

A Discipline-specific Approach to Research Data Management: Towards an Archaeological Data Protocol

Peter Doorn, DANS-KNAW

Science Europe Webinar: Aligning Research Data Management Across
Europe – The Next Step (27 January 2021)

Twitter: @pkdoorn @danskaw

Representing ARIADNE(plus) and DANS

Objectives of ARIADNEplus:

- Integrate archaeological data infrastructures across Europe
- Support FAIR data sharing in archaeology
- Promote good practices in archaeological data management
- > 40 partners from 23 European countries
- ARIADNEplus started January 2019, 48 months (follow-up of ARIADNE project 2014-2018)

<https://ariadne-infrastructure.eu>

DANS (Data Archiving and Networked Services):

- Dutch national centre of expertise and repository for research data.
- One of the largest collections of archaeological data sets in the world (> 95,000 datasets)
- Institute of [KNAW](#) (Royal Netherlands Academy of Arts and Sciences) and [NWO](#) (Dutch Research Council).

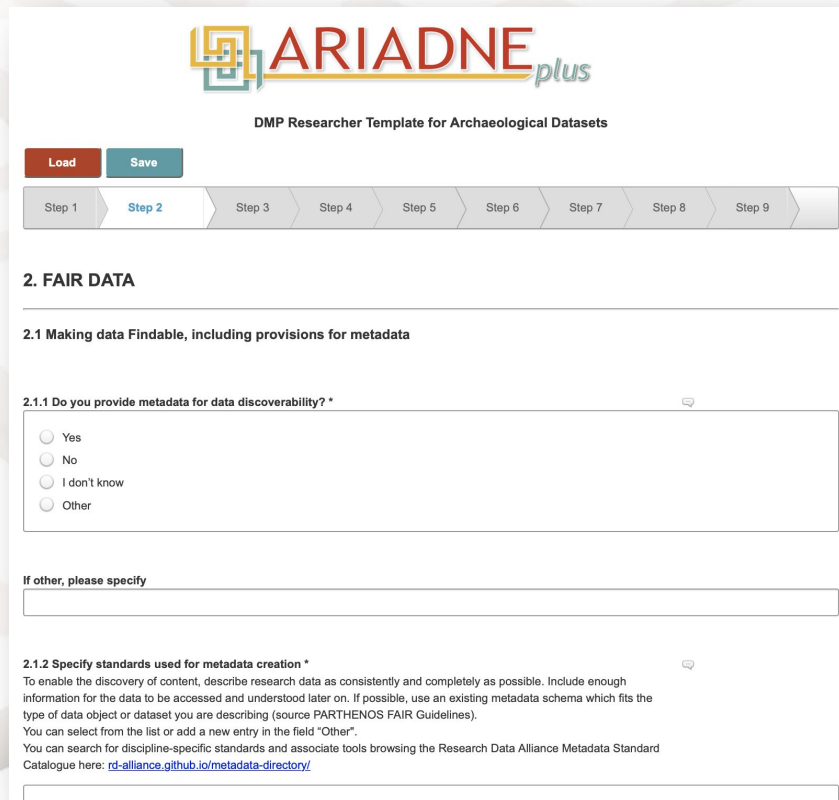
<https://dans.knaw.nl/en>

Discipline-specific data management

ARIADNEplus:

H2020 DMP requirements ⇒ ARIADNEplus DMP template:

<http://vast-lab.org/dmp/ariadneplus/form/>



The screenshot shows the ARIADNEplus DMP template form. At the top, there is the ARIADNEplus logo and the text "DMP Researcher Template for Archaeological Datasets". Below this are "Load" and "Save" buttons. A progress bar shows steps 1 through 9, with "Step 2" highlighted. The main section is titled "2. FAIR DATA" and contains two sub-sections: "2.1 Making data Findable, including provisions for metadata" and "2.1.2 Specify standards used for metadata creation".

