Data Archiving and Networked Services

DANS

Simplifying research data management by using “Domain Data Protocols”

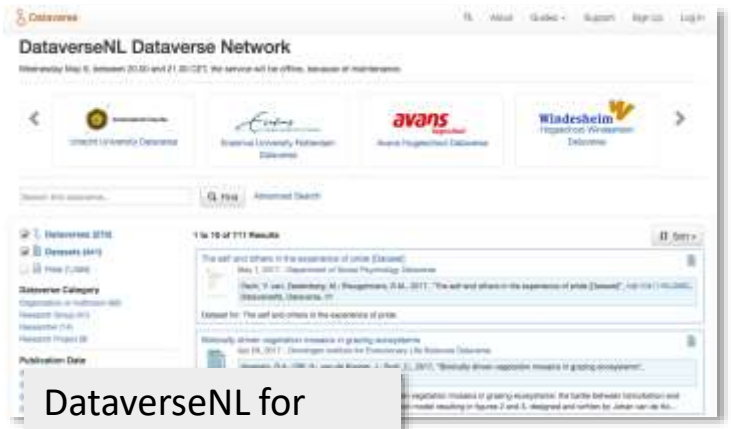
René van Horik

Workshop TAIL Project / November 22 2019 / Aveiro

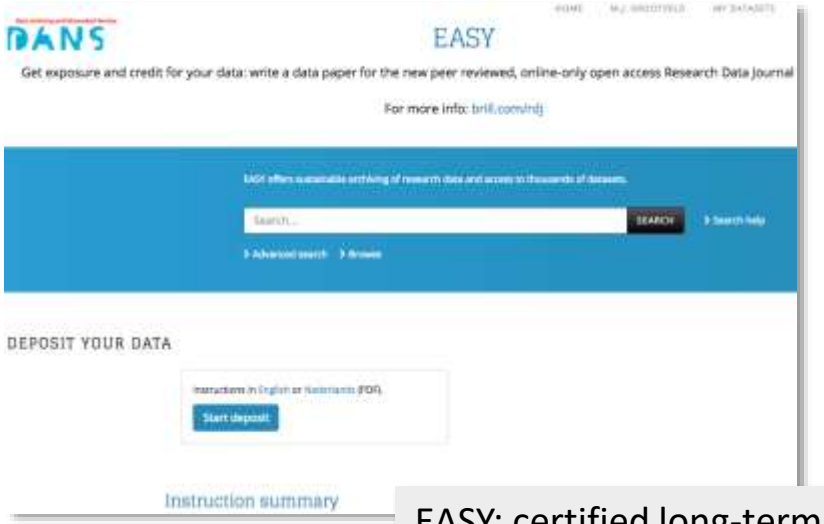
DANS is about *keeping* data FAIR

https://dans.knaw.nl

DANS services



DataverseNL for short- and mid-term data storage



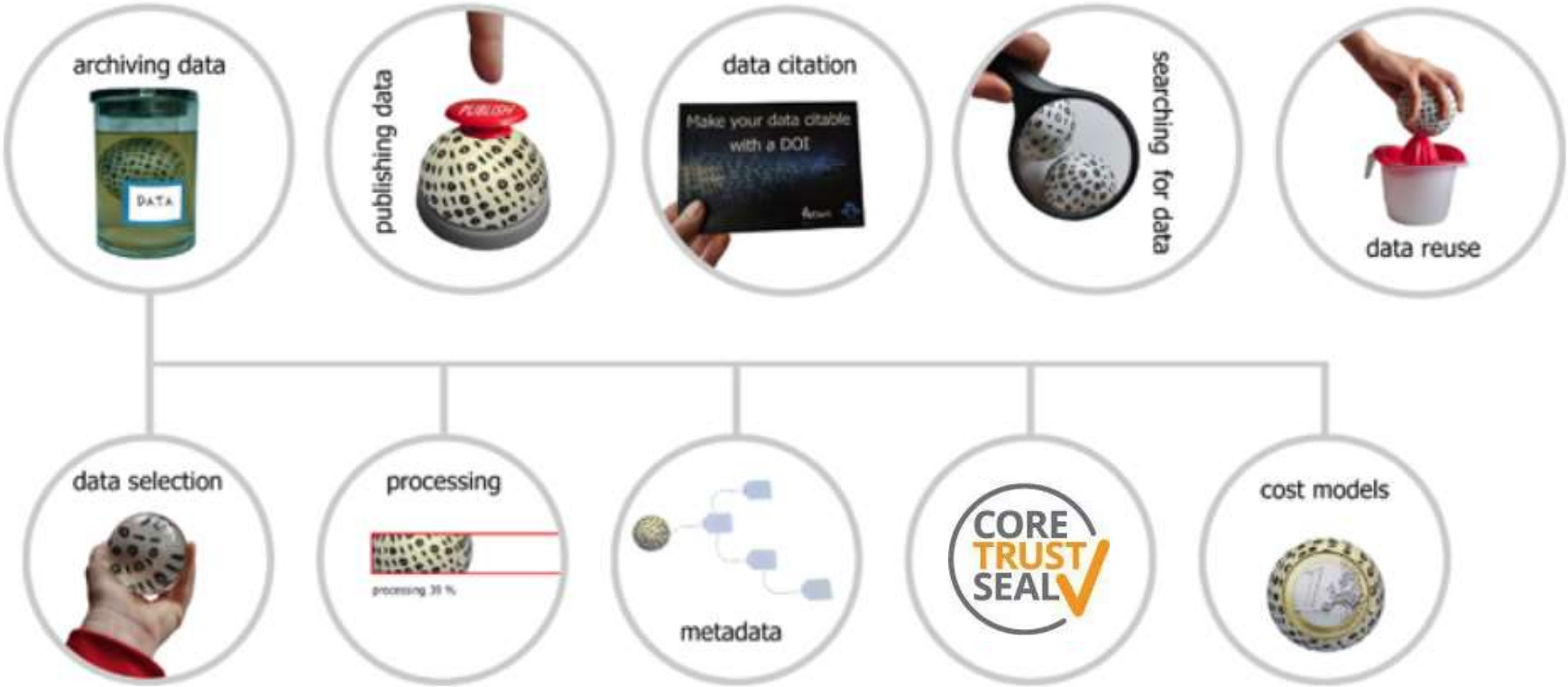
EASY: certified long-term Electronic Archiving System for self-deposit

NARCIS: Gateway to scholarly information in the Netherlands

Centre of expertise

Essentials 4 Data Support is an introductory course for those people who (want to) support researchers in storing, managing, archiving and sharing their research data.

