



# Research Data Lifecycle and Machine-actionable Data Management Plans

Tomasz Miksa - TU Wien & SBA Research

Kevin Ashley - Digital Curation Centre, UK

[10.5281/zenodo.1066109](https://doi.org/10.5281/zenodo.1066109)

# Agenda

- › you learn about us
- › we learn about you
- › setting the scene for exercise 1
- › exercise 1 - stakeholders needs within research data lifecycle
- › coffee break
- › setting the scene for exercise 2
- › exercise 2 - flow of information between stakeholders
- › Wrap-up

# you learn about us

# What is RDA

- › Research Data Alliance
  - › community-driven organization
  - › 6,000 members from 130 countries
  - › different stakeholders
- › Plenary meetings
- › Interest Groups (IGs)
  - › Active DMPs
- › Working Groups (WGs)
  - › DMP Common Standards



# Active DMPs Interest Group

## › Goals

- › specifying practical tools and services to support
  - › data creators and their organisations in managing and making data re-usable
  - › data managers and funders in administering and monitoring

## › Group collaborates with other groups on:


- › certification
- › domain repositories
- › metadata
- › preservation e-infrastructures.

› 255 participants from all over the world

› Specific challenges addressed by Working Groups

# DMP Common Standards WG

- Result of a consultation made by Active DMPs IG
- Focus on machine-actionable DMPs
  - living documents
  - automate data management
    - collect information from systems
    - trigger actions in systems
  - facilitate validation
- Launched in October 2017
- 58 members
  - Australia
  - Asia
  - Africa
  - Europe
  - North America



The screenshot shows the RIO website interface. At the top, there is a navigation bar with links for 'Articles', 'About', 'About Person(s)', 'Books', 'E-Books', 'Blog', and 'Journals'. Below this, a header section identifies the document as a 'Workshop Report' with the title 'Machine-actionable data management plans (maDMPs)' and a DOI of 'https://doi.org/10.3897/rio.3.e13086'. The authors listed are 'Stephanie Simms, Sarah Jones, Daniel Mletchen, Tomasz Miksa'. The 'Abstract' section states that the report presents outputs from the International Digital Curation Conference 2017 workshop on machine-actionable data management plans. The 'Keywords' section includes terms like 'interoperability', 'persistent identifiers', 'research data management', 'research workflows', 'data discovery', 'research infrastructure', 'research policy', and 'research funding'. The 'Date and place' section notes the workshop took place on February 20 in Edinburgh, UK. The 'Aims of the workshop' section describes the goal of understanding stakeholder requirements and developing specific use cases for machine-actionable data management plans.

<https://doi.org/10.3897/rio.3.e13086>

# DMP Common Standards - Outputs

- **Common data model for machine-actionable DMPs**
  - to model information from standard DMPs
  - NOT a template
  - NOT a questionnaire
  - modular design
    - core set of elements
    - domain specific extensions
- **Reference implementations**
  - ready to use models
    - JSON, XML, RDF, etc.
- **Guidelines for adoption of the common data model**
  - requirements for supporting systems
  - pilot studies



## DMP Common Standards WG

Home » Working And Interest Groups » Working Group » DMP Common Standards WG

WG

### Group details

**Status:** Recognised & Endorsed

**Secretariat Liaison:** Lynn Yarmey

**TAB Liaison:** Wenbo Chu



WGs Getting started (~0-6 months after RDA endorsement)

— History —

The need for establishing this working group was articulated during the 9th plenary meeting in Barcelona during the Active DMPs IG session. The discussion was framed by a white paper by Simms et al. on machine-actionable data management plans (DMPs). The white paper is based on outputs from the IDCC workshop held in Edinburgh in 2017 that gathered almost 50 participants from Africa, America, Australia, and Europe. It describes eight community use cases which articulate consensus about the need for a common standard for machine-actionable DMPs (where machine actionable is defined as "information that is structured in a consistent way so that machines, or computers, can be programmed against the structure")

The specific focus of this working group is on developing common information model and specifying access mechanisms that make DMPs machine-actionable. The outputs of this working group will help in making systems interoperable and will allow for automatic exchange, integration, and validation of information provided in DMPs, for example, by checking whether a provided PID links to an existing dataset, if hashes of files match to their provenance traces, or whether a license was specified. The common information models are NOT intended to be prescriptive templates or questionnaires, but to provide re-usable ways of representing machine-actionable information on themes covered by DMPs.

The vision that this working group will work to realise is one where DMPs are developed and maintained in such a way that they are fully integrated into the systems and workflows of the wider research data management environment. To achieve this vision **we will develop a common data model with a core set of elements**. Its modular design will allow customisations and extensions using existing standards and vocabularies to follow best practices developed in various research communities. We will **provide reference implementations of the data model using popular formats**, such as JSON, XML, RDF, etc. This will enable tools and systems involved in processing research data to read and write information to/from DMPs. For example, a workflow engine can add provenance information to the DMP, a file format characterization tool can supplement it with identified file formats, and a repository system can automatically pick suitable content types for submission and later automatically identify applicable preservation strategies.

