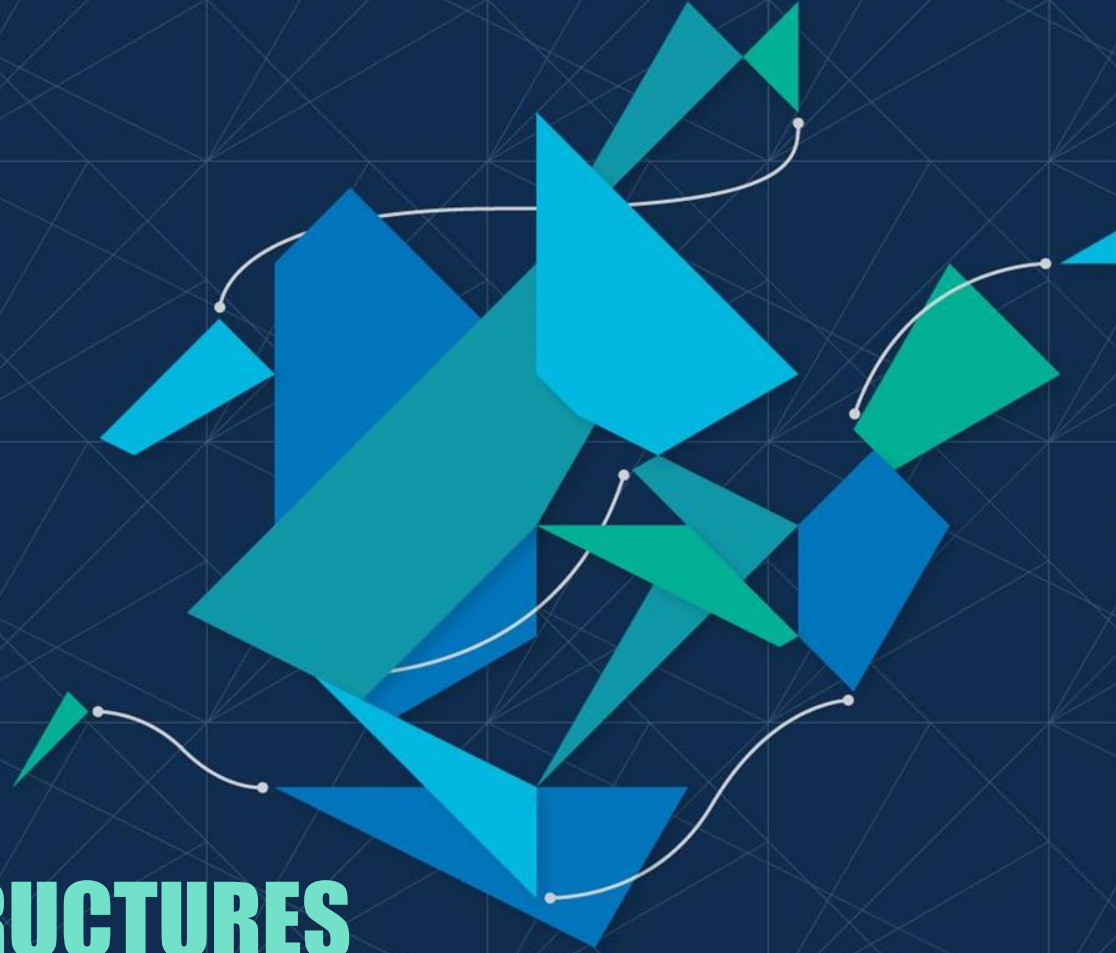


Digital researchers and data experts

We create digital tools  
to explore academic  
research in new ways.



# RESEARCH INFRASTRUCTURES

Digital researchers and data experts

We create digital tools  
to explore academic  
research in new ways.

**Dr. James Smithies and Dr. Arianna Ciula**

Director / Deputy Director of King's Digital Lab

@jamesmithies @ariciula

kdl-info@kcl.ac.uk



## Multiple components of open infrastructures

Open standards (e.g. potential technical interoperability)

Open data (governance and access policies)

Open source (code, software etc.)