2.1 Making data Findable, including provisions for metadata

2.1.1 Do you provide metadata for data discoverability? *

Yes
 No
 I don't know
 Other

If other, please specify

2.1.2 Specify standards used for metadata creation *

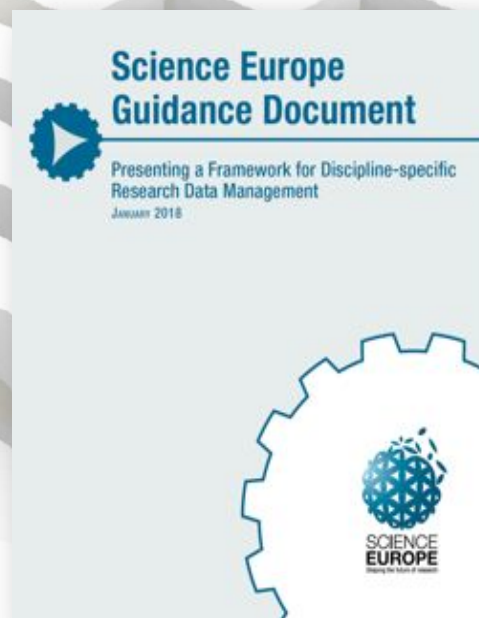
To enable the discovery of content, describe research data as consistently and completely as possible. Include enough information for the data to be accessed and understood later on. If possible, use an existing metadata schema which fits the type of data object or dataset you are describing (source PARTHENOS FAIR Guidelines). You can select from the list or add a new entry in the field "Other". You can search for discipline-specific standards and associate tools browsing the Research Data Alliance Metadata Standard Catalogue here: rd-alliance.github.io/metadata-directory/

Science Europe / NWO / DANS:

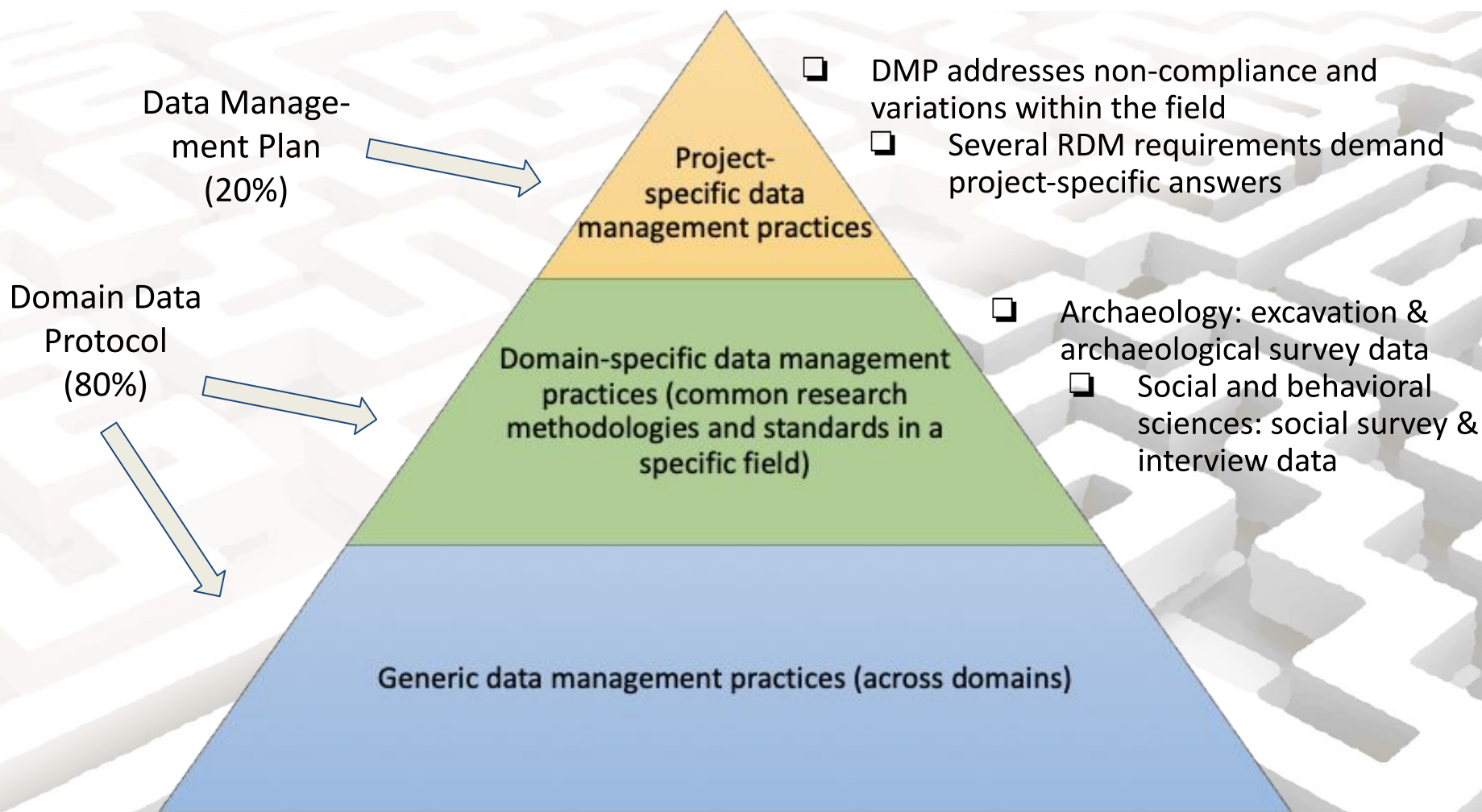
Science Europe report on Discipline-specific RDM and harmonization of RDM

