



Science & Technology
Facilities Council

Open it right: what is wrong about Open Data and what we can do about it

Vasily Bunakov
Science and Technology Facilities Council
United Kingdom



What is STFC ?

BIS | Department for
Business Innovation & Skills



Arts & Humanities
Research Council



Science & Technology
Facilities Council

- Grant awarding function
- Giving access to national and overseas facilities
- Large multi-disciplinary computing department



Science & Technology
Facilities Council

ENGAGE project: promotion of Open Data through collaborative data curation



An Infrastructure for Open, Linked Governmental Data Provision
towards Research Communities and Citizens

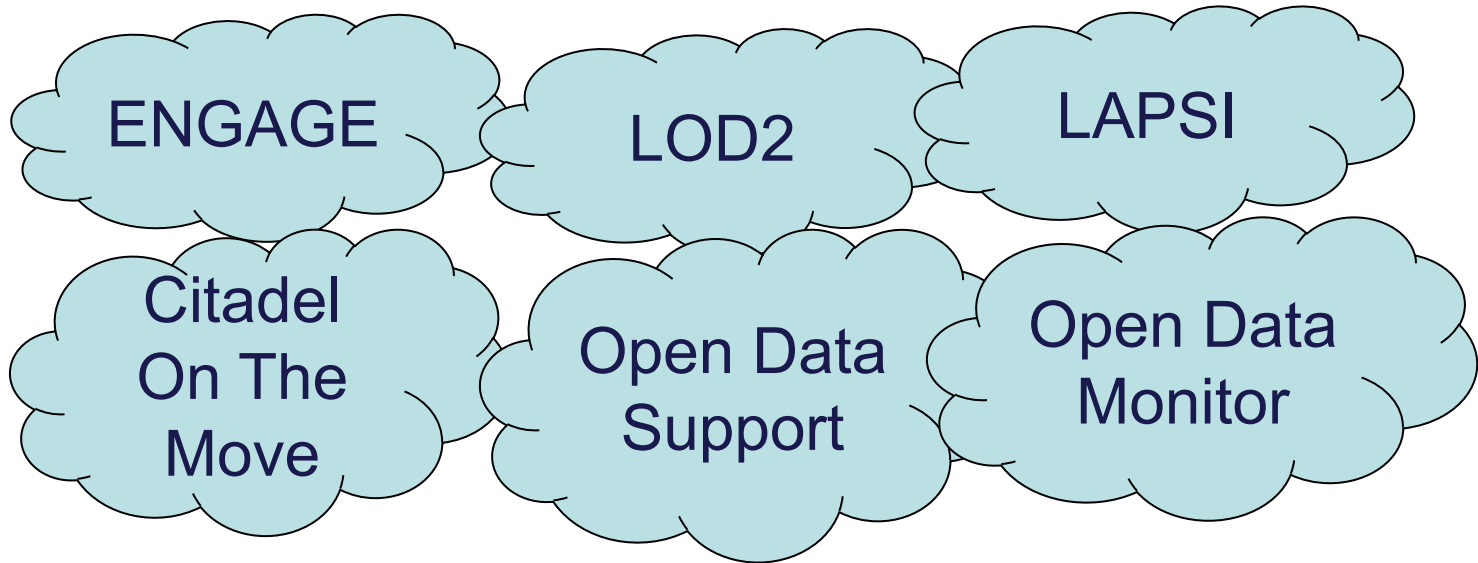


Project: www.engage-project.eu

Platform: www.engagedata.eu



A spell of doubt...



Is there room for yet another
(good) EU project on Open Data?



