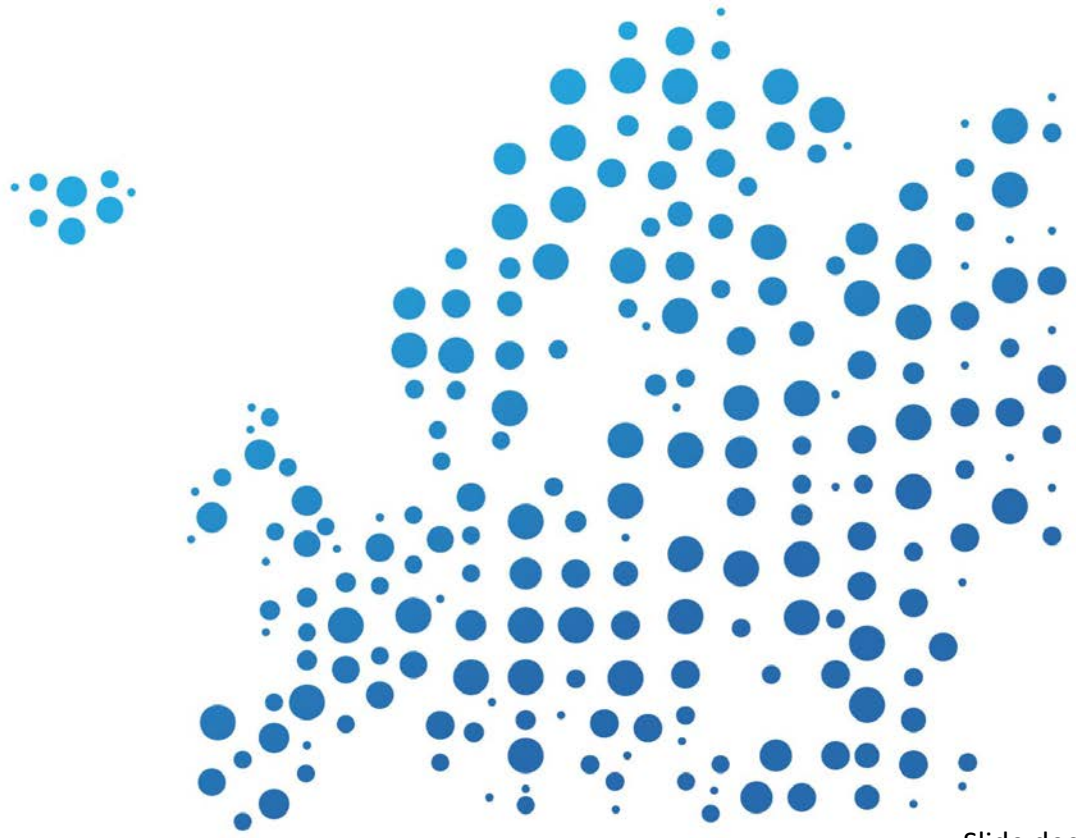
WHY SHOULD WE CARE ABOUT FAIR?

The 15th Munin Conference on Scholarly Publishing
19TH NOVEMBER 2020 | #MUNIN2020



