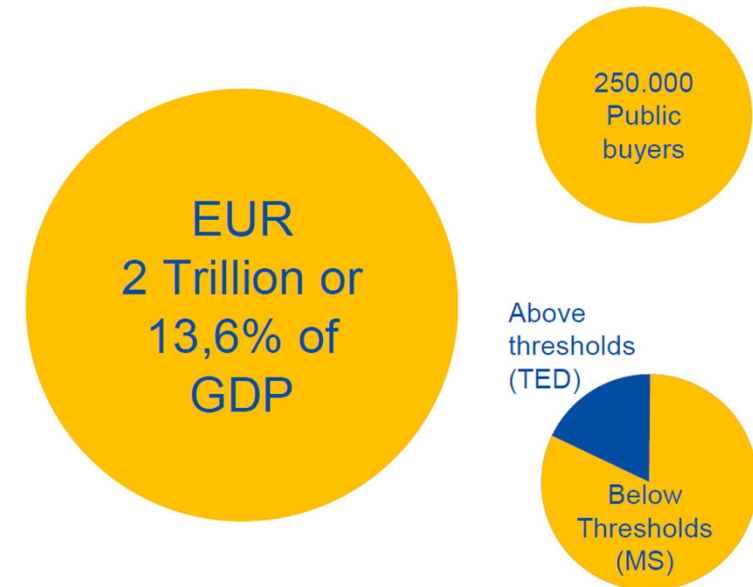
Daniel Arosa Otero

PPDS – The challenge

Public procurement data in Europe is produced by all public authorities with the capacity to perform purchases. Data is generated at the most decentralised level (buyer) but consumed and used at different levels (governments in the different administration levels but also other stakeholders such as academia).

Public procurement data bears several defies:

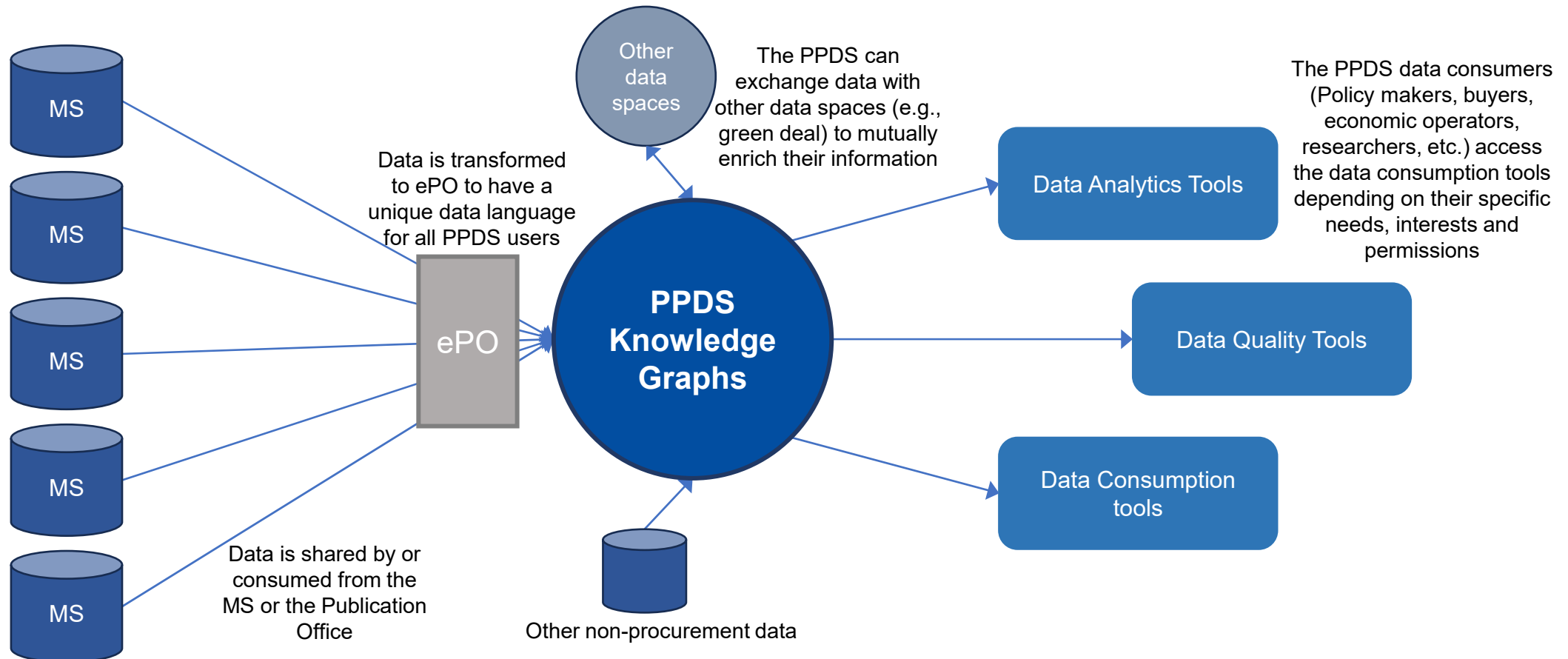
- Divergence in data standards used by the different public administrations
- Vast amount of data producers and consumers
- Lack of a consolidated access to all public procurement data at EU level beyond the mandatory requirement