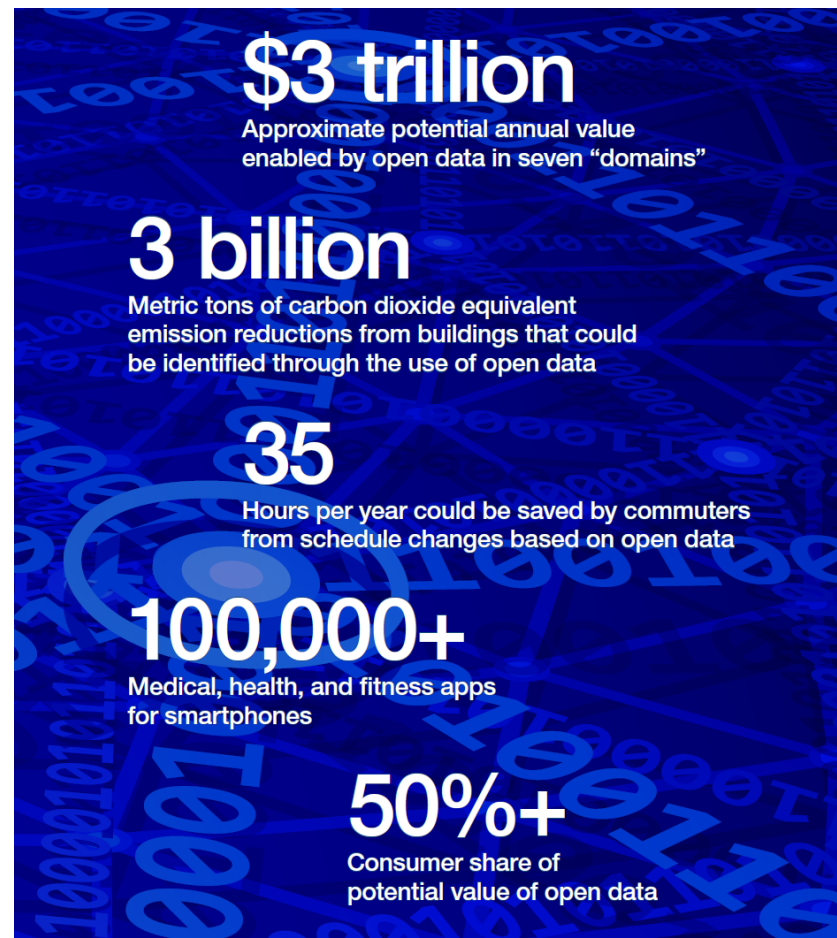
Open data potential

“...opening up public sector data could generate economic gains around €40 billion a year...”

Neelie Kroes, Vice President of the EC and Commissioner for the Digital Agenda, September 2012

“The total direct and indirect economic gains from PSI application and use across the whole EU27 economy would be in the order of €140 billion annually.”