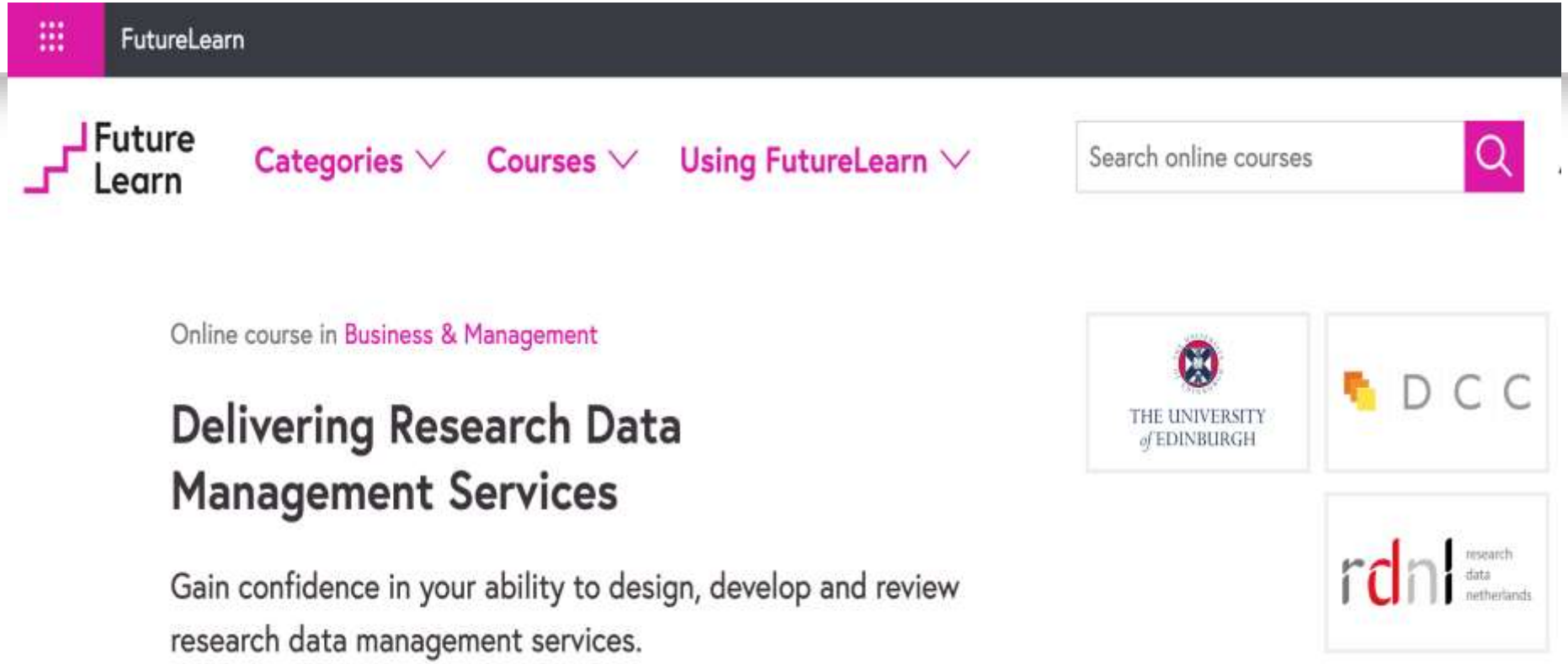
rdnl research data netherlands



Learning objectives

https://datasupport.researchdata.nl/

RDNL = DANS, 4TU.ResearchData & SURFsara



https://www.futurelearn.com/courses/delivering-research-data-management-services

EOSC-hub

EOSC-hub brings together multiple service providers to create the Hub: a single contact point for European researchers and innovators to discover, access, use and reuse a broad spectrum of resources for advanced data-driven research

https://www.eosc-hub.eu/

Advanced Computing for Research

EUDAT Collaborative Data Infrastructure

INDIGO - DataCloud
Better Software for Better Science

EOSC-hub receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 777536.

EOSC-hub and DMP

- TASK 11.2 -> Data Management Planning
  - Advocacy
  - Guidance
  - Support

Practical challenges for researchers in data sharing

