

# Digital Management of Science Data (in the Humanities)

The Digitalisation of Research | Academic Alumni Forum

University Leipzig  
28<sup>th</sup> of November 2019

Felix Rau and Patrick Helling

1



# Agenda

- What is this Research Data everybody is talking about?
- And what is Research Data Management all about?
- Open Science – Open Access – Open Data
- The Data Management Plan
- Metadata
- Repositories



# Mad Minute

## Who are you?

Please give us a short briefing about

- You
- Your background
- Your research
- Your hopes, needs, fears and dreams  
(in relation to this workshop)





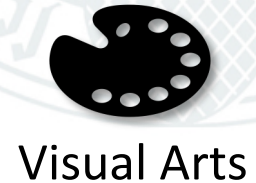
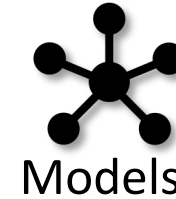
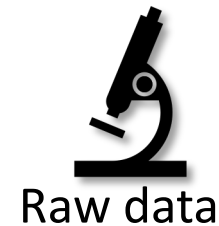
# Research data



[https://www.ulb.tu-darmstadt.de/media/ulb/pflicht\\_epublishing/bilder\\_10/Titelbild\\_Forschungsdatenmanagement\\_529x0.png](https://www.ulb.tu-darmstadt.de/media/ulb/pflicht_epublishing/bilder_10/Titelbild_Forschungsdatenmanagement_529x0.png)

# Digital research data

What is digital research data?



# Research data!?

“Research data is any information that has been collected, observed, generated or created to validate original research findings. Although usually digital, research data also includes non-digital formats such as laboratory notebooks and diaries.”

University of Leeds

“Research data, unlike other types of information, is collected, observed, or created, for purposes of analysis to produce original research results.”

University of Edinburgh

“Research data is defined as recorded factual material commonly retained by and accepted in the scientific community as necessary to validate research findings; although the majority of such data is created in digital format, all research data is included irrespective of the format in which it is created.”

Engineering and Physical Sciences Research Council (EPSRC)

# Research data!?

“Research data is any information that has been **collected, observed, generated or created** to validate original research findings. Although usually digital, research data also includes non-digital formats such as laboratory notebooks and diaries.”

University of Leeds

“Research data, unlike other types of information, is **collected, observed, or created, for** purposes of analysis to produce original research results.”

University of Edinburgh

“Research data is defined as **recorded factual material** commonly retained by and accepted in the scientific community as necessary to validate research findings; although the majority of such data is created in digital format, all research data is included irrespective of the format in which it is created.”

Engineering and Physical Sciences Research Council (EPSRC)

# Research data!?

“Research data is any information that has been **collected, observed, generated or created to validate original research findings**. Although usually digital, research data also includes non-digital formats such as laboratory notebooks and diaries.”

University of Leeds

“Research data, unlike other types of information, is **collected, observed, or created, for purposes of analysis to produce original research results**.”

University of Edinburgh

“Research data is defined as **recorded factual material** commonly retained by and accepted in the scientific community as necessary **to validate research findings**; although the majority of such data is created in digital format, all research data is included irrespective of the format in which it is created.”

Engineering and Physical Sciences Research Council (EPSRC)





# Research data types!?

- documents, spreadsheets
- laboratory notebooks, field notebooks, diaries
- questionnaires, transcripts, codebooks
- audiotapes, videotapes
- photographs, films
- test responses
- slides, artefacts, specimens, samples
- collections of digital outputs
- data files
- database contents (video, audio, text, images)
- models, algorithms, scripts
- contents of an application (input, output, logfiles for analysis software, simulation software, schemas)
- methodologies and workflows
- standard operating procedures and protocols
- ...

# Research data management Why?



10

RDM in the Humanities – Academic Alumni Forum



# A typical monday at a university...

I am reading an article in a scientific journal. The topic is about my research focus and I would like to have the data that is related to the article, but I can't find it online.

→ I contact the main author (1)

# A typical monday at a university...

The data is stored on a USB stick that is in the possession of the main author. She just moved to a new house and is still living out of boxes. The USB stick is in one of the boxes. Nobody knows in which one...

→ I need to be patient

# A typical monday at a university...

7 month later...

I finally received the USB stick and can access the data. However, the data is stored in a specific data format that can't be read by a common program.

→ I need to ask the main author, if she has the data in a different format (2)

# A typical monday at a university...

Unfortunately the data is only available in this specific format. In addition, the USB stick is the only copy of the data. The main author has used a special commercial program to create, analyze and use the data. The company went bankrupt four years ago. The software is no longer available.

Fortunately I know an old and wise professor at my institute who still has a copy of the special software on his old computer that still works. Finally I can decrypt the data, but I don't understand the logic behind the data.

→ I need to talk to the main author again to find out what the individual acronyms and descriptions stand for, which categorize the data. It's the (3)rd time... slowly but surely she is getting annoyed.

# A typical monday at a university...

The main author can help me with some of my questions. However, the research is older than I am. She can't remember most of the description patterns anymore. In addition, there is no documentation of the data and the information in the article I read is not sufficient.

I'm lucky though. The co-author knows the logic behind the description patterns. He was a PhD student and responsible for the data in the project. Unfortunately two years ago he moved back to China.

→ I would like to get in touch with the co-author in order to finally understand the data.

# A typical monday at a university...

The main author is no longer in contact with the co-author. She only knows that the co-author quit his job at the university, moved back to China and has been working as a farmer ever since. His name is Sam-Lee...

→ ...?



# A typical monday at a university...

I am reading an article in a scientific journal. The topic is about my research focus and I would like to have the data that is related to the article, but I can't find it online. I contact the main author. The data is stored on a USB stick that is in the possession of the main author. She just moved to a new house and is still living out of boxes. The USB stick is in one of the boxes. Nobody knows in which one. I need to be patient. 7 month later... I finally received the USB stick and can access the data. However, the data is stored in a specific data format that can't be read by a common program. I need to ask the main author, if she has the data in a different format (2) Unfortunately the data is only available in this specific format. In addition, the USB stick is the only copy of the data. The main author has used a special commercial program to create, analyze and use the data. The company went bankrupt four years ago. The software is no longer available. Fortunately I know an old and wise professor at my institute who still has a copy of the special software on his old computer that still works. Finally I can decrypt the data, but I don't understand the logic behind the data. I need to talk to the main author again to find out what the individual acronyms and descriptions stand for, which categorize the data. It's the (3)rd time... slowly but surely she is getting annoyed. The main author can help me with some of my questions. However, the research is older than I am. She can't remember most of the description patterns anymore. In addition, there is no documentation of the data and the information in the article I read is not sufficient. I'm lucky though. The co-author knows the logic behind the description patterns. He was a PhD student and responsible for the data in the project. Unfortunately two years ago he moved back to China. I would like to get in touch with the co-author in order to finally understand the data. The main author is no longer in contact with the co-author. She only knows that the co-author quit his job at the university, moved back to China and has been working as a farmer ever since. His name is Sam-Lee...?

# A typical monday at a university...

I am reading an article in a scientific journal. The topic is about my research focus and I would like to have the data that is related to the article, but **I can't find it** online. I contact the main author. The data is stored on a USB stick that is in the possession of the main author. She just moved to a new house and is still living out of boxes. The USB stick is in one of the boxes. Nobody knows in which one. I need to be patient. 7 month later... I finally received the USB stick and can access the data. However, the data is stored in a specific data format that **can't be read by a common program**. I need to ask the main author, if she has the data in a different format (2) Unfortunately the data is only available in this specific format. In addition, **the USB stick is the only copy of the data**. The main author has used a special commercial program to create, analyze and use the data. The company went bankrupt four years ago. The software is no longer available. Fortunately I know an old and wise professor at my institute who still has a copy of the special software on his old computer that still works. Finally I can decrypt the data, but **I don't understand the logic behind the data**. I need to talk to the main author again to find out what the individual acronyms and descriptions stand for, which categorize the data. It's the (3)rd time... slowly but surely she is getting annoyed. The main author can help me with some of my questions. However, the research is older than I am. She can't remember most of the description patterns anymore. In addition, there is no documentation on the data and the information in the article I read is not sufficient. I'm lucky though. The co-author knows the logic behind the description patterns. He was a PhD student and responsible for the data in the project. Unfortunately two years ago he moved back to China. I would like to get in touch with the co-author in order to finally understand the data. The main author is no longer in contact with the co-author. She only knows that the co-author quit his job at the university, moved back to China and has been working as a farmer ever since. His name is Sam-Lee...?

# Research data management – Why?

the short truth

- **I can't find the Data**
- **A USB stick is the only copy of the data**
- **The Data can't be read by a common program**
- **I don't understand the logic behind the data**

## The **FAIR**-Principles

**F**indable

**A**ccessible

**I**nteroperable

**R**eusable

Good scientific practice!

Preservation of cultural heritage!

# Research data management – Why? the long truth

- Documentation of your own work in an uniform, comprehensible way
  - It's a simple calculation → Research = Research Data + Publication
  - Improvement of your own visibility/reputation
- Comparability to other results of research
- No need to reinvent the wheel again and again and again and agai...
- Development of new research questions
- Potential of cooperation
- Survival of your own research



# Research data management How?

an abstract overview



# Research data need to be... **stored!**

- **Bitstream-Preservation**

- **Preserving a bitstream** in its original state means that an archive is able to maintain one version of the object
- Beware of data loss or corruption induced by **preservation** actions  
→ Redundant storage, Checksums, Recopying to new media

Archives, that **only** provide Bitstream-Preservation are called **dark archives**.



# Research data need to be... readable! (1)

- Widespread formats generated by widespread application programs are better suited for long-term archiving.
  - Image files → Tagged Image File Format (TIFF, TIF)
  - Audio files → Waveform Audio File Format (WAV)
  - Text → Plain Text Document (TXT)
  - Text → Portable Document Format/A (PDF/A)
  - Documents/Text → Extensible Markup Language (XML)

+ iterative control of formats and if applicable: migration of format!

# Research data need to be... readable! (2)

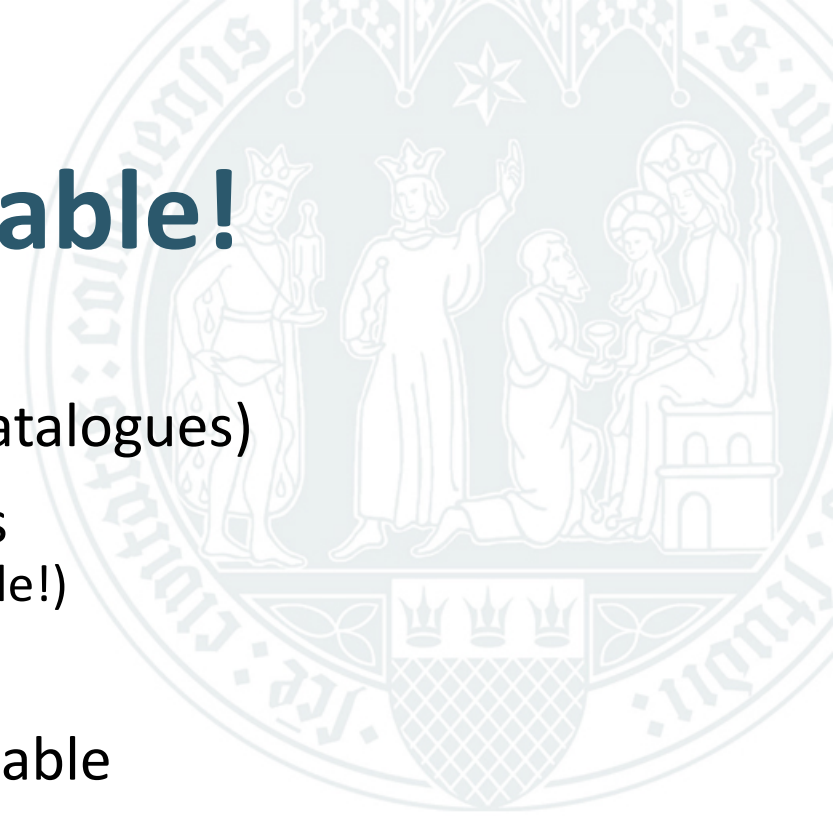
Comprehensive description of the genesis and properties of data

- Administrative Metadata
  - administrative information about data, e.g. how the data were created
- Descriptive Metadata
  - individual aspects or additional information about data for discovery and identification
- Structural Metadata
  - information about structures, methods, types etc.
- Bibliographic Metadata
  - information for representation of data in online-catalogues



# Research data need to be... **reusable!**

- For the ability of using data, it must be findable (online-catalogues)
- Easy and clear access to data/documentated rules of access (in contrast to a dark archive, which does not promise to be accessible!)
- Data must be quotable, uniquely identifiable and addressable  
→ Persistent Identifier



# Research data need to be... **analyzable!**

- Preservation of software

Research data is often generated with the help of tools and applications

It is not enough to name a used tool within a documentation. You need to (1) save the version of a tool you used and to (2) provide an environment in which this version can be used independently.

- Preservation of presentation systems

Research data is often exclusively accessible via a specific presentation layer

If software is the only access layer to research data, it is important to (1) store this software in a modular way and to (2) keep it permanently accessible/reusable.

And again...

## The **FAIR**-Principles

**F**indable

**A**ccessible

**I**nteroperable

**R**eusable



# The problems do exist

**1<sup>st</sup> Place (68%):** Further maintenance of the website(s) after the end of a research project is not guaranteed.

**2<sup>nd</sup> Place (60%):** obsolete data formats can no longer be read with current software.

**3<sup>rd</sup> Place (45%):** Data is difficult to find

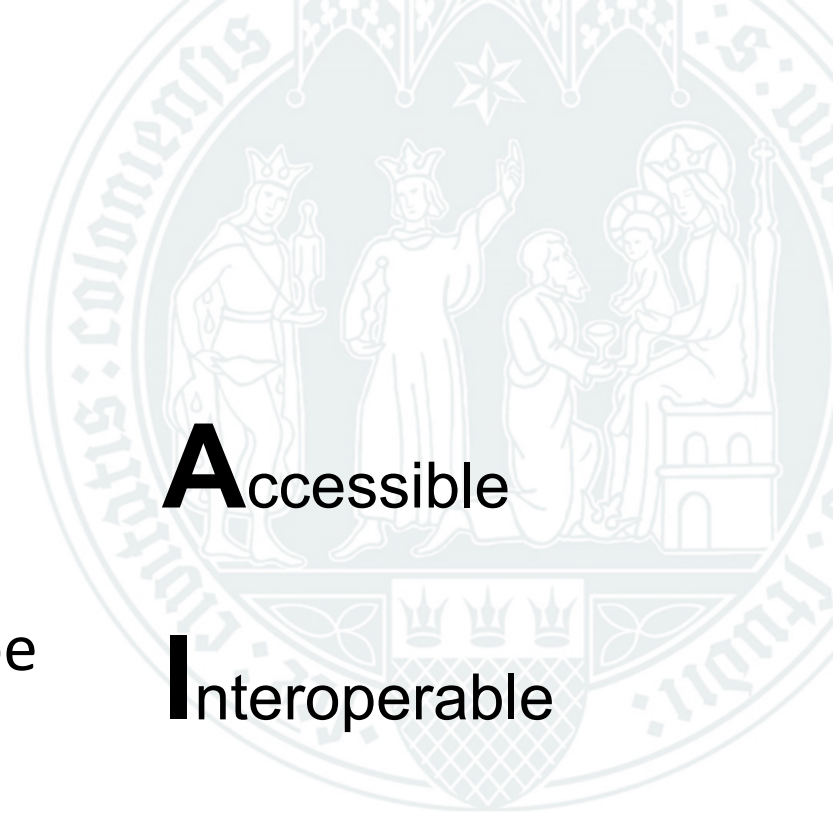
**4<sup>th</sup> Place (41%):** Data is documented insufficiently and therefore no longer reliably interpretable

**A**ccessible

**I**nteroperable

**F**indable

**R**eusable

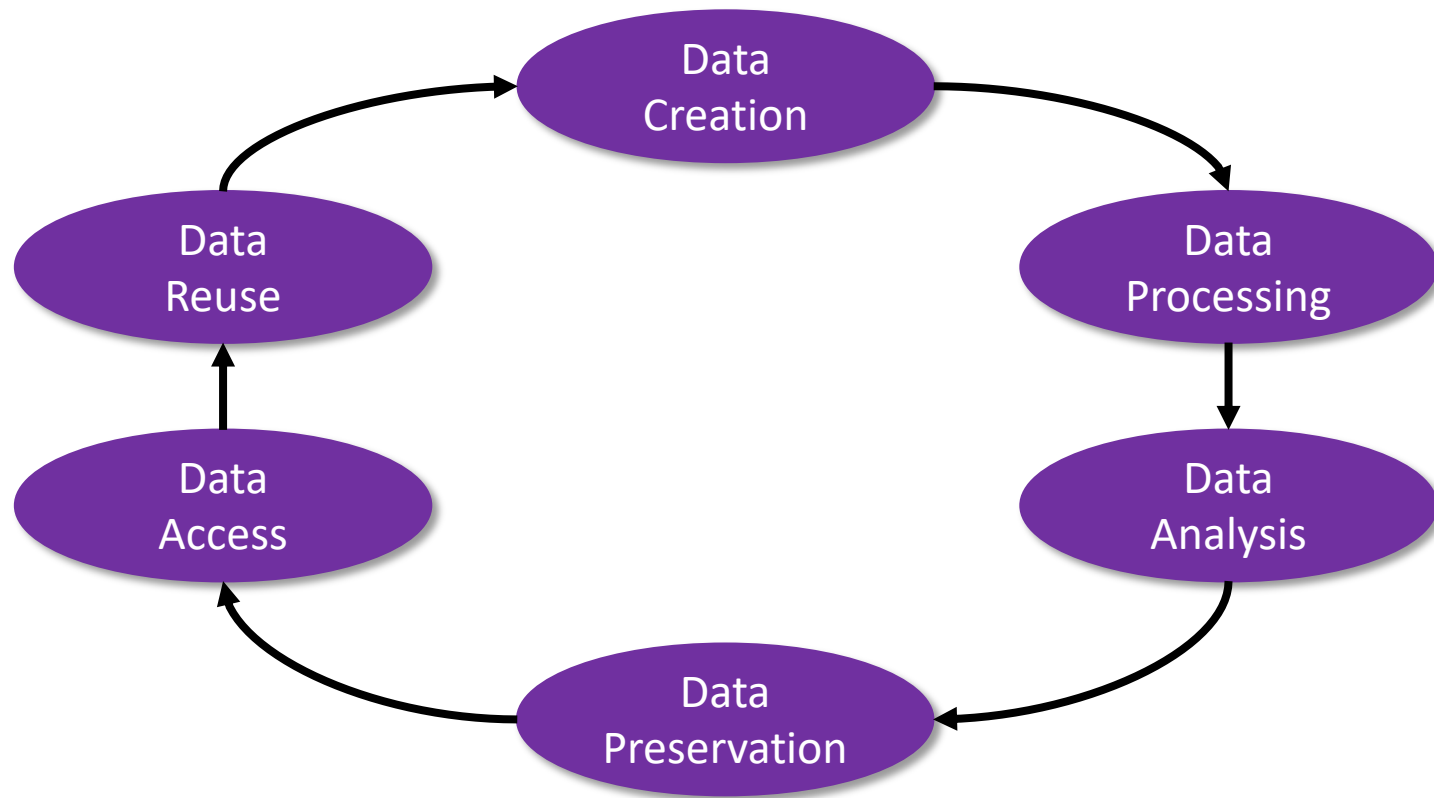


# Take home message

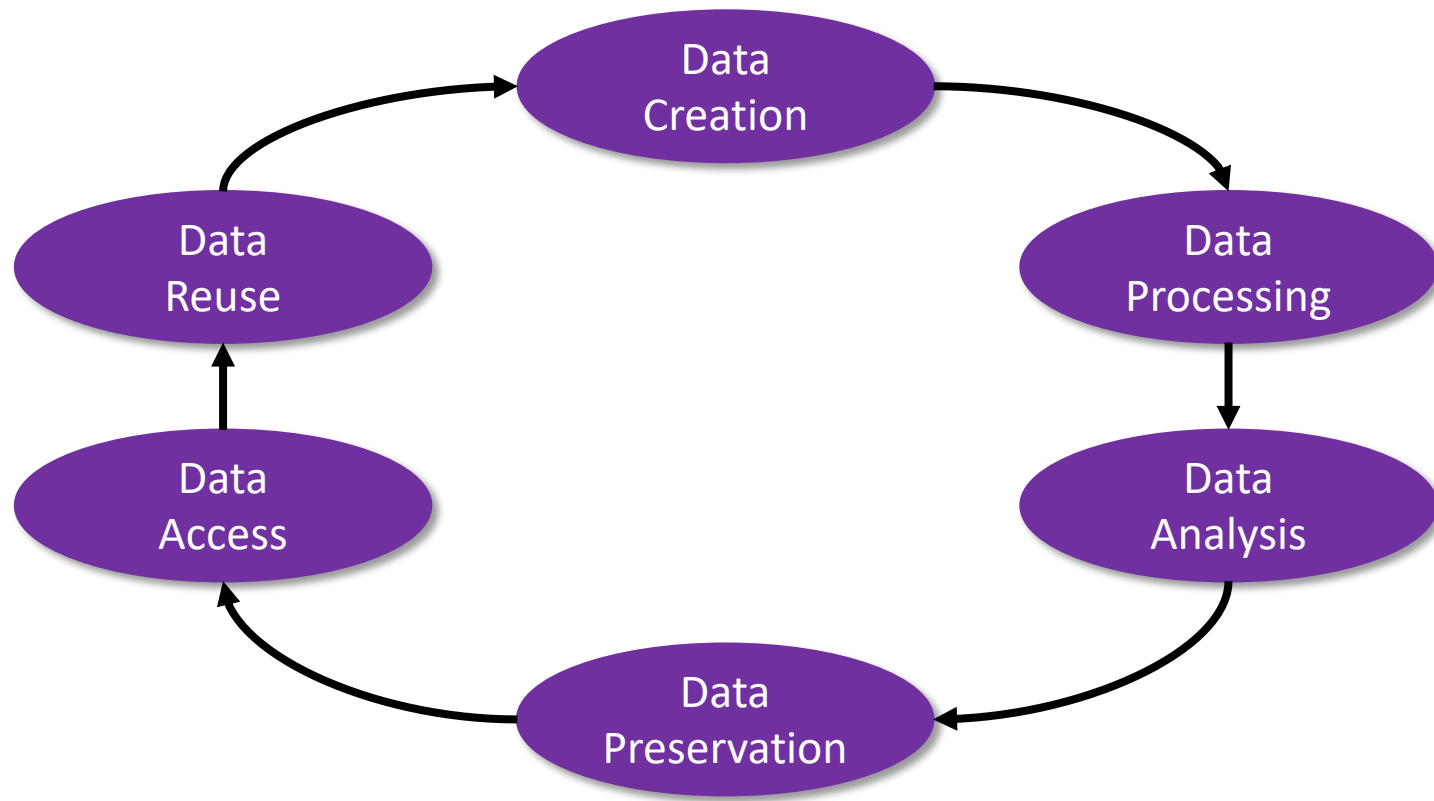
→ Make Data **F**indable, **A**ccessible, **I**nteroperable and **R**eusable!



