Clean and relevant open public data
How difficult can it be?

*CORDIS - CO*mmunity *R*esearch & *D*evelopment *I*nformation *S*ervice

@the Publications Office of the European Union (OP)

Baya Remaoun - Open Data Fest – 2 June 2017 – Caltanissetta, Italy

What's CORDIS?

- CORDIS is the European Commission's primary public repository and portal for *(almost)* all EU-funded research projects and results

- Commission (Research & Innovation DG) service operated at the OP (interinstitutional)

- OP more data oriented (metadata registry, controlled vocabularies, thesauri, open data portal etc.)

- Over 114 000 projects from FP1 to Horizon 2020

- During FP4-FP7: web services, applications, other databases, calls

- Horizon 2020: return focus to dissemination of research results

What's on CORDIS?

Two main pillars:

- **Data**: open and reusable access to public information on EU-funded projects and results held by the European Commission, e.g. project factsheets, report summaries

  - *Human resources: 1 business + 1 technical ± 4 external contract*

- **Editorial**: multilingual value-added products to support dissemination and exploitation of EU-funded research results, e.g. "Results in Brief", "Results Packs", *research*eu* magazines
  - *Human resources: 4 business ± 20 external contract*

Target audiences:

- Enterprise and industry – new products and technologies

- Researchers – scientific breakthroughs

- Policy makers – guidelines and recommendations

What's *not* on CORDIS?
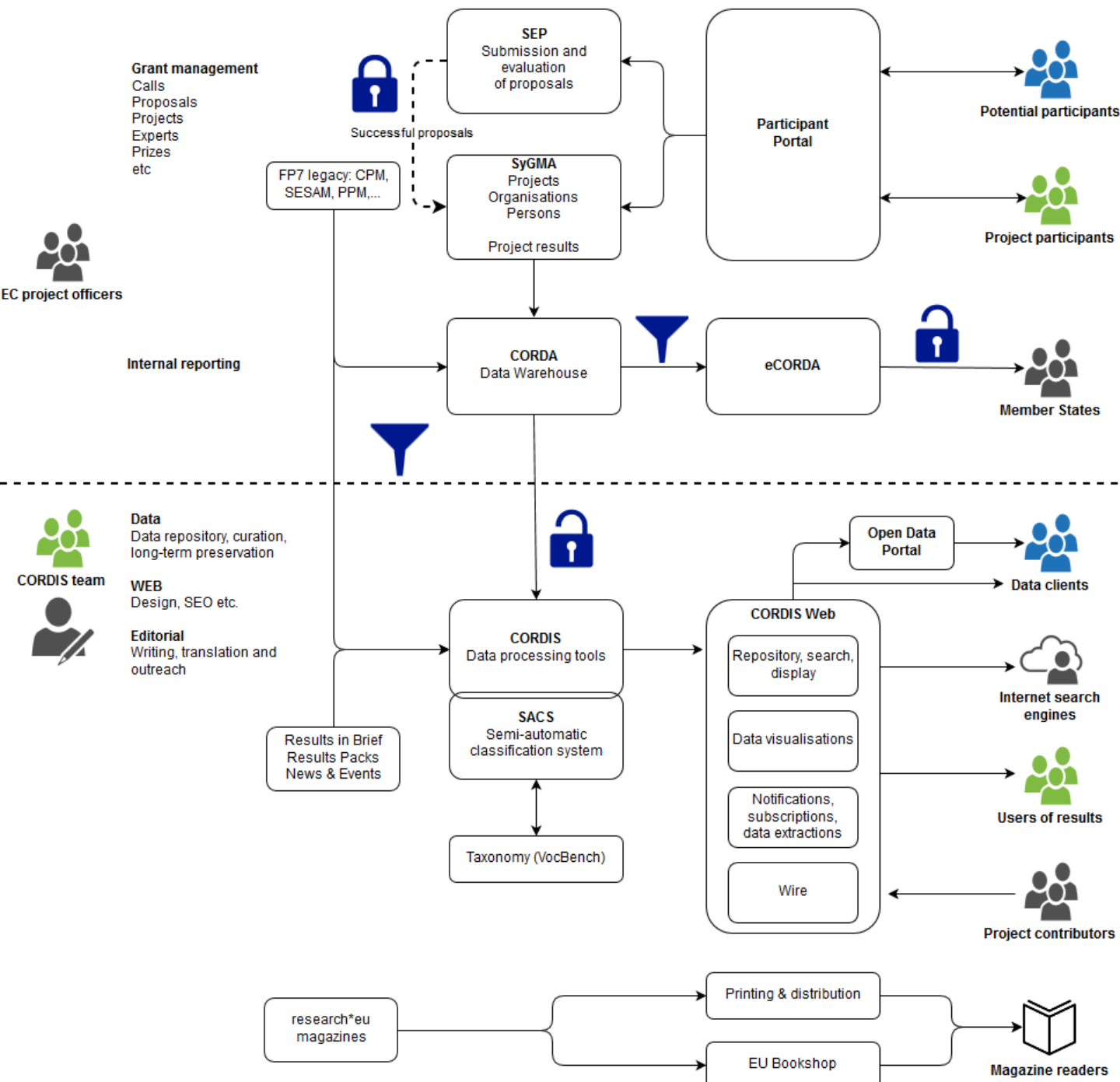
After two decades of web activity...