Springer Nature have published the results of a survey of >7,700 researchers worldwide, looking at data sharing during publication

Importance of data discoverability

76%

of respondents highly rate the importance of their data being discoverable: **most popular ranking was 10/10**

Main challenge to data sharing is **organising data in a presentable and useful way**

Almost half of all respondents **(46%)** said that **organising data** was a challenge, followed by **confusion around copyright (37%)** and **not knowing where to share data (33%)**

Majority of researchers share their data in some way

63% of respondents stated that they generally **submitted data files** as **supplementary information,** deposited the files in a **repository,** or both

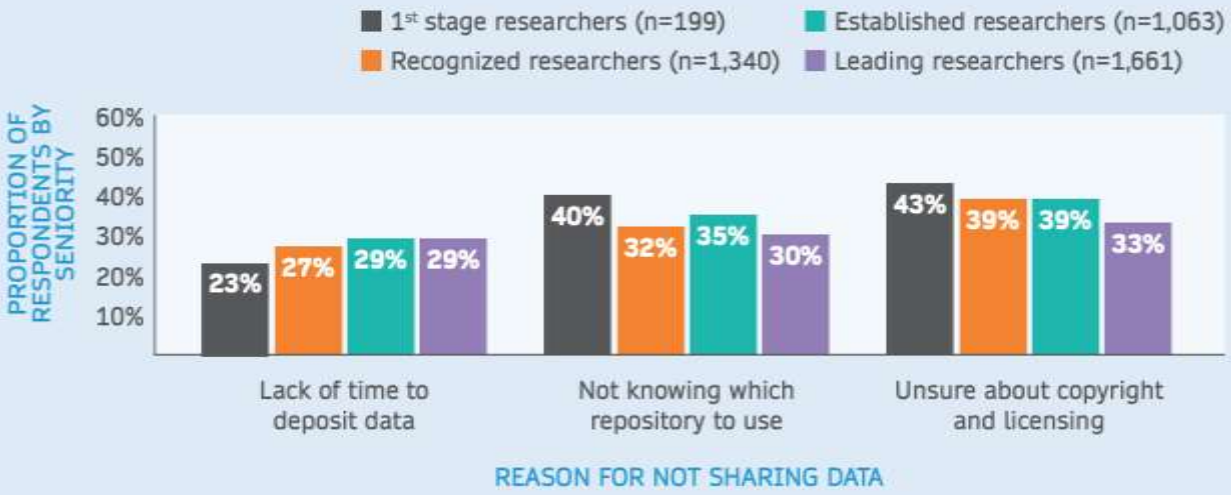
Source: Astell, Mathias; Admin, Springer Nature (2018): Infographic - Practical challenges for researchers in data sharing. Figshare. Journal contribution. https://doi.org/10.6084/m9.figshare.5996786.v4