# A wild ride through the research data life cycle



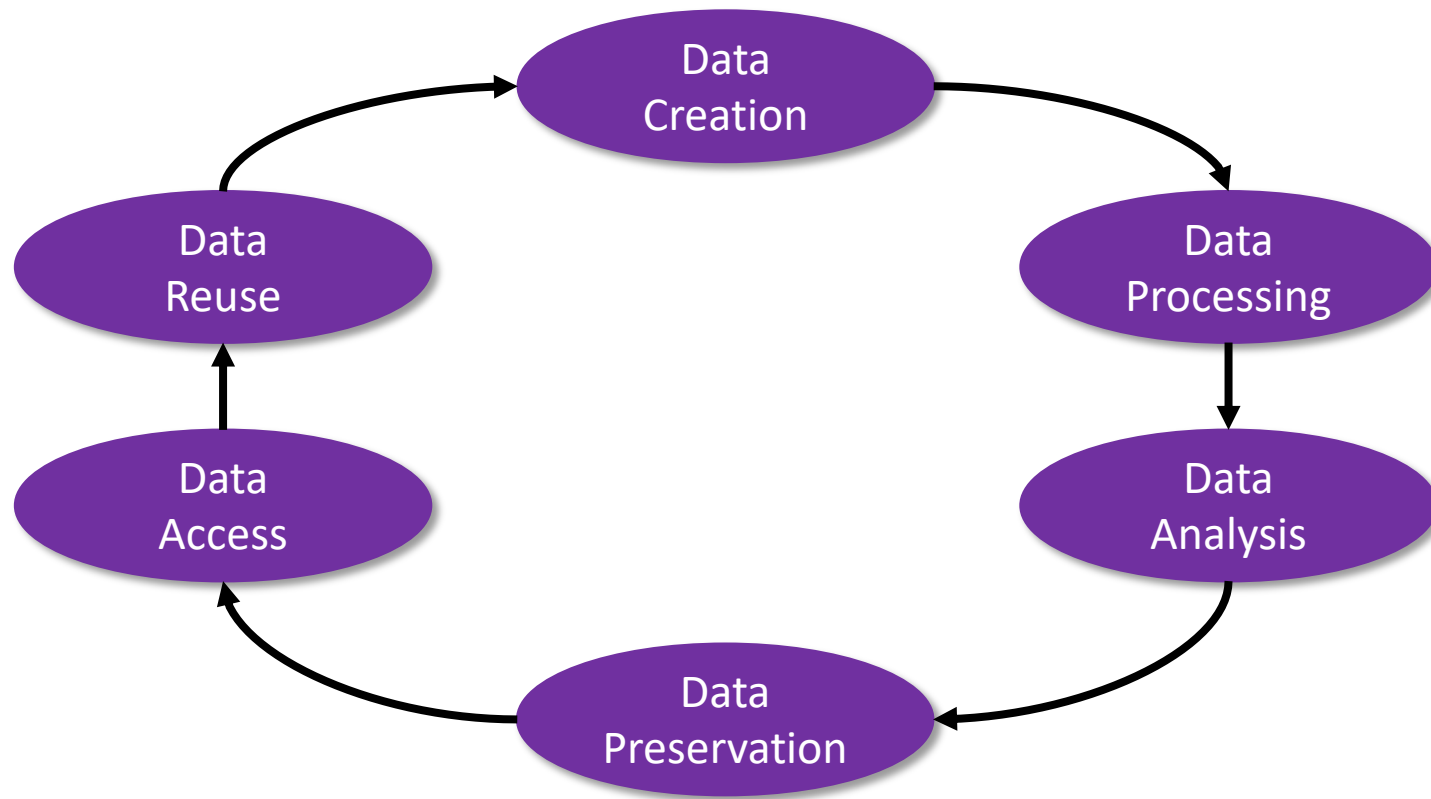
# A wild ride through the Research data life cycle



Research Data Management:

Handling of data that has been generated within scientific projects over the entire research data life cycle.

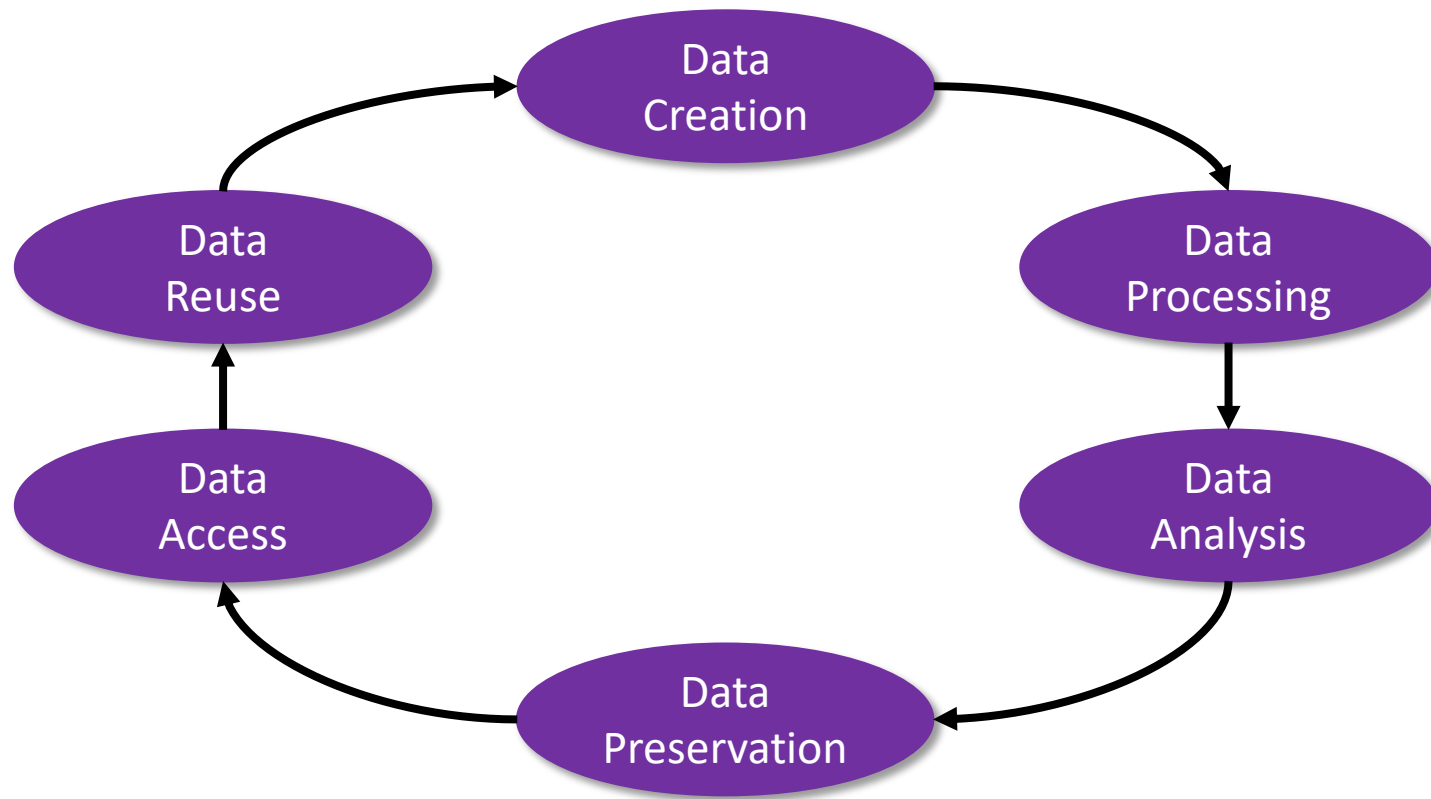
# A wild ride through the research data life cycle



- What data is expected?
- Which data will be reused?
- What should happen to the data?  
→ Data Management Plan

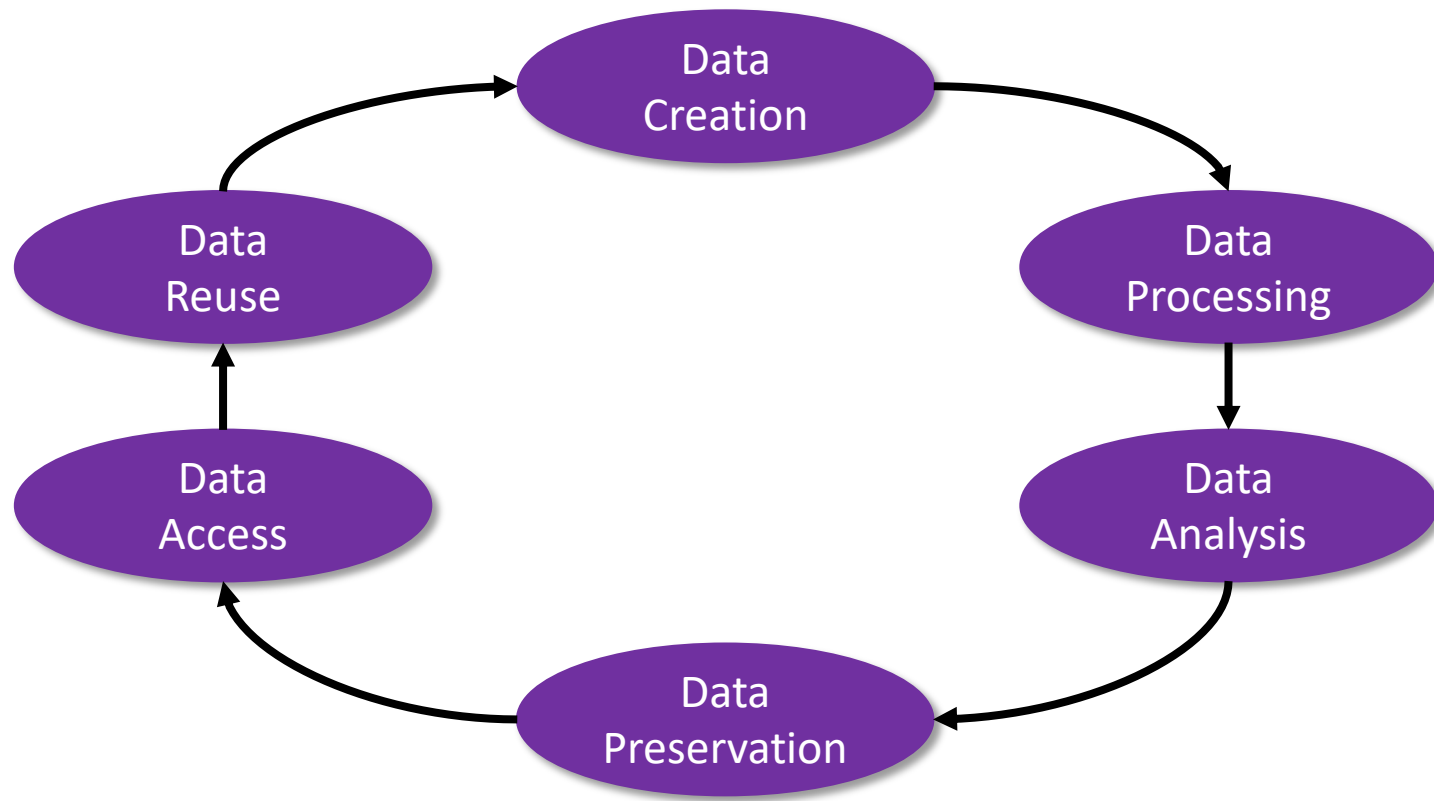


# A wild ride through the research data life cycle



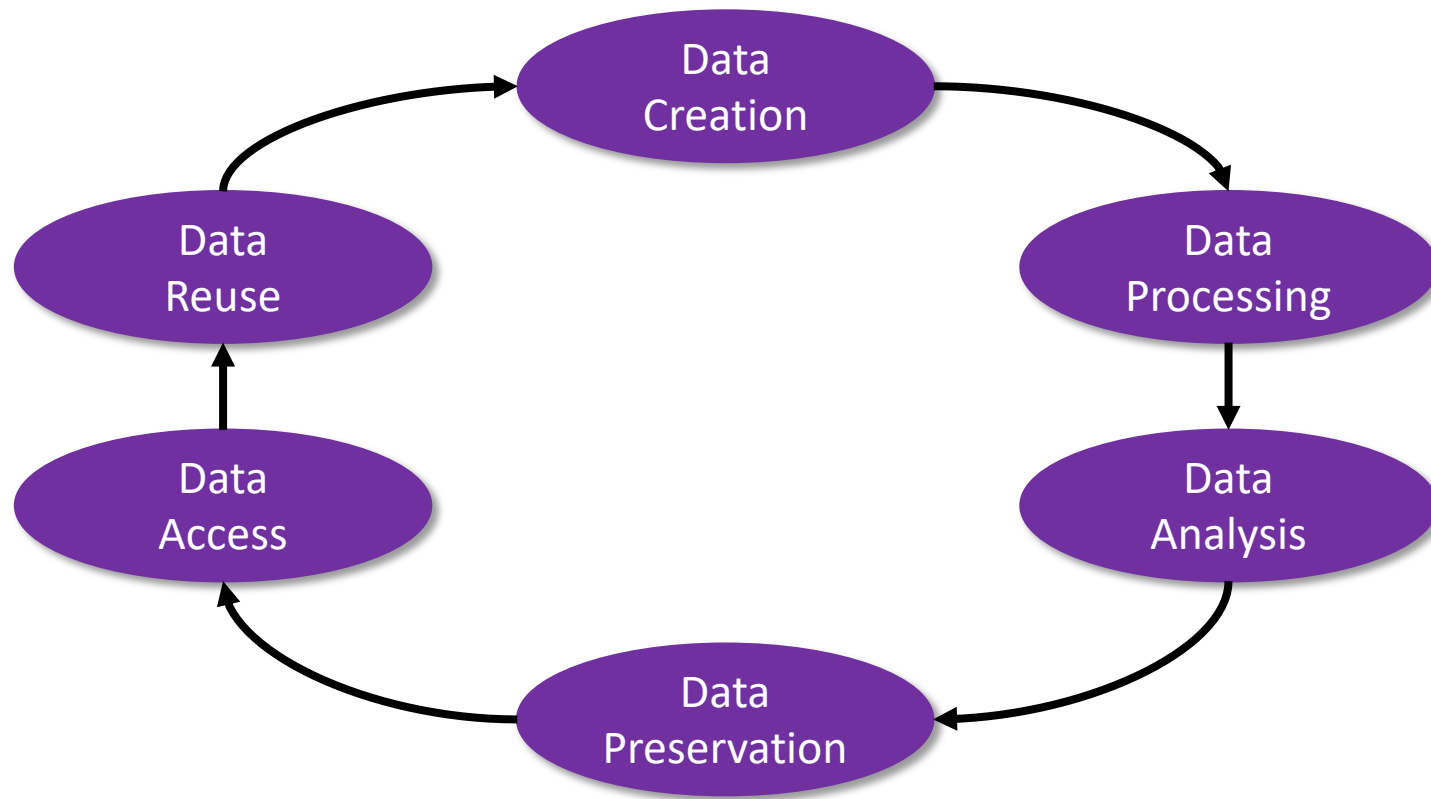
- Which data/metadata format will be used?
- How will the data be organized?

# A wild ride through the research data life cycle



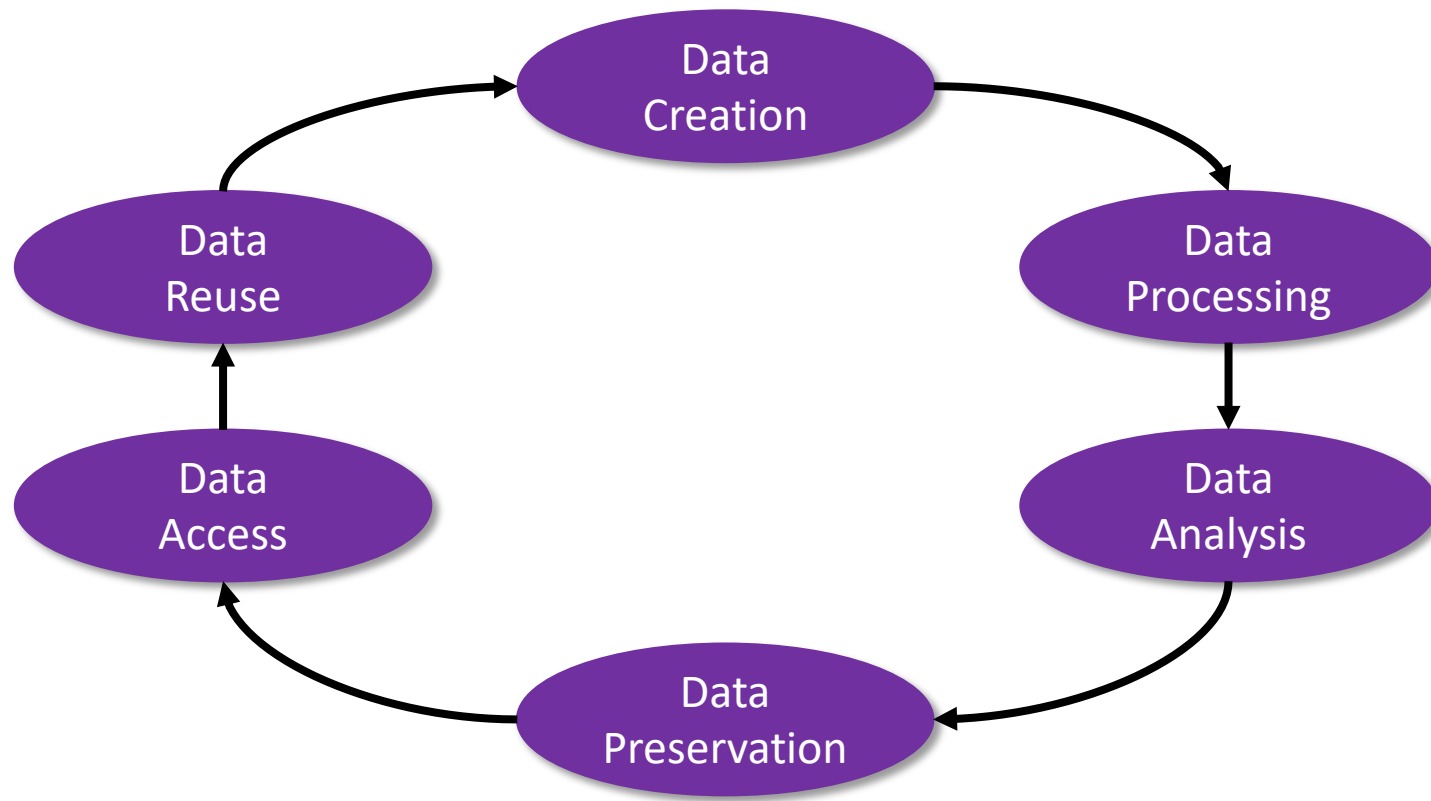
- Which new data formats will be produced?
- Which tools will be used?

