Research data management and the FAIR Data Pilot

RCSI Open Science week, 23 Oct 2019

Annalisa Montesanti, PhD

Programme Manager - Health Research Careers

HRB Health Research Board

Research. Evidence. Action.

I will discuss

- The ambition - Why
- Facilitating the transition to open and FAIR data – a step-wise approach
- (HRB) FAIR DMP Pilot – building awareness and expertise
- An international collaborative effort: The Funder Implementation Study

The ambition

Ensuring that its funded research can have the **greatest possible impact** by having more

- Openness and transparency
- Reproducibility and replicability
- Acceleration of knowledge discovery and its application
- Discoverability and accessibility
- Interoperability across different domains and disciplines
- (Re)Usability

Towards a more open research environment

Many funders and other stakeholders have now embraced

- Open science
- DORA San Francisco Declaration for researchers assessment
- Open access (moving to Plan S)
- Open data and data sharing
- Data management and stewardship as intrinsic to good research practice
- FAIR principles - BUT what do they really mean in practice?

How to align and implement them in practice?

Research data

- Research data are an important and expensive output of the scholarly research process, across all disciplines.
- They are an essential part of the resources necessary to evaluate research results, and to reconstruct the events and processes leading to them.
- Their value increases when they are aggregated into collections and become more available for re-use to establish reproducibility and to address new and challenging research questions.
- There is a growing international consensus on the need to comply with good practice in the use and reuse of research data with the ultimate goal of preserve and share research data in a manner that maximises their long-term value.
- Without proper data stewardship, the value of data risks to be greatly diminished and underutilised;
  - Poor data management is perhaps the single largest source of research waste:
  - Data are often lost or discarder or not properly managed at the end of the study;

https://www.youtube.com/watch?v=N2zK3sAtr-4

Increasing the value of research data: DM&S

There is now more emphasis on good data management and stewardship before starting a research project.

- A data management plan (DMP) is a formal document describing how research data will be managed and documented throughout a research project and the terms regarding the subsequent deposit of the data with a data repository for long-term management and preservation.
- It is supposed to be a living document which update during the project.
- Data management is an ongoing process and planning in the early stages makes the whole endeavour easier.
- Good research data management & stewardship is not a goal in itself, but rather the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse.

FAIR principles

SCIENTIFIC DATA

Comment | OPEN | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao & Barend Mons ✉ - Show fewer authors

Scientific Data **3**, Article number: 160018 (2016) | Download Citation ⬇

HRB Health Research Board

hrb.ie

What is FAIR

Data and services that are

- **findable,**
- **accessible,**
- **interoperable,**
- **re-usable**

both for machines and for people.”

What is FAIR

WHAT IS FAIR DATA?

FAIR Data aims to support existing communities in their attempts to enable valuable scientific data and knowledge to be published and utilised in a 'FAIR' manner.

Findable - (meta)data is uniquely and persistently identifiable. Should have basic machine readable descriptive metadata.

Accessible - data is reachable and accessible by humans and machines using standard formats and protocols.

Interoperable - (meta)data is machine readable and annotated with resolvable vocabularies/ontologies.

Reusable - (meta)data is sufficiently well-described to allow (semi)automated integration with other compatible data sources.

How FAIR are your data?

Findable

It should be possible for others to discover your data. Rich metadata should be available online in a searchable resource, and the data should be assigned a persistent identifier.

- ☐ A persistent identifier is assigned to your data
- ☐ There are rich metadata, describing your data
- ☐ The metadata are online in a searchable resource e.g. a catalogue or data repository
- ☐ The metadata record specifies the persistent identifier

Accessible

It should be possible for humans and machines to gain access to your data, under specific conditions or restrictions where appropriate. FAIR does not mean that data need to be open! There should be metadata, even if the data aren't accessible.