Practical challenges for researchers in data sharing

Springer Nature have published the results of a survey of >7,700 researchers worldwide, looking at data sharing during publication

Lack of time VS. lack of knowledge

Time becomes more of an issue and knowledge less of an issue as **researchers become more senior**



What can be done to increase data sharing?

Improving **education and support** on good data management, particularly at early stages of researchers' careers

Creating **faster, easier routes for sharing data** - making data easily accessible and usable by researchers

Source: Astell, Mathias; Admin, Springer Nature (2018): Infographic - Practical challenges for researchers in data sharing. Figshare. Journal contribution. https://doi.org/10.6084/m9.figshare.5996786.v4

Data Management Plan

A DMP is a brief plan to define:

- how the data will be created
- how it will be documented
- who can access it
- where it will be stored
- whether it will be shared
- where it will be preserved and published

DMPs are sometimes submitted as part of grant applications, sometimes afterwards, but they are useful whenever researchers are creating data.

Research Data Management

- Increasingly demanded by research funders, universities and other academic organizations in the context of Open Science
- Many researchers view it as yet another bureaucratic hurdle
- Varying requirements by different organizations drive researchers into despair
- Nevertheless, good practices in data management make sense:
  - Precaution against data fraud and sloppiness
  - Makes research (data collection) process more transparent
  - Makes research data more easy to share

Research Data Management: Chaos or Harmony?

Who requires RDM?

- funders: national and international, public and private
- research infrastructures: national and international (e.g. ESFRI)
- universities, RPOs
- journals (DAP)

What is required?

- which/how many criteria?
- FAIR principles?
- retention period?
- during/after research?
- data sharing?
- eligible for funding?
- including software?
- recommended repository/datacentre/archive?

How do they require it?

- via project proposal
- during evaluations
- via code of conduct
- via DMP tool

How detailed?

- data management plan
- data paragraph
- once or periodic updates
- template: narrative/checkboxes

How strict?

