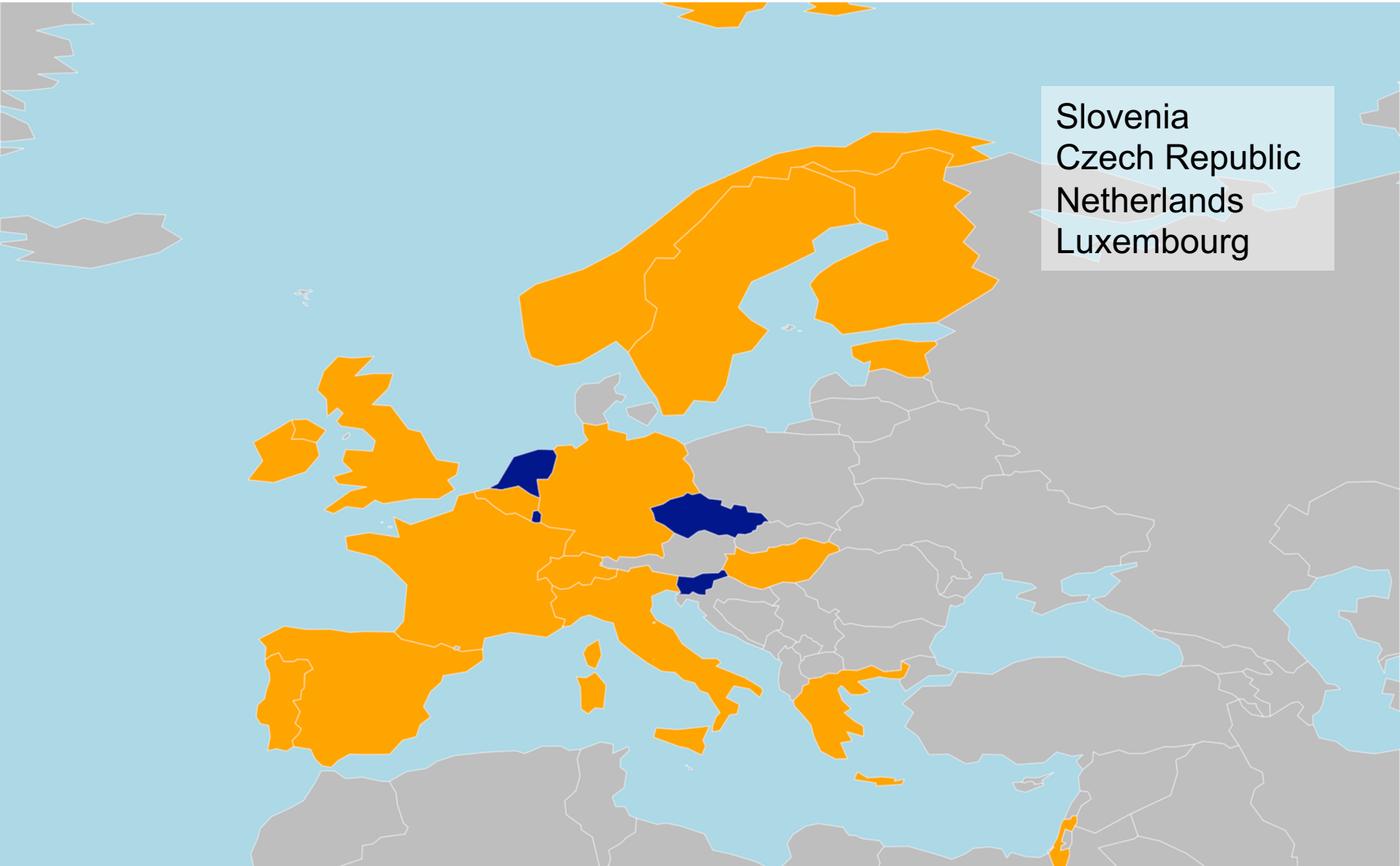
# Implementation of Data Management Plans and Data Stewardship in practice