# A wild ride through the research data life cycle



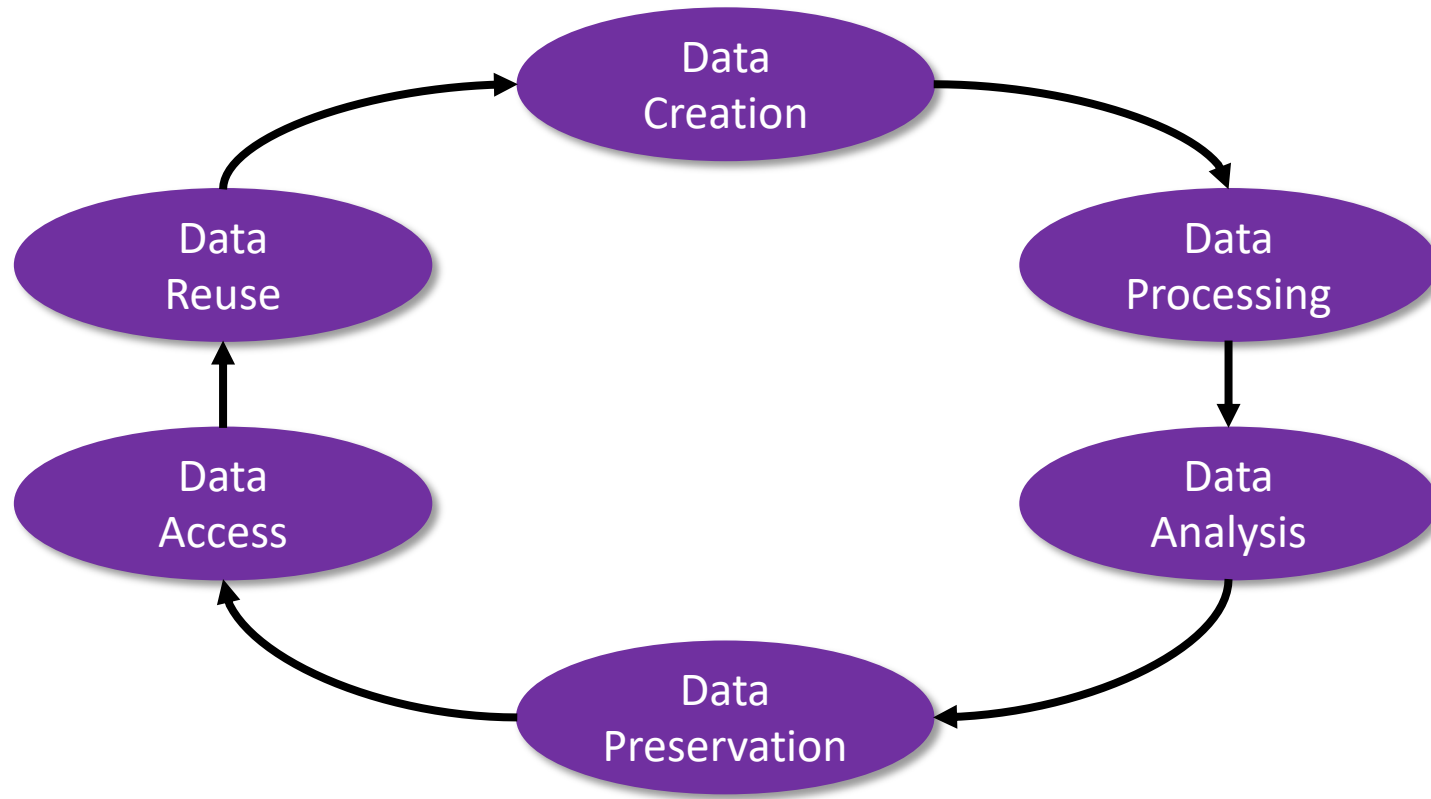
- Which data should be archived (long-term)?
- Where should the data be archived?  
→ Conditions: data formats, metadata, subject-specific or generic...

# A wild ride through the research data life cycle



- Which data should/can/must be made accessible?
- Which forms of presentation are needed?

# A wild ride through the research data life cycle



- Who should be able to find the data?
- How can data be reused?

# Finding data



I have a research question...

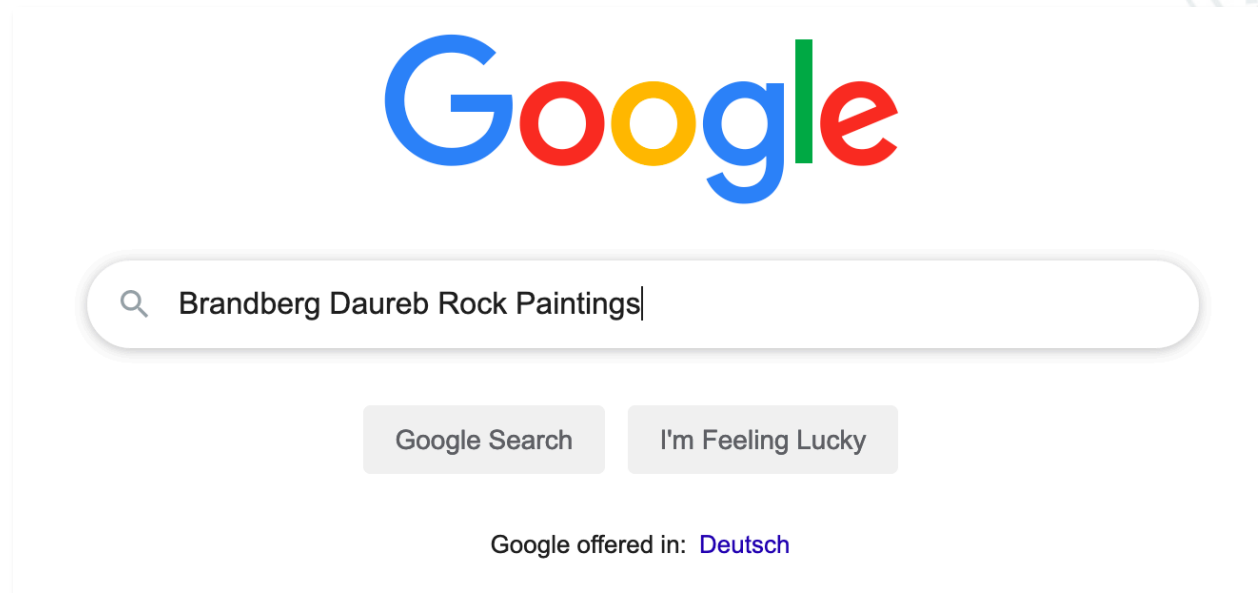
...comparing Andy Warhol's Paintings of Marilyn Monroe with the Rock Paintings at the Brandberg in Namibia!



[http://zeroequaltwo.net/wp-content/uploads/2015/07/namib\\_brandberg\\_ara\\_977x384.jpg](http://zeroequaltwo.net/wp-content/uploads/2015/07/namib_brandberg_ara_977x384.jpg)  
[https://news.artnet.com/app/news-upload/2015/08/594966\\_\\_warhol-marilyn-monroe-1\\_p-e1440069083122.jpg](https://news.artnet.com/app/news-upload/2015/08/594966__warhol-marilyn-monroe-1_p-e1440069083122.jpg)

# I have a research question...

...Comparing Andy Warhol's Paintings of Marilyn Monroe with the Rock Paintings at the Brandberg in Namibia!







Publication of the **rock art** of the **Brandberg/Daureb** as it was documented by Harald Pager († 1985) on behalf of the Institute of Prehistoric Archaeology of the ...

**Research Context:** Specific project cooperating ...

### Images for Brandberg Daureb Rock Paintings



→ [More images for Brandberg Daureb Rock Paintings](#)

Report images

### (PDF) Rock art in African Highlands, Brandberg/Daureb ...

<https://www.academia.edu> › [Rock\\_art\\_in\\_African\\_Highlands\\_Brandberg\\_...](#) ▼

**Rock art** in African Highlands, **Brandberg/Daureb**, Namibia – Painters of a Prehistoric Hunter-Gatherer world. Tilman Lenssen-Erz. Reprint from Bubenzer, O., ...

### (PDF) Brandberg-Daureb Database on Rock Art | Oliver ...

<https://www.academia.edu> › [Brandberg-Daureb\\_Database\\_on\\_Rock\\_Art](#) ▼

Tilman Lenssen-Erz & Oliver Vogels **Brandberg-Daureb Database on Rock Art** Contents Structure and Meaning of the Database .

### Daureb/Brandberg (Namibia) - Creap

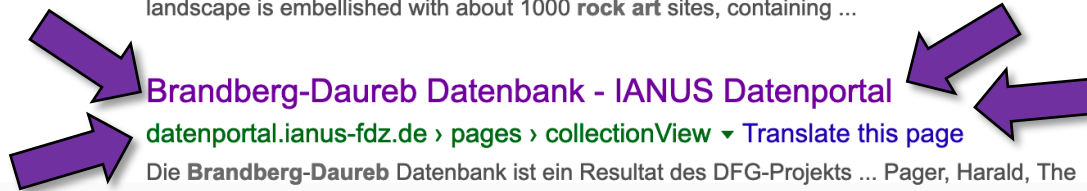
[www.creap.fr](http://www.creap.fr) › [Sites étudiés](#) ▼

The **Daureb** (or **Brandberg** in its better known colonial name) is an isolated mountain ... This landscape is embellished with about 1000 **rock art** sites, containing ...

### [Brandberg-Daureb Datenbank - IANUS Datenportal](#)

[datenportal.ianus-fdz.de](https://datenportal.ianus-fdz.de) › [pages](#) › [collectionView](#) ▼ [Translate this page](#)

Die **Brandberg-Daureb** Datenbank ist ein Resultat des DFG-Projekts ... Pager, Harald, The **Rock Paintings** of the Upper **Brandberg**, Part I, Amis Gorge.



# Brandberg-Daureb Datenbank

Dezember 2017



Projektübersicht

**Metadaten**

Daten

Startseite

Über das Datenportal

## Kontaktperson(en)



Dr. Tilman Lenssen-Erz (GND · ORCID)

Universität Köln  
Institut für Ur- und Frühgeschichte  
Forschungsstelle Afrika  
Jennerstraße 8  
50823 Köln

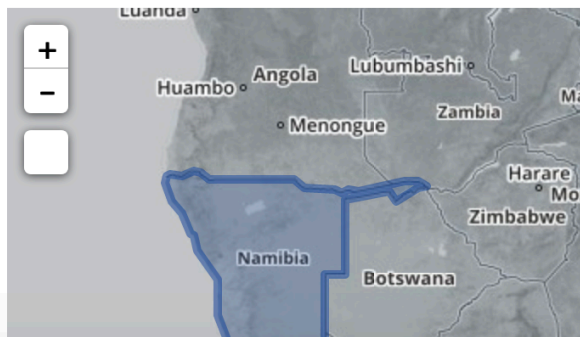
E-Mail: [lenssen.ertz@uni-koeln.de](mailto:lenssen.ertz@uni-koeln.de)

## Primärforscher

Dr. Tilman Lenssen-Erz (GND · ORCID)

Oliver Vogels (GND)

## Lokalisierung



## Informationen beziehen sich auf

Digitale Datensammlung · Analoges Archivmaterial · (Forschungs-)Projekt

## Zusammenfassung

Die Brandberg-Daureb Datenbank ist ein Resultat des DFG-Projekts (Deutsche Forschungsgemeinschaft) 'Felsbilder im Hohen Brandberg', das in der Felsbildforschung einen neuen Standard in Dokumentation und Publikation gesetzt hat. Diese weltweit in ihrer Vollständigkeit und Genauigkeit einzigartige Dokumentation wurde zwischen 1989 und 2006 nahezu vollständig vom Heinrich-Barth-Institut e.V. (Köln) in sechs Katalogen publiziert. Zusätzlich wurden alle Bilder sowie ca. 4700 Szenen anhand eines neuartigen, an sprachwissenschaftlichen Methoden orientierten Aufnahmeverfahrens (Lenssen-Erz 2001) erfasst und als Katalog in Buchform offen zugänglich gemacht (Pager 1989-2006). Der zu diesen Publikationen gehörige digitale Brandberg-Daureb Datenbestand von 671 Fundstellen (ausgenommen ist die noch unpublizierte Numas Schlucht) ist hier nun vollständig verfügbar.

Der Brandberg-Daureb Datenbestand enthält:

- Die vollständige Tabelle über 39.075 Einzelfiguren von 652 Fundstellen, aufgenommen von H. Pager
- Zusätzliche räumliche Informationen zu den vollständigen 840 Fundstellen, aufgenommen von T. Lenssen-Erz (Tilman Lenssen-Erz)

## Schlagwörter

### Fachdisziplinen

Allgemein: Archäologie ([iDAI.thesaurus](#)) · Ethnologie ([iDAI.thesaurus](#)) · Geowissenschaften ([iDAI.thesaurus](#))

### Gegenstand

Allgemein: Prähistorische Kunst · Felsmalereien · Fundstelle · Felsbilder ([iDAI.thesaurus](#)) · Gravuren ([iDAI.thesaurus](#))

### Zeitstellung

# Brandberg-Daureb Datenbank

Dezember 2017



[Projektübersicht](#)

[Metadaten](#)

[Daten](#)

[Startseite](#)

[Über das Datenportal](#)

## Kontaktperson(en)

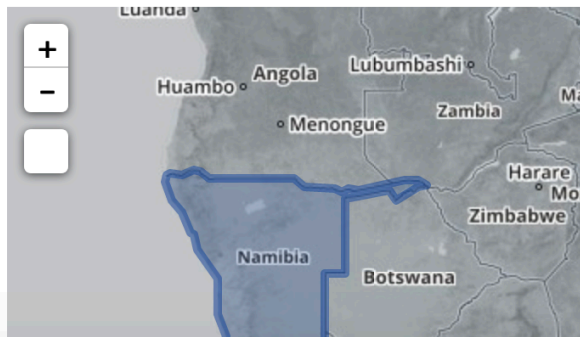


Dr. Tilman Lensen-Erz (GND · ORCID)  
 Universität Köln  
 Institut für Ur- und Frühgeschichte  
 Forschungsstelle Afrika  
 Jennerstraße 8  
 50823 Köln  
 E-Mail: [lensen.ertz@uni-koeln.de](mailto:lensen.ertz@uni-koeln.de)

## Primärforscher

Dr. Tilman Lensen-Erz (GND · ORCID)  
 Oliver Vogels (GND)

## Lokalisierung



## Schlagwörter

### Fachdisziplinen

Allgemein: Archäologie ([iDAI.thesaurus](#)) · Ethnologie ([iDAI.thesaurus](#)) · Geowissenschaften ([iDAI.thesaurus](#))

Speziell: Afrikanische Archäologie ([iDAI.thesaurus](#)) · Geoarchäologie ([iDAI.thesaurus](#))

### Gegenstand

Allgemein: Prähistorische Kunst · Felsmalereien · Fundstelle · Felsbilder ([iDAI.thesaurus](#)) · Gravuren ([iDAI.thesaurus](#))

Speziell: San · Wildbeuter · pastro-forager · Jäger-Sammler · Abris ([iDAI.thesaurus](#))

### Zeitstellung

Allgemein: Steinzeit ([iDAI.chronontology](#)) · Holozän ([iDAI.chronontology](#))

Speziell: Late Stone Age (LSA) ([iDAI.chronontology](#))

Projektdauer: 1977 - 2006

Datenerzeugung: 1977 - 1985: Arbeiten (Anfertigung von Photographien und Zeichnungen) vor Ort am Brandberg durch Harald Pager · 1989 - 2006: Auswertung und Publikation der Arbeiten von Harald Pager durch das Heinrich-Barth-Institut an der Universität Köln · 2012 - 2014: Digitalisierung der analogen Datenbestände von Harald Pager durch das AAARc - Forschungsstelle Afrika an der Universität zu Köln.

### Methode

Allgemein: Oberflächenbegehung (Survey) ([iDAI.thesaurus](#)) · Analoge Dokumentation ([iDAI.thesaurus](#)) · Vermessung und Geodäsie ([iDAI.thesaurus](#)) · Fundanalyse ([iDAI.thesaurus](#))

Speziell: Analoge Fotografie ([iDAI.thesaurus](#)) · Analoge zeichnerische Dokumentation ([iDAI.thesaurus](#)) · Archäologischer Survey (intensiv) ([iDAI.thesaurus](#)) · Topografische Aufnahme ([iDAI.thesaurus](#)) · Topographische Untersuchung ([iDAI.thesaurus](#)) · Felsbildforschung

## Brandberg-Daureb Datenbank



Dezember 2017

Projektübersicht

Metadaten

Daten

Startseite

Über das Datenportal

## Verzeichnisstruktur

Brandberg-Daureb-Database (3 Ordner   21 Dateien)
└─ Entries-and-Vocabulary (0 Ordner   5 Dateien)
└─ Maps (0 Ordner   4 Dateien)
└─ Tables (0 Ordner   6 Dateien)

## Kontaktperson(en)



Dr. Tilman Lenssen-Erz (GND · ORCID)

Universität Köln  
 Institut für Ur- und Frühgeschichte  
 Forschungsstelle Afrika  
 Jennerstraße 8  
 50823 Köln

E-Mail: [lenssen.ertz@uni-koeln.de](mailto:lenssen.ertz@uni-koeln.de)

## Identifikation &amp; Zitation

IANUS Sammlung 2017-00012  
 DOI 10.13149/r5f.3fyhip-x

Zitierhinweis

## Lizenzhinweis



## Daten



Entries\_and\_Vocabulary\_in\_BDDB.pdf

321,5 KB

Download



Entries\_in\_Table\_of\_Paintings.xlsx

69,4 KB

Download



Entries\_in\_Table\_of\_Scenes.xlsx

66,6 KB

Download



Vocabulary\_in\_Table\_of\_Paintings.xlsx

43,6 KB

Download

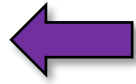
Entries\_in\_Table\_of\_Paintings.xlsx

Dateigröße: 69,4 KB

Lizenzhinweis



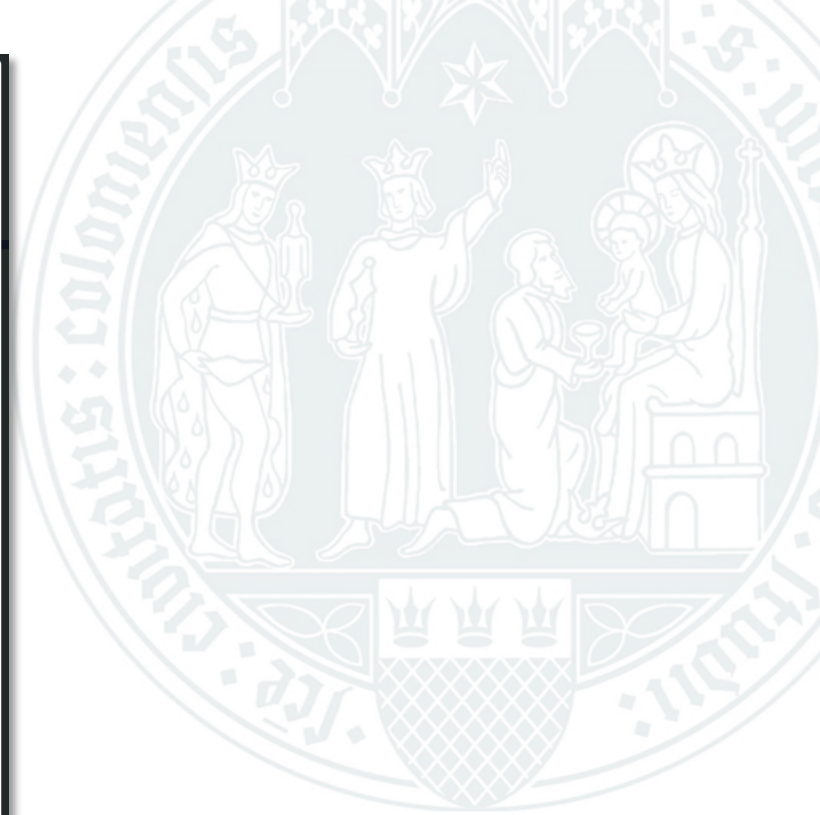
Lizenzierung: CC-BY-SA 3.0



[Download file](#)



