Data strategies for Research Infrastructures

Introduction

19 February 2015, ESO headquarters, Munich

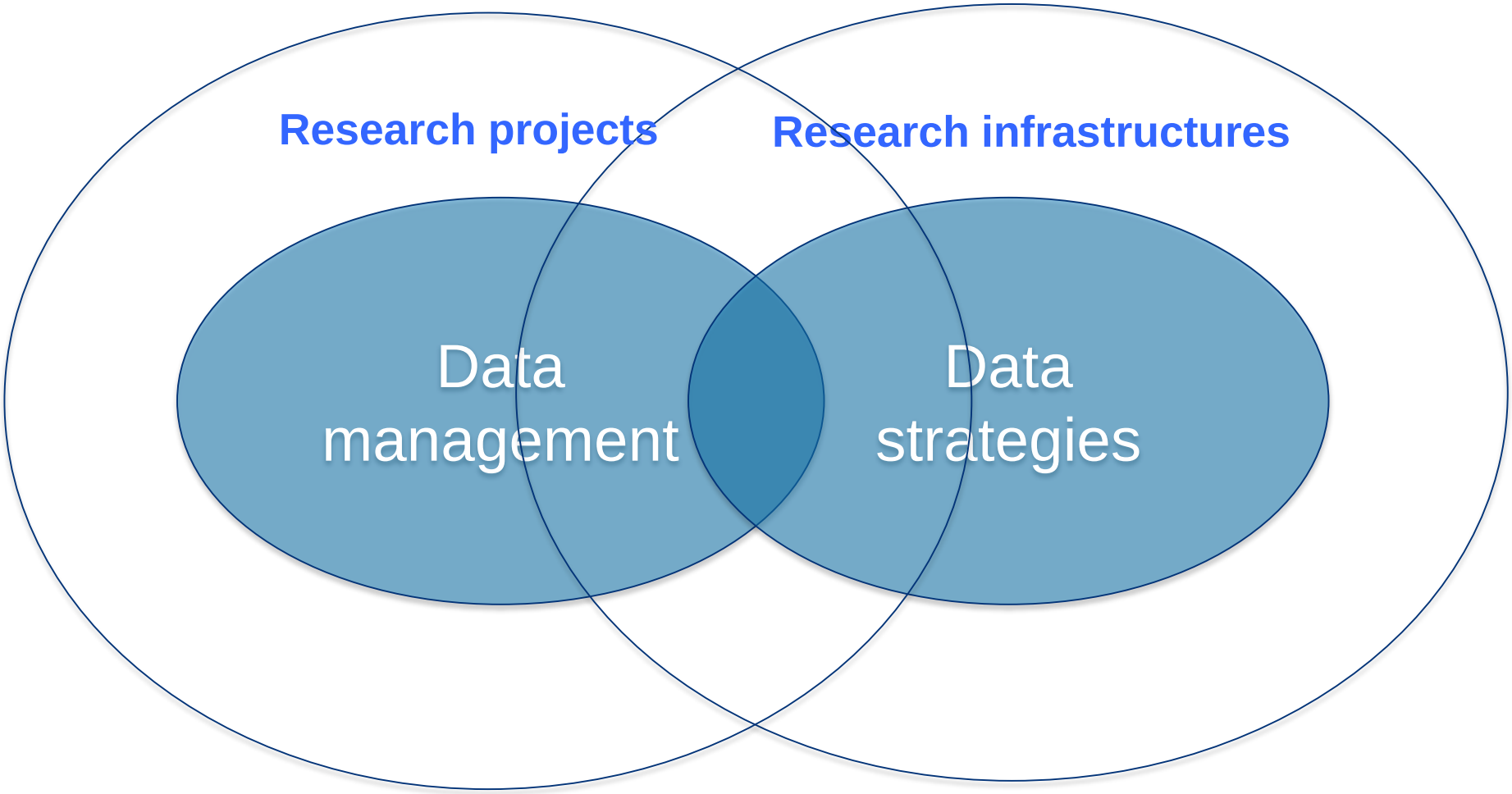
Stephanie Suhr

BioMedBridges

What this workshop is about..