Communication from the Commission to the European Parliament, December 2011



McKinsey&Co., October 2013



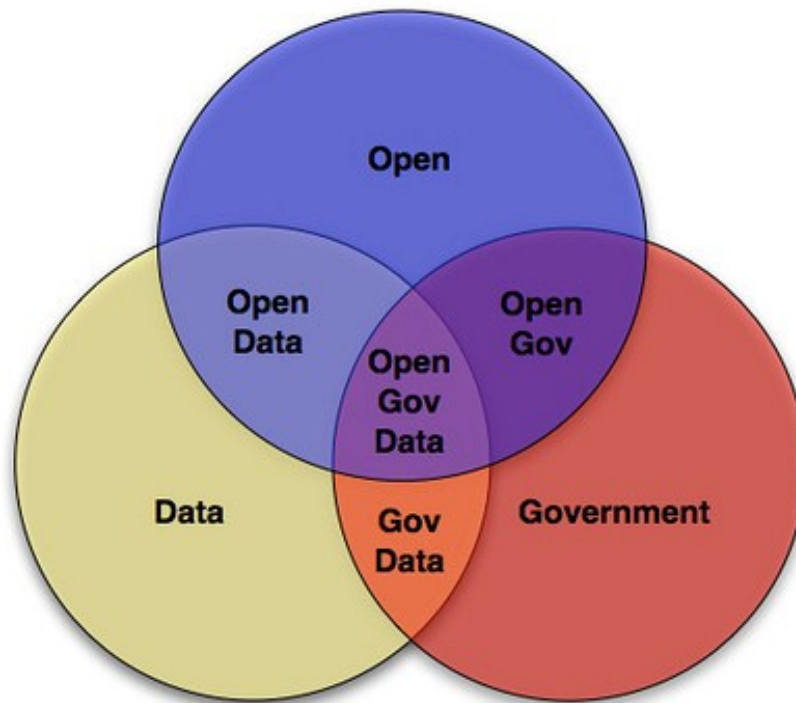
Keys to Open Data benefits



- Enough data
- Enough openness
- Enough quality

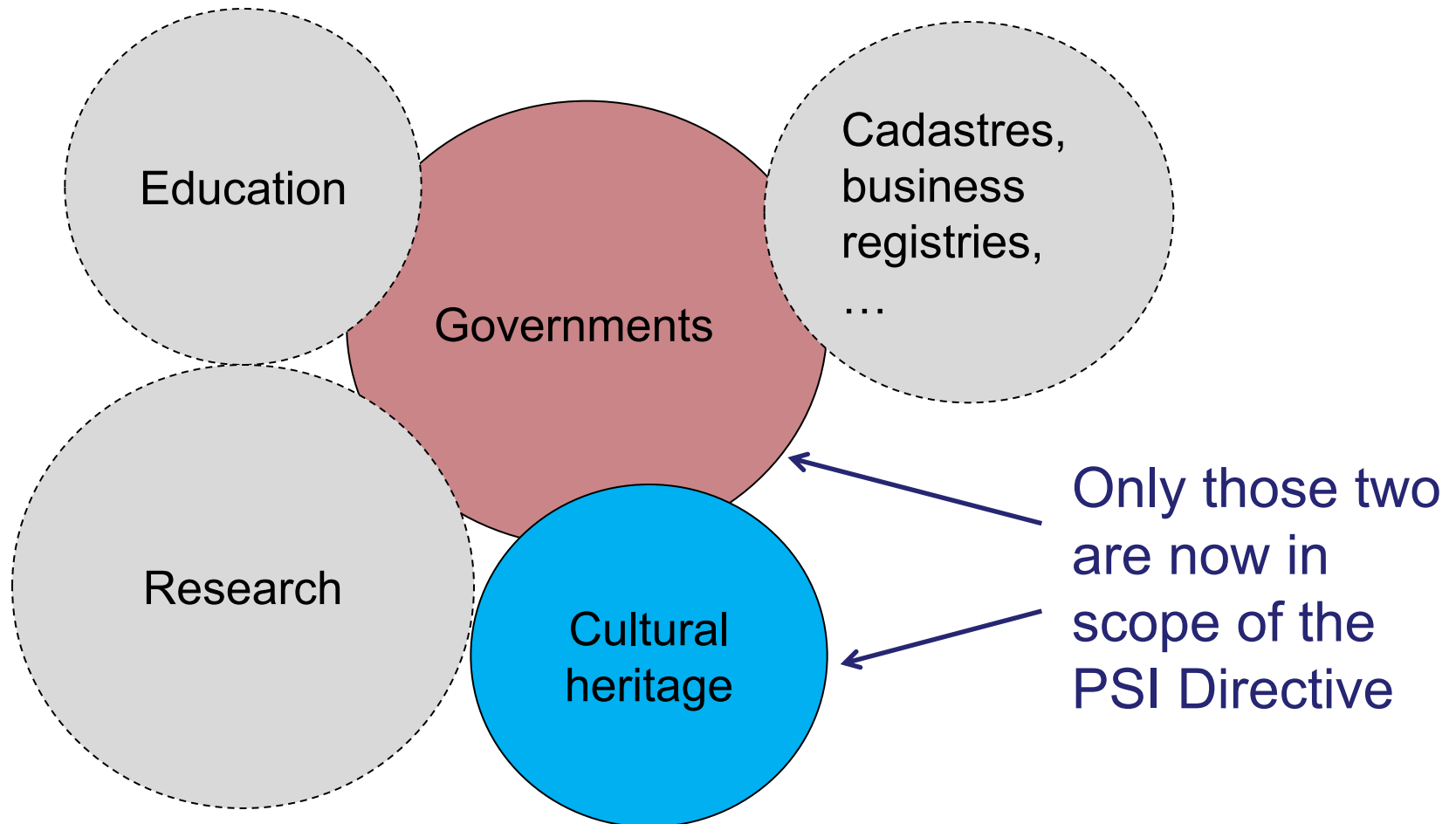


Open / Government / Data





Public Sector: not governments only





Example: Is environmental data of public interest?