No preview available



Automatisches Speichern AUS Entries\_in\_Table\_of\_Paintings-1

Start Einfügen Zeichnen Seitenlayout Formeln Daten Überprüfen Ansicht

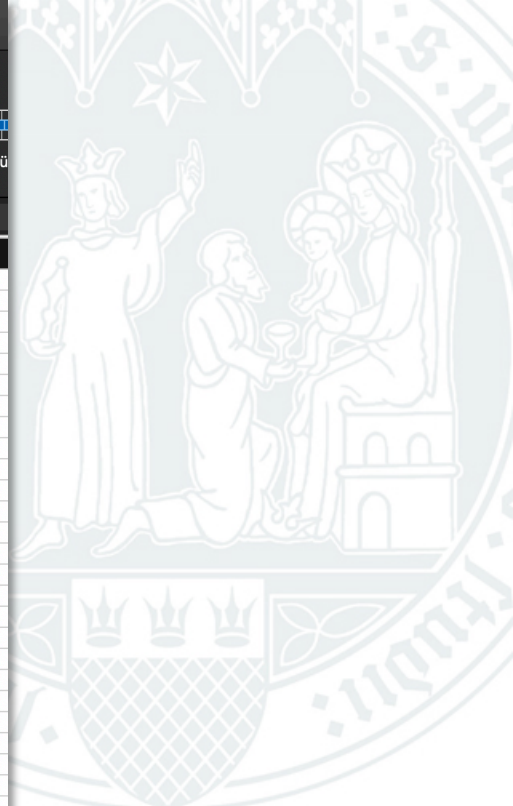
Calibri (Textkörper) 11 Textumbruch Standard Bedingte Formatierung Als Tabelle formatieren Zellenformatvorlagen Einfügen

F17

	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	Entries in S	Frequency																		
2	2 arrows	1																		
3	2 dots	1																		
4	2 lines	11																		
5	2 strokes	1																		
6	3 lines	1																		
7	aloe	2																		
8	aloe (XL)	1																		
9	aloe dichot	4																		
10	animal	1070																		
11	animal (ma	1																		
12	animal (XL)	5																		
13	animal (XS)	62																		
14	animal hea																			
15	animal trac	25																		
16	antbear	2																		
17	antelope	586																		
18	antelope (f	1																		
19	antelope (n																			
20	antelope (X	9																		
21	antelope (X	7																		
22	antelope (y	1																		
23	antelope h	2																		
24	arch	4																		
25	arrow	36																		
26	arrows	41																		
27	arrows \& t	1																		
28	baboon	11																		
29	bag	266																		
30	bag (XL)	2																		
31	bag (XS)	9																		
32	bag \& bow	1																		
33	bag \& clut	1																		
34	bag \& quiv	1																		
35	bags	7																		
36	bat-eared fr	3																		
37	beak	1																		
38	bee	17																		
39	bees	2																		
40	bird	98																		
41	bird (XS)	12																		
42	bird legs	1																		
43	black licher	2																		
44	blot	230																		
45	blots	66																		
46	blots \& lin	1																		
47	blots \& str	1																		
48	bovid	163																		
49	bovid (mal	1																		
50	bovid (XL)	3																		
51	bovid (XS)	2																		
52	bow	141																		
	bow (XL)	1																		
	bow \& arr	9																		

Now I can analyze and use this data for my own research (following the given licenses).

Special Features Static Dynamic Supplementary Body Posture Transitive Action Applicative Action Direct Object Goal (Indirect Object) +



dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est

Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut

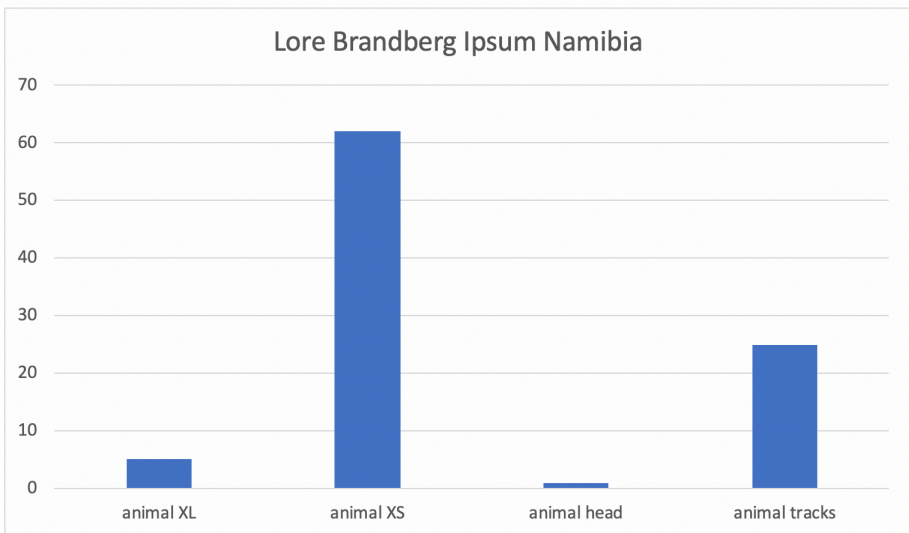
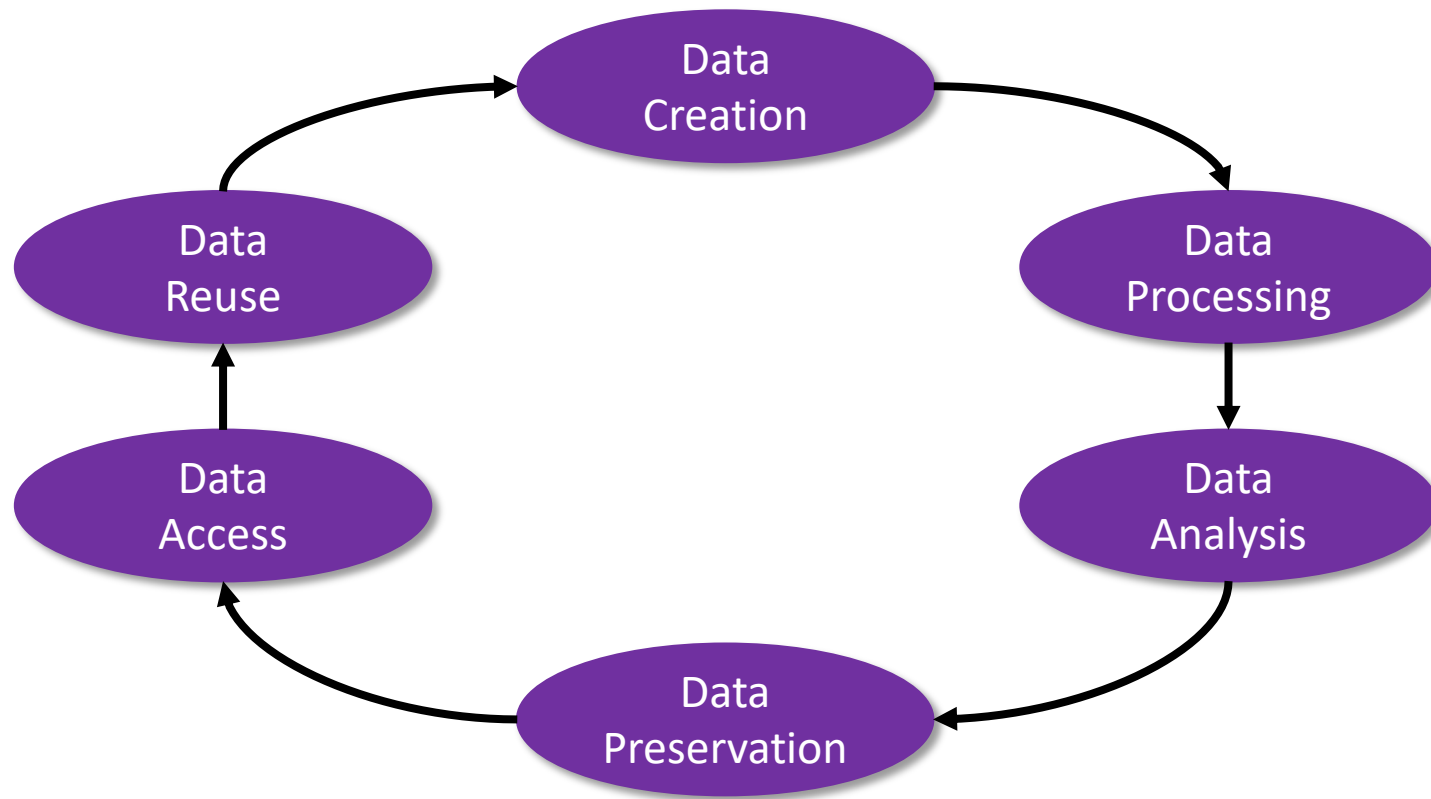


Figure 1: Lore Brandberg Ipsum Namibia (Lenssen-Erz, 2016).

labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut



# What happened?



- We had a research question and were searching for research data
- We were able to find data, because someone published it

In our case:

- Tilmann Lenssen-Erz published (a lot of) his data about rock paintings at the Brandberg in Namibia
- The Data is accessible via IANUS, a domain-specific repository for archaeological data



# Publishing data





Browse

Search on figshare...

Log in

Sign up

**store, share, discover research**

get more citations for all of the outputs of your academic research  
over 5000 citations of figshare content to date

ALSO FOR INSTITUTIONS & PUBLISHERS

*"figshare wants to open up scientific data to the world"* **WIRED**

The background figure: Merged NavCam images of Rosetta... by K.-Michael Aye in Planetary Science



simplifying your research workflow

needed to publish ●

Title

Rock Art and Andy Warhol - A Comparison

Authors

Patrick Helling

Search co-authors by name, full email or ORCID. Hit enter after each.

Categories

Archaeological Science, Art

Item type

Preprint

Keyword(s)

Andy Warhol Rock Art Namibia Brandberg Art Archaeology

Add keywords for easy discovery. Hit enter after each

Description

The Concept of Iteration in Rock Art and Paintings of Andy Warhol.

B I U x<sub>2</sub> x<sup>2</sup>

Funding

Search grant by name/number or add your own

+ Add another grant

Tips

you can still drag more file(s) on the page or browse

Preview item (private)

Edit timeline



References

10.1543/data.234565

Licence (what's this?)

CC BY 4.0

This item is a draft (metadata required for publication missing)

Apply embargo

Make file(s) confidential

Generate private link

DOI Reserve Digital Object Identifier

Delete item

Cancel

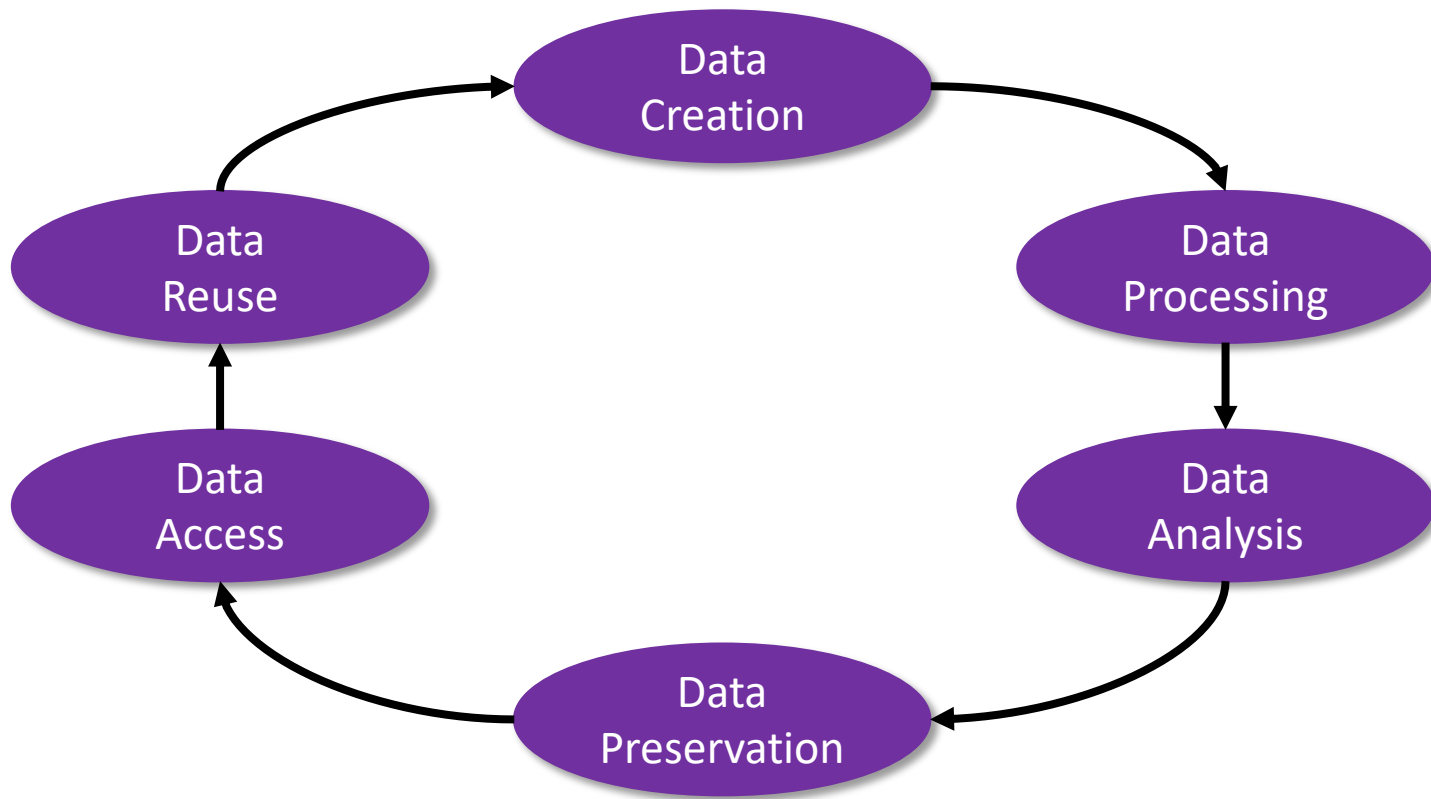
DOI



Publish

Save changes

# What happened?



- The paper (and the dataset) get(s) a DOI from figshare, can be found via the search function and is linked to the dataset we produced
- Behind the scenes: The metadata is also sent to a number of other aggregating services, e.g. DataCite
- The paper and the dataset are now permanently available via the DOI

# Practical Exercise

- 1. Try to find Data for your own research**
- 2. Try to find out, where you can publish your data (ideally domain-specific)**

Document your search

- How did you do your research?
- Where did you find something?
- What did you find?
- What was difficult?
- ...

Short presentation of your results!



# Take home message

The data life cycle has been already encountered by each of us in many situations of our daily work. It is the fundamental principle that underlies many scientific steps.

We must be a comprehensive part of the data life cycle to support science, handle data in the sense of good scientific practice and profit from/promote (excellent) scientific progress.

That means: **Do not just use data, make data available as well.**

Oh and:

→ Make Data **F**indable, **A**ccessible, **I**nteroperable and **R**eusable!

# Open science



55

RDM in the Humanities – Academic Alumni Forum



# Open science

Making scientific research, e.g. publications, data, physical samples, and software accessible to everyone.

## Six Principles of Open Science

Open Data

Open Source

Open Methodology

Open Peer Review

Open Access

Open Educational Resources





# Open science

Making scientific research, e.g. publications, data, physical samples, and software accessible to everyone.

## Six Principles of Open Science

### Open Data

Open Source

Open Methodology

Open Peer Review

### Open Access

Open Educational Resources



# Open science

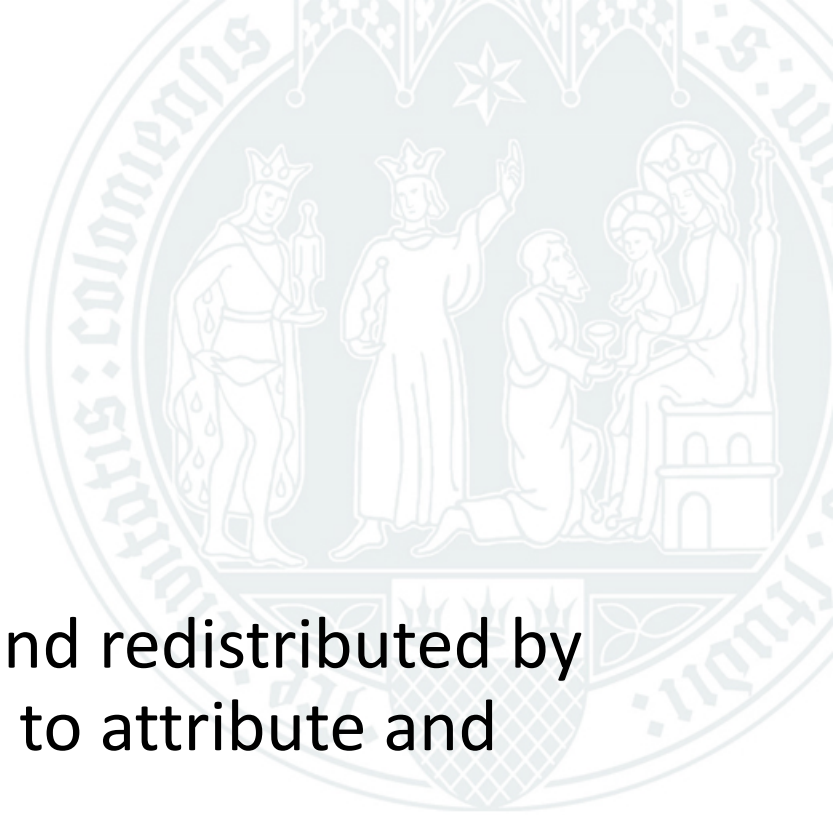
Open data



# Open data

Open data is data that can be freely used, reused and redistributed by anyone – subject only, at most, to the requirement to attribute and sharealike.

Open Data Handbook



# European Commission Horizon 2020

## OPEN ACCESS TO RESEARCH DATA

Research data is information (particularly facts or numbers) collected to be examined and considered, and to serve as a basis for reasoning, discussion or calculation.

Open access to research data - the **right to access and reuse digital research data** under the terms and conditions set out in the Grant Agreement.

As far as possible, projects must then take measures to enable third parties to access, mine, exploit, reproduce and disseminate (free of charge for any user) this research data.

**One straightforward and effective way of doing this is to attach Creative Commons Licences (CC BY or CC0) to the data deposited.**

Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020

# Open science

Open access



# Roads to open access

**The gold road to Open Access** → primary publication of scientific and scholarly works as articles in OA journals, as OA monographs, or as contributions to openly accessible edited volumes or conference proceedings.

**The green road to Open Access** → practice of providing OA to a version of work published in a closed-access journal or with a closed-access publisher by depositing it in an openly accessible institutional or disciplinary repository.



# Open access

Open Access to scientific and scholarly literature means its free availability on the public internet permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself.

