Open Science and Sharing Research Data: Towards European Guidelines on RDM procedures

Organised by Science Europe and the
Netherlands Organisation for Scientific Research (NWO)
Brussels, 30 January 2018

Introduction to the Concept of Domain Protocols and State of Affairs

Peter Doorn

Chair of the Science Europe Working Group on Research Data

Data Archiving and Networked Services (DANS), The Netherlands

@dansknaw @pkdoorn

Science Europe WG Research Data

Until 2016, the SEWGRD worked on basic aspects of research data, such as:

- Funding of data management and infrastructures: https://goo.gl/eokd1j
- Legal aspects related to copyright and Text and Data Mining (TDM)
- Common data terminology: http://sedataglossary.shoutwiki.com/wiki/Main_Page

Since summer 2016 the Working Group has focused on Domain Protocols for Research Data Management

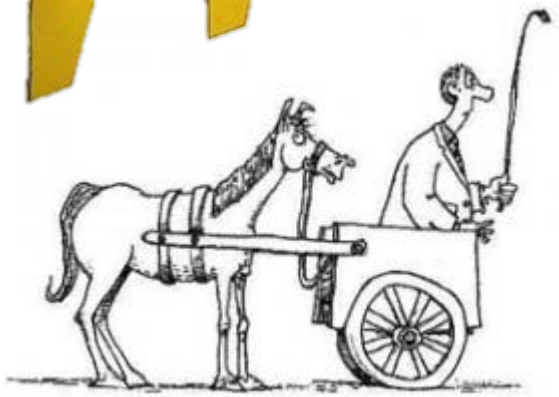
Growing demands for Data Management Plans

- A growing number of SE Member Organisations have formulated policies, requirements, templates, etc. for Research Data Management (RDM) and Data Management Plans (DMP)
- The practices and cultures of data stewardship and data sharing vary among and within domains and communities, often depending on methodologies and nature of data collected/processed

Many researchers and communities support data management planning

- However, the rewards (citations) are modest while the costs of data management may be high
- Cost of data management need to be incorporated in research projects
- Data management is a basic quality control mechanism in research and not a formality
- The involvement of research communities is vital for the success of RDM policies

What we try to avoid:



One size of data management doesn't fit all: a domain-oriented approach

Specialized data management practices are in use by different disciplines and communities.

A "bottom-up" approach complementing the "top-down" requirements, involving research communities, is needed:

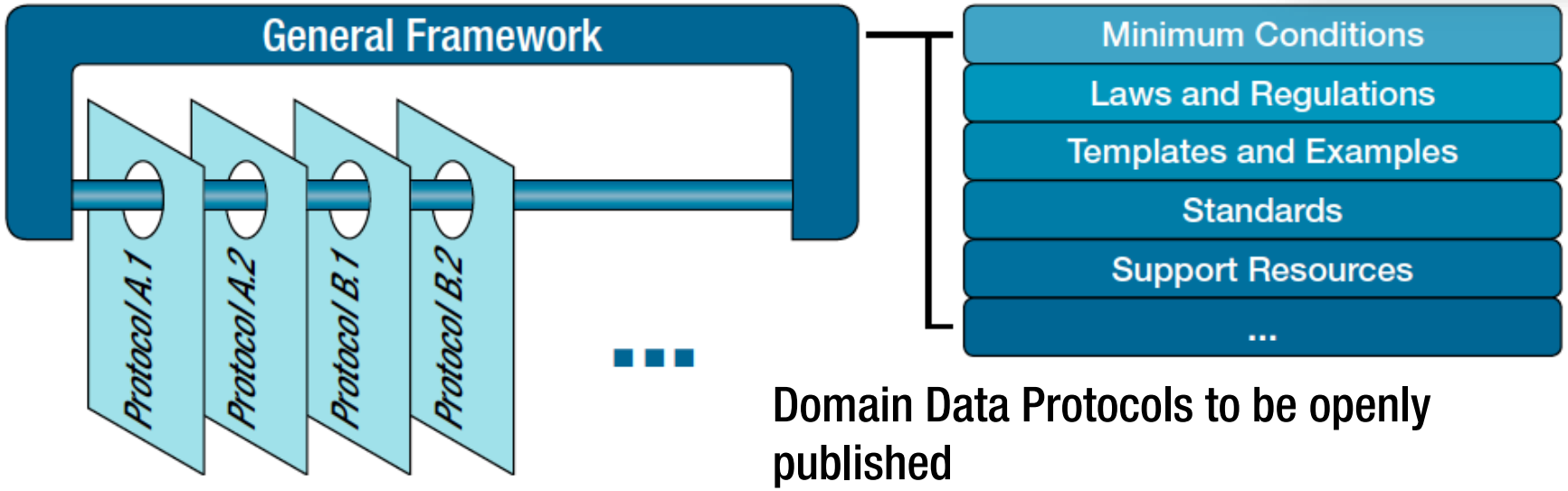
- Will be more suitable to community needs
- Will get better acceptance/adoption by communities