Roland Krause
ELIXIR-LU

ECCB 2018

# Workshop organization

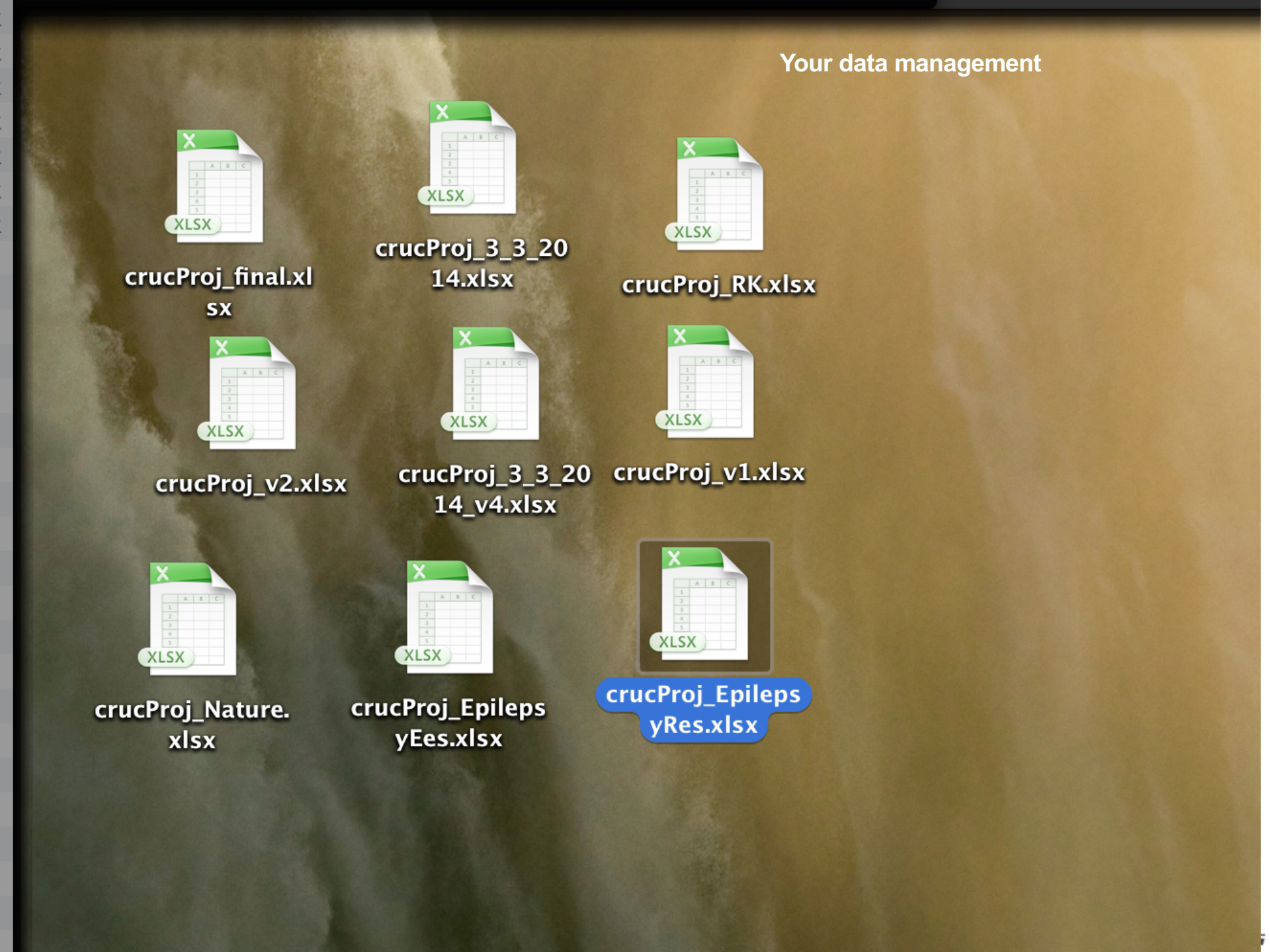


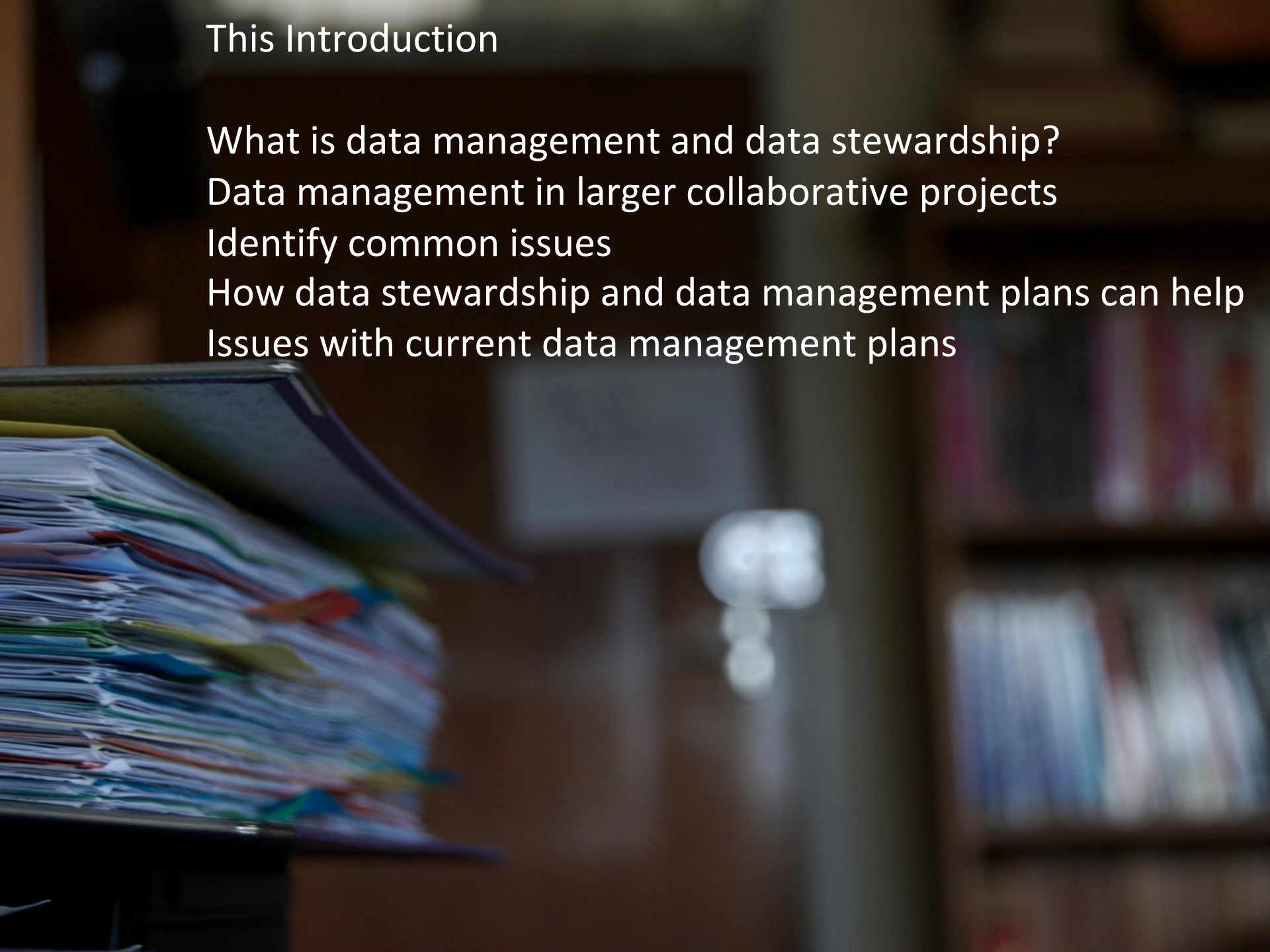
# Dr Bruce Banner aka The Incredible Hulk

Unrealistic depiction of a scientist

- … can throw cars around and resist gun shots
- Dr Bruce Banner has 7 PhDs
- Stores all research results in databases

Your data management





This Introduction

What is data management and data stewardship?
Data management in larger collaborative projects
Identify common issues
How data stewardship and data management plans can help
Issues with current data management plans