- obligatory
- (strong) recommendation
- advice

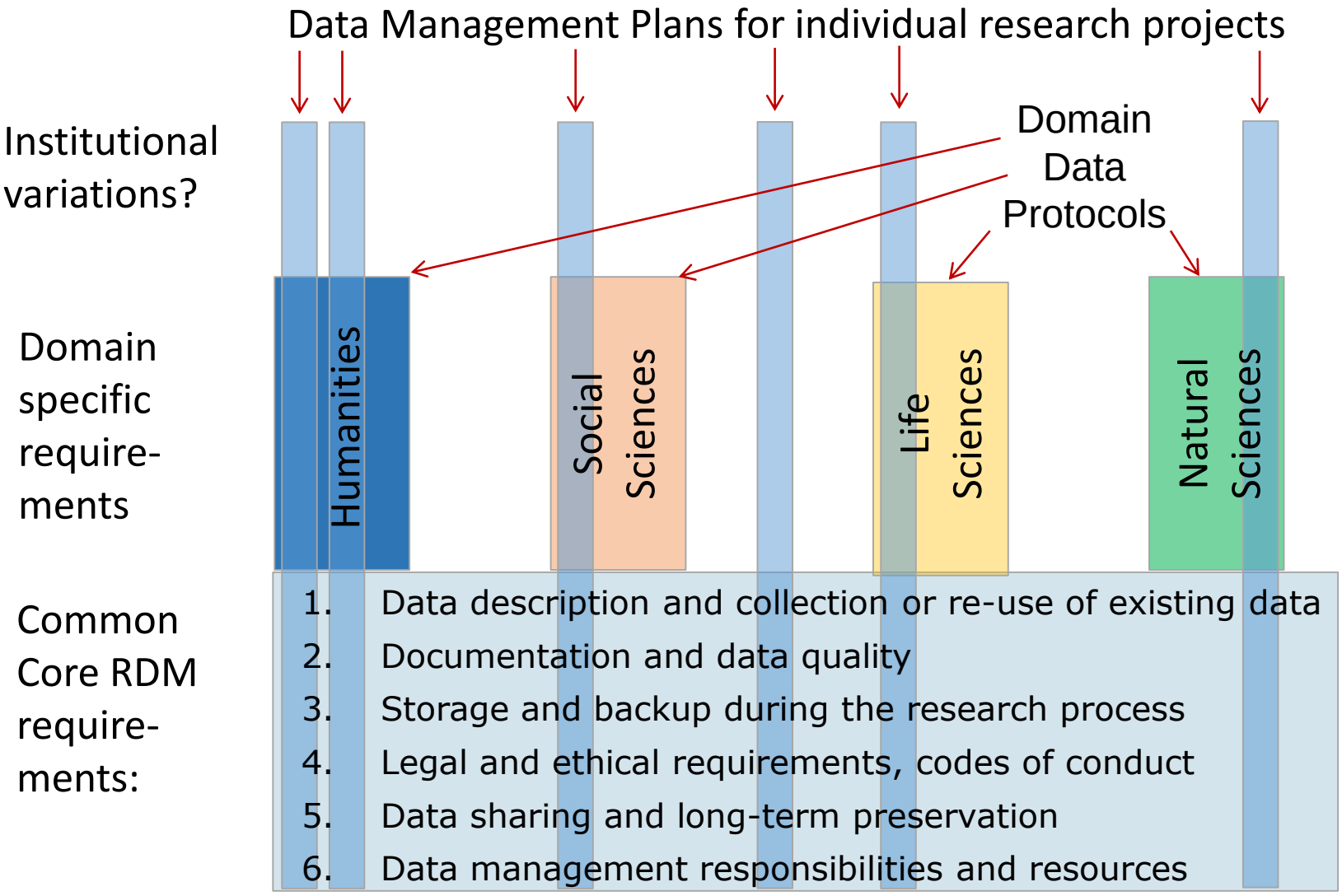
Summary of RDM situation

- Agreement on the overall policy aims of data management
- Many details differ at the level of the execution

It makes sense to:

- Distinguish between core DMP requirements on which we can all agree, and domain-specific requirements
- Minimize all other (institutional, national) variation

Domain Data Protocols



Our resources

18.01.2018

Guidance Document Presenting a Framework for Discipline-specific Research Data Management

Research organisations and funders increasingly ask researchers to create Data Management Plans for their work and proposals. A lack of standardisation means that these can be time-consuming to create and difficult to compare and evaluate. Science Europe presents a framework for the creation of domain-specific protocols that can be used as standardised templates, reducing the administrative burden on both researchers, research organisations, and funders.

https://www.scienceeurope.org/our-resources/guidance-document-presenting-a-framework-for-discipline-specific-research-data-management