Open licences (e.g. Creative Commons - see more at <http://opendefinition.org/>)

Open access (to scholarly research)

Affordances of open approaches to technology for the Humanities

Assessment of interlocking aspects of open technical ecosystems (e.g. ethical issues)

See Smithies (2017), *The Digital Humanities and the Digital Modern*. Chapter 5: Towards a Systems Analysis of the Humanities (pp. 113-151).

## Developed by FORCE 11

### **Findable**

e.g. metrics: unique and persistent identifiers; rich metadata

### **Accessible**

e.g. metadata stay when data deleted

### **Interoperable**

e.g. FAIR metadata

### **Re-usable**

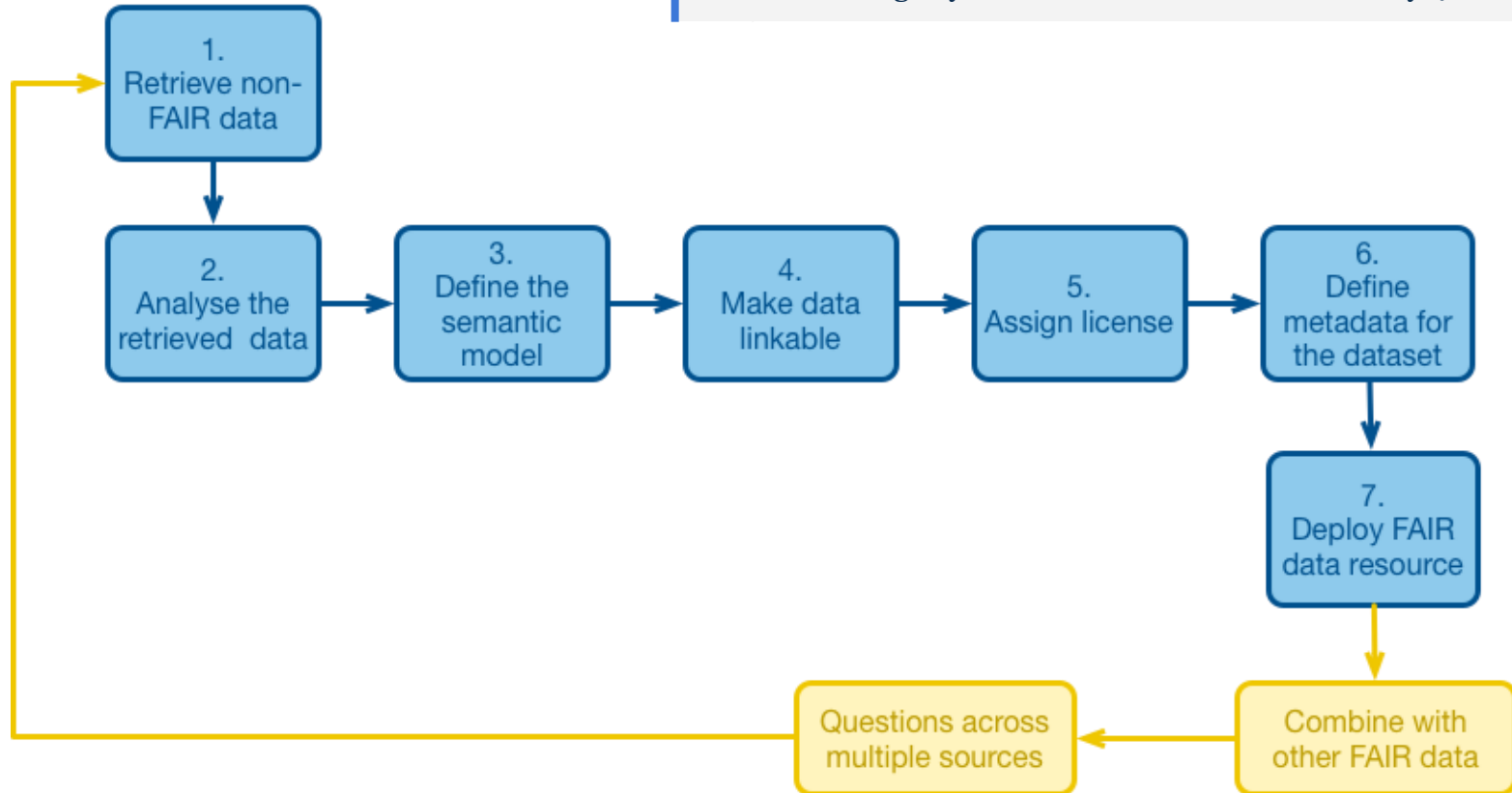
e.g. clear data use license

Metrics changing and evolving

Wilkinson et al. (2016), [The FAIR Guiding Principles](#); van Erp (2018), [Testing the FAIR metrics on data catalogs](#); Parthenos Project (undated), [The FAIR Principles](#).

## FAIRification Process

Go Fair (undated), [Fairification Process](#)  
(The image by [GO FAIR](#) is licenced under CC by 4.0)



## Good Practice

Across **research lifecycle** >> planning, collecting, storing, using, sharing and preserving research data

Across **types of data** >> qualitative, quantitative, factual/non-factual, numerical, textual, audio-visual ...

Costs and time

File formats



Time

Metadata (technical, administrative, descriptive, use, preservation) >> systematic, controlled values, consistency, standards