- ☐ Following the persistent ID will take you to the data or associated metadata
- ☐ The protocol by which data can be retrieved follows recognised standards e.g. http
- ☐ The access procedure includes authentication and authorisation steps, if necessary
- ☐ Metadata are accessible, wherever possible, even if the data aren't

Interoperable

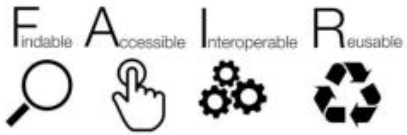
Data and metadata should conform to recognised formats and standards to allow them to be combined and exchanged.

- ☐ Data is provided in commonly understood and preferably open formats
- ☐ The metadata provided follows relevant standards
- ☐ Controlled vocabularies, keywords, thesauri or ontologies are used where possible
- ☐ Qualified references and links are provided to other related data

Reusable

Lots of documentation is needed to support data interpretation and reuse. The data should conform to community norms and be clearly licensed so others know what kinds of reuse are permitted.

- ☐ The data are accurate and well described with many relevant attributes
- ☐ The data have a clear and accessible data usage license
- ☐ It is clear how, why and by whom the data have been created and processed
- ☐ The data and metadata meet relevant domain standards



'How FAIR are your data?' checklist, CC-BY by Sarah Jones & Marjan Grootveld, EUDAT. Image CC-BY-SA by SangyaPundir

HRB Health Research Board

FAIR vs Open: a misconception

FAIR does not mean open:

- Open data should be available to everyone to access, use, and share, without licences, copyright, or patents. It is expected that open data at most should be subject to attribution/share-alike licenses.

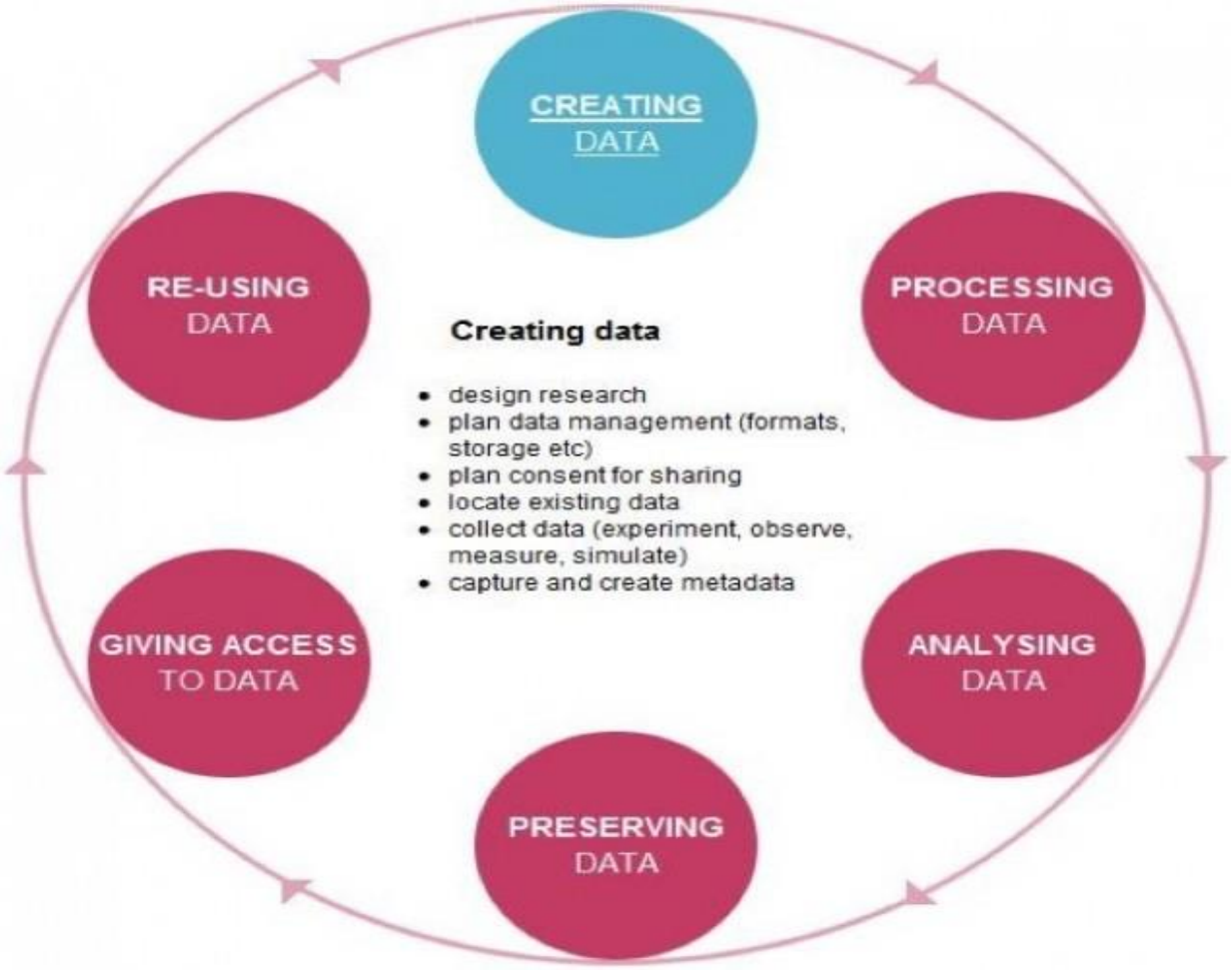
FAIR data uses “**Findable**” where data should be able to be found by appropriate people at appropriate times. This can include shared folders, drives, private databases, public databases or more. It really depends on what part of the data life cycle the data is currently in.