Raw Data | Data.Gov | Open Federal Data - Mozilla Firefox

File Edit View History Bookmarks Tools Help

DATA Raw Data | Data.Gov | Open Fe... +

data.gov https://explore.data.gov/catalog/raw?sortBy=most_accessed&sortPeriod=month ☆ Google

An Official Web Site of the United States Government Sign Up


 **DATA.GOV**
EMPOWERING PEOPLE

HOME INTERACTIVE DATASETS RAW DATA APPS GEODATA COMMUNITY METRICS OPEN DATA SITES GALLERY WHAT'S NEW

Raw Data

Use the catalog below to access U.S. Federal Executive Branch datasets. Click on the name of a dataset to view additional metadata for that dataset. By accessing the data catalogs, you agree to the Data Policy. The Raw Data catalog provides an instant download of machine readable, platform-independent datasets.

« Go back to the main data catalog.


RAW DATA



Search

View Types

- Datasets
- External Datasets
- Files and Documents
- Filtered Views
- Charts
- Maps

Browse Raw Datasets

Most Accessed This month

	Name	Popularity	Type
1.	Worldwide M1+ Earthquakes, Past 7 Days Geography and Environment ANSS, geologist, plate, real time, environment, ... Real-time, worldwide earthquake list for the past 7 days	162,348 views	
2.	CMS Medicare and Medicaid EHR Incentive Program, electronic health record products used for attestation Science and Technology electronic health record, ... Data set merges information about the Centers for Medicare and Medicaid	5,504 views	



1st Challenge / Opportunity: Bring more types of PSI data

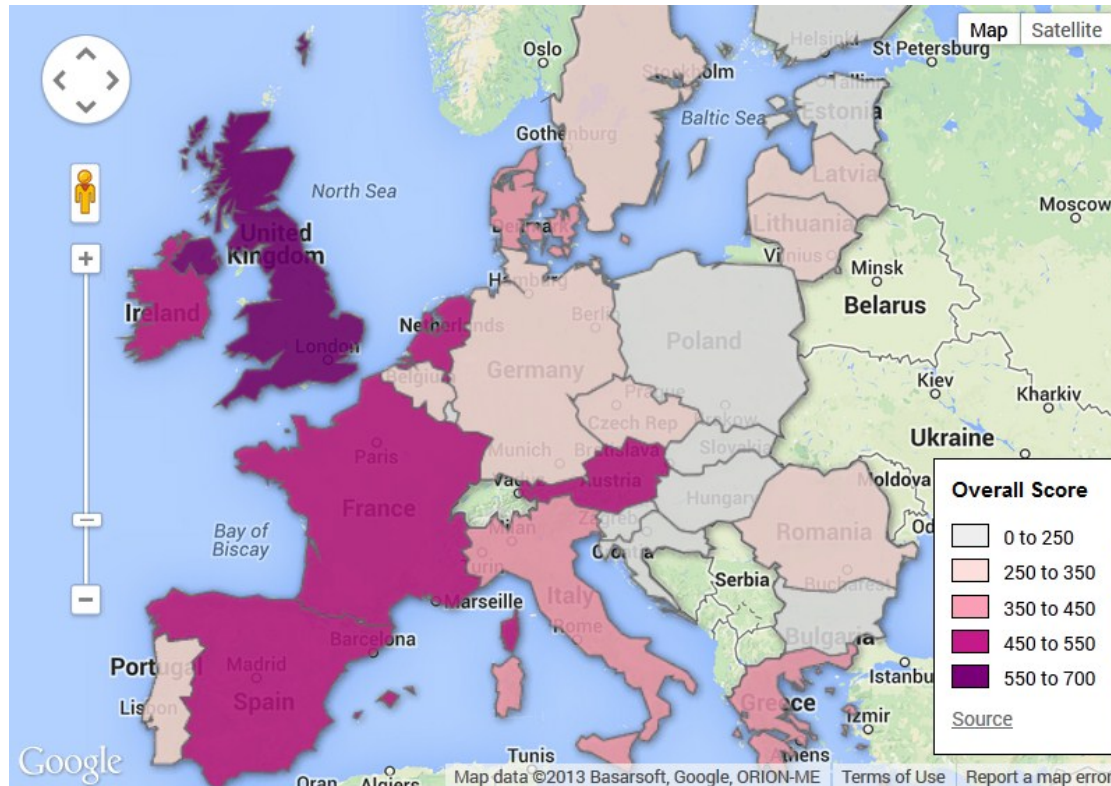
Good candidates:

- Environmental data
- Social research data
- Computer simulations

**Everything that can support
(better) business decisions**



Openness indicators: top down approach



<http://epsiplatform.eu/content/european-psi-scoreboard>



2nd Challenge / Opportunity: bottom-up Indicators

How you measure:

- Data size (volume, variety, dynamics)
- Data openness
- Data quality

Having the Commission opinion is all right:
now what Open Data practitioners think?



Along with Indicators, we need considering Facilitators

Examples:

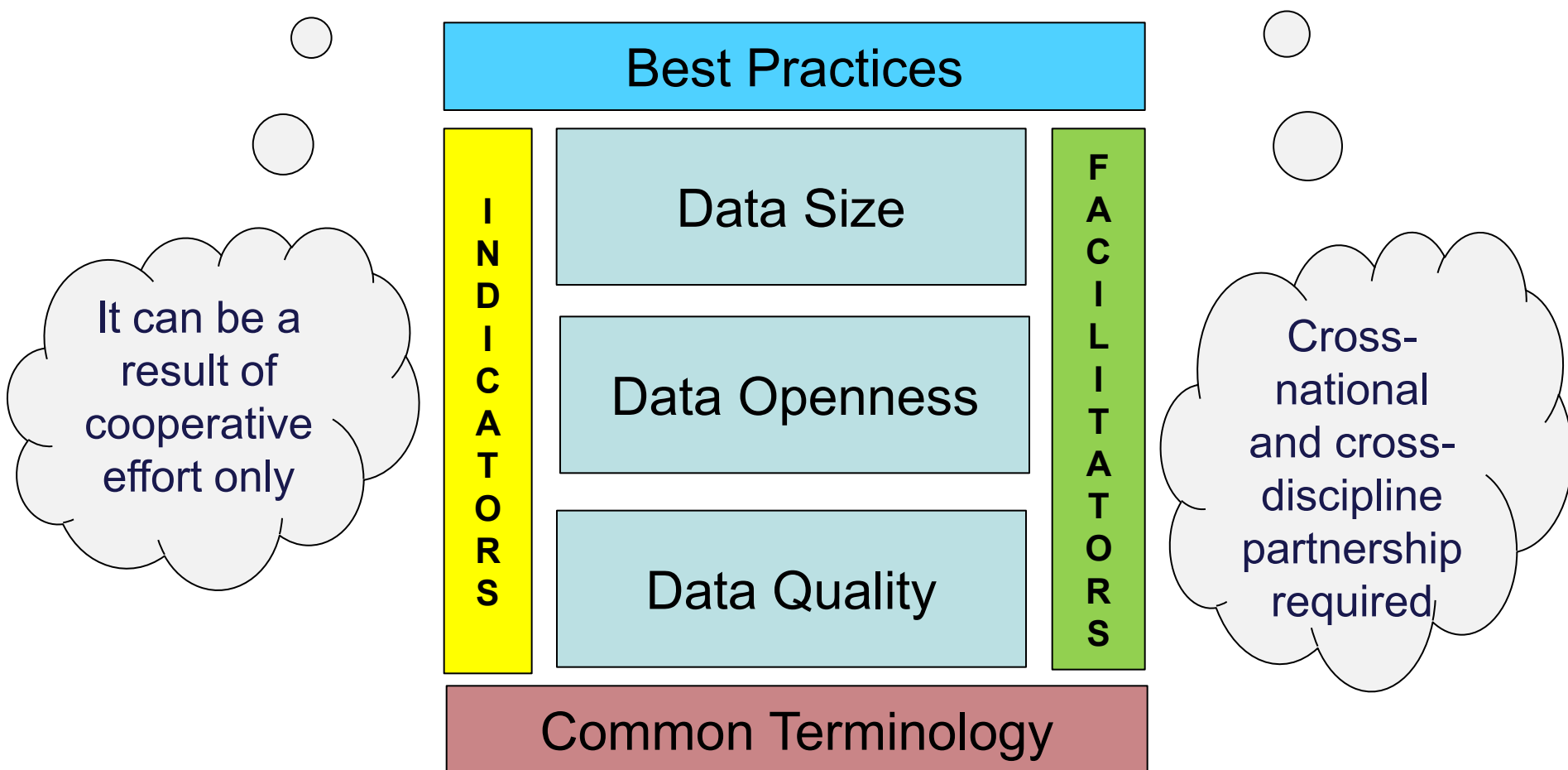
- Data licences, including machine-executable
- Reference information: vocabularies, ontologies, ...
- APIs and best practices of providing them
- Data provenance, workflows
- Persistent Identifiers