### DMP Common Standards WG

**Status:** Recognised & Endorsed

**Chair(s):**

..

**Group Email:**

**Secretariat Liaison:** Lynn Yarmey

**TAB Liaison:** Wenbo Chu

Public - accessible to all site users

[Join Group](#)

Index

Add new content

[Click here to create a wiki index for this group.](#)

[Group Mailing list Archive](#)

### Case Statement

RDA WG DMP Common Standards Case Statement  
Comments 2

### Outputs & Recommendations

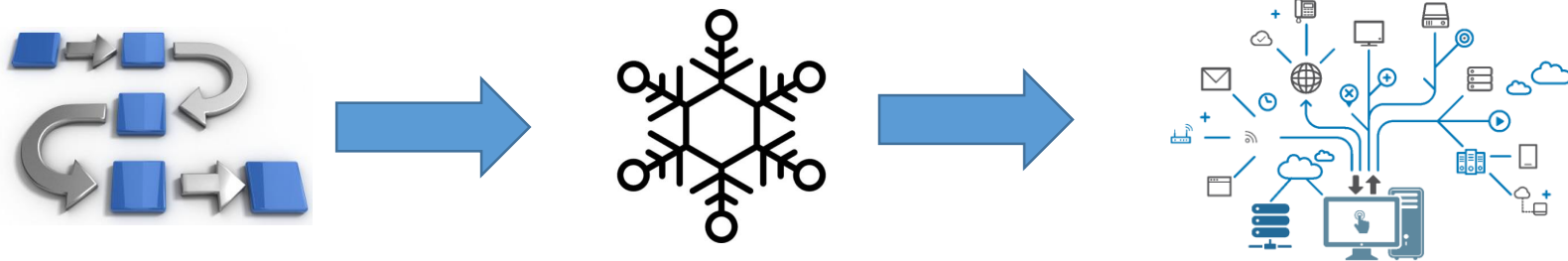
[Create new Outcomes](#)

<https://www.rd-alliance.org/groups/dmp-common-standards-wg>



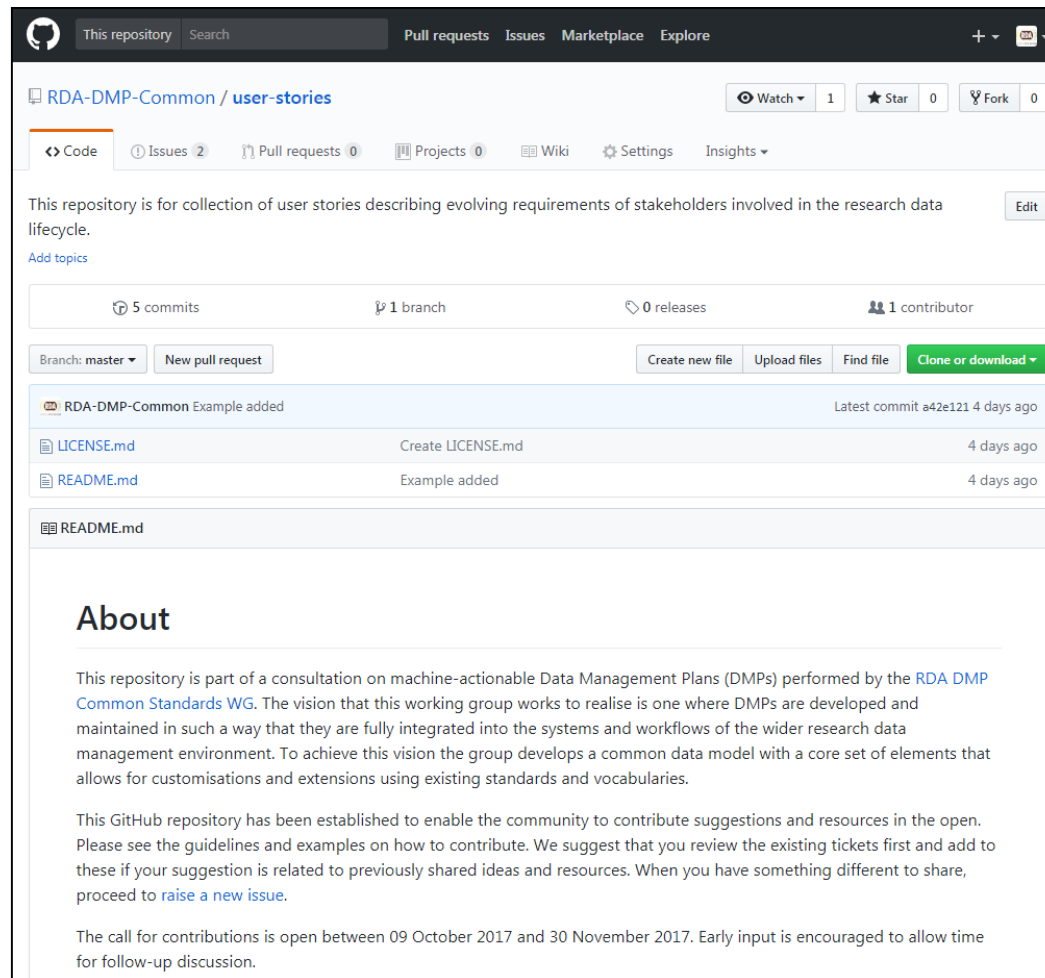
# DMP Common Standards - activities

- Open consultation on GitHub
  - define user stories for machine-actionable DMPs
  - define expectations
  - everyone can contribute
    - write new user stories
    - review existing
- Workshop(s)



