FAIR data, the SSH, CO-OPERAS

Elena GIGLIA
Porto, Sept. 17 2019
elena.giglia@unito.it

WELCOME TO THE **CO-OPERAS** WORKSHOP

PLEASE RAISE YOUR HANDS...
-RESEARCHERS
-LIBRARIANS
-RESEARCH MANAGERS
-...

… a bit of context

OPERAS:
40 partners
16 countries
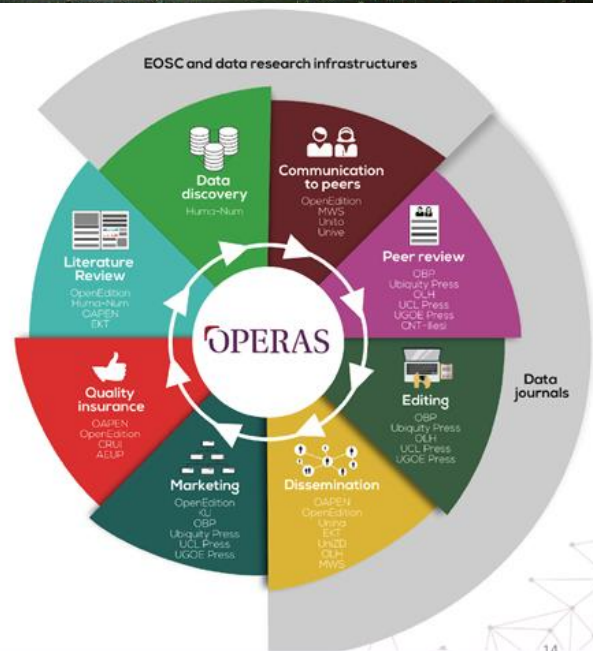
-TAKING CARE OF THE WHOLE CYCLE, INVOLVING ALL THE PLAYERS
-TAKING BACK CONTROL OVER SCHOLARLY COMMUNICATION

OPERAS: STRATEGY

- integration rather than fragmentation
- coordination rather than competition
- nurturing the players to increase the overall quality of the system

https://www.hirmeos.eu/

Open Access, Monographs, Science

HIRMEOS - High Integration of Research Monographs in the European Open Science infrastructure

OPEN SOURCE TOOLS CAN BE REUSED

The HIRMEOS Project
from HIRMEOS

https://vimeo.com/260765668

...leveraging on HIRMEOS

IDENTIFICATION
(DOI, FundRef, ORCID)

entity-fishing Service
Implementations on Publishing Platforms

Named Entity Recognition and Disambiguation
Entity extraction and disambiguation is the task of determining the identity of entities mentioned in a text against a knowledge base. The identification and resolution of named entities like person-name, location, etc. provides many practical applications, e.g. possibility to extract lists of people, to map different texts, to generate timelines and to provide an enhanced search. This is of great importance not only for research but also for the publishing process.

INRIA Entity-Fishing and its Advantages
Entity-Fishing, the NERD implementation developed by INRIA, is a service available within the DARIAH-EU infrastructure and used by the HIRMEOS partners to enrich Open Access digital monographs published on five digital platforms.
- No requirement for expertise in knowledge engineering
- Multilingual support (currently it supports 5 languages)
- Standard REST API

Entity phishing

ENTITY
RECOGNITION

Annotation Service
for Digital Monographs

Why Open Annotation
Open Annotation is essential in nearly any part of the research lifecycle. It enables organization and collaboration atop research materials; inline peer review; augmentation of articles with additional information, links, images or videos; elaboration around citations; content corrections or updates, and has extensive use cases in the teaching and learning space

The Hypothes.is Tool
The Hypothes.is Annotation Tool allows annotations at a sentence or phrase level, such as criticism or notes on news, blogs, scientific articles, books, terms of use, campaign initiatives, legislative procedures, and more. The tool is based on an open source JavaScript library and annotation standards developed by the World Wide Web Consortia (W3C)

Annotation

ANNOTATION

**Peer-Review
Certification System**

There is a prototype of a peer-review certification system in testing. Easy to understand icons display which kind of peer-review process has taken place to increase the confidence in the quality of the OA monographs published on the HIRMEOS partner platforms.
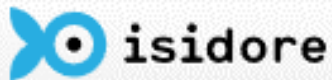
Certification

Welcome!