Dr. Natalie Harrower

Director, Digital Repository of Ireland
Royal Irish Academy

@natalieharrower | @dri_ireland
www.allea.org
@ALLEA_academies

My Background

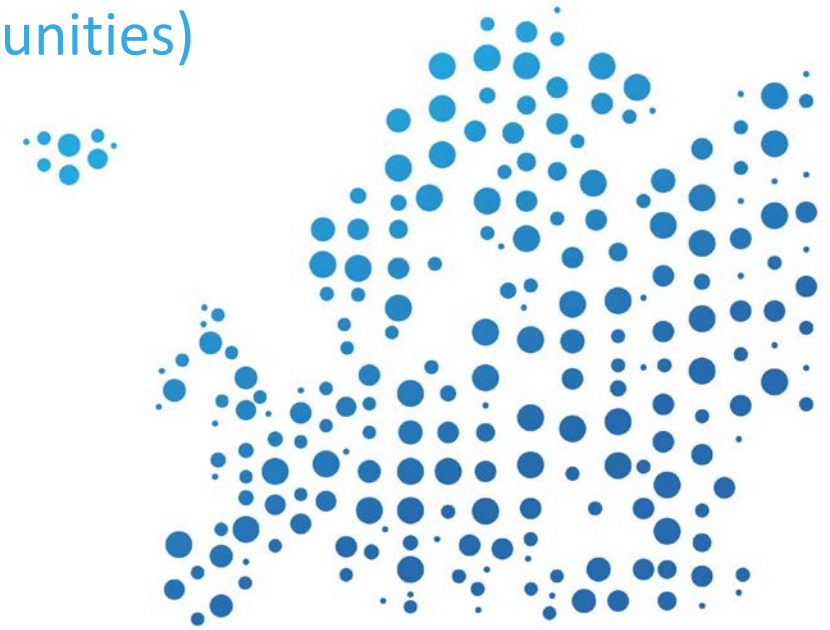
Chair, ALLEA e-Humanities working group (2015-2020) (Academy research)

EC FAIR Expert Group *Turning FAIR into reality* (Defining the plan)

European Open Science Cloud FAIR working group (Communities)

Director, Digital Repository of Ireland (HSS infrastructure)

Humanities Scholar (Research background)