FAIR data uses "**Accessible**" to mean accessible by appropriate people, at an appropriate time, in an appropriate way.

- Data can be FAIR when it is private, when it is accessible by a defined group of people, or when it is accessible by everyone (open data).

FAIR data uses “**Interoperable**” to refer more to how the data is formatted (e.g. standard formatting), whether the software for interpreting/interrogating/using the data is available (e.g. freely, with a license etc.)

The data cycle



HRB facilitating the transition to FAIR data in health research

Funders' approaches

To pr ave
(or ar r
DMP

- Fu ith
  th rable
  ar

Many

HRB FAIR DMP pilot

Main stakeholders
- GFISFO office
- Researchers
- Data stewards
- VP/Deans research
- Research offices
- Irish funders
- NORF
- Other funders
- Science Europe
- DCC

13 Data stewards upskilled and engaged in FAIR and DMP

Phase I
Grant Applications

Q2/Q3 2018 - Q3 2019

Phase II
DMPs completion and review

Q3/4 2019 - Q3 2020

Phase III
Post award monitoring and end of grant

Q4 2020 – 2024/2025

HRB constant support to Data Stewards and facilitating discussions and integration into the system

Phase I
Grant Applications

Before submitting a grant application, Lead Applicants liaise with data stewards in Host Institution

They complete the
(1) *Outline for Data stewardship and*
(2) FAIR Data Management costs

Peer-Review and panel review stage

Application unsuccessful

31 Applications successful eligible for pilot

Award offered and
Contract negotiated

Awards start at different time (Sept 2019- April 2020

Outline on FAIR Data Management and Stewardship (500 words limit).