# Data stewardship vs data management

**Management**
Local, manipulating data
Data flow within the project

**Examples**
File naming conventions
Collecting metadata
Back-up

How to name files
https://speakerdeck.com/jennybc/how-to-name-files

**Stewardship**
Organizing data before and beyond the project

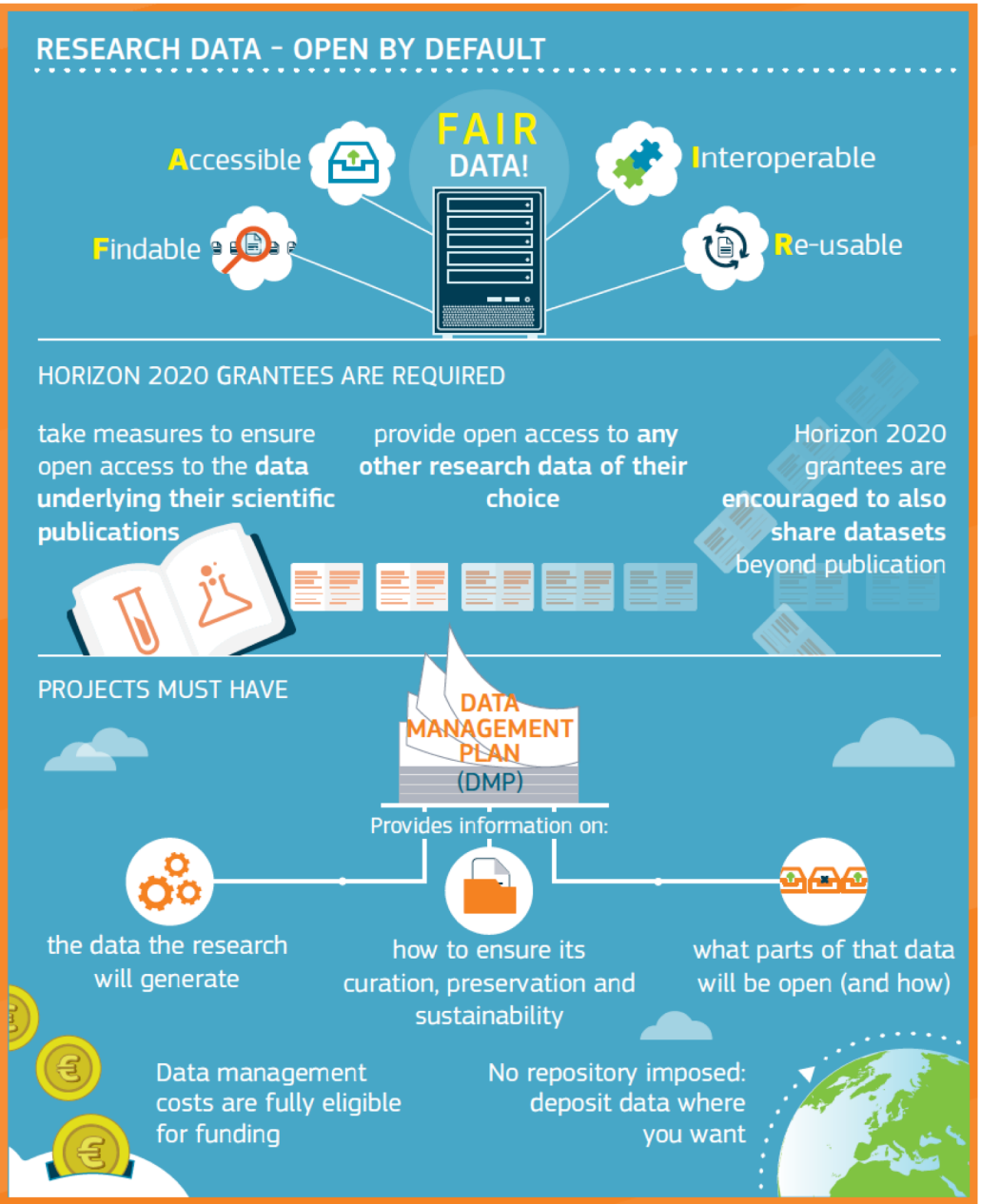
**Examples**
Choice of repositories
Decisions on meta data collection
Archival
Long-term care of valuable digital assets for additional discovery and downstream investigations
Notably includes all digital objects including documentation and analysis code

**Data stewardship wizard**

Luxembourg Centre for Systems Biomedicine

elixir LUXEMBOURG

# Do you need a data management plan?




http://datafairport.org/

Data FAIRport

Find, Access, Interoperate & Re-use Data

European Commission

http://ec.europa.eu/research/openscience

elixir LUXEMBOURG

# An Example Project: CSBdpwAP*

* Complex, sensitive, bulky data project with actual people

- Project on a complex disease
  - Human – patient – data
  - Genetic data
  - Animal models
- Public sources
- Data contributed by the project partners
- Data generated during the project on samples provided by the partners

Luxembourg Centre for Systems Biomedicine

elixir LUXEMBOURG

# Issues of data stewardship

- Data sharing across institutions
  - Data is only shared if required
  - Data deposited to public repositories is only a slice of existing
  - Data sharing agreements not in place
    - Note that GDPR requires contracts for processing of sensitive data
- Responsible principal investigators far removed from data handling
- People handling data do not sit at the Steering Committee
- Access to storage limitations
- Data is thrown together in directories without provenance
- Data is to be deposited in a complex database that never the PostDoc never finishes
- People leave the project and/or academic research

# Open data

Embargoed until publication
Only the required slice of data published
Data published without accompanying analysis questionable

# Perspectives

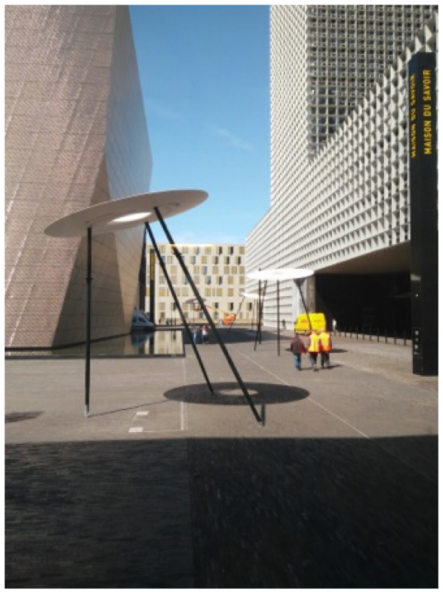
## When make data open?

- On production
- On publication
- On end of the project
- Take your filthy hands of *my* data

## Who owns data anyway?

- Principal investigator
- Institutions
- Funding bodies
- Patients donating samples

# Thank you for your attention!



roland.krause@uni.lu

@spitshine - @elixir_lu