TEXTS AS DATA:
TOWARDS A FAIRER, OPEN LANDSCAPE IN THE HUMANITIES
The road ahead today

1. A bit of context and a starting point
2. Defining «data» in the SSH
3. Issues with FAIRness
4. What’s next? The idea of «continuum»
Main purpose and objectives

The main purpose of the CO-OPERAS IN is the FAIRification of the research process and resources in the SSH, leveraging both on building services, sharing standards and on changing the communication culture in SSH. A second purpose is the contribution of CO-OPERAS network to the FAIR standards from the SSH data.

To improve Findability, CO-OPERAS IN will implement

- Identification services, already tested in the HIRMEOS project
- Metadata enrichment, with data publications crosslinking, it will promote dialogue within

THE RESEARCH INFRASTRUCTURE FOR THE HUMANITIES AND SOCIAL SCIENCES

- coordinating and federating resources
- nurturing the players
- taking care of the whole cycle

TO INCREASE THE OVERALL QUALITY OF THE ECOSYSTEM AND PROVIDE THE SERVICES RESEARCHERS NEED/DESERVE
...Houston, we have a problem...
... is «data» still a dirty word in the SSH?

Are artistic/humanities disciplines/methodologies only considered valid if we call them ‘data’? In other words, can we not accept different kinds of research methodologies as valid *on their own terms*, rather than on those imported from (or imposed by) other fields? #munin2018

In the humanities, we all use research data, although we may not be aware of it. It is like in the case of Monsieur Jourdain, the title character of Molière’s *Le Bourgeois gentilhomme*, who learnt, to his great satisfaction, that unwittingly he had been speaking prose all his life. With research data in the humanities it is exactly the same: you are using it, even if you don't know it, and once you realise it, it will affect your research workflow forever.
Data in the SSH, the pillar

We could then define data in the humanities broadly as all materials and assets scholars collect, generate and use during all stages of the research cycle. In this report we focus on digital assets.

RECOMMENDATIONS

» Think of all your research assets as research data that could be potentially reused by other scholars. Consider how useful it would be for your own work if others shared their data.
What is «data»?

- Data are never «raw» in the humanities.
- Data are always expression of a method.
- There is always a choice (methodological, epistemological, political...).
- There is always interpretation, subjectivity (data are not generated by a machine).
- Data are discussed, not taken for granted.
- Also books are data for further research/interpretation.

Data are anything you can formalize through a language.

Better «record» than data.

Data = documents
- Weak (mere registration)
- Strong (human intervention)

Data are a process, dynamic and diachronic.
...what is FAIR?

ACCESSIBLE

TRUSTED REPOSITORIES, FORMATS

ACCESSIBLE DOES NOT EQUATE TO OPEN
ACCESSIBLE = WHERE TO FIND THE DATA UNDER WHAT CONDITIONS

REUSABLE

LICENSES AND DOCUMENTATION

FINDABLE

METADATA, PERSISTENT IDENTIFIERS...

INTEROPERABLE

ONTOLOGIES, STANDARDS

TO KNOW MORE

Comment | OPEN
The FAIR Guiding Principles for scientific data management and stewardship
...why should we care about FAIR data?

We, Ministers, delegates and other participants attending the launch event of the European Open Science Cloud (EOSC):

1. Recall the challenges of data driven research in pursuing excellent science as stated in the “EOSC Declaration” signed in Brussels on 10 July 2017.

2. Reaffirm the potential of the European Open Science Cloud to transform the research landscape in Europe. Confirm that the vision of the European Open Science Cloud is that of a research data commons, inclusive of all disciplines and Member States, sustainable in the long-term.

3. Recognise that the implementation of the European Open Science Cloud is a process, not a project, by its nature iterative and based on constant learning and mutual alignment. Highlight the need for continuous dialogue to build trust and consensus among scientists, researchers, funders, users and service providers.