The Budapest Open Access Initiative (BOAI) Declaration of 14 February 2002



# Data Management Plan



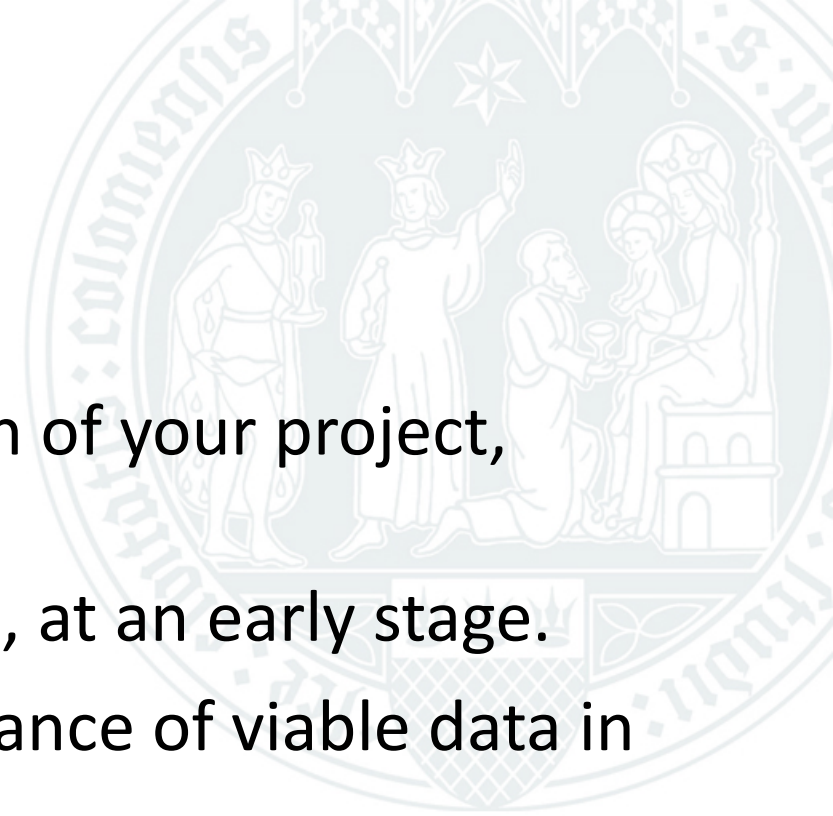
# What's a data management plan?

A data management plan (DMP) is a **written document** that describes the data you expect to acquire or generate during the course of a research project, how you will manage, describe, analyze, and store those data, and what mechanisms you will use at the end of your project to share and preserve your data.

[Stanford Libraries Data Management Plans](#)

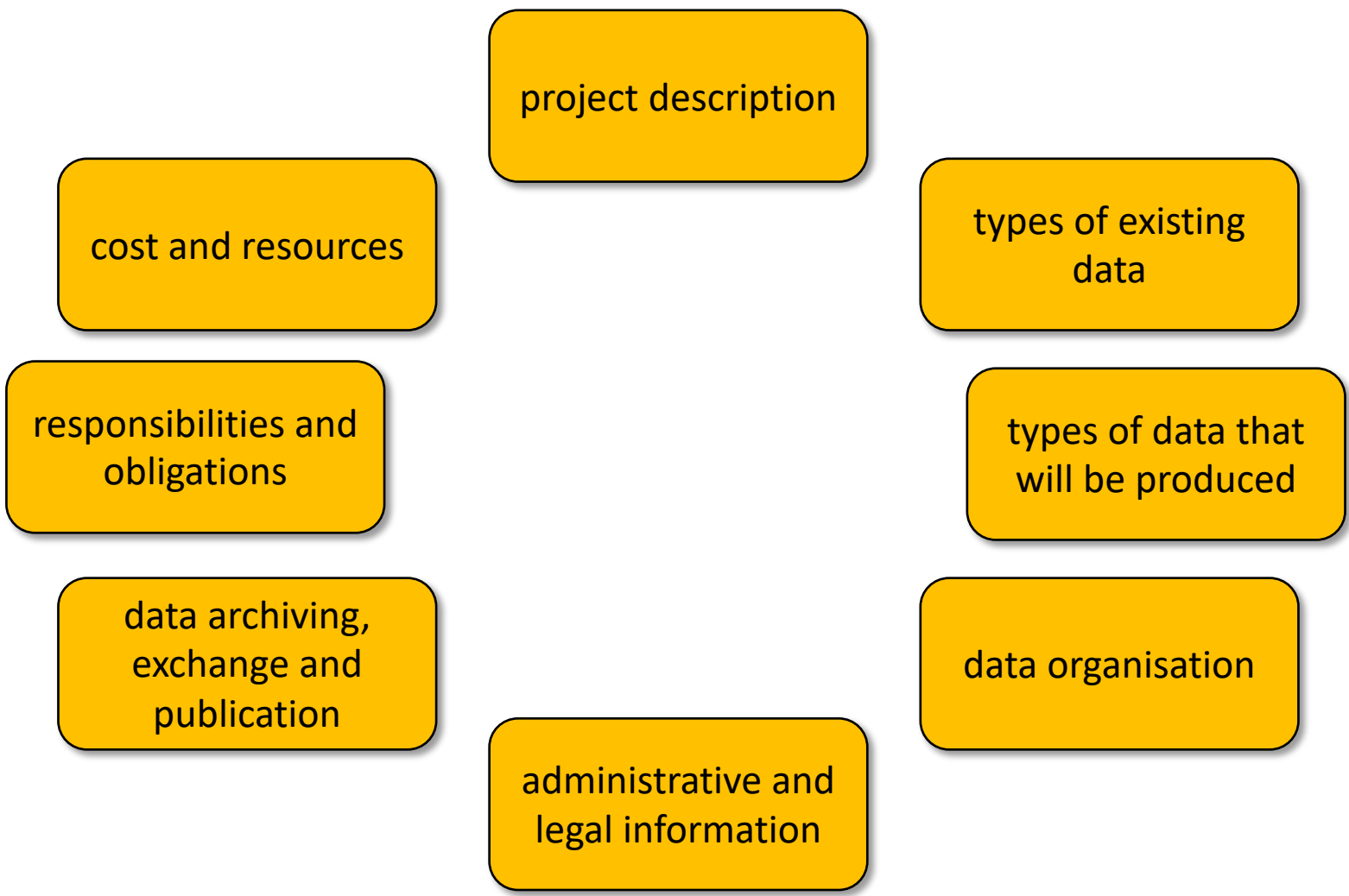
# Why? So, you ...

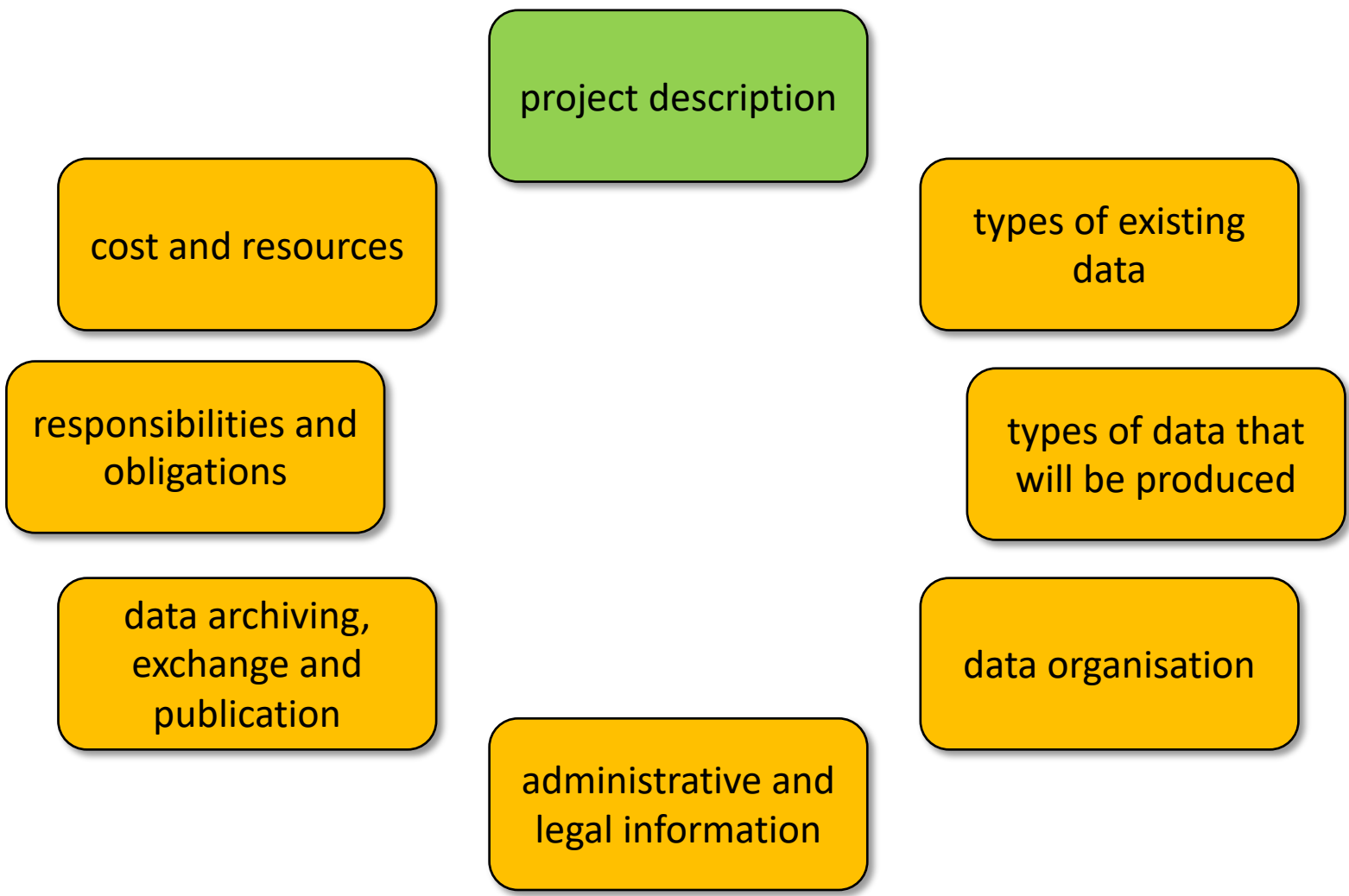
- have a structured (and data-centered) description of your project, from the start.
- can identify potential issues in data management, at an early stage.
- have a reduced risk of data loss and increased chance of viable data in 10 years time.
- have explicitly defined responsibilities.
- don't reactively or retro-actively improvise some kind of data management



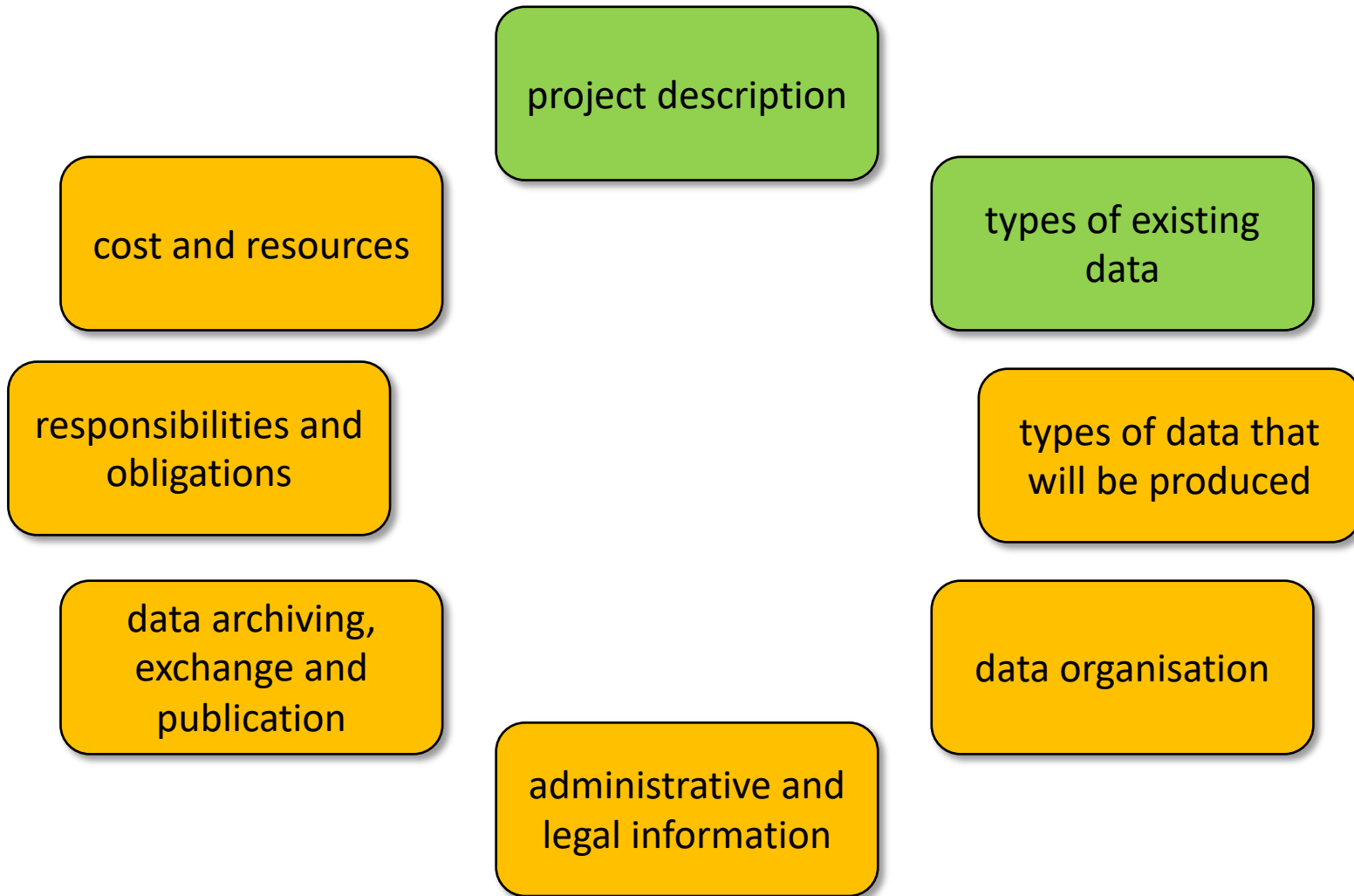
# What goes into a DMP?



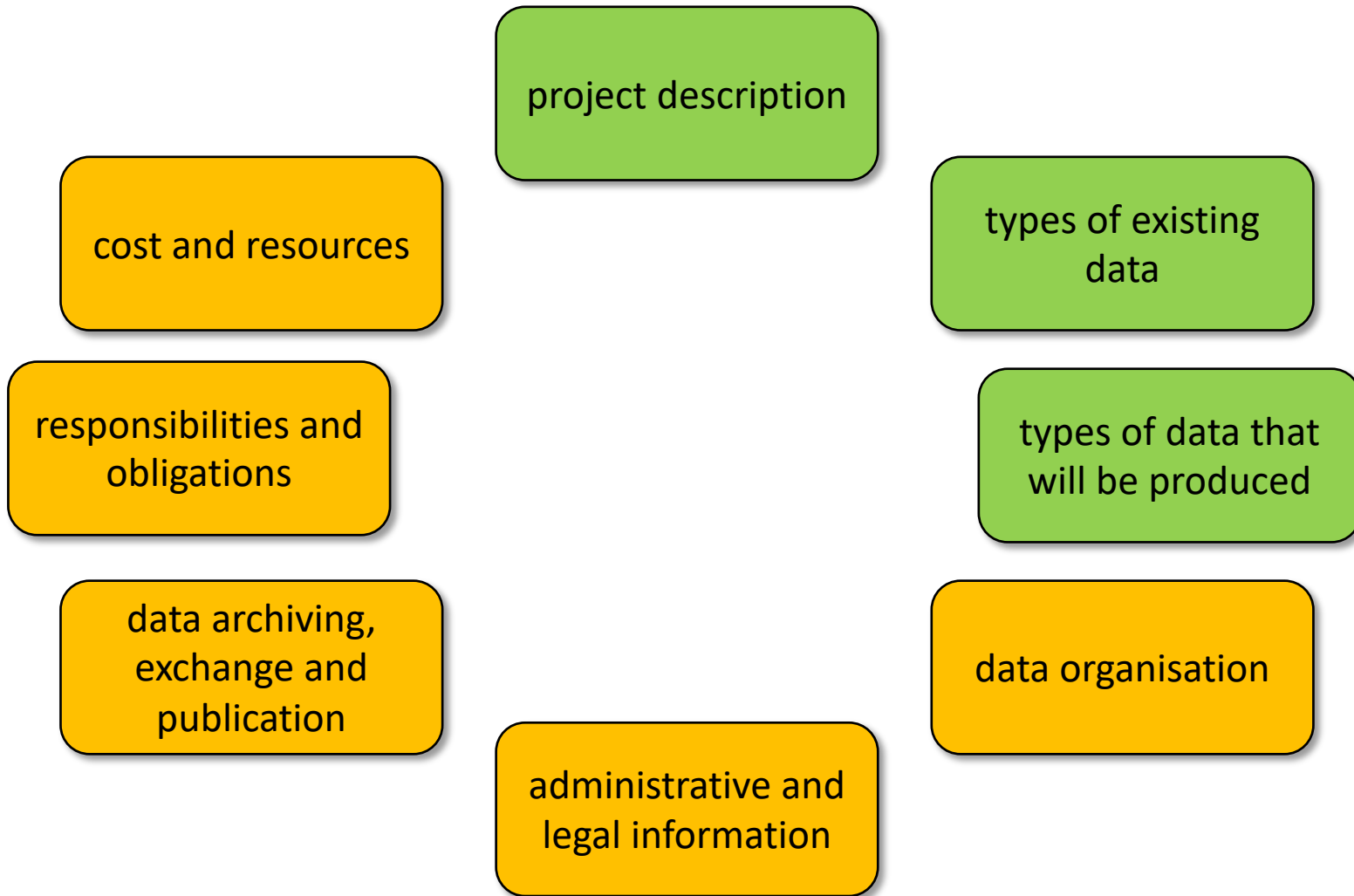




General information  
(e.g. title, objective,  
funder, project duration)



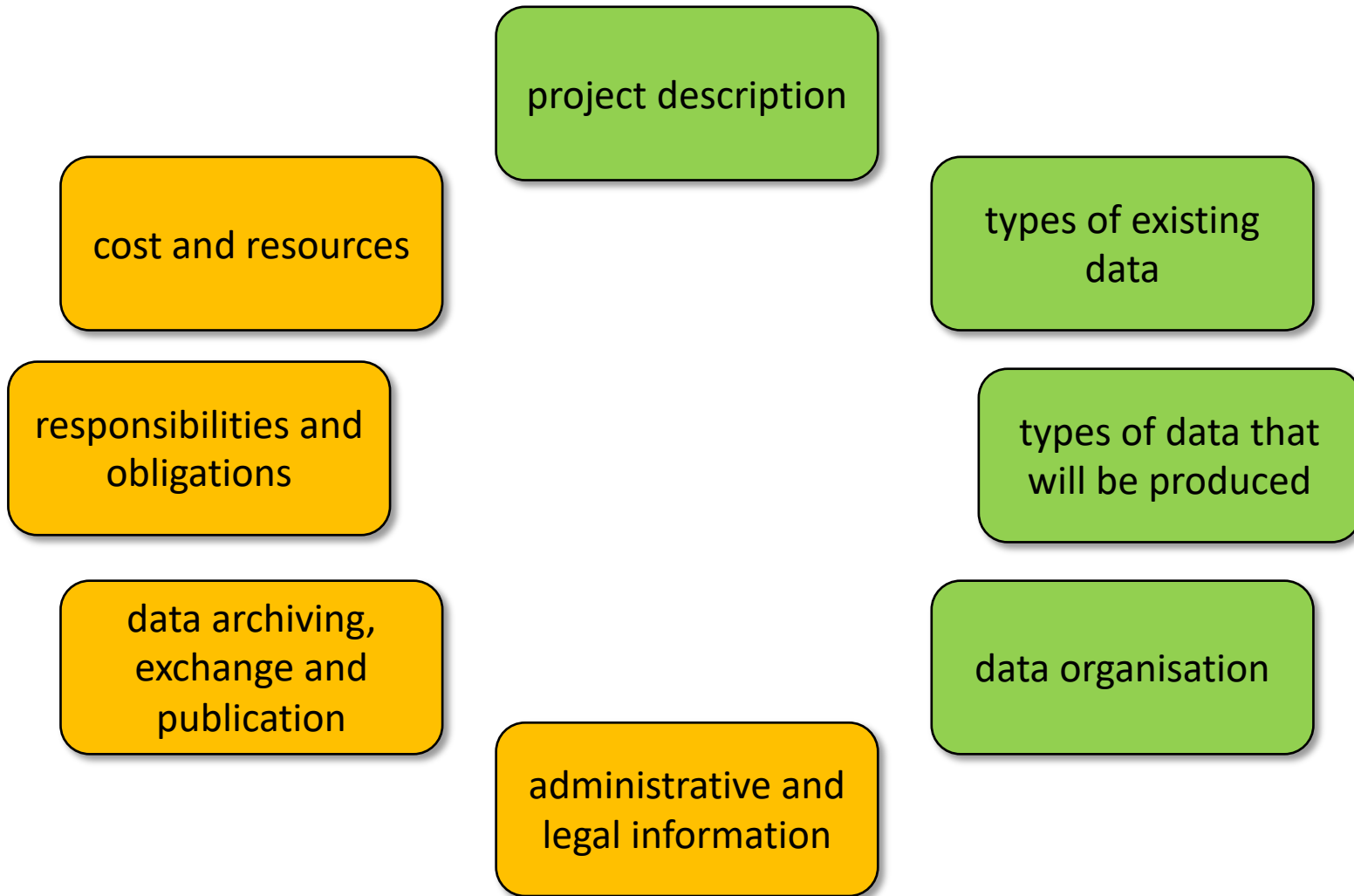
A description of existing data which will be (re-) used in the project and a description of how they will be integrated into the project