METRICS
DASHBOARD

Readership metrics

Summary	10	141,894	158
Metrics by title	Titles published	Total Access	Countries reached
Global reach			

Quarterly readership

Measure	Value
classics library: sessions	7,613
google books: views	65,634
oapen: downloads	1,390
obp html: sessions	8,830
obp pdf: sessions	51,762

OPERAS
open access in the european research area through scholarly communication

PEER REVIEW
CERTIFICATION

...3 AXES...

isidore

DISCOVER RESEARCH
(multilingual discovery tool)

OPERAS / TRIPLE 29/05/2019

OPERAS receives funding for European discovery solution TRIPLE

The European Commission will finance the project TRIPLE (Targeting Researchers through Innovative Practices and multiLingual Exploration) under the Horizon 2020 framework with approx. 5,6 million Euros for a duration of 42 months. TRIPLE will...
May 29, 2019

doab
directory of open access books

CERTIFY RESEARCH
(peer review)

hypotheses
ACADEMIC BLOGS

ACTIVATE RESEARCH
(research and society blogging)

CO-OPER

GO FAIR: a bottom-up international approach

Context of GO FAIR
Watch videos

...CO-OPERAS IS AN IMPLEMENTATION NETWORK in **GO FAIR**
TO «FAIRIFY» THE RESEARCH PROCESS IN THE SSH

IS THERE A TOOL, A METADATA SET, ANYTHING A COMMUNITY USE?
LET'S CONVERGE ON IT AND THEN BUILD UPON [LIKE THE TCP-IP]

…BUT FIRST, F

FAIR data training
Findable
Accessible
Interoperable
Reusable
FAIR for Developers
FAIR data self-assessment tool

F1. (meta)data are assigned a globally unique and ete

There are many resources created by the ARDC on the topic of metadata

- Metadata guide
- Data versioning

The ARDC has information on persistent identifiers on three different leve

- Persistent identifiers: awareness level
- Persistent identifiers: working level
- Persistent identifiers: expert level

It is also a provider of services for minting persistent identifiers of many di
of the data being identified):

- Digital Object Identifier (DOI) System for research data
- Handle minting Service (Identify My Data)
- International Geo Sample Numbers (IGSN)
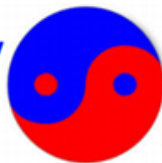
Complementary to the assignment of persistent identifiers is their proper

FAIR: what you have to do

FAIR Principles
Sci. Data 3:160018 doi: 10.1038/sdata.2016.18 (2016)

Technology

Domain-relevant content

Findable:

F1 (meta)data are assigned a globally unique and persistent identifier;

F2 data are described with rich metadata;

F3 metadata clearly and explicitly include the identifier of the data it describes;

F4 (meta)data are registered or indexed in a searchable resource;

Interoperable:

I1 (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2 (meta)data use vocabularies that follow FAIR principles;

I3 (meta)data include qualified references to other (meta)data;

Accessible:

A1 (meta)data are retrievable by their identifier using a standardized communications protocol;

A1.1 the protocol is open, free, and universally implementable;

A1.2 the protocol allows for an authentication and authorization procedure, where necessary;

A2 metadata are accessible, even when the data are no longer available;

Reusable:

R1 meta(data) are richly described with a plurality of accurate and relevant attributes;

R1.1 (meta)data are released with a clear and accessible data usage license;

R1.2 (meta)data are associated with detailed provenance;

R1.3 (meta)data meet domain-relevant community standards;

CC BY Erik Schultes

# FAIR data – priorities

1.7 Priority recommendations

1.7.1 Step 1: Define – concepts for FAIR Digital Objects and the ecosystem

» Rec. 1: Define FAIR for implementation
» Rec. 2: Implement a model for FAIR Digital Objects
» Rec. 3: Develop components of a FAIR ecosystem

In order to implement FAIR, research communities must define how the FAIR principles and related concepts apply in their context. This will differ based on the data types, the nature of research (e.g. ethical sensitivities or commercial partners) and the level of existing support for data sharing. The process of definition will help to identify points where the FAIR principles need to be supported with additional concepts and policies. To make