Requirements ⇒ Domain Data Protocols -

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



Why does a domain approach to data management make sense?



Advantages of working with domain protocol:

Provides a model DMP, answering about 80% of the RDM requirements

- For researcher:

- Much faster and easier to make a DMP
- As protocol is based on good practices of a scientific field, it has a normative and educational function

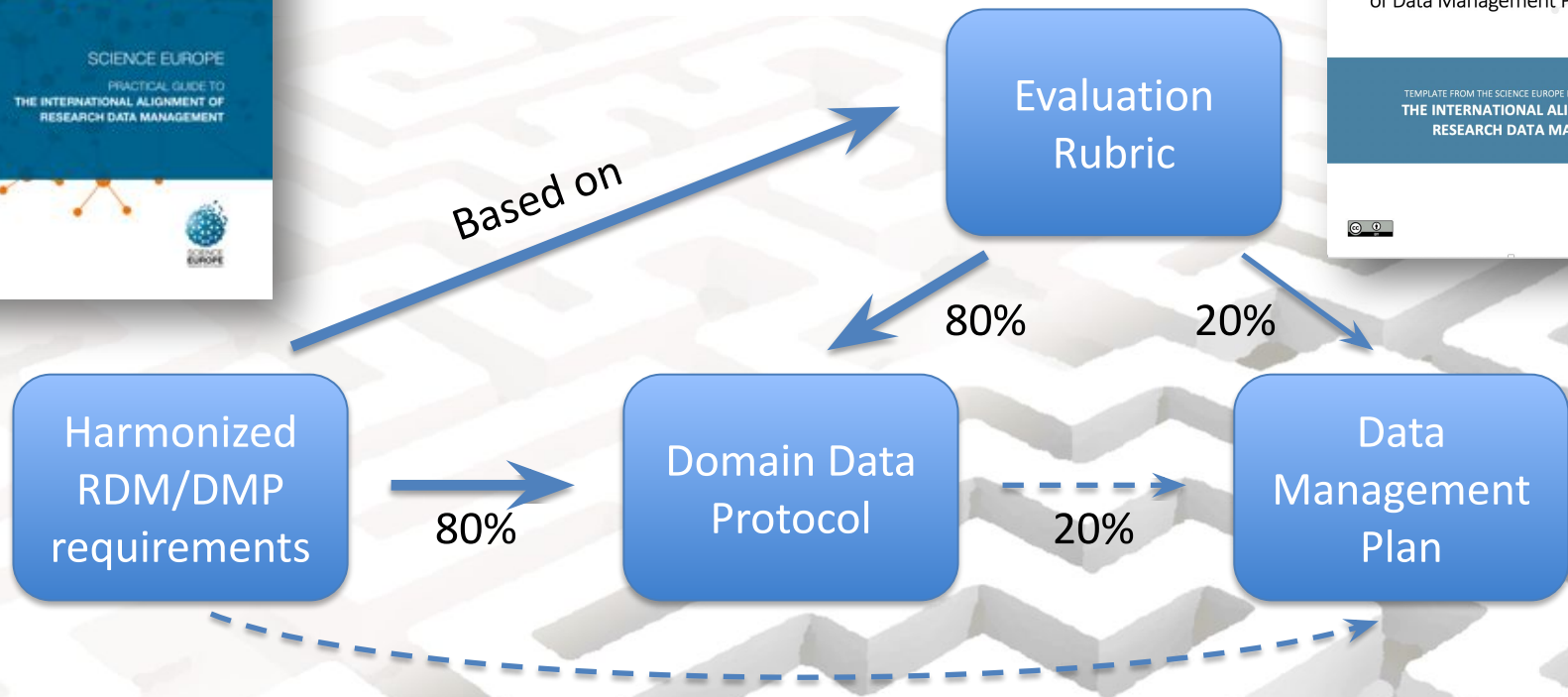
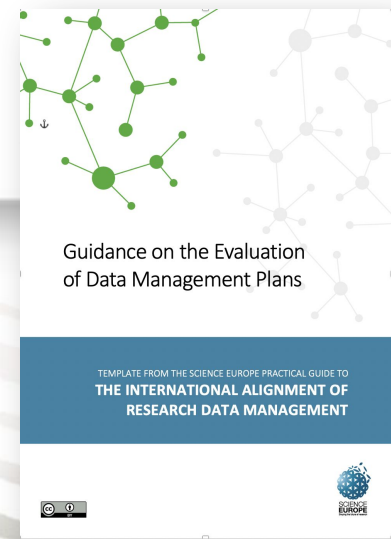
- For funder:

- much easier and quicker to evaluate
- Protocol is directly linked to harmonized Science Europe requirements and model answers of evaluation rubric (where possible)

Million Dollar Question:

How much detail is required (in a Protocol and DMP)?

How does it work?



- Domain protocol provides general replies to the questions of the RDM requirements
- Principle of comply or explain: if you do not comply, you can explain in DMP
- In general: specifications and further details to be given in DMP
- A domain protocol is not obligatory

First attempt to formulate Domain Data Protocol

- Protocol for social science surveys and interview data
- Prototype in Google Forms presented and discussed at European Survey Research Association Conference, July 2019, Zagreb



Section 1 of 5

Domain Data Protocol Social and Behavioural Sciences

Draft version 11/7/2019 - Peter Doorn - Test for ESRA Conference, Zagreb

Generally well-received

However:

- Too detailed
- Many answers still had options (non-compliance and details solved within the protocol form)
- Too similar to an extensive DMP template

<https://tinyurl.com/yxath2gg>

Second attempt in context of ARIADNEplus

- PARTHENOS project had already worked on DMP template for archaeology (Paola Ronzino – PIN, Prato, Italy)
- Archaeology Data Protocol is simplified and adapted from social science version:
 - Options moved to guidance
 - Non-compliance and details to be specified in ARIADNE DMP
 - A substantial number of articles in the protocols for social sciences and archaeology are very similar
 - Almost every article is accompanied by guidance, references to literature or online explanations, etc.
- Nevertheless, Archaeology Data Protocol still has 50 articles. However:
 - Note that this is comparable to the number of bullets in the S.E. Guidance and Evaluation Rubric
 - Whole handbooks are devoted to subjects such as privacy, quality management, data security, and research ethics; yet, the essence of such subjects is summarized in just a few protocol articles each.

