



Data friction

Achievements and pitfalls for “liquid” data

**Public sector data: still a missed opportunity?
Brussels, 04/06/2019**

Dietmar Gattwinkel
DG CONNECT, European Commission
Unit "eGovernment & Trust"
Dietmar.Gattwinkel@ec.europa.eu

The challenge

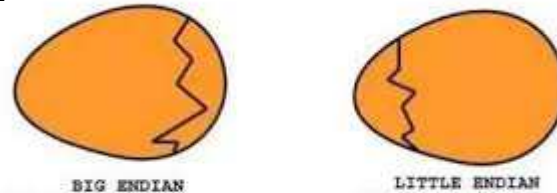
“There is only one "real world" but many different descriptions of this world depending on the aims, methodology and terminology of the observer.”

INSPIRE Data Specification on Land Cover – Technical Guidelines

Interoperability or Civil War?

(Gullivers Travels, 1726 by Jonathan Swift)

A law requiring all citizens of Lilliput to break their eggs at the little ends only. → Results in a a civil war breaking out between the “little endians” and the “big endians”.



In 1980 Danny Cohen borrowed the terms Little-Endian and Big-Endian for one of the most basic issues of data technical interoperability in data processing (byte ordering)



Open Data Principles as facilitators of interoperability

- **Data must be primary**

Data is as collected at the source, with the highest possible level of granularity, not in aggregate or modified forms.

- **Data Must Be Accessible and Access Must Be Non-Discriminatory**

Data is available to the widest range of users for the widest range of purposes, with no requirement of registration

- **Data Formats Must Be Non-Proprietary**

Data are available in a format over which no entity has exclusive control.



Simplyfying the Jigsaw

FINDABILITY AS AN INTEROPERABILITY ISSUE

Findability bottom up

Standardizing description

- *Metadata Standards – DCAT-AP and Core Vocabularies*



- *But how do we agree on keywords, let alone terms*

Keywords

inspireidentifiziert	61766
Správní jednotky	56295
Krajinný pokryv	56212
Katastrální mapa	51831
Budovy	37796
infomapaccessservice	31786
Využití území	31613
Bebauungsplan	31382
RÚIAN	30932
inspire	28612

Findability Top Down

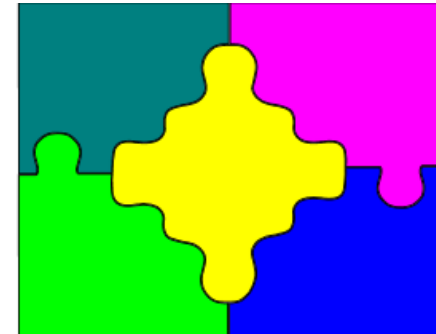
- *High value Datasets (Open Data Directive)*
- *Model data portfolios for municipalities*
Building on experience with a model data portfolio for selected Austrian cities, the Bertelsmann Foundation together with the Open Data Portal of Germany (GovData), had the KDZ – Centre for Administrative Research (Vienna) draw up catalogue of close to 300 datasets



A FRAMEWORK FOR DATA INTEROPERABILITY

Standardise Standardisation

- **eDAMIS** (electronic **D**ata files **A**dministration and **M**anagement **I**nformation **S**ystem) and **SDMX** (Statistical Data and Metadata eXchange)
- **INSPIRE**
- model repository, multilingual feature concept dictionary, and glossary
- Stakeholder engagement (consensus building process)
 - Spatial Data Interest Communities (SDIC)
 - Legally Mandated Organisations (LMO)
- Principle: Harmonize or transform?