Our resources

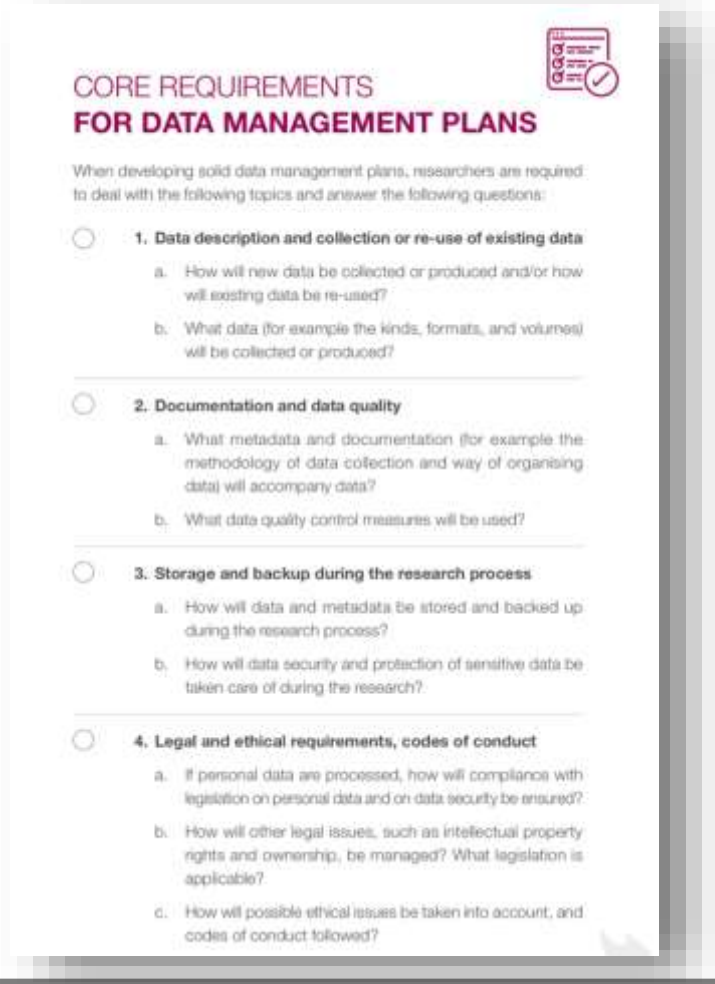
01.01.2019

Practical Guide to the International Alignment of Research Data Management

Developed by experts from Science Europe Member Organisations, this guide aims to align research data management (RDM) requirements across various research organisations. Science Europe will work to promote these requirements in order to ensure they are accepted by as many stakeholders as possible.

https://www.scienceeurope.org/our-resources/guidance-document-presenting-a-framework-for-discipline-specific-research-data-management

Aligned requirements and simplified Data Management Plan process



CORE REQUIREMENTS
FOR DATA MANAGEMENT PLANS

When developing solid data management plans, researchers are required to deal with the following topics and answer the following questions:

1. **Data description and collection or re-use of existing data**
   a. How will new data be collected or produced and/or how will existing data be re-used?
   b. What data (for example the kinds, formats, and volumes) will be collected or produced?
2. **Documentation and data quality**
   a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany data?
   b. What data quality control measures will be used?
3. **Storage and backup during the research process**
   a. How will data and metadata be stored and backed up during the research process?
   b. How will data security and protection of sensitive data be taken care of during the research?
4. **Legal and ethical requirements, codes of conduct**
   a. If personal data are processed, how will compliance with legislation on personal data and on data security be ensured?
   b. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?
   c. How will possible ethical issues be taken into account, and codes of conduct followed?

Domain Data Protocols

- to be formulated/accepted by research communities
- to be endorsed by research funders
- principle: comply or explain
- reduces need for individual data management plans
- simplifies evaluation of DMPs by funders

CORE REQUIREMENTS
FOR DATA MANAGEMENT PLANS

When developing solid data management plans, researchers are required to deal with the following topics and answer the following questions:

1. **Data description and collection or re-use of existing data**
   a. How will new data be collected or produced and/or how will existing data be re-used?
   b. What data (for example the kinds, formats, and volumes) will be collected or produced?

2. **Documentation and data quality**
   a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany data?
   b. What data quality control measures will be used?

3. **Storage and backup during the research process**
   a. How will data and metadata be stored and backed up during the research process?
   b. How will data security and protection of sensitive data be taken care of during the research?

4. **Legal and ethical requirements, codes of conduct**
   a. If personal data are processed, how will compliance with legislation on personal data and on data security be ensured?
   b. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?
   c. How will possible ethical issues be taken into account, and codes of conduct followed?

5. **Data sharing and long-term preservation**
   a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?
   b. How will data for preservation be selected, and where will data be preserved long-term (for example a data repository or archive)?
   c. What methods or software tools will be needed to access and use the data?
   d. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?

6. **Data management responsibilities and resources**
   a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?
   b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

Example on next slide

example

Topic

Requirement (question)