EPFL, [Services for Researchers](#)  
(Scroll down to “Research Data Management”).

Image from [EPFL Research Data Management FAST GUIDES, Fast Guide #1 Research Data: the basics](#)

## At KDL

RDM integral to KDL SDLC >> pre-project, foundations, evolutionary development, release and post-project

Scope and size of project >> methods

e.g. **Pre-project** (analysis inclusive of design and implementation)

1. First meeting
2. Feasibility assessment and review
  - a. Follow up meetings/remote comms
  - b. Sample data
  - c. High level MoSCoW
3. Product Quote including outline of *SLA* and *Archiving & Sustainability* options
4. Data Management Plan



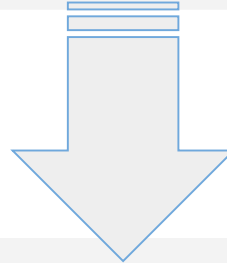
e.g. criteria

- Size, type, heterogeneity, structure and fuzziness, provenance, copyright and licensing of data
- Compatibility with software stack and workflows

## At KDL

But...

Research data emerge as project develops; very Agile cases >>  
RDM practices embedded in **Evolutionary Development**



e.g. reflect on iterations and data workflows

- What kind of data (type, format and size) are you expecting to create and/or process in your project or research?
- Where are the datasets hosted if they already exist or where are you expecting to host them, and for how long?
- Requirements with respect to data import and export functionalities? data processing, versioning, citability, compliance to domain-specific standards, data anonymization and sensitivity, re-use and licensing?



