# Studying infrastructures for open science

Teresa Gomez-Diaz

CNRS – Laboratoire d'informatique Gaspard-Monge



EGI Marketplace
EGI 2015 - Lisbon, 18-22 may 2015



# Open science: introduction

How to deal with research production, mainly articles, data, software?

- Articles are the most visible and accessible part of research.
- Software and data participate to the construction of research results published in articles.
- Research software is sometimes the "forgotten component".
- The distribution of software and data raises similar issues.

- Infrastructures make these objects visible, accessible, reusable, linked.
- Linked means to be able to retrieve (information from) any related object, once you have (information about) one.

**Wanted**: *use*, *copy*, *modify*, *redistribute*, *mine*, *link*... (note: *legal terms*)
in order to study, collaborate, validate...
but sometimes impossible without **(open) access**.

# Plan

To study infrastructures we analyze different components:

1. Design
   - Stakeholders
   - Target public requirements
2. Realisation
   - Services
   - Teams and gouvernance
   - Servers, interfaces (web sites...)
3. Evaluation
   - Scope
   - Scientific information
   - A successful infrastructure?

# 1.- Design: ***stakeholders*** (1/2)

- ***Designers***: projects, scientific communities, institutions or labs...
  - decide goals and objects to deal with
  - decide target public
  - study target public requirements
  - propose services
  - have landscape knowledge
  - find funders

- ***Funders***: research institutions, governments, EC, fundations, industry, scientific publishers...
  - establish mission (with designers)
  - provide political and scientific support
  - provide funding, resources
  - participate in evaluation
  - establish free/open access and other policies
  - avoid duplication of efforts and funding

# 1.- Design: ***stakeholders*** (2/2)

- ***Target public(s)***:
  - a scientific community
  - a research institution, a laboratory
  - several scientific communities (cross disciplinary exploration)
  - other infrastructures
  - SMEs, industry
  - society at large

- ***Infrastructure teams*** (see Infrastructure realisation)

# 1.- Design: target public requirements (1/2)

- **Researcher: user level**
  - formation, support, acquire best practices
  - how to find existing production
  - access to other experts skills
  - share own experiences
  - monitoring technology advances

- **Researcher: producer level**
  - all the above plus:
  - evaluation, recognition
  - distribution of own production (technical, legal issues)
  - promotion (scientific, technology transfert)

# 1.- Design: target public requirements (2/2)

- **Research institutions**
  - visibility, accessibility of the production
  - patrimonial management
  - evaluation and quality of the production
  - establish free/open access and other policies

- **Scientific community**
  - all the above plus:
  - specific ethical issues (for ex. medical studies)

# 2.- Infrastructure realisation: services

Range of possible services:

- metadata publication, links to related authors and objects
- search, mining, retrieval interfaces
- feed back tools: metrics, comments...
- publication of reviewed descriptions of research production (notices)
- peer review procedures for scientific publishing
- discovery, testing interfaces for software and data
- object deposit, preservation, permanent links...
- support on licensing, guidelines, best practices
- development, collaborative and social networking tools
- HPC, grid, cloud, networking services
- training, workshops

New services can be added following evolutions in the public requirements.

# 2.- Infrastructure realisation: teams and gouvernance

- whole internal team: can include scientists, computer engineers, librarians, funders, users and other experts
- gouvernance bodies: manages funding, resources, leads the technical team, decides evolutions and roadmap, organises the collaboration network, is the guarantee of the policies' application
- technical team: technical management and support, ensures the quality of service, answers demands of others teams
- scientific and expert team: include users, organise the scientific information, has editorial responsabilities, warrants the quality of the scientific information, manages the peer review services (if any)
- users' committee: propose evolutions, mendings, new services, gives feedback and evaluation

Come and participate (?)
The challenge is to build the architecture of participation.

# 2.- Infrastructure realisation: servers, interfaces

Include all the technical activities:

- provide services
- 7/7, 24/24
- quality of service
- technical evolutions
- software and other components
- monitoring tools

# 3.- Evaluation: scope, scientific information

By externals: evaluation also happens each time a potential user appears on the horizon.

Scope

- objects
- services
- target communities

Scientific information

- theme classification
- keywords (several levels)
- updating procedures
- reviewed
- publication workflows

# 3.- Evaluation: a successful infrastructure?

- Evaluation criteria
  - quality of information
  - quality of service
  - adopted by target public, well acknowledged
  - gouvernance
  - political and financial support
  - sustainability
  - collaboration/interaction with other infrastructures
- Sustainability: it is not only an economic issue
  - team, its organization, its gouvernance
  - funding
  - adoption by target public
- Links and collaboration with other infrastructures
  - interoperability
  - develop common standards
  - coordination, common strategy

# Conclusion

Research production (articles, software, data, and many others) spreads research knowledge.

Infrastructures are tools to help and build Open science, they make research production visible, accessible and reusable.

Improving understanding of infrastructures helps to improve Open science functioning.

**Goal:** to get to researchers "everyday life" and make software, data and other scientific objects shared and reused.

# References


- S. Andreozzi, O. Appleton, S. Coelho, T. Ferrari, S. Holsinger, Y. Legré, "The Open Science Commons paper", (01/2015, CC-BY-NC-SA), http://go.egi.eu/oscwp
- M. Assante, L. Candela, D. Castelli, P. Manghi, P. Pagano, "Science 2.0 Repositories: Time for a Change in Scholarly Communication", D-Lib Magazine, vol. 21 (1/2) article n. 4, http://www.dlib.org/dlib/january15/assante/01assante.html
- Juliana Freire, "A Provenance-Based Infrastructure for Creating Reproducible Papers", Workshop on Reproducible Research, 07/2011, https://www.youtube.com/watch?v=YHc0gNIgfdM
- Fred Friend, "A good future for university repositories", Repositories Support Project event Implementing funders' open access policies, 05/2014, http://www.rsp.ac.uk/events/implementing-open-access-fundersu2019-policies/
- Hong, Neil Chue, "Dealing with software: the research data issues", Dealing with Data Conference, SSI, 08/2014, https://www.era.lib.ed.ac.uk/handle/1842/9402
- Victoria Stodden, "Reproducible Research: A Digital Curation Agenda", 7th IDCC, 12/2011, https://www.youtube.com/watch?v=zJ-UxoiGSJM https://web.stanford.edu/~vcs/talks/IDCC-Dec62011-STODDEN.pdf
- And many consulted platforms...