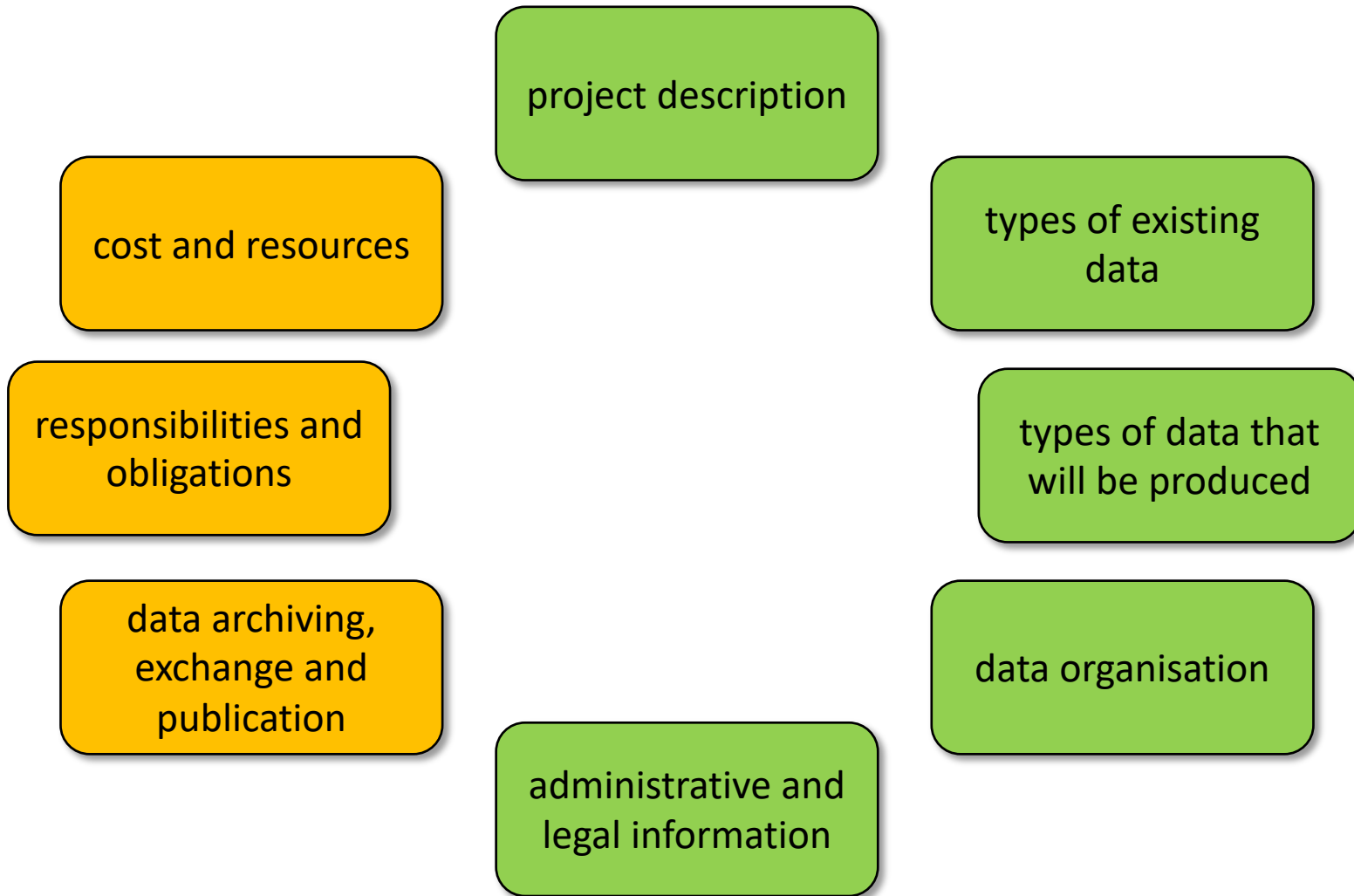
A description of the data types and data formats that will be used and produced.

Additional information on data generation and quality assurance (such as documentation and validation) should be given as well.





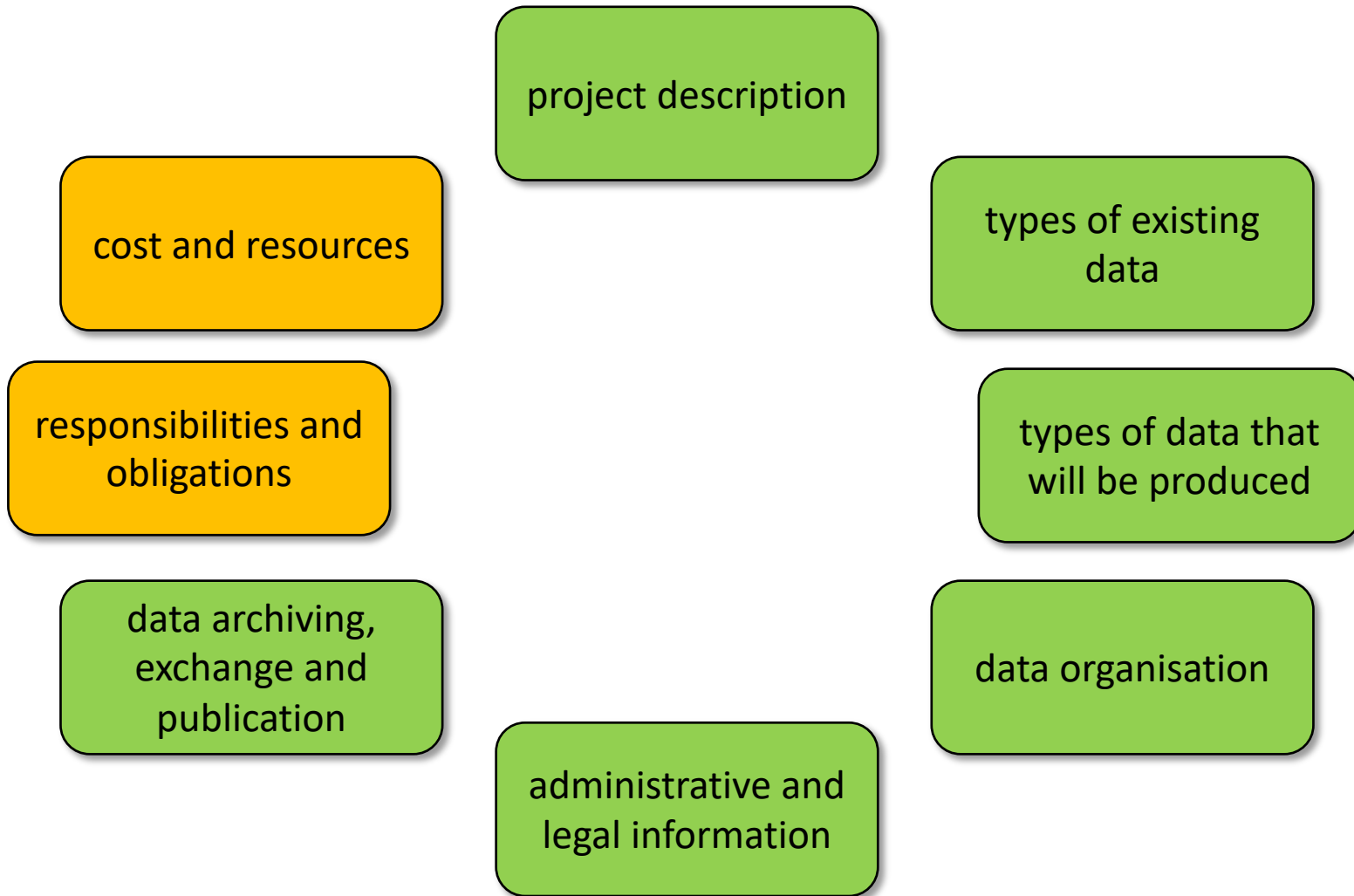
Information on consistent data management guidelines within the project (e.g., data backup, file naming, synchronization, versioning and set ups supporting collaborative work).



Requirements of funding bodies or relevant legal regulations

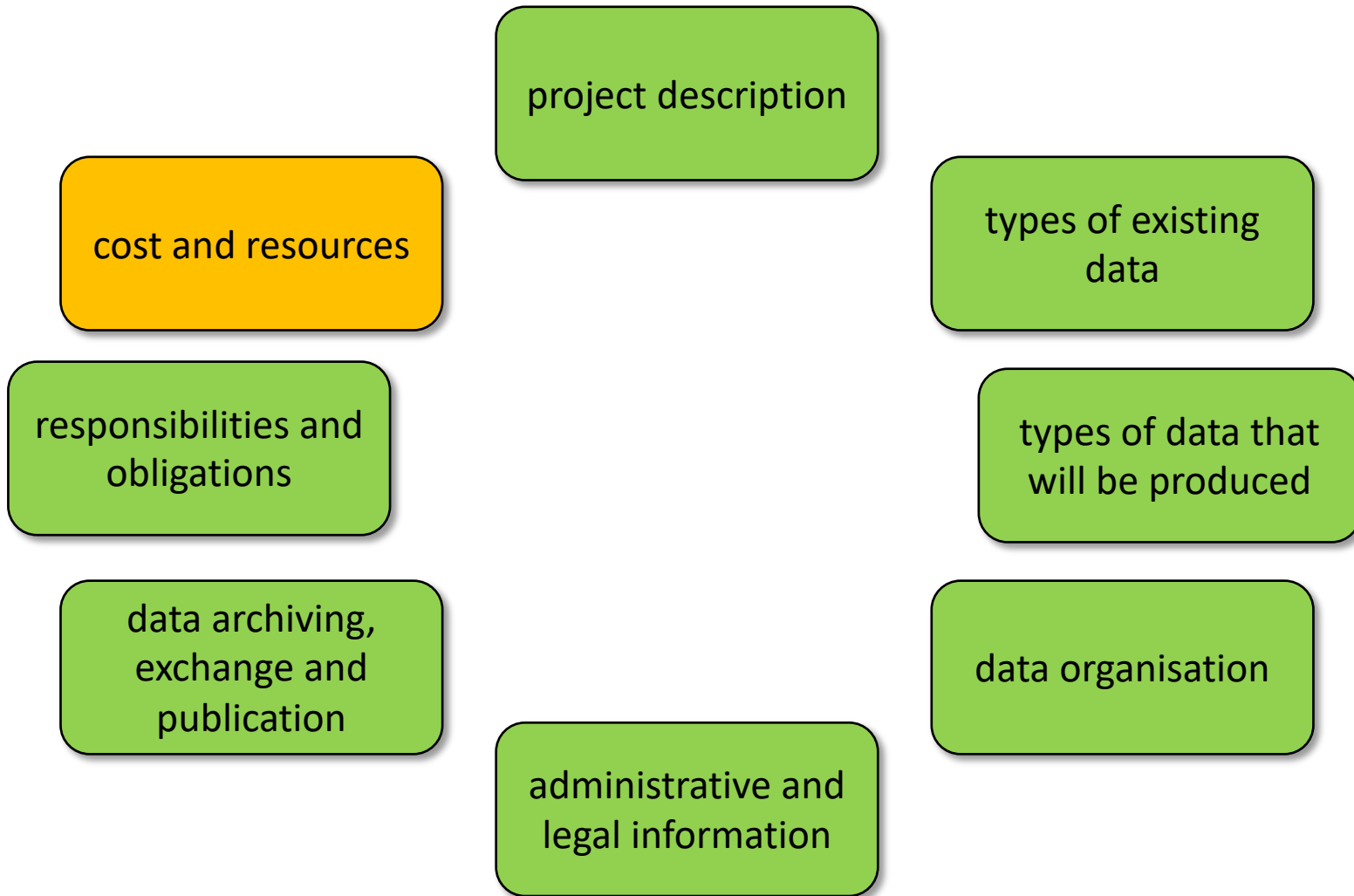
Access and usage policies (access rights, usage restrictions),

Data protection and backup

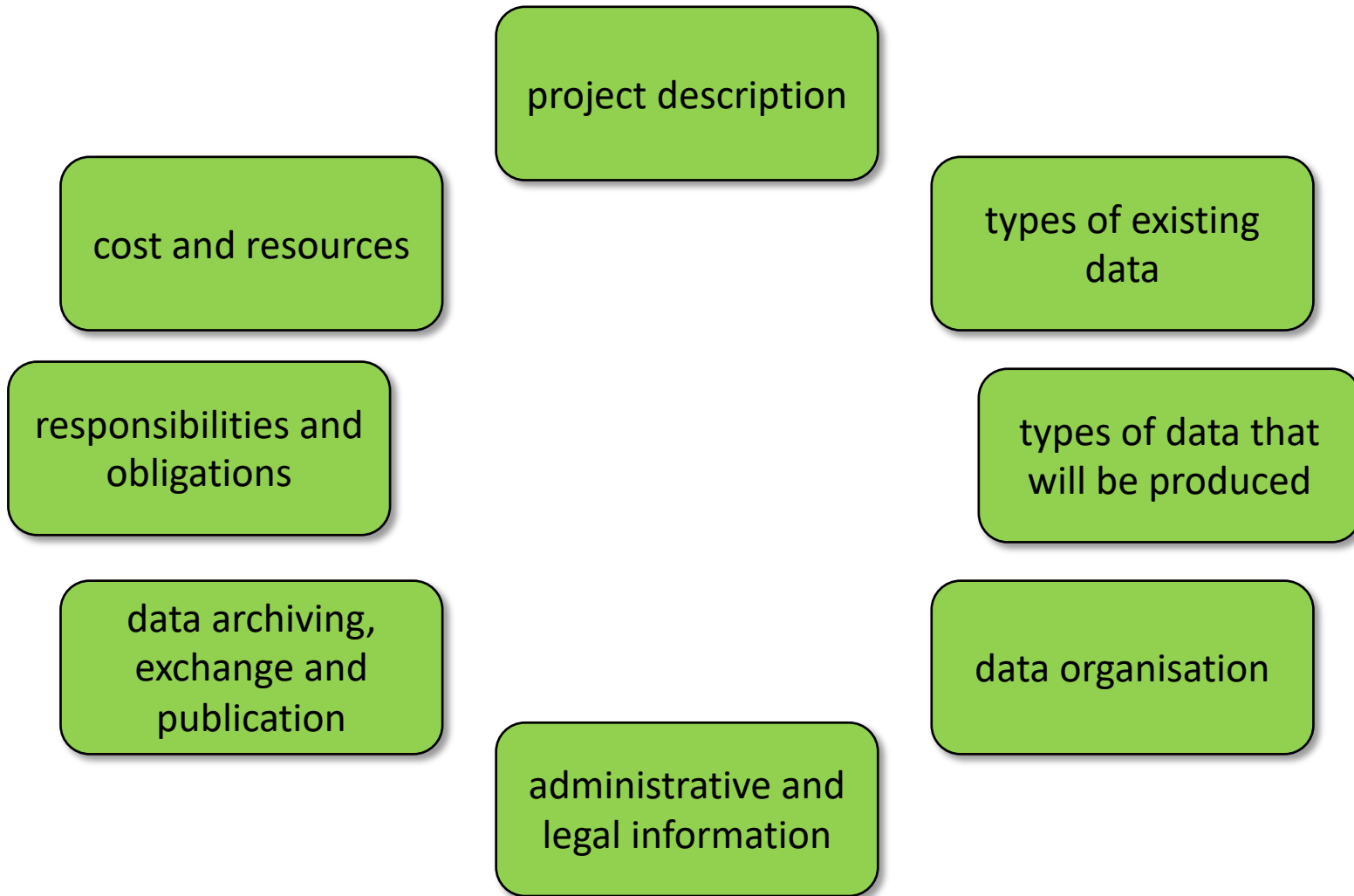


What (types of) data are exchanged? How will the exchange be realised?

Information about planned interoperability with data services that are established in your community .

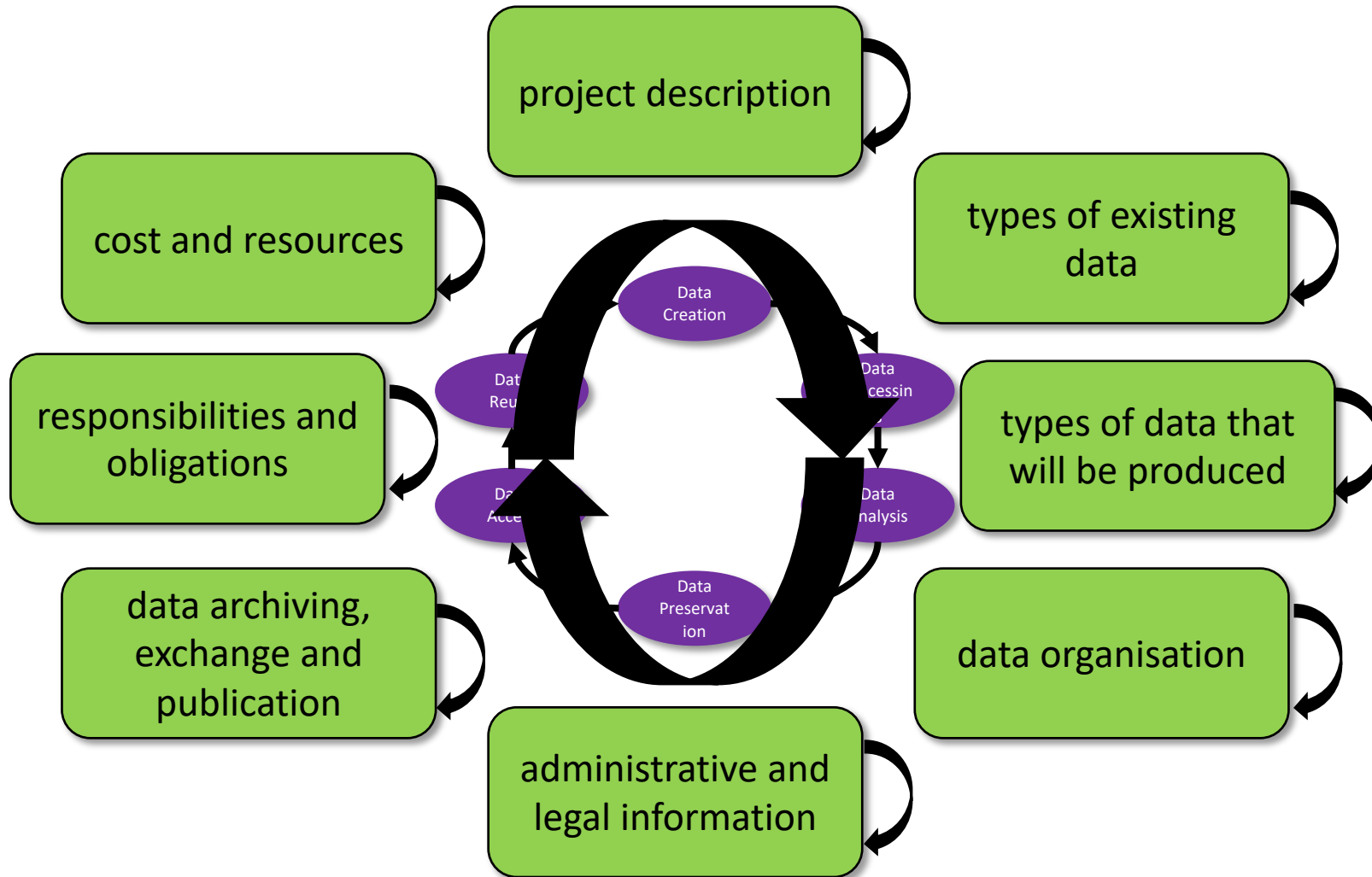


Who is responsible for data management in your project. How are the responsibilities distributed?



Description of cost and required person months for implementing the data management plan, including costs for data curation, creation of metadata, archiving etc.