## Data Management Plans

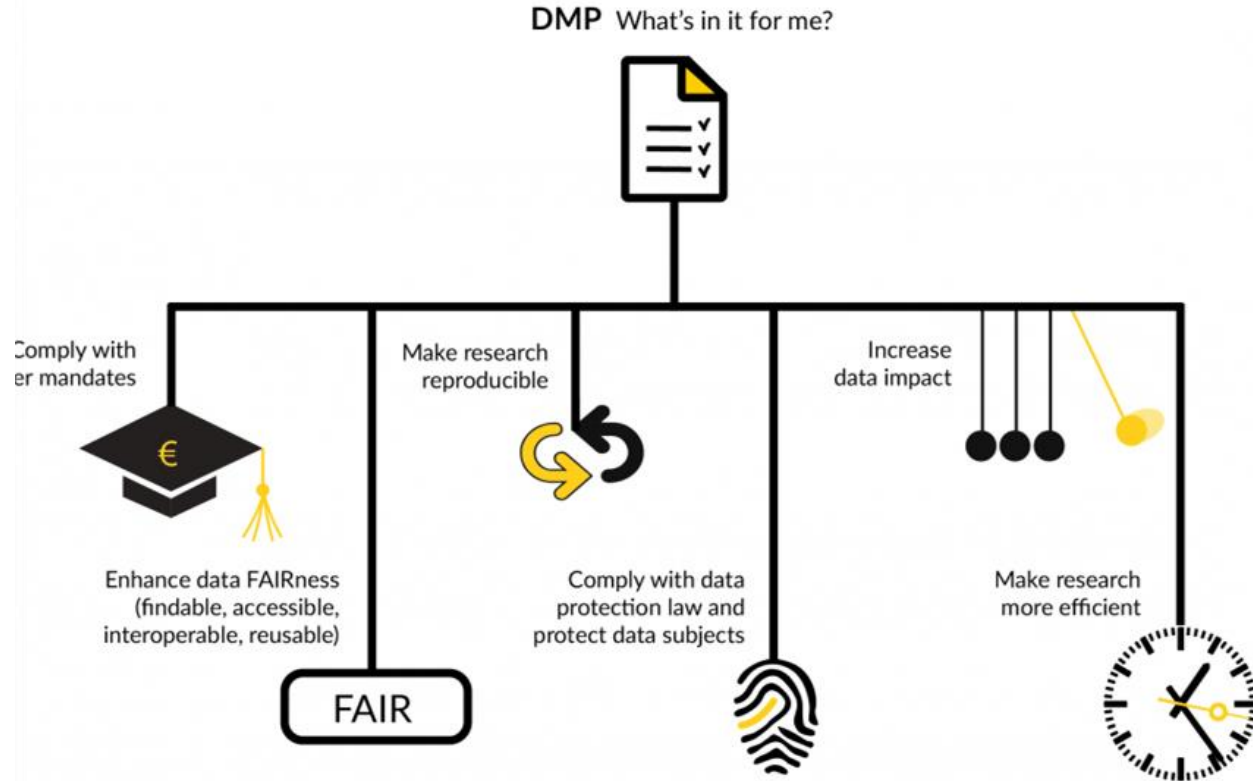


Image from Utrecht University's guide to [Data Management Planning](#).

## Data Management Plans

1. Briefly introduce the types of data the research will create. Why did you decide to use these data types?
2. Give details on the proposed methodologies that will be used to create the data. Advise how the project team selected will be suitable for the data/digital aspects of the work, including details of how the institution's data support teams may need to support the project
3. How will the data be stored in the short term?
  - a. What backup will you have in the in-project period to ensure no data is lost?
4. How the data will be stored in the long term
  - a. Where have you decided to store it, why is this appropriate?
  - b. How long will it be stored for and why?
  - c. Costs of storage – why are these appropriate? Costs related to long term storage will be permitted providing these are fully justified and relate to the project Full justification must be provided in Justification of Resources (JoR)
5. How the data will be shared and the value it will have to others
  - a. How the data will enhance the area and how it could be used in the future?
  - b. Releasing the data – advise when you will be releasing and justify if not releasing in line with AHRC guidelines of a minimum of three years. If the data will have value to different audiences, how these groups will be informed?
  - c. Will the data need to be updated? Include future plans for updating if this is the case.
  - d. Will the data be open or will you charge for it? Justify if charging to access the data
  - e. Financial requirements of sharing – include full justification in the JoR
6. Ethical and Legal considerations
  - a. Any legal and ethical considerations of collecting the data
  - b. Legal and ethical considerations around releasing and storing the data – anonymity

From AHRC UKRI Guides: [Data Management Plan](#)

## Overview of General Data Protection Regulation (GDPR)

EU Regulation (2016/679 of 27 April 2016) on the protection of natural persons with regard to the processing of personal data and the free movement of such data

On 25 May 2018 it became directly applicable in all the EU Member States and replaced the Personal Data Directive of 1995 >> supersedes national rules

Personal Data → “any information relating to an identified or identifiable natural person (‘data subject’)”