- Participant Portal: H2020 (FP7) calls, documents, registration, proposal submission, evaluation, experts, grant management...
  - Partners: will move to Participant Portal

- Europa and co: web pages for policy, themes, general public
  - Digital Transformation ongoing

- Output of research grants : (open access) publications, research data, project websites, grey literature...
  - not owned by EC so we can only capture links

- Certain parts of research programmes, e.g. JRC, EIT, some JUs

How is CORDIS managed?

- H2020 Common Support Centre in DG Research & Innovation (RTD): provides guidance and coordination, Partnership Agreement with OP

- CSC represents the "research family":
  - DGs: AGRI, CNECT, EAC, ENER, GROW, HOME, MOVE, RTD
  - Agencies: EASME, ERCEA, INEA, REA, GSA
  - Joint Undertakings: BBI, Clean Sky, ECSEL, FCH, IMI, SESAR, Shift2rail

From internal EC systems to users

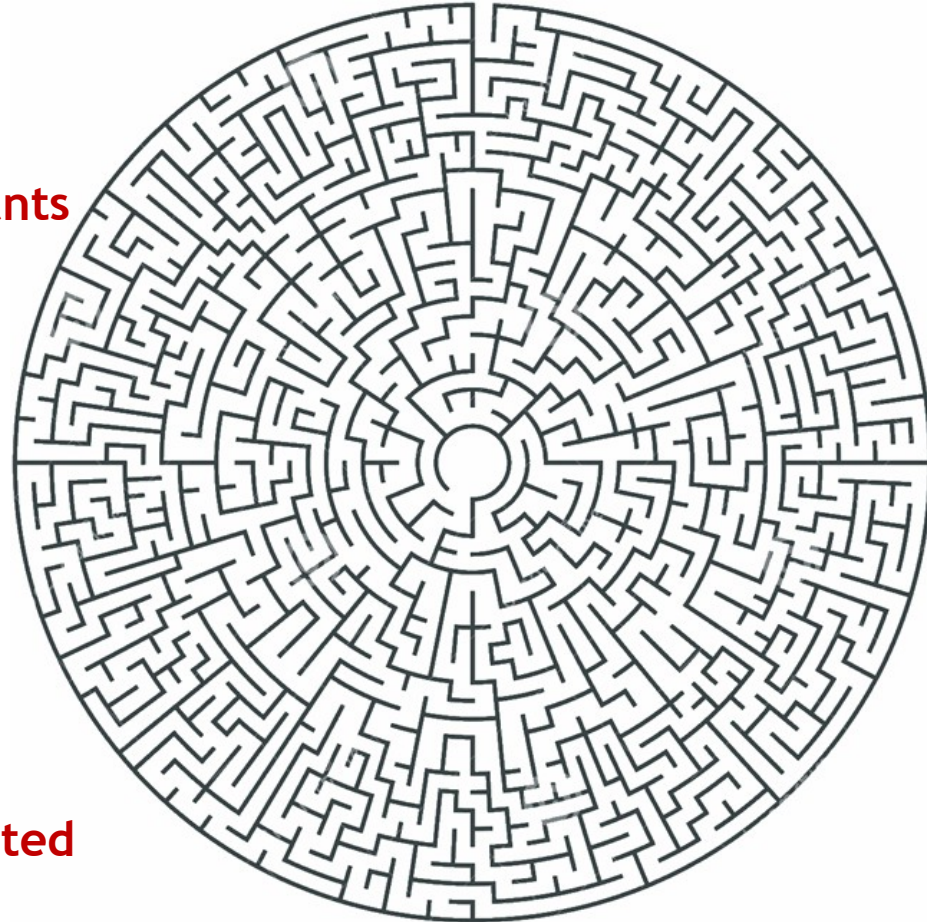# CORDIS evolution

Until 2011/2012

Mostly entry point for **project participants**

Focus on:

- Calls
- Reference documents
- Policy documents and webpages
- Partners
- Etc.

Multiple applications/databases

➔ Data as support only - **taken for granted**
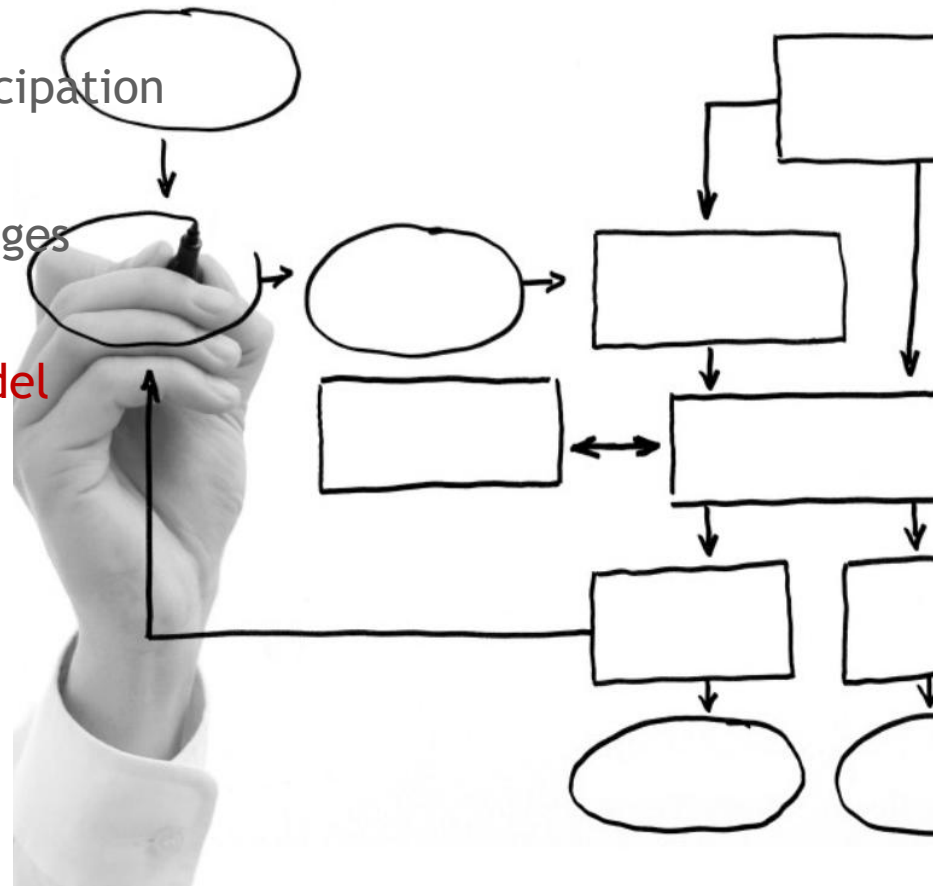
CORDIS evolution

After 2011/2012:

Information/applications on project participation moved to Participants portal

Thematic and other 20+ years of static pages were archived (completed in 2017)

Migration of technical platform, data-model and redevelopment of web applications