Standards as Lingua Franca

Translation of Data Requests in TOOP

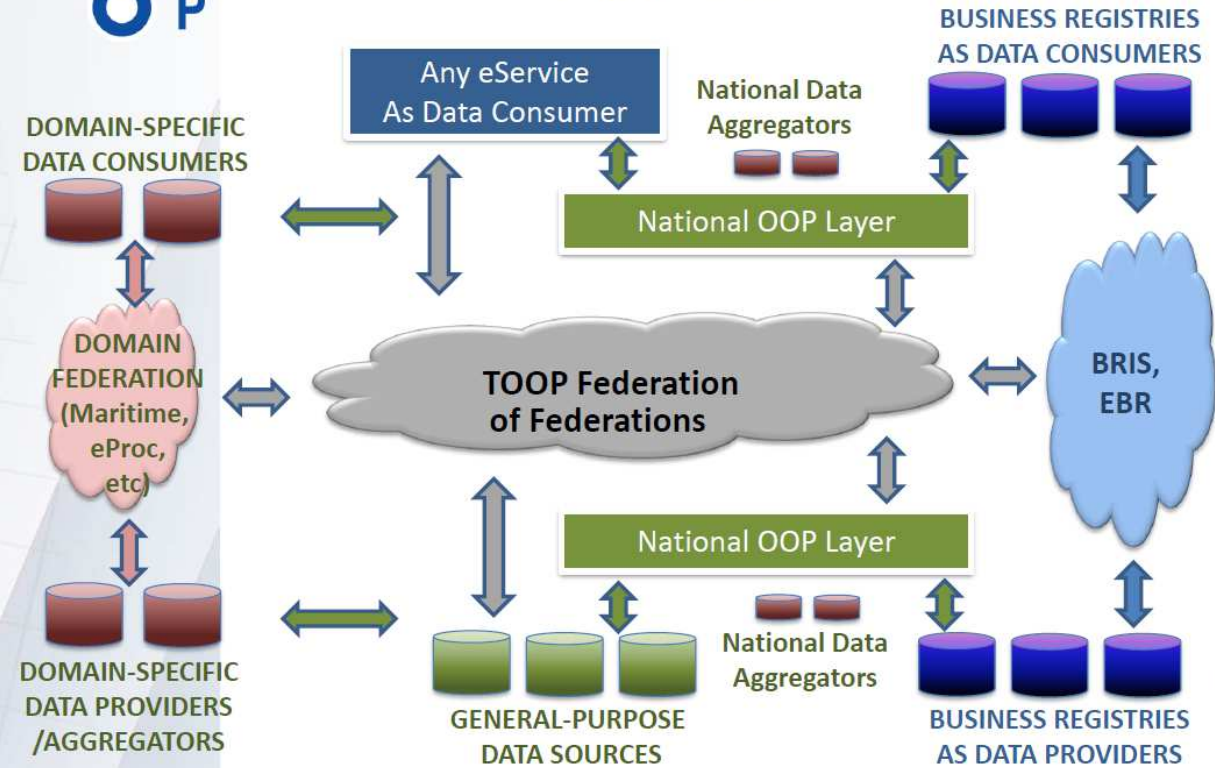


How TOOP Pilots fit into the big picture



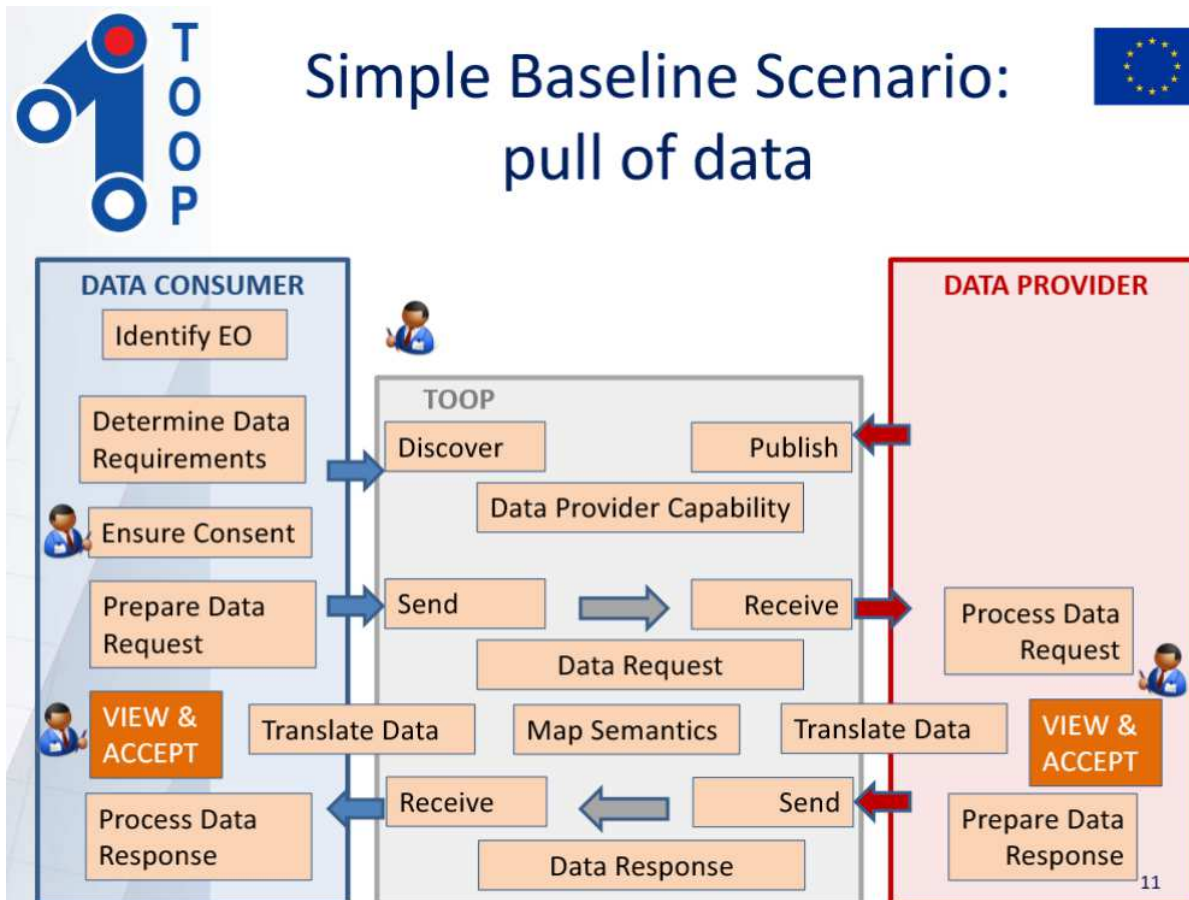
European Standards as Intermediaries

Reduces necessary interfaces by $(n^2-3n)/2$
Uncouples levels of interoperability through federation