However:

- Terms of reference and guidelines are needed, to ensure legal compliance, comparability, procedures and basic quality standards
- This implies that research funders and performing institutions are to align their core RDM requirements

Actively involve communities in formulating RDM good practices

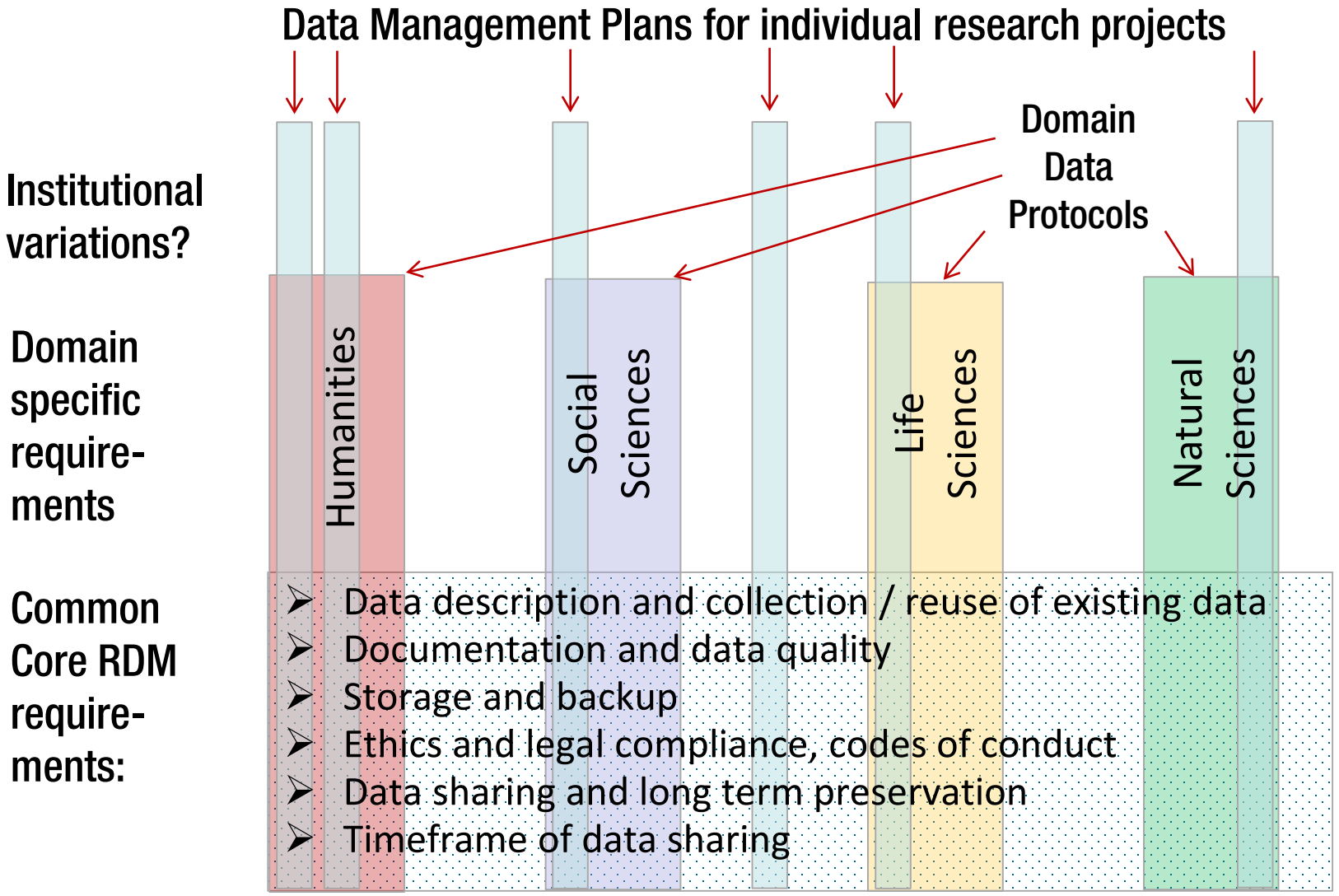
Science Europe M.O.'s to align RDM requirements and endorse Data Protocols Framework (Terms of Reference for Domain Protocols)



Domain Data Protocols to be openly published

Report by Aerts & Doorn (2016): "A Conceptual Approach to Data Stewardship and Software Sustainability": http://goo.gl/ycj8QH

Common core and domain specific requirements for DMP's



Authorship of protocols: at which level of granularity?