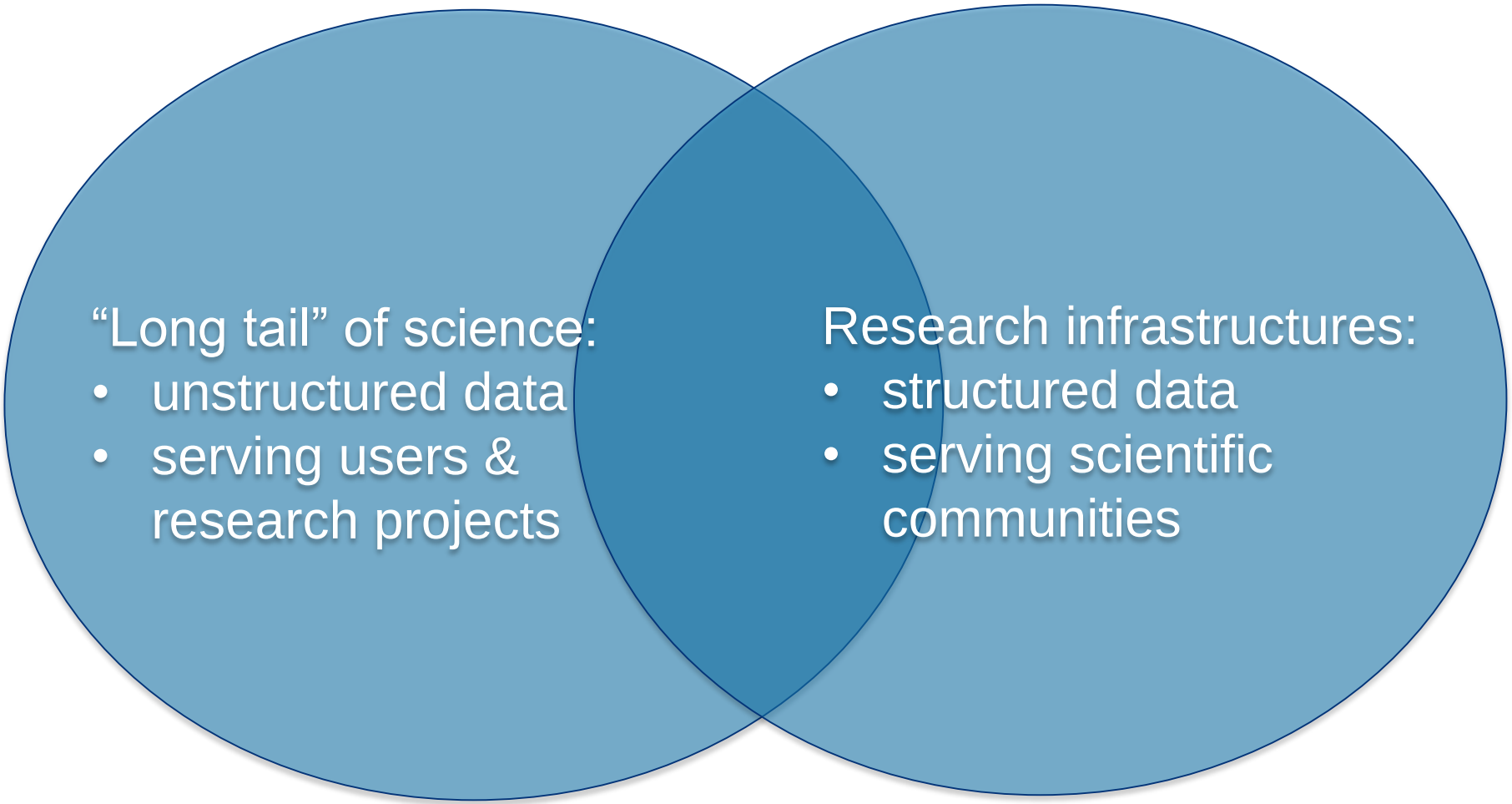
…and what it isn't!

This is a simplification!



This is a simplification!

RI data strategy?...



RI data strategy?...

→ The RI **data strategy** sets the framework for users/projects to develop **data management** plans

Type of RI?

- **Data producing** and **data managing** RIs

(note that RIs can have both roles!)