@natalieharrower

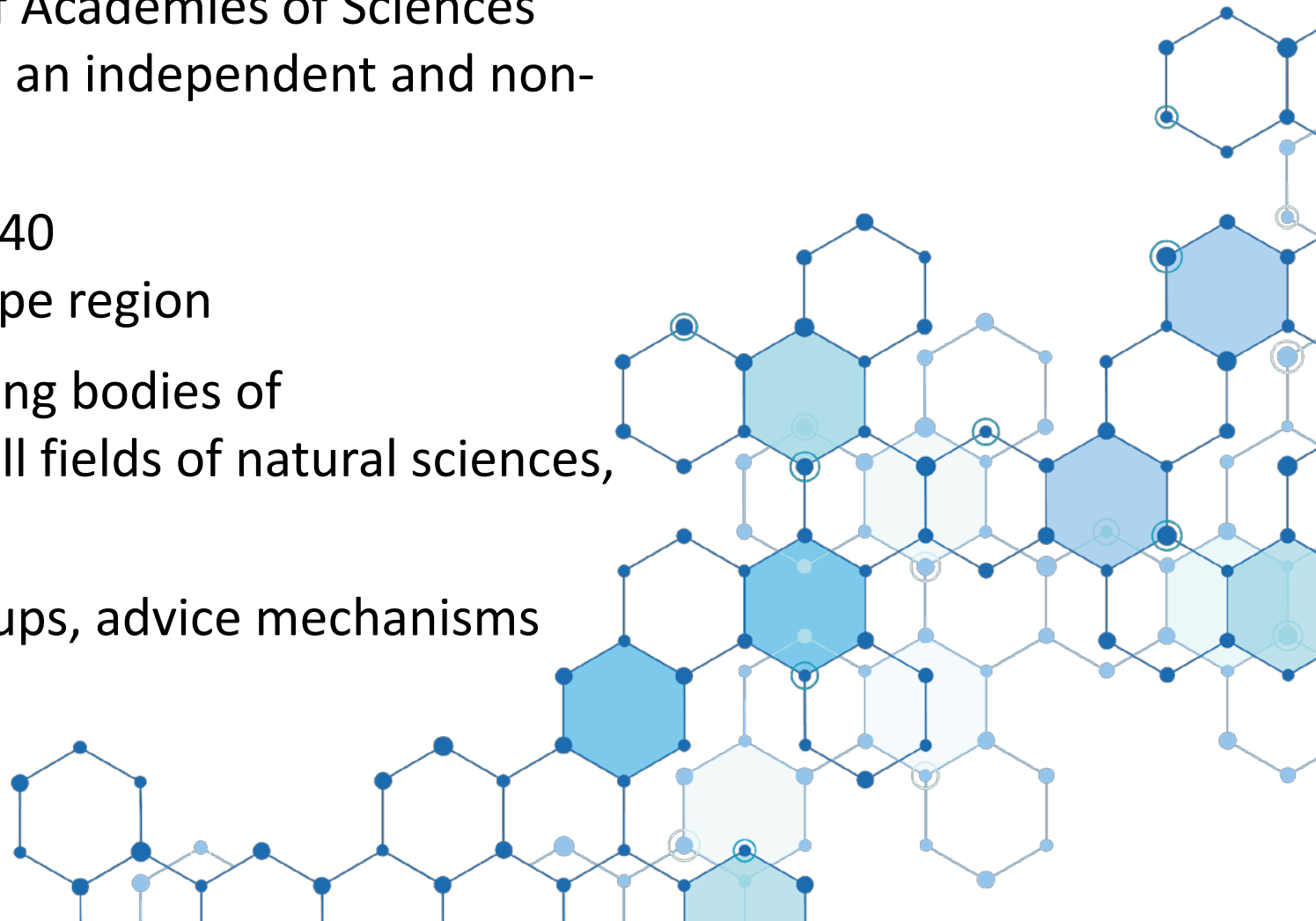
ALLEA and its members

ALLEA is the European Federation of Academies of Sciences and Humanities, founded in 1994 as an independent and non-profit scientific association

More than 50 academies from over 40 countries across the Council of Europe region

Member academies are self-governing bodies of distinguished scholars drawn from all fields of natural sciences, social sciences and the humanities.

ALLEA has a number of working groups, advice mechanisms



Aims of the E-Humanities Working Group

- Raising awareness of priorities and concerns in the digital humanities
- Digital best practices for the Humanities