The DMP is a description of the research data management process for the whole data life cycle.



# Work with a plan!

- RDM is a dynamic process that extends over the duration of the whole research project.
- The DMP is a guideline that will evolve with the reality encounter during the project.



Make a plan!



Have a plan.





## Wizard zur Planerstellung für das Datenmanagement (experimentell)



### Administrative Daten zum Projekt

Welches Forschungsprojekt wird die Daten bereitstellen?

Was sind die Ziele des Projektes?

Wer ist verantwortlicher Projektleiter?

Wie ist der Name der Forschungseinrichtung bzw. Universität?

Wer sind die Ansprechpartner in diesem Projekt?

Finden verwandte Richtlinien der Institution Anwendung?

# RDMO



# RDMO

Research Data Management Organiser

Stockholm Public Library (Samantha Marx) / CC BY 2.0

82

RDM in the Humanities – Academic Alumni Forum



## Welcome

DMPonline helps you to create, review, and share data management plans that meet institutional and funder requirements. It is provided by the Digital Curation Centre (DCC).

Join the growing international community that have adopted DMPonline:



17,622 Users



203 Organisations



23,083 Plans



89 Countries

Some funders mandate the use of DMPonline, while others point to it as a useful option. You can [download funder templates](#) without logging in, but the tool provides tailored guidance and example answers from the DCC and many research organisations. Why not sign up for an account and try it out?

Sign in
Create account

**\* Email**

**\* Password**

Forgot password?

Remember email

Sign in

- or -

Sign in with your institutional credentials



## Research on DMPs

Project Details

Plan overview

Write Plan

Share

Download

**\* Project title**

Research on DMPs

 mock project for testing, practice, or educational purposes**Funder**

National Science Foundation (USA)

**Grant number**

e.g. 123456

**Project abstract****ID**

48553

### Select Guidance

To help you write your plan, DMPonline can show you guidance from a variety of organisations.

Select up to 6 organisations to see their guidance.

 Digital Curation Centre

Find guidance from additional organisations below

[See the full list](#)[Save](#)

... and for real?



# DMP Templates from Organizations and Funders

- your uni might have guidelines and templates
  - Universität Leipzig *Grundsätze für das Management von Forschungsdaten an der Universität Leipzig* (17. April 2019)  
[https://www.uni-leipzig.de/fileadmin/ul/Dokumente/190424\\_Grunds%C3%A4tze\\_Forschungsdatenmanagement.pdf](https://www.uni-leipzig.de/fileadmin/ul/Dokumente/190424_Grunds%C3%A4tze_Forschungsdatenmanagement.pdf)
- your funder might have guidelines and templates
  - DFG: *Umgang mit Forschungsdaten* (DFG-Leitlinien zum Umgang mit Forschungsdaten)  
[https://www.dfg.de/foerderung/antrag\\_gutachter\\_gremien/antragstellende/nachnutzung\\_forschungsdaten/](https://www.dfg.de/foerderung/antrag_gutachter_gremien/antragstellende/nachnutzung_forschungsdaten/)
  - for the US: [https://dmptool.org/public\\_templates/](https://dmptool.org/public_templates/)

# NEH/NSF

## NATIONAL SCIENCE FOUNDATION (NSF): NSF-SBE: SOCIAL, BEHAVIORAL, ECONOMIC SCIENCES

### ROLES AND RESPONSIBILITIES

The DMP should outline the rights and obligations of all parties to the roles and responsibilities in the management and retention of research data. It should also consider changes to roles and responsibilities that will occur, should a principal investigator or co-PI leave the institution or project. Any costs should be explained in the Budget Justification pages.

Guidance:

- [Data Management for SBE Proposals and Awards PDF](#)
- [NSF Proposal & Award Policies & Procedures Guide \(PAPPG\)](#)
- [NSF Frequently Asked Questions \(FAQs\) for Public Access](#)

### EXPECTED DATA

The DMP should describe the types of data, datasets, physical collections, software, digital instruments, and other materials to be produced in the course of the project. It should then describe the expected types of data to be retained.

Guidance:

- [Data Management for SBE Proposals and Awards PDF](#)
- [NSF Proposal & Award Policies & Procedures Guide \(PAPPG\)](#)
- [NSF Frequently Asked Questions \(FAQs\) for Public Access](#)

### PERIOD OF DATA RETENTION

SBE is committed to timely and rapid data destruction. However, it recognizes that types of data can vary widely and that acceptable norms vary by scientific discipline. It is strongly committed, however, to the underlying principle of timely access, and all costs should access how this will be met in the DMP statement.

Guidance:

- [Data Management for SBE Proposals and Awards PDF](#)
- [NSF Proposal & Award Policies & Procedures Guide \(PAPPG\)](#)
- [NSF Frequently Asked Questions \(FAQs\) for Public Access](#)

### DATA FORMAT AND DISSEMINATION

The DMP should describe data formats, media, and dissemination approaches that will be used to make data and materials available to others. Policies for public access and sharing should be described, including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements.

Research centers and those partnerships with industry or other user communities must also address how data are to be shared and merged with partners, center members, and other institutional stakeholders.

Guidance:

- [Data Management for SBE Proposals and Awards PDF](#)
- [NSF Proposal & Award Policies & Procedures Guide \(PAPPG\)](#)
- [NSF Frequently Asked Questions \(FAQs\) for Public Access](#)

### DATA STORAGE AND PRESERVATION OF ACCESS

The DMP should describe physical and cyber resources and facilities that will be used for the effective preservation and storage of research data. These can include third party facilities and repositories.

Guidance:

- [Data Management for SBE Proposals and Awards PDF](#)
- [NSF Proposal & Award Policies & Procedures Guide \(PAPPG\)](#)
- [NSF Frequently Asked Questions \(FAQs\) for Public Access](#)

### ADDITIONAL POSSIBLE DATA MANAGEMENT REQUIREMENTS

More stringent data management requirements may be specified in particular NSF solicitations or results from data policies established at the PI's home institution. Additional requirements will be specified in the program solicitation and award conditions. Principal Investigators to be supported by such programs must discuss how they will meet these additional requirements in their Data Management Plans.

Guidance:

- [Data Management for SBE Proposals and Awards PDF](#)
- [NSF Proposal & Award Policies & Procedures Guide \(PAPPG\)](#)
- [NSF Frequently Asked Questions \(FAQs\) for Public Access](#)

# NEH ODH (Office of Digital Humanities) NSF SBE (Social, Behavioral, Economics)

1. Roles and Responsibilities
2. Expected Data
3. Period of Data Retention
4. Data Format and Dissemination
5. Data Storage and Preservation of Access
6. Additional Possible Data Management Requirements (NSF SBE)





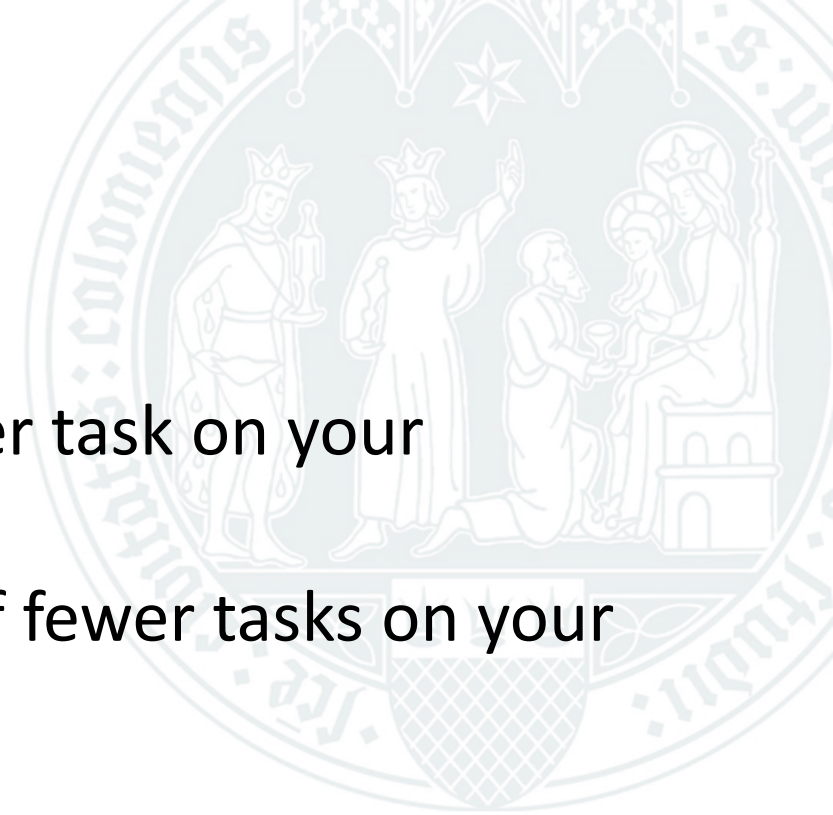
# Take home message

Using a data management plan now means another task on your agenda

Using a data management plan now means a lot of fewer tasks on your agenda later

Oh and beside that:

→ Make Data **F**indable, **A**ccessible, **I**nteroperable and **R**eusable!



# Metadata



# FAIR principles

## To be Findable:

- (meta)data are assigned to globally unique and eternally persistent identifier.
- Data are described with rich metadata.
- (meta)data are registered or indexed in a searchable resource
- Metadata specify the data identifier.

## To be Accessible:

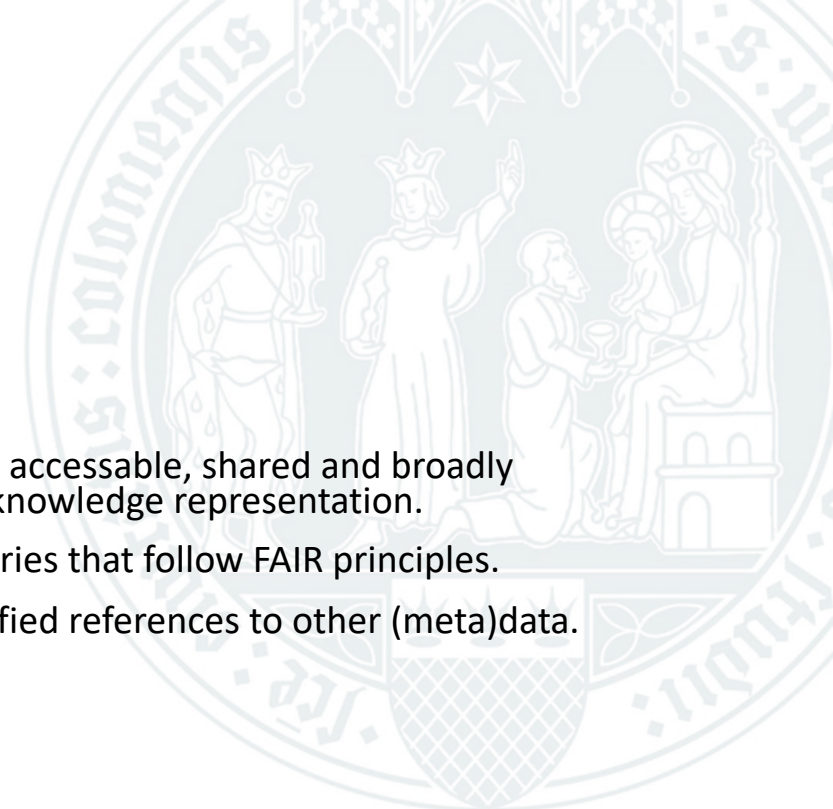
- (meta)data are retrievable by their identifier using a standardized communications protocol.
- The protocol is open, free and universally implementable.
- The protocol allows for an authentication and authorization procedure, where necessary.
- Metadata are accessible, even when the data are no longer available.

## To be Interoperable:

- (meta)data use a formal, accessible, shared and broadly applicable language for knowledge representation.
- (meta)data use vocabularies that follow FAIR principles.
- (meta)data include qualified references to other (meta)data.

## To be Re-usable:

- (meta)data have a plurality of accurate and relevant attributes.
- (meta)data are released with a clear and accessible data usage license.
- (meta)data are associated with their provenance.
- (meta)data meet domain-relevant community standards.



# FAIR principles

## To be Findable:

- **(meta)data** are assigned to globally unique and eternally persistent identifier.
- Data are described with rich **metadata**.
- **(meta)data** are registered or indexed in a searchable resource
- **metadata** specify the data identifier.

## To be Accessible:

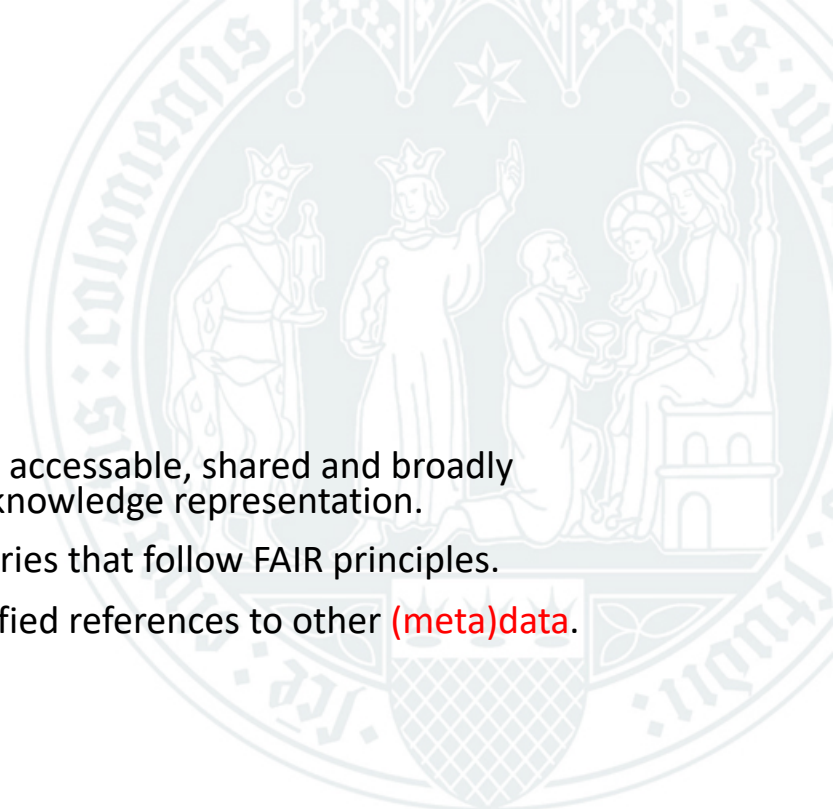
- **(meta)data** are retrievable by their identifier using a standardized communications protocol.
- The protocol is open, free and universally implementable.
- The protocol allows for an authentication and authorization procedure, where necessary.
- **Metadata** are accessible, even when the data are no longer available.

## To be Interoperable:

- **(meta)data** use a formal, accessible, shared and broadly applicable language for knowledge representation.
- **(meta)data** use vocabularies that follow FAIR principles.
- **(meta)data** include qualified references to other **(meta)data**.

## To be Re-usable:

- **(meta)data** have a plurality of accurate and relevant attributes.
- **(meta)data** are released with a clear and accessible data usage license.
- **(meta)data** are associated with their provenance.
- **(meta)data** meet domain-relevant community standards.



# What is Metadata?

- “Information about information“  
(W3C Generic Resources and Web Metadata)
- “Data about Data“  
(CESSDA Training)
- “Data that provides information about other data“  
(Merriam-Webster)
- “Data that describes and gives information about other data.“  
(Oxford English Dictionary)



# Metadata is a love letter to the future!



1-48 of 214 results for Clothing, Shoes & Jewelry : Men : Clothing : Reds : "nike" Sort by: Featured

- Department**
  - < Any Department
  - < Clothing, Shoes & Jewelry
  - < Men
  - Clothing**
    - Shirts
    - Jackets & Coats
    - Active
    - Swim
    - Underwear
    - Socks
- Avg. Customer Review**
  - ★★★★★ & Up
  - ★★★★☆ & Up
  - ★★★☆☆ & Up
  - ★★☆☆☆ & Up
- Amazon Fashion**
  - Top Brands
- Brand**
  - Nike
  - Nike Golf
- Price**
  - Under \$25
  - \$25 to \$50
  - \$50 to \$100
  - \$100 to \$200

\$ Min \$ Max Go
- New Arrivals**
  - Last 30 days
  - Last 90 days
- Special Sizes**
  - Big & Tall
- Color**
  - < Clear
  - Color selection icons

Price and other details may vary based on size and color



NIKE Men's Dri-FIT Cotton 2.0 Tee ★★★★★ < 416



Nike Men's NSW Club Pant Open Hem ★★★★★ < 17



Nike Men's NSW Club Jogger ★★★★★ < 26



Nike Dri-FIT Men's Training T-Shirt ★★★★★ < 57



NIKE Men's Pull Over Hoodie ★★★★★ < 50



Nike Men's Sportswear Club Shorts ★★★★★ < 551



NIKE Mens Legend Short Sleeve Tee ★★★★★ < 1,021 \$26<sup>99</sup> Ships to Germany



Nike Men's NSW Club Crew ★★★★★ < 11



Nike Men's Dry Tee ★★★★★ < 60



Nike Men's Legend Dri-Fit Shirt ★★★★★ < 261 \$19<sup>73</sup> \$22.90 Ships to Germany



Nike Sportswear Men's T-Shirt ★★★★★ < 22

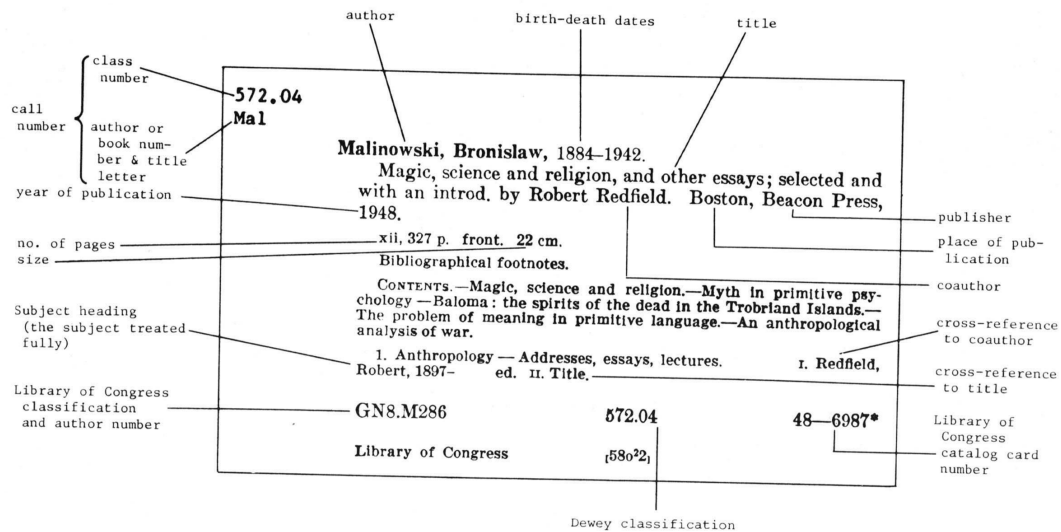


Nike Men's 7" Challenger Short ★★★★★ < 113

Best Seller







<b>Personal name:</b>	Malinowski, Bronislaw, 1884-1942.
<b>Title:</b>	Magic, science and religion and other essays
<b>Published/Created:</b>	Boston, Beacon Press, 1948.
<b>LCCN Permalink:</b>	<a href="https://lcn.loc.gov/48006987">https://lcn.loc.gov/48006987</a>
<b>Description:</b>	xii, 327 p. front. 22 cm.
<b>LC classification (full):</b>	GN8 .M286
<b>LC classification (partial):</b>	GN8
<b>Related names:</b>	Redfield, Robert, 1897-ed.
<b>Contents:</b>	Magic, science and religion.—Myth in primitive psychology.—Baloma: the spirits of the dead in the Trobriand Islands.—The problem of meaning in primitive language.—An anthropological analysis of war.
<b>Subjects:</b>	Anthropology.
<b>Notes:</b>	Bibliographic footnotes.
<b>LCCN:</b>	48006987
<b>Dewey class no.</b>	572.04
<b>Type of material:</b>	Book

Remember ...

# FAIR

**Findable**, Accessible, Interoperable, and Reusable



November 30, 2018

Lesson Open Access

# Arts and Humanities Research Council Data Management Plan Rubric

Donaldson, Mary; Higman, Rosie

This rubric is designed as a checklist or marking aid for those reviewing data management plans for submission to the Arts and Humanities Research Council (AHRC). The Data Management Plan should outline the project's approach to managing data. It is mandatory to include for all Leadership Fellows, Research Grants and Follow on Funding applications but is not required for Research Networking.