4. Highlight that Europe is well placed to take a global leadership position in the development and application of cloud services for Science. Reach out to new and open to the world, reaching out over time to other parts of the world.

5. Recall that the Council of Europe has established a roadmap and the federated

9. Call for the European Open Science Cloud to provide all researchers in Europe with seamless access to an open-by-default, efficient and cross-disciplinary environment for storing, accessing, reusing and processing research data supported by FAIR data principles.

Science Cloud a reality, hinting at the need to further strengthen the ongoing dialogue across institutions and with stakeholders, for a new governance framework to be launched in Vienna, on 23 November 2018.
Preliminary issues

**IN WHICH STEP AND HOW SHOULD THE FAIR PRINCIPLES BE APPLIED?**

**IT’S TIME CONSUMING, AND THERE IS NO INCENTIVE OR REWARD**

**WE NEED TO PRESERVE THE SPECIFICITY OF HOW WE DO RESEARCH IN THE HUMANITIES**

**METADATA IN WHICH LANGUAGE? NATIONAL, SAME AS DATA, ENGLISH?**

**LACK OF DATA AND METADATA SKILLS AMONG RESEARCHERS**

**ABSOLUTE NEED OF A REGISTRY OF EXISTING TOOLS**

**ARE WE, THE RESEARCHERS, WHO USE THESE RESOURCES WITHOUT ASKING HOW DATA WERE CREATED OR DESCRIBED, PART OF THE PROBLEM?**

**Services and tools need to be sustainable**

---

*Awareness of the FAIR principles and willingness to adopt them is not sufficient to transform data practices in any discipline. The paradigm shift requires effort, and this effort, which impacts on many roles in the research and higher education sectors, requires incentives, support, and recognition for adoption to be successful.*

*ALLEA report*
Specific issues - overview

- Metadata are so tailored on a specific project that interoperability is almost impossible.
- Complexity of corpora (they have many layers).
- Almost complete lack of metadata standards.
- Maintain the richness of metadata – if you lose the context in the SSH, you lose everything.
- Lack of unique identifiers.
- Data in the humanities often have a legal uncertainty.
- Lack of single point of access to text, data.
- Multilingualism.
- Absolute need of ontologies (philosophical job, not I.T.).
- Need of legal advice on licenses.
Why is FAIR an issue in the SSH?

- SSH ARE FRAGMENTED IN MANY DISCIPLINES, USUALLY GROUNDED IN REGIONAL, NATIONAL AND LINGUISTIC COMMUNITIES
- HETEROGENEITY OF SSH DATA (VARIETY OF DATA TYPES, FORMATS, LANGUAGES, METADATA)
OPERAS approach

CONSIDER ALL THE RESEARCH DIGITAL OUTPUTS AS ELEMENTS OF THE SCHOLARLY COMMUNICATION LIFECYCLE

SSH COMMUNITY NEEDS TO BE STRUCTURED AROUND SHARED EXPERTISE AND PRACTICES

THE FAIR PRINCIPLES ARE ONE OF THE MOST VALUABLE TOOLS AS THEY CAN BE BROADLY APPLIED AND WIDELY SHARED
...working for you...

- DISCIPLINARY WORKSHOPS TO SHARE BEST PRACTICES
- WORKSHOPS ON FAIRIFICATION OF REAL, EXISTING DATABASES
- «FAIR IMPLEMENTATION PROFILE» IN EACH DISCIPLINE / EACH TYPE OF RESEARCH OUTPUT
- WORKING IN SYNERGY WITH PROJECTS LIKE PARTHENOS
- EXPLORE A MINIMAL METADATA SET, EXPANDABLE [TO BE ADDED TO FAIRSHARING]
...with a vision

PROVIDING THE BUILDING BLOCKS OF A FAIR CONTINUUM FOR SSH RESEARCHERS, FROM DATA PRODUCTION AND MANAGEMENT, TO RESULTS DISSEMINATION AND PUBLICATION
... thank you and... call us!