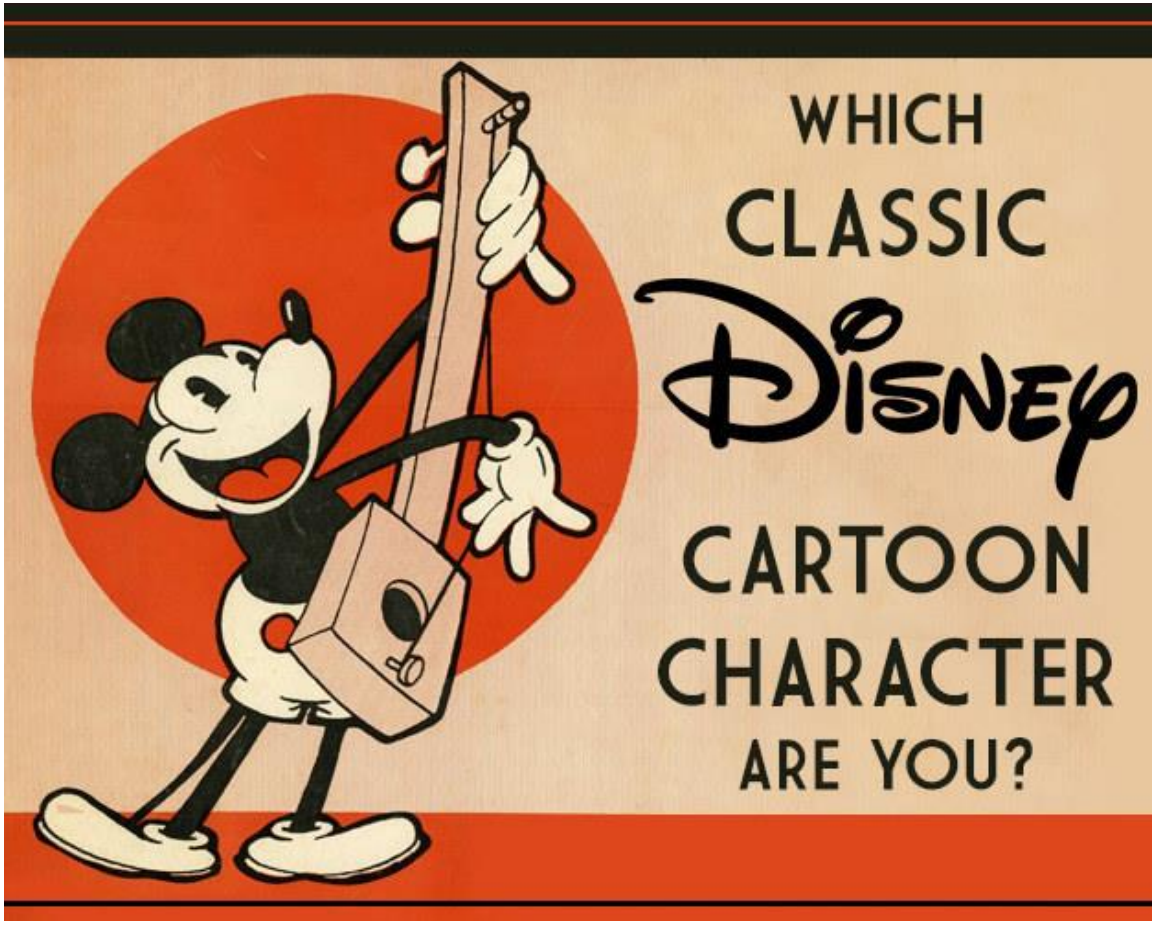
Principles of data management and sharing at European Research Infrastructures doi:10.5281/zenodo.8304

Data producing RIs

- Policies and processes for raw data and in some instances also for processed (annotated/curated) data
- Users:
  - Researchers who produce data
  - Researchers who use data deposited at the RI
- Data are produced by researchers in the context of funded research projects
  - → data belong to the user!
  - RIs do not assert any IP rights over the data produced, but may impose policies before and/or after data is produced

Data managing RIs

- May have policies for:
  - raw and processed (annotated/curated) data, including privacy and data security policies where applicable
  - results derived of deposited data (research results gained from the re-use of deposited data)
- Manage data produced elsewhere
- Users:
  - Researchers who have deposited their own data with the RI
  - and/or those who are interested in re-use or repurposing of data deposited by other researchers


WHICH
CLASSIC
Disney
CARTOON
CHARACTER
ARE YOU?

What kind of RI are you?

- Example: biobanking data (BBMRI)
  - Thousands of European biobanks
  - most have their own data management, archival and access infrastructure
  - ensure data protection and necessary consent of the participants/donors
  - data sharing policies within BBMRI are very diverse
  - closely tied to the informed consent given by the data subjects to the individual biobanks
  - Different national legislation in the countries where the biobank is located apply!

Key question:

- *Is the provision of data (part of) the **core business** of the research infrastructure?*
  - if yes: need to assess and address details of data-related needs of the user community!
  - if no: an extreme option is to send users home with their data… (Note: they will then likely need to take it to another infrastructure!)

Data strategies: factors

- Scientific
  - e.g. data reproducibility, uniqueness, value of processed and/or raw data
- Technical
  - storage, network, computation...
- Financial
  - Cost of data storage, transfer, reproduction

Data strategies: factors

- Political
  - drivers e.g. from funding bodies, large organisations, national interests
- Legal/ethical/formal
  - e.g. need to store patient data in country of origin, requirements from journal publishers, data management plans, etc.
- Social
  - data sharing mentality of the community in question, how they do science - e.g. standards, best practices

- **15:45-16:45 Flash presentations**

- Voluntary!
- 5 minutes each
- Present:
  - Any data-related solutions / considerations / developments in your RI that are relevant
  - or pose questions that have come up during the workshop to the other participants

- 16:45-17:30 Wrap-up discussion

Questions?
Comments?

Please ask!