Structured content – ongoing work:

➜Understand & document

➜Obtain additional data

➜Re-structure

The EU Open Data Portal was created in 2012 (1st CORDIS dataset: 2013)

CORDIS constraints

- Internally Commission treats each FP separately
    - ➜CORDIS needs to harmonise over all FPs

- Source systems not conceived to collect data to be exposed externally
    - ➜Data quality issues

- OP outside the "research family"
    - ➜Less understanding of business needs and political context

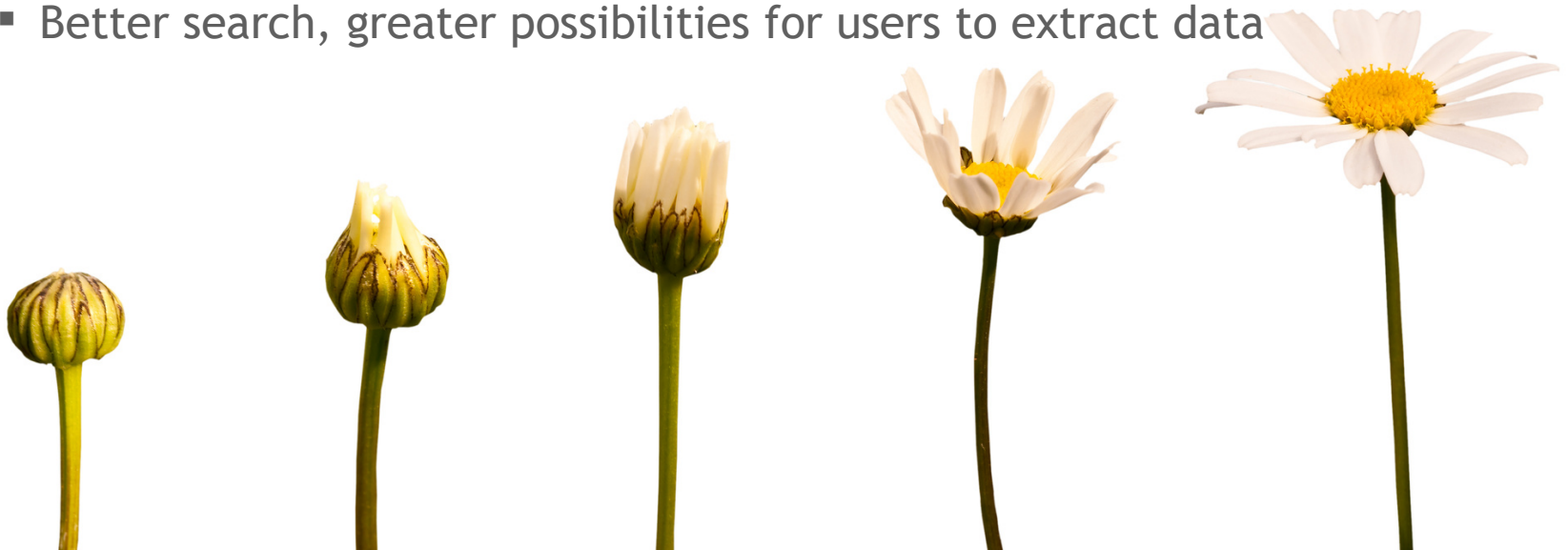# In the meantime at the European Commission...

- EC open data policy became important around 2011, but...
  - EC data operations and quality were/are still taken for granted

- Simplification of the procedures for participation and grant management ➜ H2020

- Corporate data, information and knowledge management policy at the European Commission decided in 2016 (C(2016) 6626 final) to:
  - Improve information retrieval and delivery
  - Work together and share information and knowledge
  - Maximise use of data for better policy-making
  - Create a culture of knowledge sharing and learning

What's next for CORDIS?

- New call for tender: specific lot for data management

- Better communication with data sources raising awareness

- Thematic taxonomy (text mining, align with EuroVoc) and Linked Open Data

- Better search, greater possibilities for users to extract data

Thank you for your attention!

Any questions...?

cordis@publications.europa.eu

@CORDIS_EU

cordis.europa.eu