- Several ESFRI ERICs are well placed
- Rely on existing work as much as possible rather than asking communities to start from scratch
- Think modular - the detail can vary according to need:
  - Even a very generic protocol or 'model DMP' will be helpful
  - You don't have to oblige anything or anybody: Researchers still write their individual DMPs, motivating where they deviate from the norm/protocol in their field (comply or explain principle)
  - Communities will decide on the detail that they find useful
  - There may be alternative DDPs for different purposes (depending on size of project, type & volume of data, etc.) within one domain
- Approach "volunteers" from different domains to kick-off the process (proof of concept)

Selection of proof-of-concept communities for domain data protocols

Domain	Community
1. Humanities (general)	DARIAH
2. Humanities – Archaeology	PARTHENOS - ARIADNE
3. Linguistics - Language data	CLARIN
4. Social Sciences - Survey research	CESSDA
5. Social & Behavioural Sciences – Psychology	Psychology departments and associations
6. Social Sciences - Ageing Studies	SHARE and TILDA
7. Life Sciences - Bio-informatics	ELIXIR
8. Plant Science	ERA-CAPS (former Working Group on RDM)
9. Climate Research	ICOS / ENVRI+

General reactions from communities

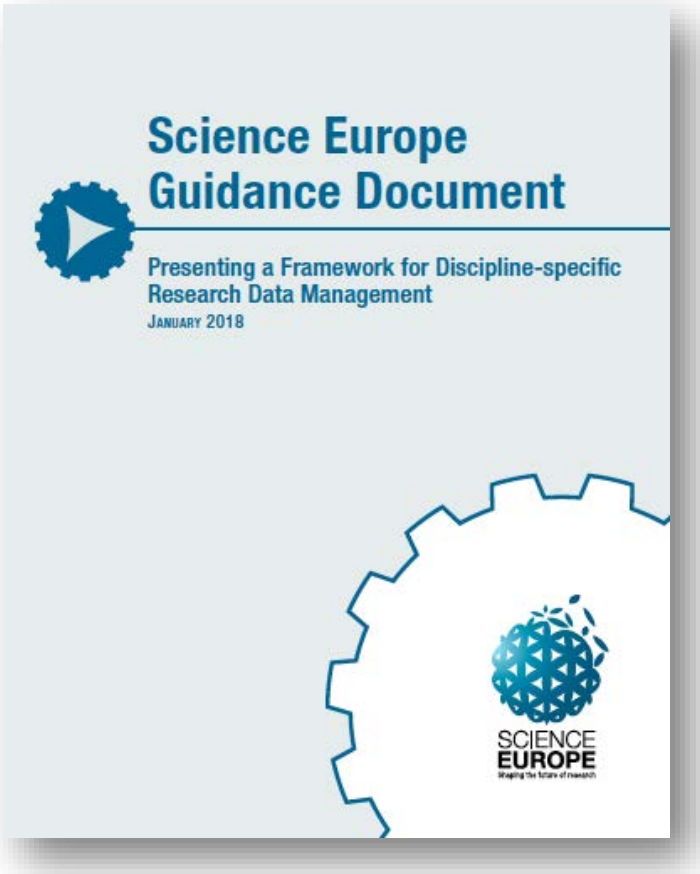
- Almost all reactions positive, general interest of communities to cooperate with the S.E. initiative
- Several are already working towards this direction:
  - Plant Science, Climate Research: **Data policies**
  - Life Sciences, Bio-informatics: **RDM Recommendations and guidelines**
  - Humanities: **Detailed RDM template**
  - Psychology (NL): **Data storage guideline**
  - . . .
- Domain approach fits in with other DMP developments and activities (RDA, Force11, DCC, etc.)

Summary: the advantages of this approach

- Counter different RDM requirements from funder to funder, from university to university, from institute to institute
- Active involvement of scientific domains and scholarly communities increases acceptance and usefulness of RDM
- Less work for researchers proposing projects by accepting domain protocol as part of DMP
- Provision to researchers of a learning vehicle on RDM practices in their field, thus raising the general quality level of data management
- Reduced DMP processing costs and burdens for funders and researchers, and more focus on and better assessment of deviating RDM solutions

Current status and next steps

- Develop and publish exemplar protocols
- Seek acceptance by communities
- Endorse protocols as basic/generic DMPs for domains



DANS is about keeping data FAIR

Data Archiving and Networked Services
DANS