- Open Research and FAIR data in the Humanities
- International scientific collaboration
- https://allea.org/e-humanities/

DOI: https://doi.org/10.7486/DRI.tq582c863

Why create this report? And how?

- To address the principles of Open Research/Open Science in relationship to **data management** from a **Humanities** perspective
- To identify **challenges**, and provide **recommendations** for humanities researchers on steps to take in making data **FAIR**
- To build these recommendations **collectively**
- **Open consultation** process included DARIAH workshop, received 200+ comments and several long submissions

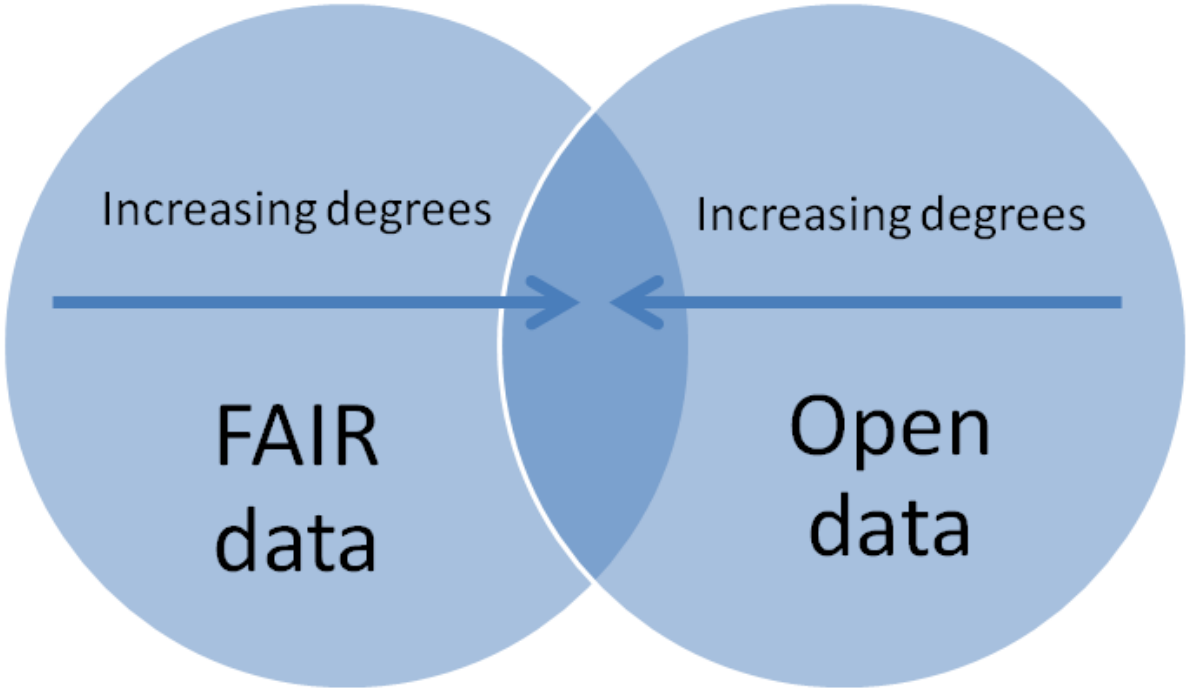
What is FAIR?