- Data description and collection or reuse of existing data
- Documentation and data quality
- Storage and backup
- Ethical and legal compliance, codes of conduct
- Data sharing and long-term preservation
- Data management responsibilities and resources

Eligible costs in funding budget

6. FAIR Data Management Costs	**Applicable to LA from institutions participating to the HRB Pilot only: Costs related to data management, FAIRification, storage and archiving of research data in line with best practice of data management and stewardship and the FAIR principles incurred during the lifetime of the project should be included..**

People	Staff time per hour for data collection, data anonymisation,
	staff time per hour for data management/stewardship support, training, etc
Storage and computation	cloud storage, domain hosting charge
Data access	secondary data access, costs for preparing data for sharing (eg anonymisation)
Deposition and reuse	costs for depositing research data and metadata in an open access data repository
	e.g. defining semantic models, making data linkable, choosing the licence, defining metadata for dataset, deploying/publishing
Others	Please further explain
Notes	The HRB is currently not covering the cost of long term preservation of data
	This list is not exhaustive and aims to provide examples only of eligible costs

Phase II
DMPs completion

31 Awards

Tools:
E. g. DMPonline (science Europe), DS wizard,

Lead Applicant is working closely with data stewards at HI and prepare full DMP

31 DMPs+ declarations submitted to HRB

HRB initial review of DMP

International panel to review a sample of DMPs

Data stewards Provide feedback to HRB and Panel on process, tools, challenges and facilitators