Example: first Science Europe RDM requirement (1a)

Science Europe Requirement	Science Europe Guidance	“Sufficiently addressed” according to Science Europe Evaluation Rubric
1a. How will new data be collected or produced and/or how will existing data be reused?	<ul style="list-style-type: none">• Explain which methodologies or software will be used if new data are collected or produced.• State any constraints on reuse of existing data if there are any.• Explain how data provenance will be documented.• Briefly state the reasons if the reuse of any existing data sources has been considered but discarded.	<ul style="list-style-type: none">• gives clear details of where the existing data come from and how new data will be collected or produced. It clearly explains methods and software used.• explains, if existing data are reused, how these data will be accessed and any constraints on their reuse.• explains clearly, if applicable, why new data must be collected, rather than re-using existing data.

Archaeology Data Protocol 1a: 5 articles

Nr.	Protocol article	To be specified in DMP	Maps to Ariadne DMP	Protocol Guidance
1a.1	The research team has checked whether previous data on the subject of the research exist that can be reused. If this is the case, such data will be reused, respecting licenses and applicable constraints, including intellectual property rights.	<ul style="list-style-type: none"> Specify pre-existing data that will be reused If applicable, Explain why pre-existing data will not be reused 	8	
1a.2	The provenance of reused data will be duly described and in publications the original data owners will be gracefully acknowledged.		9	<ul style="list-style-type: none"> Provenance provides insight into where the original data come from and by whom, when and how they were created. On data citation see: https://datacite.org/cite-your-data.html
1a.3	Possible additions and corrections to existing data will be made available under the same conditions as apply to the original data.		13	This is called a "share alike" license. For further information see: https://creativecommons.org/licenses/by-sa/4.0/
1a.4	The motivation for the collection of new data is derived from the research objectives, which are described in the research proposal.	Details of where data come from (provenance) and/or how data will be collected or produced	6	
1a.5	The data collection uses methods and tools appropriate for the research design, and will follow good practices in archaeology.	Specify data collection methods and tools	9, 14	These methods and tools are further described in section 2b. and 4c. Archaeology Data Service / Digital Antiquity Guides to Good Practice: https://guides.archaeologydataservice.ac.uk/g2gpwiki/

Full Protocol: <https://tinyurl.com/y3mlgtoh>

Conclusions

Conclusions:

- Many Protocol articles appear to be discipline-independent.
- Guidance is usually more domain specific, especially with respect to
 - Domain standards and norms, codes of conduct
 - Importance of privacy and ethics related to personal information
 - Importance of specific data types and formats
 - Role of sampling in research
 - Domain repositories
- It is easier to comply with the protocol and less work on a DMP is needed
 - During research: if institutional RDM policies and support are in place
 - After research: If data is shared via a certified Trusted Digital Repository

Next Steps

Align domain protocols with RDM requirements and evaluation rubric

- Funders further implement S.E. RDM requirements and evaluation rubric
 - Question: will Horizon Europe follow the S.E. requirements?
- In progress: acceptance and implementation of domain protocol by ARIADNE partners and wider archaeological community
- Consultation of ARIADNE with funding organisations \Rightarrow endorsement by funders of ARIADNE domain protocol and connected DMP template
- It is straightforward to formulate Protocols for other domains (at least in the social sciences and humanities) after the first one will be in operation

THANK YOU! Comments and Questions welcome

ARIADNE is a project funded by the European Commission under the H2020 Programme, contract no. H2020-INFRAIA-2018-1-823914.

Contact: peter.doorn@dans.knaw.nl

www.ariadne-infrastructure.eu

www.dans.knaw.nl

Draft Archaeology Domain Protocol: <https://tinyurl.com/y3mlgtoh>