Findable **A**ccessible **I**nteroperable **R**eusable

- Open Access to Publications: longstanding element in Open Scholarship
- Argument: Data and other components of the research process should also be made available.
- Promotes better research, openness, transparency, efficiency, impact, integrity, exposure, impact, and return on investment
- *The FAIR Guiding Principles for scientific data management and stewardship*: https://doi.org/10.1038/sdata.2016.18

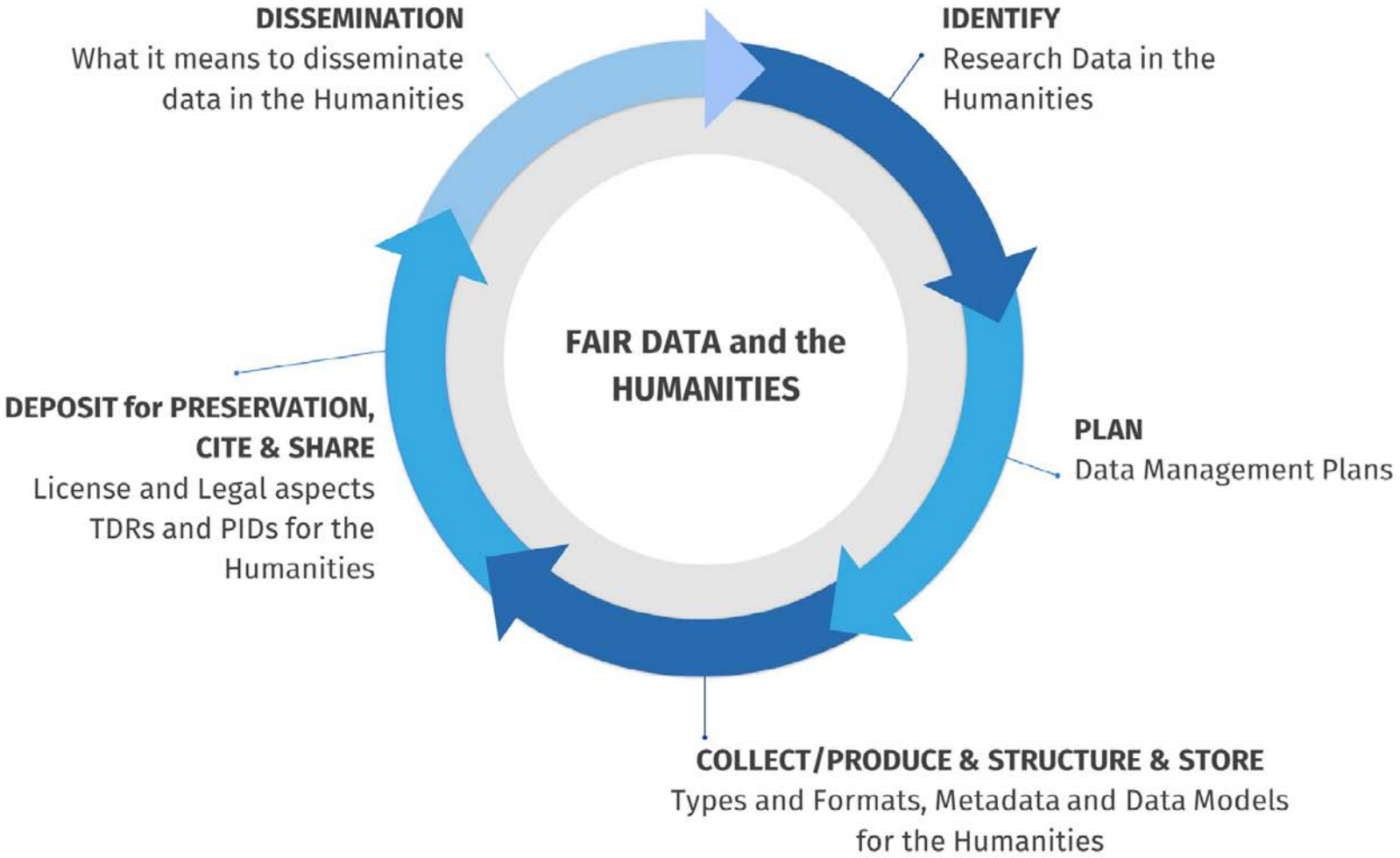
FAIR vs Open

- Data can be FAIR or Open, both or neither
- The greatest potential reuse comes when data are both FAIR and Open
- Even internal or restricted data will benefit from being FAIR, and there are legitimate reasons for restriction which vary by discipline



Report Structure

- Sections ordered around the research data lifecycle
- Introduction of the data sharing issue & challenges to humanities data
- Recommendations for researchers
- Suggestions for further reading