3 STORAGE AND BACKUP DURING THE RESEARCH PROCESS

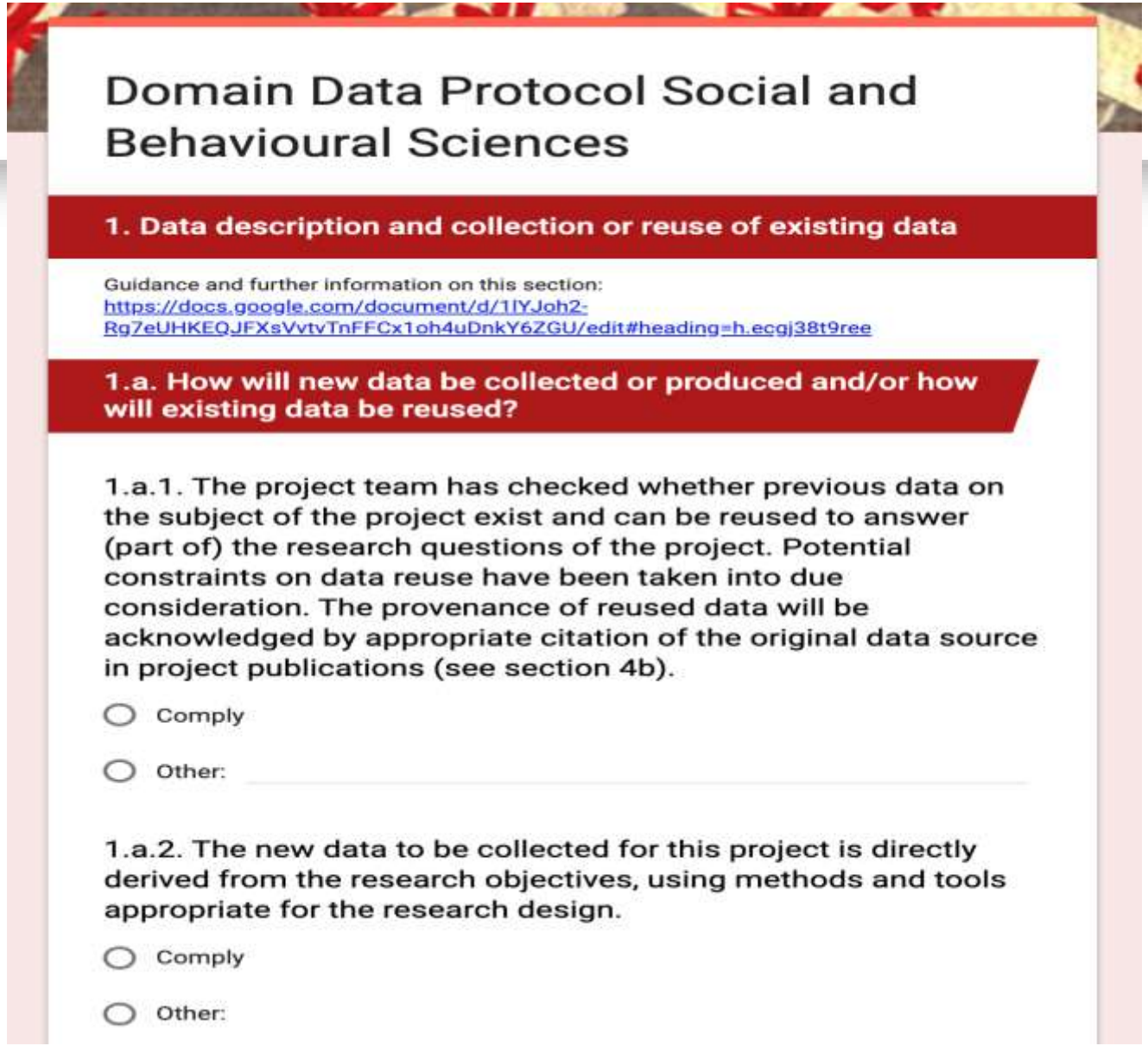
3a

How will data and metadata be stored and backed up during the research?

- Describe where the data will be stored and backed up during research activities and how often the backup will be performed. It is recommended to store data in least at two separate locations.
- Give preference to the use of robust, managed storage with automatic backup, such as provided by IT support services of the home institution. Storing data on laptops, stand-alone hard drives, or external storage devices such as USB sticks is not recommended.

Guidance

Protocol for Social Science Data



Domain Data Protocol Social and Behavioural Sciences

1. Data description and collection or reuse of existing data

Guidance and further information on this section:
https://docs.google.com/document/d/1IYJoh2-Rg7eUHKEQJFXsVvtvTnFFCx1oh4uDnkY6ZGU/edit#heading=h.ecgj38t9ree

1.a. How will new data be collected or produced and/or how will existing data be reused?