**Facilitators should be mapped to the
Quantity/Openness/Quality, and to the Indicators**



Public Sector Open Data

Reference Architecture





What the heck is “Reference Architecture”?

“A template solution for an architecture for a particular domain” –
http://en.wikipedia.org/wiki/Reference_architecture

In our case, it is a collaboratively created, well-structured, and re-usable corpus of knowledge about Open Data publishing in Public Sector – across different nations and Public Sector Information providers: public administration, research, ...

Think about it as a (version 0.1) template for Enterprise Architecture http://en.wikipedia.org/wiki/Enterprise_architecture – only Enterprise here is A) of a specific, loosely coupled nature, and B) specialized in one area: Open Data publishing



Sounds like a Horizon 2020 project

Why is it timely?

Lots of national, regional and domain-specific Open Data initiatives; lack of well-structured shared expertise

Scope of it?

Reference Architecture for Public Sector Open Data

SEE PREVIOUS SLIDES

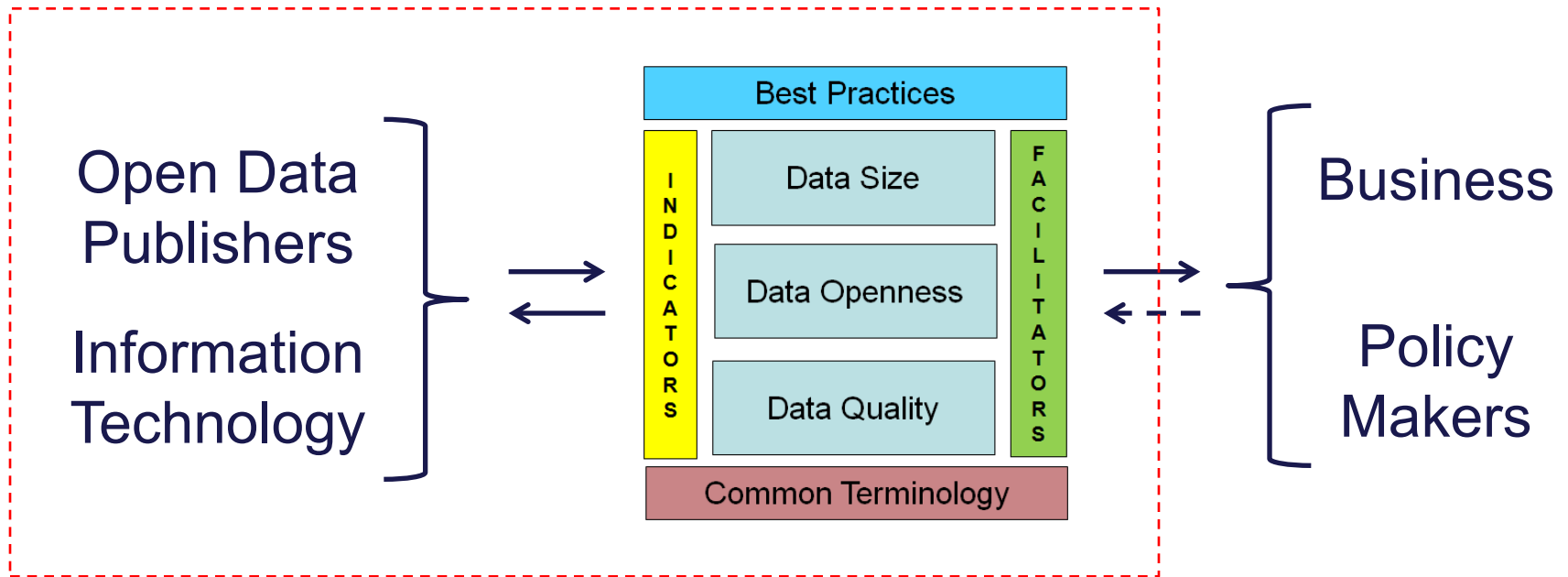
Value of it?