It is an updated version of the original rubric created by Sarah Jones, Digital Curation Centre, to reflect the changes in the data management requirements of the AHRC:

<http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/AHRC%20DMP%20Compliance%20rubric.pdf>

Preview

Page: 1 of 6 Automatic Zoom

Arts and Humanities Research Council Data Management Plan Rubric

Authors:  
Rosie Higman, [rosie.higman@manchester.ac.uk](mailto:rosie.higman@manchester.ac.uk), University of Manchester  
Mary Donaldson, [mary.donaldson@glasgow.ac.uk](mailto:mary.donaldson@glasgow.ac.uk), University of Glasgow

Purpose:  
This rubric is designed as a checklist or marking aid for those reviewing data management plans for submission to the Arts and Humanities Research Council (AHRC). The Data Management Plan should outline the project's approach to managing data. It is mandatory to include for all Leadership Fellows, Research Grants and Follow on Funding applications but is not required for Research Networking.

Files (663.3 kB)

Name	Size
Arts and Humanities Research Council Data Management Plan Rubric.pdf	663.3 kB

md5:61668c889b7e9800a94e2b87e966d5dc

Citations

Show only:  Literature (0)  Dataset (0)  Software (0)  Unknown (0)  Citations to this version

No citations.

857 views 589 downloads

See more details...

9

Tweeted by 11

See more details

Indexed in

OpenAIRE

Publication date: November 30, 2018

DOI: [10.5281/zenodo.1745533](https://doi.org/10.5281/zenodo.1745533)

Keyword(s): research data management, humanities, arts, data management plans

Communities: Research data management (RDM) open training materials, Research Data Network

License (for files): Creative Commons Attribution 4.0 International

Versions

Version 1 10.5281/zenodo.1745533

Cite all versions: Cite all versions by using the DOI 10.5281/zenodo.1745532. This DOI represents all versions and will always resolve to the latest one. Read more.

Share

AddThis

Cite as

Donaldson, Mary, & Higman, Rosie. (2018, November). Arts and Humanities Research Council Data Management Plan Rubric. Zenodo. <http://doi.org/10.5281/zenodo.1745533>

title

authors

description

publication date

persistent identifier

keywords

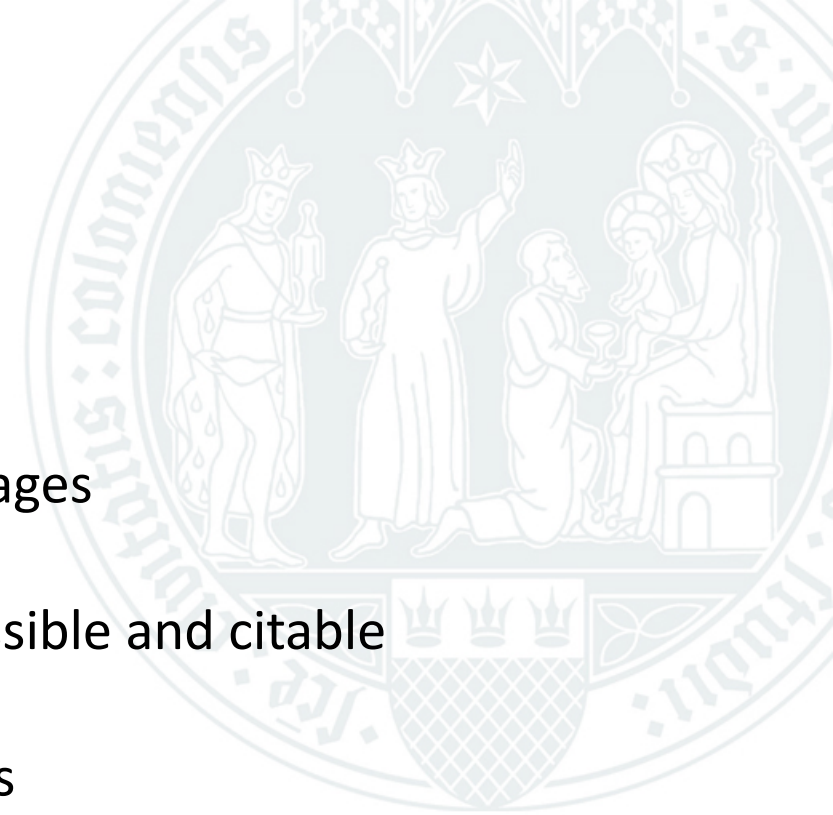
license

version

file size

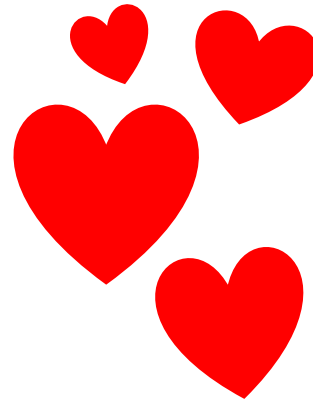
# Metadata formats

- Dublin Core
  - minimalistic standard for digital documents such as web pages
- DataCite
  - compact format with the goal to make research data accessible and citable
- LIDO
  - metadata standard used for describing museum collections
- CMDI
  - metadata framework used in the European research infrastructure CLARIN
- Schema.org
  - a schema for structured data on the web (e.g. used by Google Dataset Search)



10

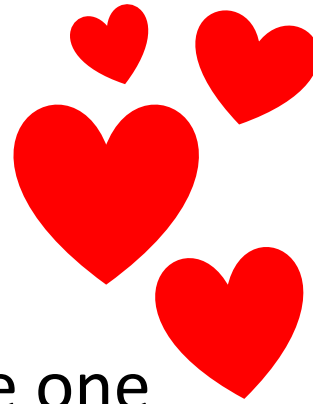
# Speed metadating



- model yourself (15 min)
  - describe yourself as a person with 15 to 20 self defined elements
  - the resulting description should be a model that describes you, unambiguously
- Result presentation:
  - present yourself to the group
  - what was your strategy finding/defining elements? What problems did you encounter?
  - Can you spot similar elements in the descriptions of the other participants? If yes, note them down.

10

# schema.org



compare the model you built with the one from <https://schema.org/Person>

- How close are your description and schema.org?
- Did you find elements that you didn't think of and that would have been useful.



# Take home message

Without metadata the life as we know it would not be possible:

How would you buy a red pullover online? Impossible!

aka

How would you develop research questions and do research? Nearly impossible!

Describe your data with the same love you describe your research in a journal article

And do not forget:

→ Make Data **F**indable, **A**ccessible, **I**nteroperable and **R**eusable!

# Repositories





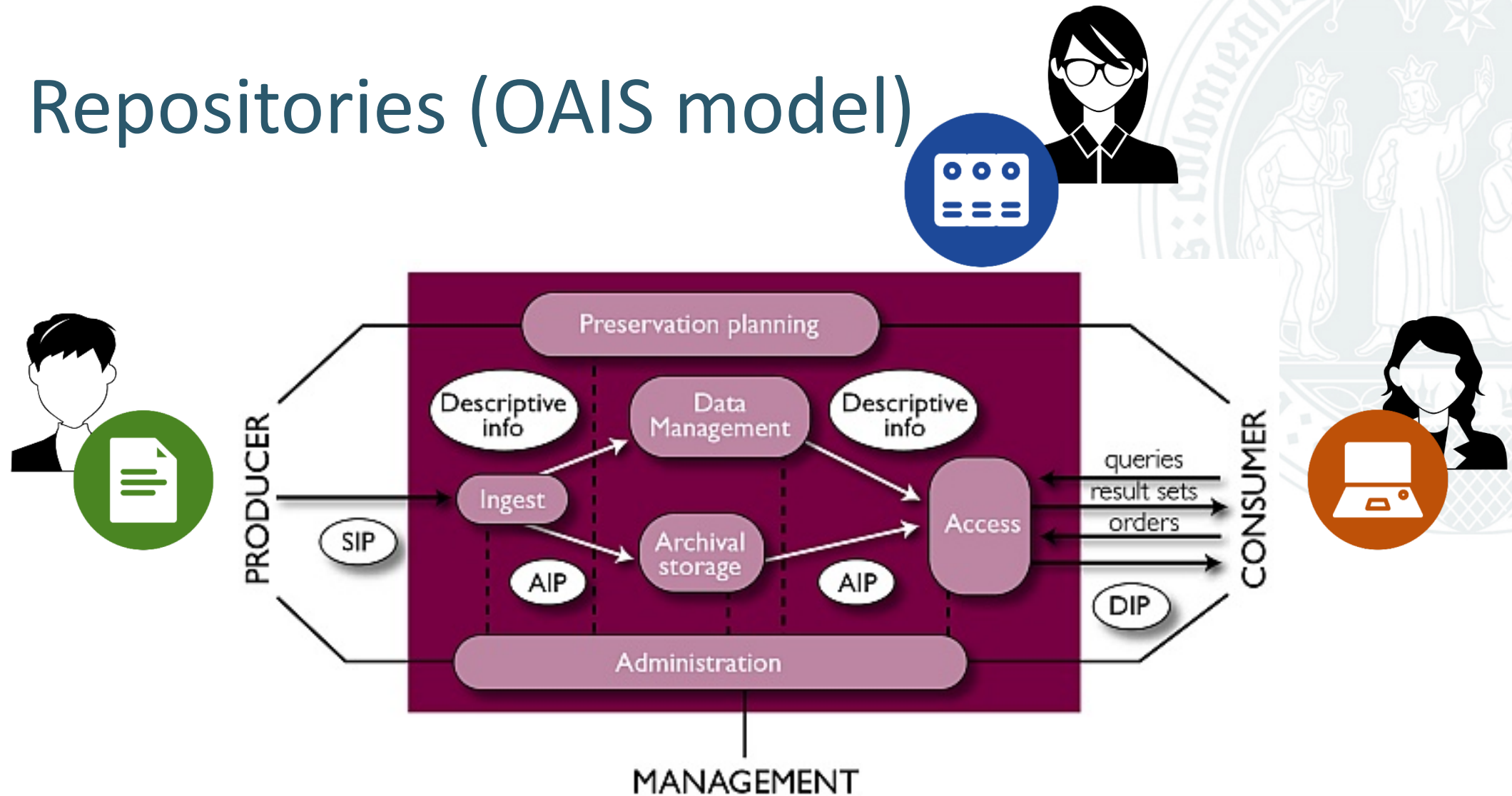
# A definition

- Definition: Repository (aka Data Repository or Digital Data Repository) is a **searchable and queryable** interfacing entity that is able to **store, manage, maintain and curate Data/Digital Objects**. A data repository provides a service for human and machine to make data discoverable/searchable through collection(s) of metadata.
- Explanation: A data repository returns data sets with appropriate features and/or the bit stream/dynamic data objects instantiating a data/digital object if a persistent identifier is being issued. A repository should have a globally unique identifier that refers to it and an URL allowing access to the repository.  
Repositories store data and can also store its associated metadata. Some repositories may be specialized to store metadata. New Collections (aggregations) are, or can be, built from repository data for analysis purposes. New PIDs are required for such collections.

Research Data Alliance (RDA) Term Definition Tool

<https://smw-rda.esc.rzg.mpg.de/index.php/repository>

# Repositories (OAIS model)



# OAIS model

Open Archival Information System

## INGEST

- acceptance of a “submission information package (SIP)” → data with metadata
- verification of archival suitability, completeness, integrity
- extraction of descriptive information (metadata)
- generation of the “archival information package (AIP)”

## STORAGE

- physical reception of AIPs (backup, redundant storage, integrity check, ...)

## MANAGEMENT

- curation of metadata, organisation of access



# OAIS model

Open Archival Information System

## **ACCESS**

- acceptance of requests
- dependent on access conditions
- creation of a “dissemination information package (DIP)”

## **PRESERVATION PLANNING**

- management of (technical) change
- preservation of the AIPs through migration and emulation
- preservation of integrity

## **ADMINISTRATION**

- technical administration (configuration, access rights, interfaces)



# FAIR Principles

## To be Findable:

- (meta)data are assigned to globally unique and eternally persistent identifier.
- data are described with rich metadata.
- (meta)data are registered or indexed in a searchable resource
- metadata specify the data identifier.

## To be Accessible:

- (meta)data are retrievable by their identifier using a standardized communications protocol.
- the protocol is open, free and universally implementable.
- the protocol allows for an authentication and authorization procedure, where necessary.
- metadata are accessible, even when the data are no longer available.

## To be Interoperable:

- (meta)data use a formal, accessible, shared and broadly applicable language for knowledge representation.
- (meta)data use vocabularies that follow FAIR principles.
- (meta)data include qualified references to other (meta)data.

## To be Re-usable:

- (meta)data have a plurality of accurate and relevant attributes.
- (meta)data are released with a clear and accessible data usage license.
- (meta)data are associated with their provenance.
- (meta)data meet domain-relevant community standards.

# FAIR Principles

## To be Findable:

- (meta)data are assigned to globally unique and eternally persistent identifier.
- data are described with rich metadata.
- (meta)data are registered or indexed in a searchable resource
- metadata specify the data identifier.

## To be Accessible:

- (meta)data are retrievable by their identifier using a standardized communications protocol.
- the protocol is open, free and universally implementable.
- the protocol allows for an authentication and authorization procedure, where necessary.
- metadata are accessible, even when the data are no longer available.

## To be Interoperable:

- (meta)data use a formal, accessible, shared and broadly applicable language for knowledge representation.
- (meta)data use vocabularies that follow FAIR principles.
- (meta)data include qualified references to other (meta)data.

## To be Re-usable:

- (meta)data have a plurality of accurate and relevant attributes.
- (meta)data are released with a clear and accessible data usage license.
- (meta)data are associated with their provenance.
- (meta)data meet domain-relevant community standards.

# Requirement FAIR (1)

## Findable:

**F1. (meta)data are assigned a globally unique and eternally persistent identifier.**

Repositories offer lasting discoverability of research data and other digital objects with the help of persistent identifiers

- Handle
- DOI
- URN-NBN

Prefix                      Suffix

doi:10.2314/529827735

# Intermission PIDs

doi:10.2314/529827735



<https://doi.org/10.2314/529827735>



<https://a-repository.org/path/to/file>

<https://new-repository.com/new-path/to/old-file>





# Requirements FAIR (2)

## Findable

**F3. (meta)data are registered or indexed in a searchable resource.**

Repositories have a search page and distribute metadata via OAI-PMH metadata harvesting (and other interfaces)



November 30, 2018

Lesson Open Access

# Arts and Humanities Research Council Data Management Plan Rubric

Donaldson, Mary; Higman, Rosie

This rubric is designed as a checklist or marking aid for those reviewing data management plans for submission to the Arts and Humanities Research Council (AHRC). The Data Management Plan should outline the project's approach to managing data. It is mandatory to include for all Leadership Fellows, Research Grants and Follow on Funding applications but is not required for Research Networking.

It is an updated version of the original rubric created by Sarah Jones, Digital Curation Centre, to reflect the changes in the data management requirements of the AHRC:

<http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/AHRC%20DMP%20Compliance%20rubric.pdf>

Preview

Page: 1 of 6 Automatic Zoom

Arts and Humanities Research Council Data Management Plan Rubric

Authors:  
Rosie Higman, [rosie.higman@manchester.ac.uk](mailto:rosie.higman@manchester.ac.uk), University of Manchester  
Mary Donaldson, [mary.donaldson@glasgow.ac.uk](mailto:mary.donaldson@glasgow.ac.uk), University of Glasgow

Purpose:  
This rubric is designed as a checklist or marking aid for those reviewing data management plans for submission to the Arts and Humanities Research Council (AHRC). The Data Management Plan should outline the project's approach to managing data. It is mandatory to include for all Leadership Fellows, Research Grants and Follow on Funding applications but is not required for Research Networking.

Files (663.3 kB)

Name	Size
Arts and Humanities Research Council Data Management Plan Rubric.pdf	663.3 kB

md5:61668c889b7e9800a94e2b87e966d5dc

Citations

Show only:  Literature (0)  Dataset (0)  Software (0)  Unknown (0)  Citations to this version

No citations.

857 views 589 downloads

See more details...

9

Tweeted by 11

See more details

Indexed in

OpenAIRE

Publication date: November 30, 2018

DOI: [10.5281/zenodo.1745533](https://doi.org/10.5281/zenodo.1745533)

Keyword(s): research data management, humanities, arts, data management plans

Communities: Research data management (RDM) open training materials, Research Data Network

License (for files): Creative Commons Attribution 4.0 International

Versions

Version 1 10.5281/zenodo.1745533

Cite all versions: Cite all versions by using the DOI 10.5281/zenodo.1745532. This DOI represents all versions and will always resolve to the latest one. Read more.

Share

AddThis

Cite as

Donaldson, Mary, & Higman, Rosie. (2018, November). Arts and Humanities Research Council Data Management Plan Rubric. Zenodo. <http://doi.org/10.5281/zenodo.1745533>

title  
authors  
description

publication date  
persistent identifier  
keywords  
license

version  
file size

# Requirements FAIR (3)

## Accessible

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

- There are mechanisms that make sure that sensitive data can only be accessed by authorised persons
- The repository implements federated AAI, Shibboleth

https://idp.rrz.uni-koeln.de/idp/profile/SAML2/Redirect/SSO:jsessionid=0859D95085991

Universität zu Köln  
Shibboleth

Anmelden bei Gigamove - RWTH Aachen

Benutzername  
jblumtri

Passwort  
••••••••

Anmeldung nicht speichern

Lösche die frühere Einwilligung zur Weitergabe Ihrer Informationen an diesen Dienst.

Anmeldung

Webbasierter Dienst um größere Dateien im Internet auszutauschen. Verwendete Attribute: - eduPersonPrincipalName: für die Zuordnung von

- Studierende (einschließlich Gasthörer) loggen sich mit Ihrem Studierenden-Account des Rechenzentrums ein
- Lehrende und Beschäftigte der UzK loggen sich mit Ihrem Personal-Account des Rechenzentrums ein
- In beiden oben genannten Fällen ist keine gesonderte Registrierung notwendig! Bei Account-Problemen hilft das Rechenzentrum.

> Passwort vergessen?

> Hilfe benötigt?

# Requirements FAIR (4)

## Interoperable

### 13. (meta)data include qualified references to other (meta)data.

- Information is linked to other resources and norm data
- e.g. all persons are identified by an ORCID, ISNI or GND and are via these identifiers connected to other resources



# Requirements FAIR (5)

## Reusable

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

- Metadata and the landing page of the data specify clear access conditions  
→ In particular the repositor offers contracts, licences, and codes of conducts



# Find a repository...

**exercise**

I am scientist and comparing religion. I finished my project in which I collected contemporary data. Now I want to

- archive my data, but nobody should have access to it
- instead my metadata should be findable in an persistent way

---

I am scientist in media studies and I wrote three outstanding and groundbreaking papers, that should be published digitally. I want to

- make my papers accessible to my research community
- publish my papers under an open licence

# Find a repository



I have a lot of audiodata that documents the language of indigenous people. It is the only record of the language, so the audiodata needs to be archived as cultural heritage.

- I am using the CMDI-Metadata standard for describing mydata
  - I have a (real) informed consent and I am allowed to make the data accessible via the internet
- 

I did excavations in Africa and now I have digital pictures of shards. In addition, I have several .csv-sheets, that describe the shards. A metadata-scheme is existing as well.

- All data should be published together (at one place)
- The data should be accessible for my research community directly

# Find a repository...



I have a lot of audio data that document the language of indigenous people. It is the only record of this endangered language, so the audio data need to be archived as cultural heritage.

- I am using the CMDI metadata standard for describing my data
- I have (real) informed consent and I am allowed to make the data accessible via the internet

---

I did excavations in Africa and now I have digital pictures of pottery shards. In addition, I have several csv-sheets, that describe these shards. A metadata scheme is existing as well.

- All data should be published together (in one place)
- The data should be accessible for my research community directly



# Find a repository...

- Comparative religious studies: Contemporary data
- Media studies: Outstanding and groundbreaking paper
- Linguistics: Archiving audio data
- Archaeology: Pictures of and csv-sheets about shards

→ check out [re3data.org](https://re3data.org) for the latest trendy repositories of the season





# Take home message

Storing your data is easier than you thought...

...and finding a domain-specific repository that fits can be more difficult than we thought.

But in the end of space and time, there is always zenodo.org!

Oh and one last time:

→ Make Data **F**indable, **A**ccessible, **I**nteroperable and **R**eusable!

# Thank you for your attention

...now go home and make your data FAIR!

[doi:10.5281/zenodo.3556389](https://doi.org/10.5281/zenodo.3556389)