- DEFINE FAIR FOR IMPLEMENTATION
- IMPLEMENT A MODEL FOR FAIR DIGITAL OBJECTS
- DEVELOP COMPONENTS OF FAIR ECOSYSTEM

FAIR data: the ideal FAIR object
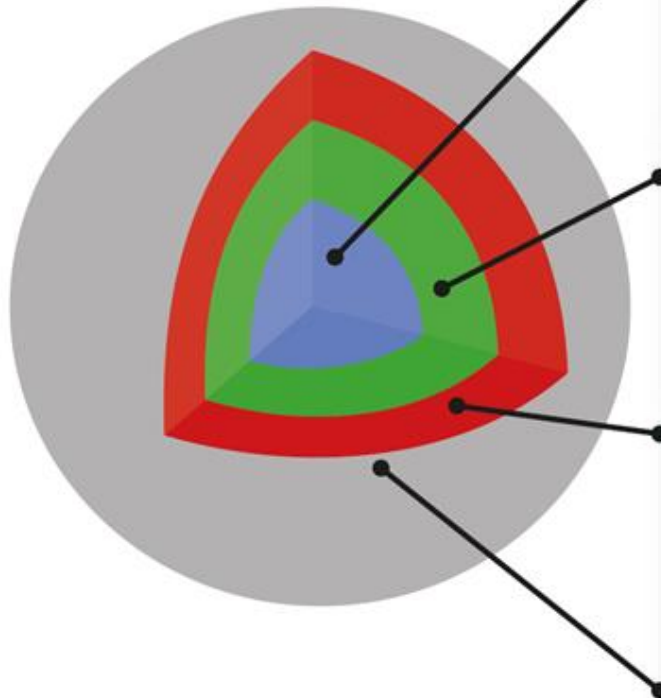
DIGITAL OBJECT
Data, code and other research outputs
At its most basic level, data or code is a bitstream or binary sequence. For this to have meaning and to be FAIR, it needs to be represented in standard formats and be accompanied by Persistent Identifiers (PIDs), metadata and documentation. These layers of meaning enrich the object and enable reuse.

IDENTIFIERS
Persistent and unique (PIDs)
Digital Objects should be assigned a unique and persistent identifier such as a DOI or URN. This enables stable links to the object and support citation and reuse to be tracked. Identifiers should also be applied to other related concepts such as the data authors (ORCIDs), projects (RAIDs), funders and associated research resources (RRIDs).

STANDARDS & CODE
Open, documented formats
Digital Objects should be represented in common and ideally open file formats. This enables others to reuse them as the format is in widespread use and software is available to read the files. Open and well-documented formats are easier to preserve. Data also need to be accompanied by the code use to process and analyse the data.

METADATA
Contextual documentation
In order for Digital Objects to be assessable and reusable, they should be accompanied by sufficient metadata and documentation. Basic metadata will enable data discovery, but much richer information and provenance is required to understand how, why, when and by whom the objects were created. To enable the broadest reuse, they should be accompanied by a plurality of relevant attributes and a clear and accessible usage license.

ECD, 2015

SSH: fragmentation...

MULTILINGUALISM

fragmentation of **research fields**, across **many disciplines** and subdisciplines, usually grounded in **regional, national and linguistic specific communities**

SCARCE DISCOVERY AND REUSE

fragmentation of SSH **data** across **different types, formats, languages, disciplines**

SPECIFIC DATA CURATION

fragmentation of **small and smart data** which need to be precisely **qualified, described, managed and curated**

…BUT, IN THE FIRST PLACE… WHAT IS «DATA» IN THE SSH?

# …that's why we need specific paths and tools to make data FAIR

ACCESSIBLE

- metadata integration into a single point of access and research

- best practices on data preservation shared between different repositories and platforms

FINDABLE

- persistent Identifiers assigned to SSH data, as implemented for the publications in the HIRMEOS project

- metadata enrichment

- discovery tool enhanced with controlled vocabularies supporting multilingualism

INTEROPERABLE

- common adoption of standards as piloted in the HIRMEOS project

- conversion tool for publishing open standards

- mapping to ontologies and controlled vocabularies (FAIR sharing registries)