Panel feedback provided to HRB

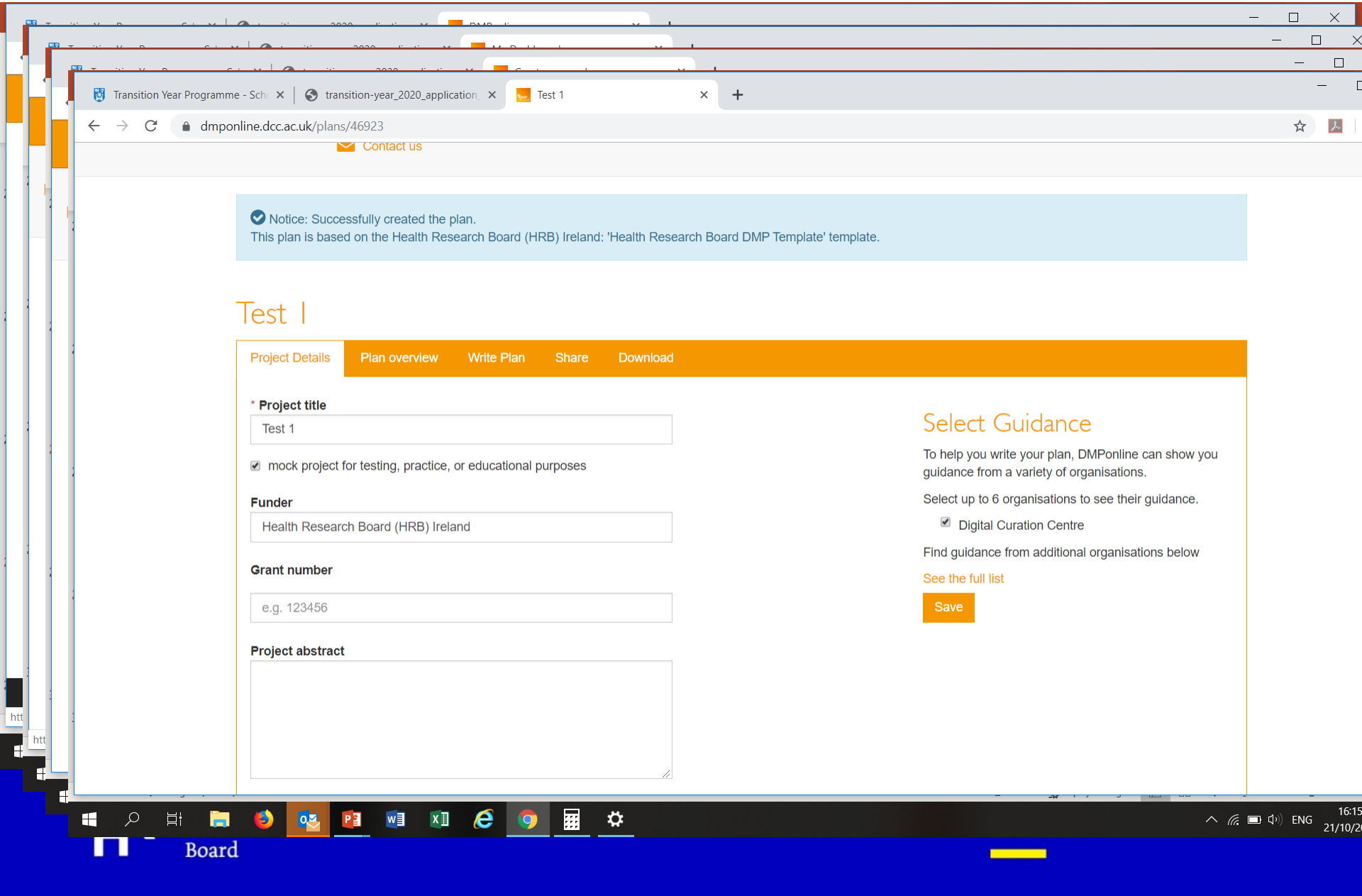
Report disseminated

Reflections and revisions

HRB data policy
And implementation refined

Collaboration with Digital Curation Centre on DMPs

- HRB currently using DMP online to host DMP template for researchers
- In discussion with other Irish research funders about creating a national platform for DMP


Transition Year Programme - Sch
transition-year_2020_application_
Test 1
dmponline.dcc.ac.uk/plans/46923
Contact us
Notice: Successfully created the plan.
This plan is based on the Health Research Board (HRB) Ireland: 'Health Research Board DMP Template' template.
Test 1
Project Details
Plan overview
Write Plan
Share
Download
* Project title
Test 1
mock project for testing, practice, or educational purposes
Funder
Health Research Board (HRB) Ireland
Grant number
e.g. 123456
Project abstract
Select Guidance
To help you write your plan, DMPonline can show you guidance from a variety of organisations.
Select up to 6 organisations to see their guidance.
Digital Curation Centre
Find guidance from additional organisations below
See the full list
Save
ENG
Board