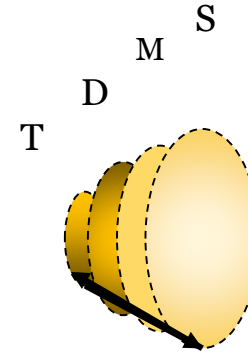
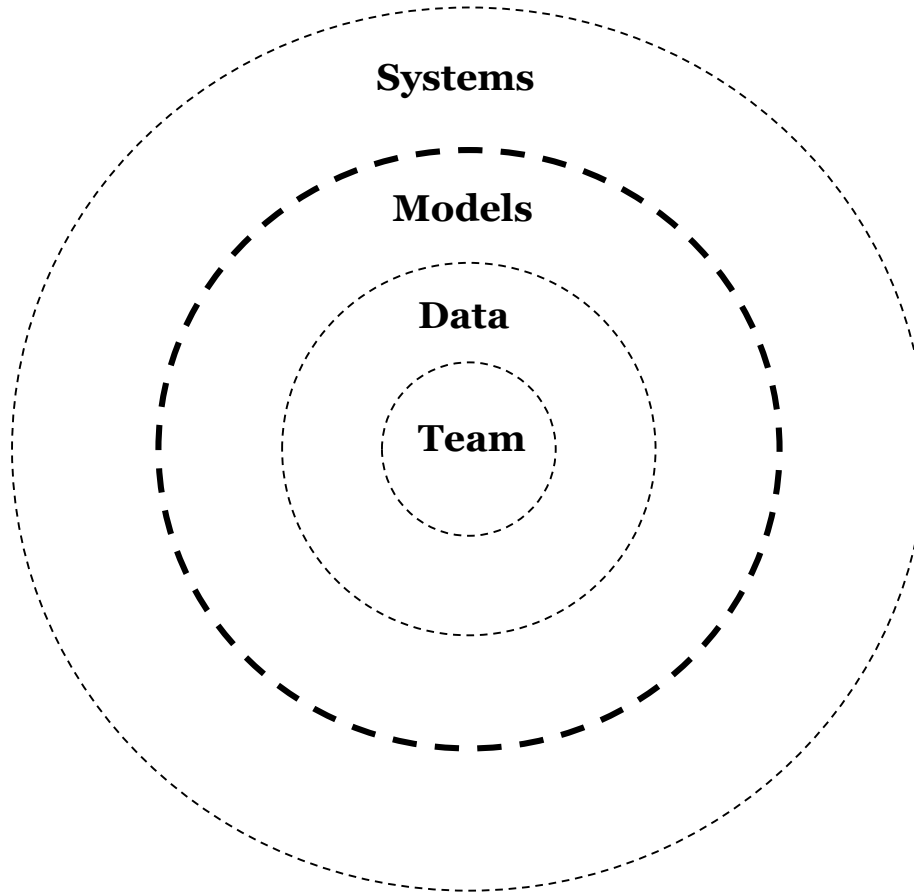
- Informed consent required to process personal data (stricter protection for sensitive data)
- Rights of data subjects e.g. to information, access, rectification, erasure, restriction
- Obligations of data controllers e.g. implementing “privacy by design and by default”, keeping a record of processing activities or carrying out a data protection impact assessment
- Exceptions for research or archiving in the public interest >> leeway for Member States

## GDPR in a Research & Archiving Context

Data Controllers and Processors >> Code of Conduct

Legal but also **Technical** and **Ethical** perspectives at stake

European Parliament (2016), [General Data Protection Regulation](#).



Co-constitution  
(bidirectional  
influence &  
dependencies)

Ciula and Smithies (forthcoming),  
Sustainability and modelling at  
King's Digital Lab.

## Action, Creativity, Dependencies, Short lifetime

**Versioning** e.g. use of Git within RSE team

**Sharing** e.g. GitHub

**Describing** e.g. README documentation

**Licensing** e.g. Open licence such as MIT

**Publishing** e.g. Zenodo

See:

Katz (2016), [Software vs Data](#)

Katz (2017), D.S. [Software Citation Implementation Working Group](#)

Smith et al. (2016), [Software Citation Principles](#)

## Citable as any other Research Product?

Industry standards and RSE best practices

Credits >> e.g. CASRAI [CRediT](#)

Format

Including authors, version and repository

Digital researchers and data experts

We create digital tools  
to explore academic  
research in new ways.

**Dr. James Smithies and Dr. Arianna Ciula**

Director / Deputy Director of King's Digital Lab

@jamesmithies @ariciula

kdl-info@kcl.ac.uk