Standards as Lingua Franca

Translation of Data Requests in TOOP



- Reuses CEF Digital BB **eDelivery** for LookUp and Transport
- Automatic Semantic Mapping

APIs as enablers for Data Interoperability (APIs4DGovStudy)



API State of play

- Assess digital government **APIs landscape, opportunities, current trends and horizons**

API Added value

- Identify the **key enablers, drivers, barriers, potential risks and mitigates**
- Identify **socio-economic impacts** of data sharing, data-driven government services and APIs

API Framework proposal

- Identify relevant **ICT standards and specifications**, select **best practices** and develop **recommendations**
- Provide a set of **policy recommendations** and propose **domains and thematic areas** to focus on



Specification v3.0

- The specification contains a standard, programming language-agnostic interface description for REST APIs, which allows both humans and computers to discover and understand the capabilities of a service.
- For instance, OAS v3.0 is currently implemented by the Netherlands' Cadastre, Land Registry and Mapping Agency (Kadaster) and the Netherlands Chamber of Commerce (Kamer van Koophandel). Also, OAS 3.0 has been implemented by different suppliers and vendors having a dominance over their market including the founders of OAI, Microsoft and Google.



Specification v3.0

Market acceptance

Coherence: no existing European standard or technical specifications being under consideration to become a European standard

Open Process

- **Openness**
- **Consensus**
- **Transparency**

Maintenance: following a defined and publicly available maintenance policy

Availability: publicly and free of charge

Intellectual Property

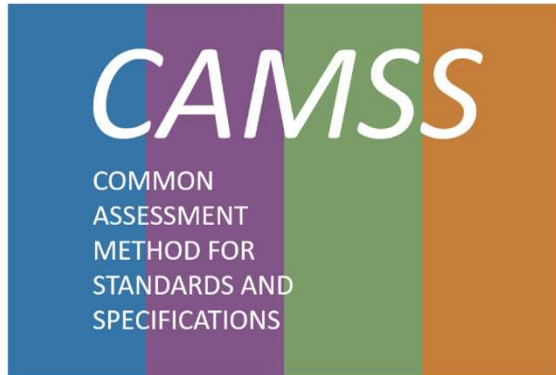
Rights (IPR): License available

Relevance

(Vendor) Neutrality and stability

Quality





Common assessment method for standards and specifications

Market acceptance

Coherence: no existing European standard or technical specifications being under consideration to become a European standard

Open Process

- **Openness**
- **Consensus**
- **Transparency**

Maintenance: following a defined and publicly available maintenance policy

Availability: publicly and free of charge

Intellectual Property Rights (IPR): License available

Relevance

(Vendor) Neutrality and stability

Quality





An ecosystem for Data Exchange in Europe

WHY NOW?

Coordinated Plan on AI

- For increased usage, data within a space should be made interoperable as much as possible, notably by agreeing on aiming for data formats that are open, FAIR, machine readable, standardised and documented, both in the interaction between public and private sectors, within sectors and across sectors.
- Creating a common European Data Space: a seamless digital area with the scale that will enable the development of new products and services based on data.





Thank you!



Follow us on  **@EU_eGov**  **EU_Public Services**



PITFALLS

Beware of “mosaicking”

- Grad student Anthony Tockar cross referenced “celebrities in taxis in Manhattan in 2013” photographs obtained through a Google Images search for “celebrities in taxis in Manhattan in 2013” (through the visible medallion number and the pickup location) with the “anonymized” taxi database, New York City had released to identify celebrity destinations and how much they paid and tipped for each fare.
-
- [\[Kalev Leetaru: The Big Data Era of Mosaicked Deidentification: Can We Anonymize Data Anymore? Forbes, 8-24-2016\]](#)



Further Examples

- Latanya Sweeney demonstrated in 2000 that 87% of the American population can be uniquely identified by a combination of just their ZIP code, gender and date of birth
- credit card data can be deanonymized at 90% accuracy using just three purchases
- a year's worth of cell phone records of 1.5 million subscribers can be rapidly reidentified using just four reference points
- Google Images searches for photographs of “celebrities in taxis in Manhattan in 2013” grad student Anthony Tockar was able to take the medallion number that is typically clearly visible in such photographs and the pickup location and time and cross-reference it against the taxi database to identify celebrity destinations