Phase III
Post award monitoring and end of grant

Monitoring of budget

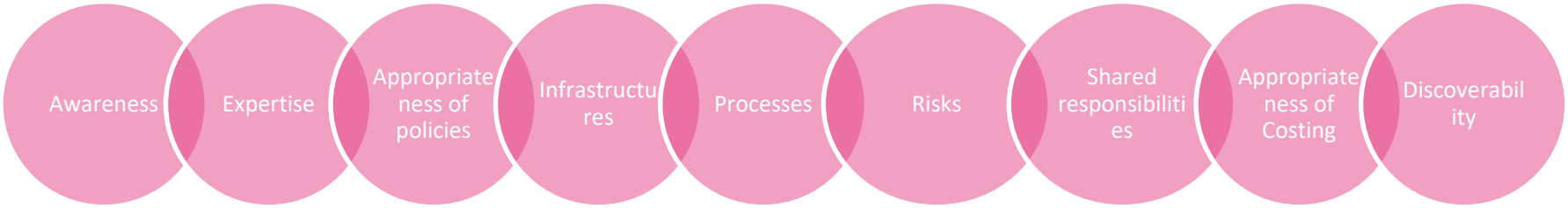
Final DMPs submitted at End of Grant

Outputs and Outcome collected

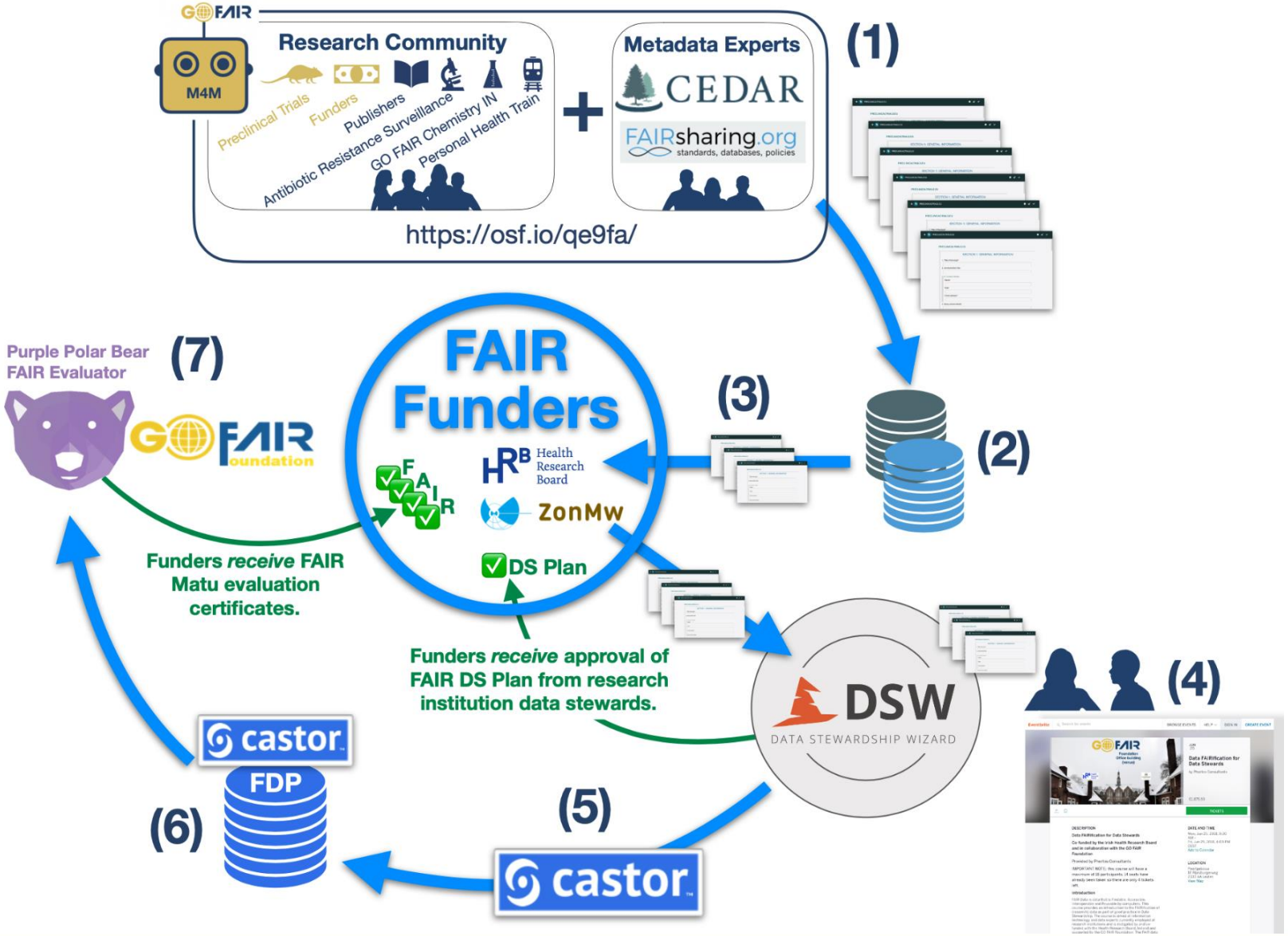
Feedback from data stewards

Lessons learnt

Outcomes – Increased…….



FAIR Funder Pilot: https://osf.io/b9fz4/



To summarise

- Research data are an important and expensive output of the scholarly research process, across all disciplines.
- Everybody has a role in ensuring that research can have the **greatest possible impact**
- Data management is an ongoing process and planning in the early stages makes the whole endeavour easier.
- Good research data management & stewardship is not a goal in itself, but rather the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse.

Thank you for listening

amontesanti@hrb.ie

hrb.ie

Find out more
about our work