The **greatest challenge** for the development of the **PPDS** is how to reach 'convergence' amongst all the stakeholders, data providers and data consumers, facilitating the access to public procurement data for end-users

PPDS – The solution

Creating a data space that **consolidates the data from every public procurement data source in Europe** by providing interoperability of the data through **common semantics of the data (ePO Ontology)**. PPDS provides **data analytical capabilities** to facilitate evidence-based decision making for buyers, economic operators and policy makers.



PPDS – Impact

The project success will be marked by the implementation and escalation of the PPDS. The long-term objective is that the PPDS becomes the 'de facto' data access point for all procurement data in the EU. Achieving an optimal development of PPDS will depend to a large extent on the participation and collaboration of stakeholders.

Policy makers

Use of the PPDS to analyse the functioning of the public procurement market and produce evidence-based policy-making in the area of procurement

Companies

Have an easy-to-use portal that gives them access to a much greater number of procurement information with better data quality

NGO/Academia/Journalists

Use the PPDS data for research and investigation purposes



Public buyers

Acquire the most comprehensive knowledge base of procurement all around Europe (and beyond)

Service providers

Use of the PPDS to create value based on the public procurement data

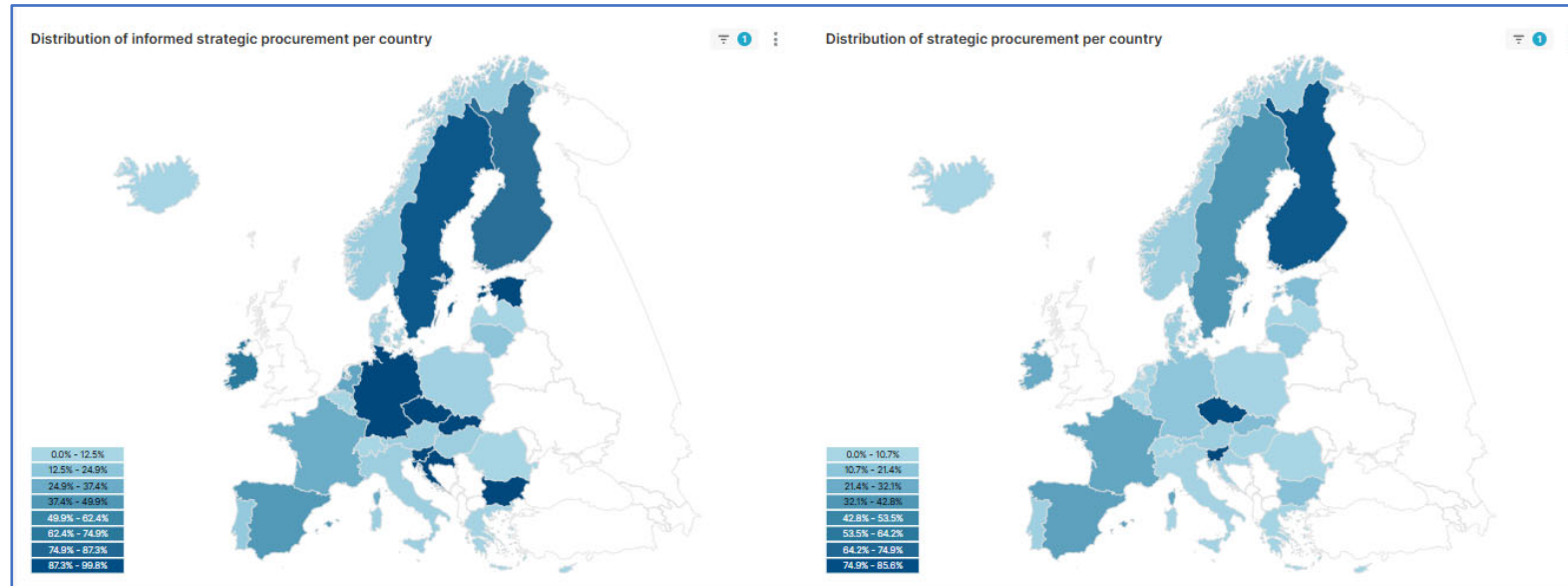
Citizens

Benefit from increased transparency on public spending from public administrations

Strategic Procurement

The public procurement notices (eForms) allow buyers to indicate if the procedure includes social, environmental or innovative objectives. PPDS uses this information to produced insights.

Which country makes use of the optional fields on strategic procurement. The darker the better.



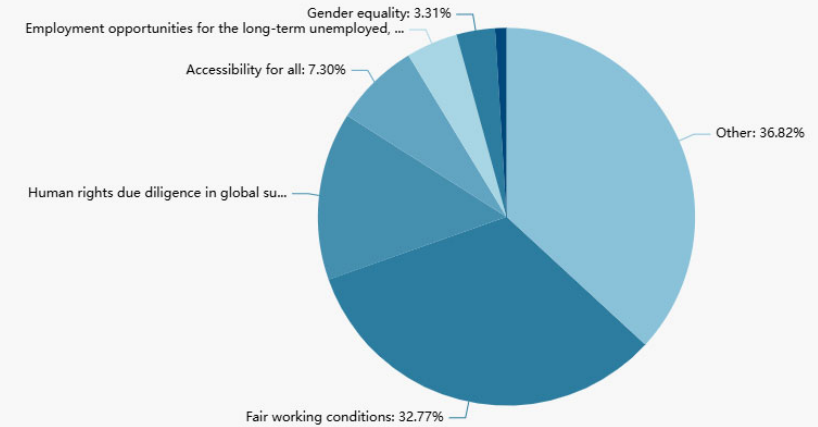
How much of the procedures are actually green, social or innovative.

Strategic Procurement

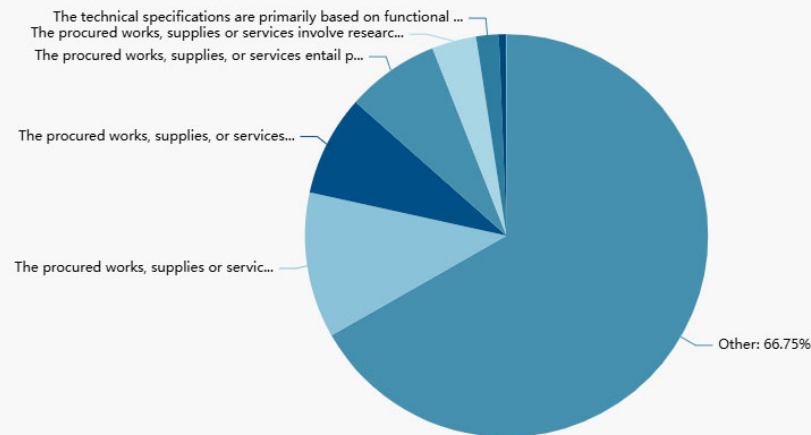
Strategic procurement focuses on using public procurement to respond to important societal challenges:

- 1) **Social objectives** (gender equality, fair working conditions, etc.)
- 2) **Innovative objectives** (inclusion of research, fostering innovation, etc.)
- 3) **Environmental objectives** (climate change mitigation actions, pollution reduction, etc.)