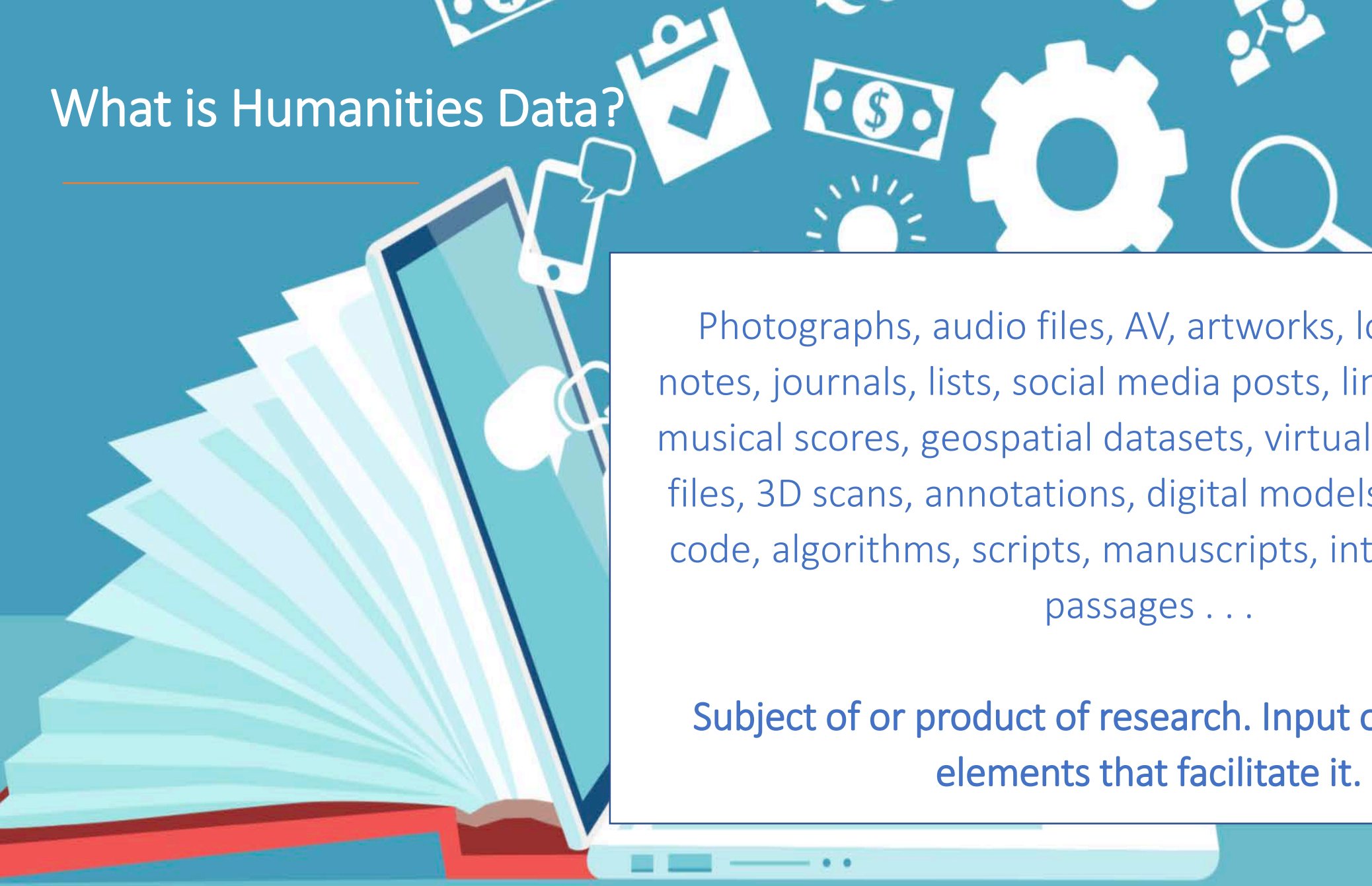
Sections

IDENTIFY: Research Data in the Humanities
PLAN: Data Management Plans (DMPs)
COLLECT, PRODUCE, STRUCTURE and STORE:
- Types and Formats
- Metadata
- Data Models

DEPOSIT, PRESERVE and SHARE:
- Legal Aspects
- Licences
- Trustworthy Digital Repositories and Persistent Identifiers