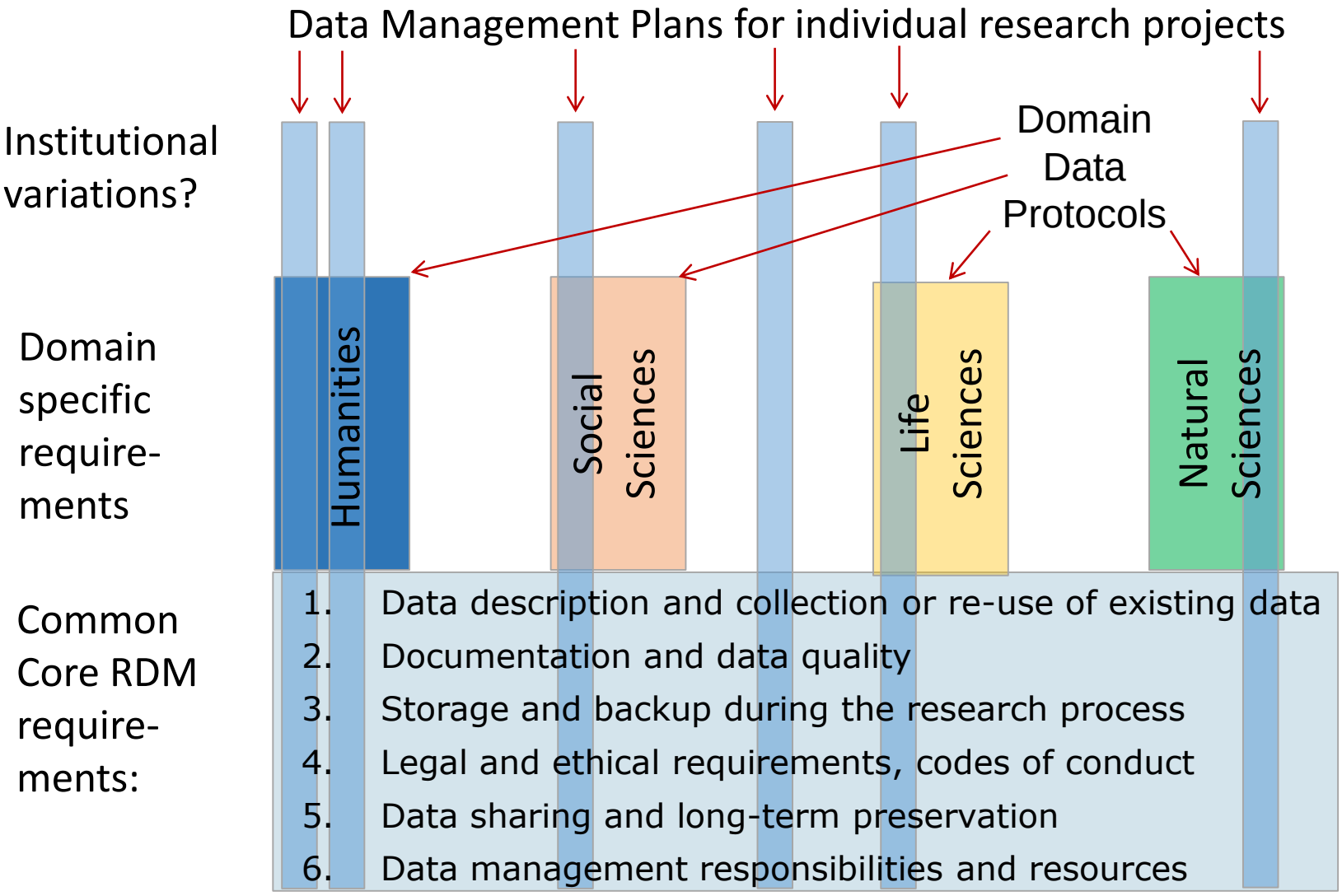
1.a.1. The project team has checked whether previous data on the subject of the project exist and can be reused to answer (part of) the research questions of the project. Potential constraints on data reuse have been taken into due consideration. The provenance of reused data will be acknowledged by appropriate citation of the original data source in project publications (see section 4b).

- ○ Comply
- ○ Other:

1.a.2. The new data to be collected for this project is directly derived from the research objectives, using methods and tools appropriate for the research design.

- ○ Comply
- ○ Other:

Domain Data Protocols



For discussion

- Is the idea of a Domain Data Protocol useful?
- Is the Protocol too long or too detailed?
- Is the language understandable (not too difficult, complicated, technical?)
- Which organization(s) could be asked to endorse it?
- Which topic needs most guidance and training?
- How to stimulate the application of Domain Data Protocols?
- What are the next steps?
  - Several Science Europe members have begun to implement the core requirements
  - Develop and publish exemplar domain protocols
  - Seek acceptance of domain protocols by communities

Thanks

Some slides are created by

- Peter Doorn
- Marjan Grootveld