- **Common knowledge base**
- **Instrument for engagement with business**
- **“Shell” for IT projects**
- **Input for policy makers**

We look for more partners



Two contributors and four beneficiaries



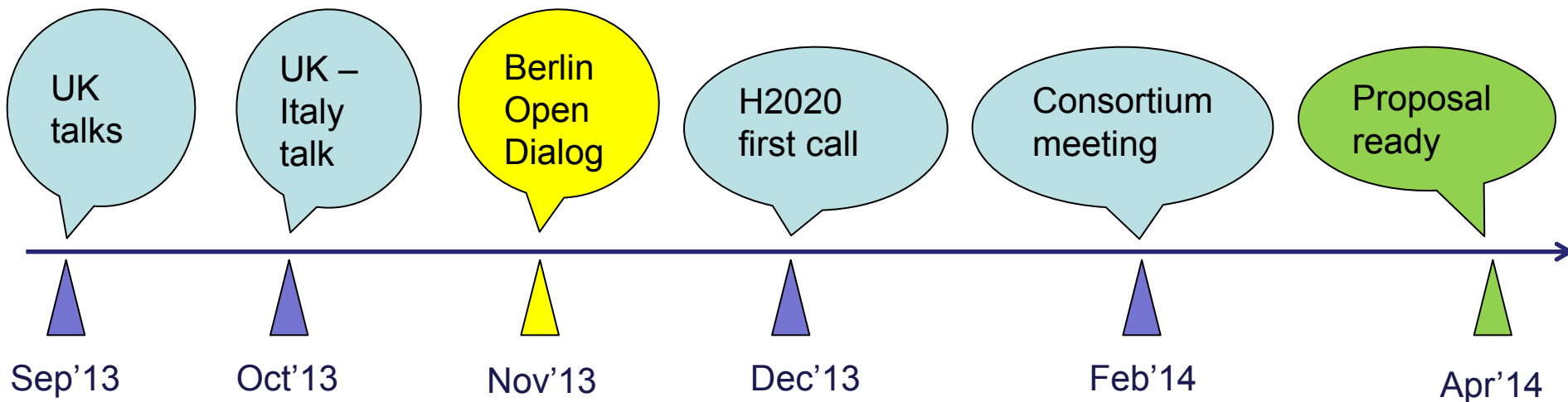
Small and Medium Enterprises may benefit
(relatively) more than Big ones



Where it fits in Horizon 2020?

“Societal challenges” pillar (~ 40% of Horizon 2020 budget)
“Inclusive, Innovative and Secure Societies” challenge

Timeline



Thank you, and get in touch!

My e-mail: vasily.bunakov@stfc.ac.uk

STFC Scientific Computing: www.stfc.ac.uk/SCD