DISSEMINATE
LEGACY DATA
CONCLUSIONS

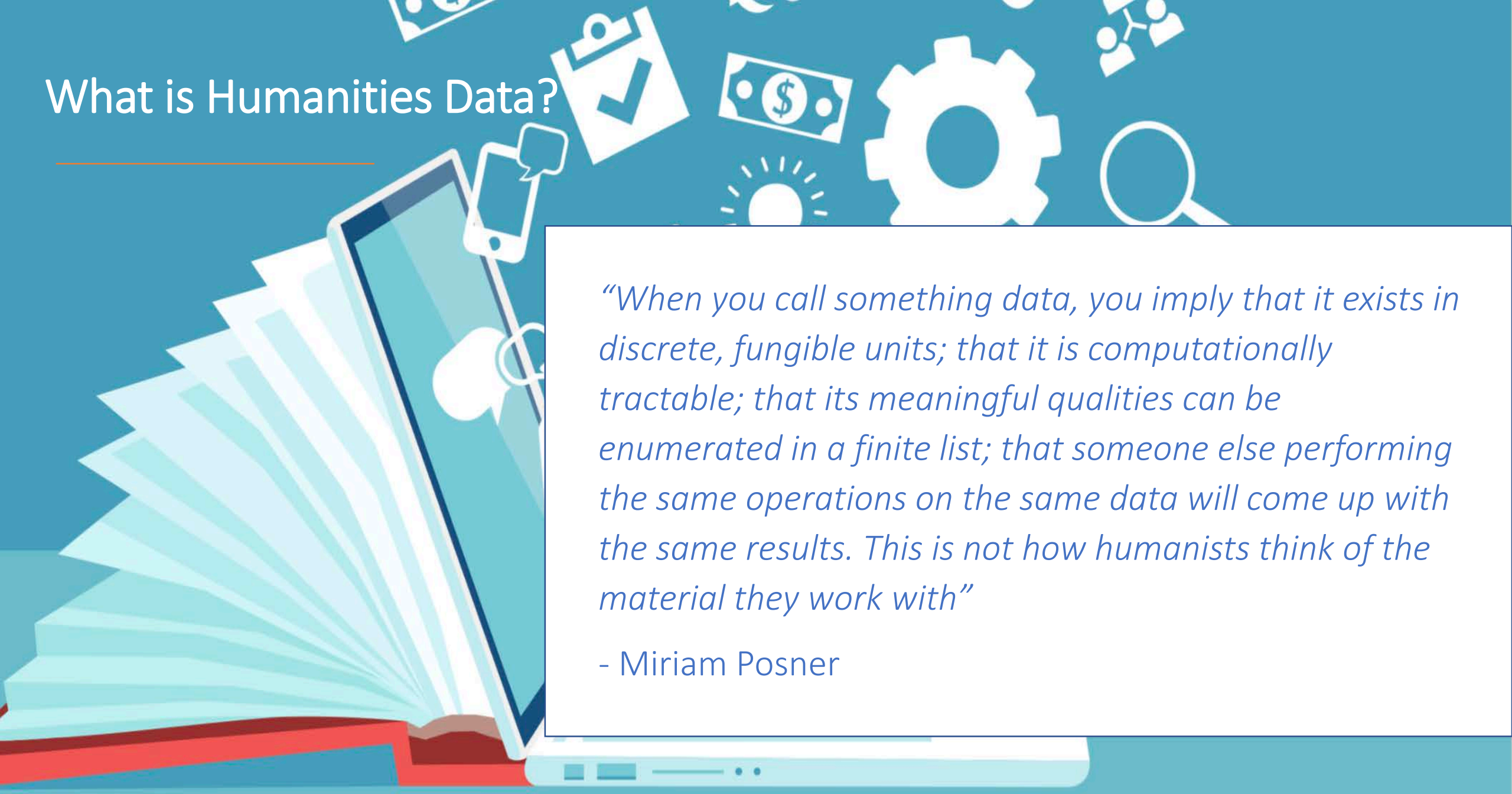
What is Humanities Data?



Photographs, audio files, AV, artworks, logbooks, field notes, journals, lists, social media posts, linguistic corpora, musical scores, geospatial datasets, virtual reality and CAD files, 3D scans, annotations, digital models, oral histories, code, algorithms, scripts, manuscripts, interviews, poetic passages . . .

Subject of or product of research. Input or output. And elements that facilitate it.

What is Humanities Data?



"When you call something data, you imply that it exists in discrete, fungible units; that it is computationally tractable; that its meaningful qualities can be enumerated in a finite list; that someone else performing the same operations on the same data will come up with the same results. This is not how humanists think of the material they work with"

- Miriam Posner

What is Humanities Data?

In the humanities, we "resist the blanket term 'data' for the very good reason that we have more and precise terminology (e.g. primary sources, secondary sources, theoretical documents, bibliographies, critical editions, annotations, notes, etc.) available to us to describe and make transparent our research processes"

- Edmond & Tóth-Czifra

What Challenges do we face? Data Silos

Silos courtesy
Ingrid Dillo, DANS

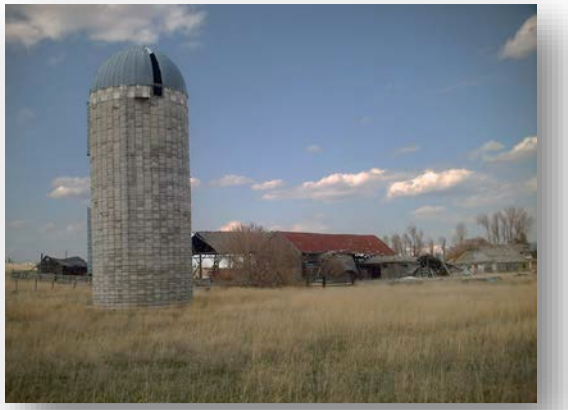
Data Archiving and Networked Services
DANS

Historical databases

Linguistic corpora

Archaeological GIS

Art image collections



Textual corpora

- Held on laptops, university servers, USB drives
- Proprietary formats
- Lacking context, metadata, licences

How do we approach these challenges?