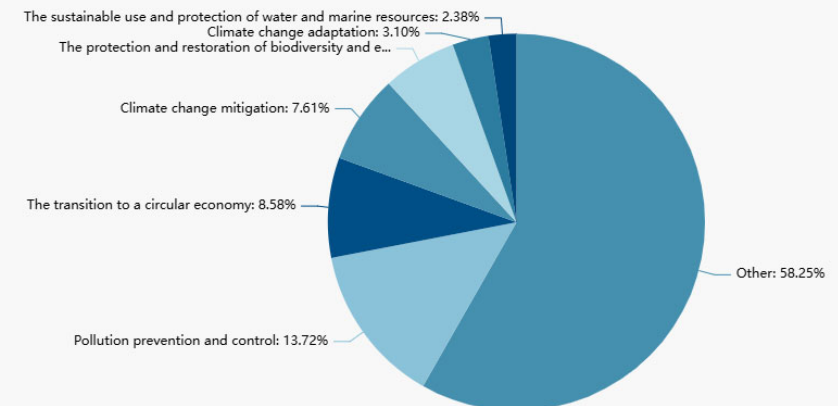
Distribution of strategic procurement indicating social objectives



Distribution of strategic procurement indicating innovative objectives



Distribution of strategic procurement indicating environmental objectives



SME participation in public procurement

The participation of Small and Medium-Sized companies in public procurement is an objective of all public institutions in the EU (and beyond). The PPDS enables the analysis of SME participation so decision-makers can implement the necessary measures to increase their participation.

40,5% of the contracts are awarded to SMEs, despite they represent 99,9% of the EU companies



Switzerland is the country where more contracts are awarded to SMEs, followed by Slovakia and Slovenia.

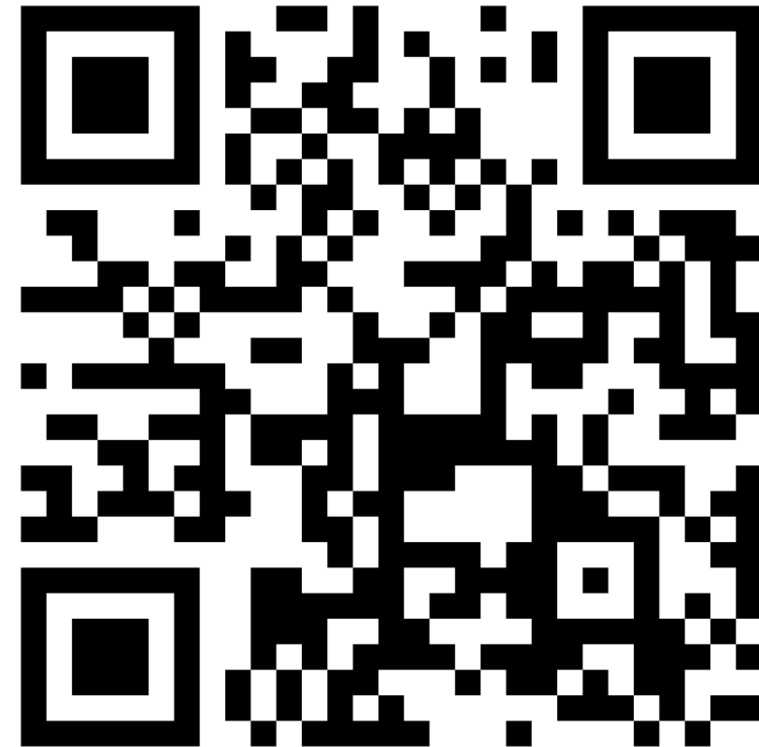
Access the PPDS website to see other use cases of public procurement data analysis

The PPDS website is the entry point to several use cases that use public procurement data:

- Measuring the data quality on public procurement
- Competition metrics (Single Bidder, Direct Awards)
- Procurement Efficiency Measurements
- Award Criteria Analysis
- SME Participation Analysis

and more!

Scan to access the PPDS website!



Additional works from NTT DATA in data spaces

NTT DATA elaborated a **White Paper** in collaboration with AWS on the **Power of Data Spaces for the Future of Digital Public Services**

The white paper elaborates on:

- Real-world examples and projects from NTT DATA and AWS like the **Public Procurement Data Space (PPDS)** and **European Mobility Data Space (EMDS)**
- Key challenges in data quality, governance, and sovereignty
- Best practices for digital service adoption
- High-level data-space architecture and core components
- 10 key lessons learned from successful implementations
- Practical guidance on hosting and securing data spaces with AWS

Scan to download the
White Paper!



Thank you very much
for your attention!

Daniel Arosa Otero
NTT DATA
Daniel.arosa.otero@nttdata.com