REUSABLE

- help spreading licensing practices

- improve the use of annotation tools

- turn texts into machine readable data to boost Text and Data Mining

- "activate research" through the COESO blog platform
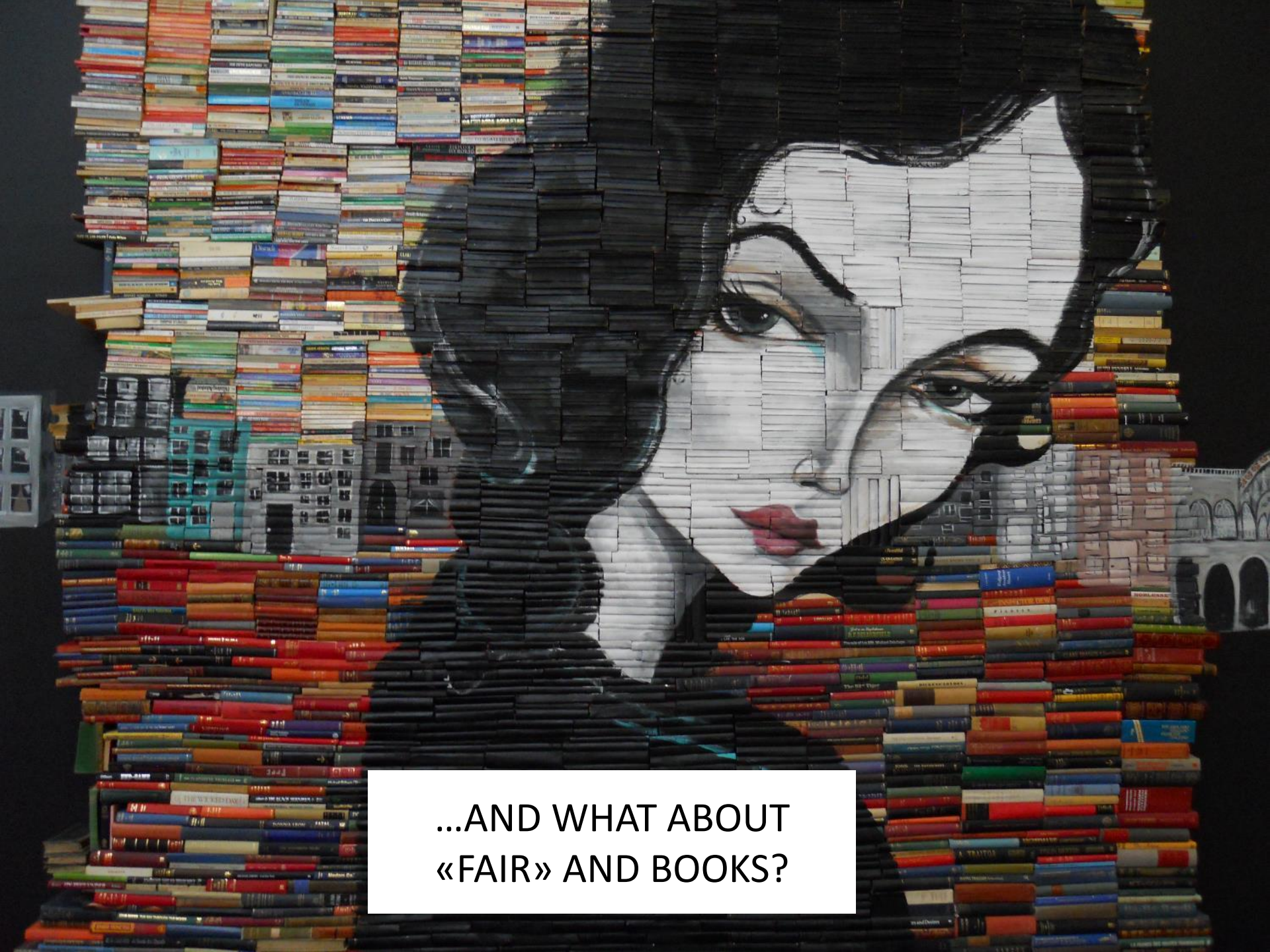
...AND WHAT ABOUT «FAIR» AND BOOKS?

AHEAD OF US

- TRIPLE: FAIR Data and discoverability (Peter Kraker)
- Is FAIR fair enough for cultural heritage? (Franco Niccolucci)
- FAIR and innovation in scholarly communication (Erzsébet Tóth-Czifra)

Building an international FAIR ecosystem in the SSH
(Ana Miguéis, Bruno Neves, Carlos Costa, Delfim Leão)

...HAVE A NICE WORKSHOP!