- Understanding what data you use and produce through research
- Proper attention to research data throughout the research lifecycle
- Planning for data management, curation, and stewardship early
- Curation as an active and ongoing process
- Curation and stewardship are both a research area in their own right, and an element of research practice – recognition & support required
- Active steps to make data intelligible to others through FAIR principles

Select Recommendations 1

IDENTIFY DATA & PLAN

All research projects deal with data, with common humanities data including corpora, images, interviews, digitised works of art. Devise a **Data Management Plan** (DMP) prior to collecting data. Choose a **template** tailored to the humanities.

METADATA

Early in the research process, aim to identify a metadata standard that is suitable for your **discipline** or domain, and one that is compatible with the **repository** in which you will deposit your dataset(s)

DATA TYPES AND FORMATS

Choose **open formats** or widely adopted formats (“file types”) where possible, and look for those widely used in your **discipline**. Consider human and machine readable outputs for sharing data (e.g. PDF vs XML)

Select Recommendations 2

LEGAL ASPECTS

Ask for **consent** to share **anonymised** data and establish transparent and well-documented anonymisation routines that consider not just direct identifiers, but also how a combination of indirect identifiers could reveal identities.

LICENCES

Identify who owns the data; You may only attribute a licence to a work of which you are the copyright holder.

Use free and standardised licences such as Creative commons (CC)

Avoid applying more restrictive licences like NC (non-commercial) or ND (no derivatives) just to be ‘on the safe side’.

TDRs and PIDs

Ideally, choose a disciplinary-specific trustworthy data repository (TDR) that has been certified, and supports the creation of persistent identifiers (PIDs) for your data

Responses to the open consultation

- 2-month period, 200+ comments received as well as several longer submissions
- Most feedback was very detailed and specific. Not ‘principled’
- Provides insight into key concerns – what needs to be addressed
  - e.g. Desire to see more exhaustive lists of research outputs to cover diversity of humanities, and full attention to software used
  - e.g. Attention to issues of consent and data anonymisation of direct and indirect identifiers
  - e.g. Awareness of physical vs digital data
  - e.g. Debates about human readable vs machine readable outputs

Dissemination, Collaboration, Partnerships … help!

- Consult registries such as re3data.org and fairsharing.org to find suitable repositories
- Set up an ORCID ID, which creates a linked profile
- CV / profile sites such as Academia.edu or Research Gate are useful for dissemination, but they are not suitable as failsafe archives.
- Galleries, Libraries, Archives, Museums now starting to work towards FAIR digital collections
- Aggregator services provide greater discoverability, e.g. Europeana
- DARIAH - European research infrastructure

Conclusions

- The FAIR principles are gaining momentum in the research sector, and funders are increasingly mandating the timely sharing of FAIR data and other outputs (eg software). They are also being built into the requirements for the European Open Science Cloud (EOSC)
- Pathways for implementing and assessing FAIRness are still being developed, and they require community tailoring
- Success of FAIR is dependent on commensurate changes in research culture: training and access to data stewards/support, changes in rewards and incentives

@natalieharrower

Thanks!

Contact

n.harrower@ria.ie
@natalieharrower
@dri_ireland

www.allea.org
@ALLEA_academies

Links

ALLEA E-Humanities working group:
https://allea.org/e-humanities/

Sustainable and FAIR Data Sharing in the Humanities:
https://doi.org/10.7486/DRI.tq582c863