# GitHub consultation

> <https://github.com/RDA-DMP-Common/user-stories>



The screenshot shows the GitHub repository page for `RDA-DMP-Common / user-stories`. The repository is currently on the `master` branch. It has 5 commits, 1 branch, 0 releases, and 1 contributor. The repository description states: "This repository is for collection of user stories describing evolving requirements of stakeholders involved in the research data lifecycle." The commit history shows three recent commits: "RDA-DMP-Common Example added" (latest commit `a42e121`, 4 days ago), "LICENSE.md" (Create LICENSE.md, 4 days ago), and "README.md" (Example added, 4 days ago). The `README.md` file is open, showing the "About" section. The text in the `README.md` describes the repository's purpose as part of a consultation on machine-actionable Data Management Plans (DMPs) performed by the RDA DMP Common Standards WG. It outlines the vision of fully integrated DMPs and provides instructions for contributing to the repository, including a call for contributions from October 9 to November 30, 2017.

# Workshop objectives

- › identify stakeholders at each lifecycle stage
  - › e.g. who can provide cost estimations when planning?
- › identify how **available** information changes over the lifetime of a DMP
  - › e.g. when exact size of data is important and when not?
- › how **need for** information changes over the lifetime of a DMP
  - › e.g. is a persistent identifier important to funders in the planning phase or when a project ends?

- › This workshop should also help you
  - › define data policies
    - › roles of stakeholders involved in data management
    - › services and infrastructure needed
    - › potential for automation
  - › find overlaps between packages and partners in the e-infrastructure's austria+ project

e-infrastructures  
austria



learn

we learn about you

# Organization

- › Universität Wien
- › Technische Universität Wien
- › The Austrian Social Science Data Archive TU
- › ZID
- › FWF - Der Wissenschaftsfonds
- › Wirtschaftsuniversität Wien
- › Medizinische Universität Wien
- › University of Innsbruck

# Role in the Organization

- › Repository Manager
- › Data processing specialist
- › Research Data Manager
- › Research support
- › Data Processing Specialist
- › Open Access Manager
- › Research Administrator
- › Research Service
- › IT services
- › Cataloguer

# Introduction round

- › Your name
- › Your location
- › Your role
- › What's the most important thing you want to get out of this meeting?





# Setting the scene

# Data Management Plans (DMPs)

- › manually created text documents
- › considered as bureaucracy
- › created too late
- › vague
- › depend on human factor
  - › scrupulousness
  - › awareness



# Data Management Plans



How to discover these tools?  
Which one do I need to use?  
Why do I have to provide the same  
information again?

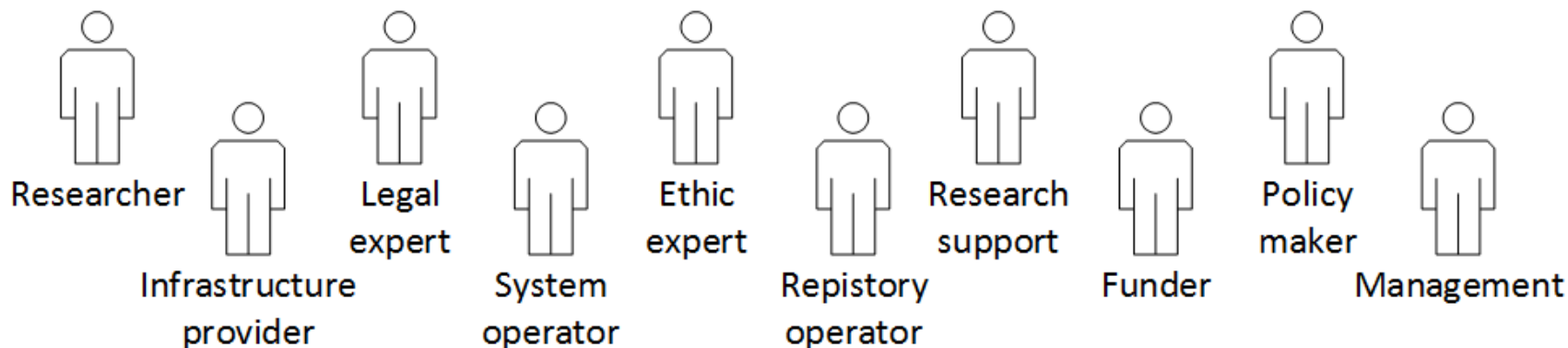


Why haven't they consulted us before?  
Who is going to pay for this?  
We don't have enough people for that!

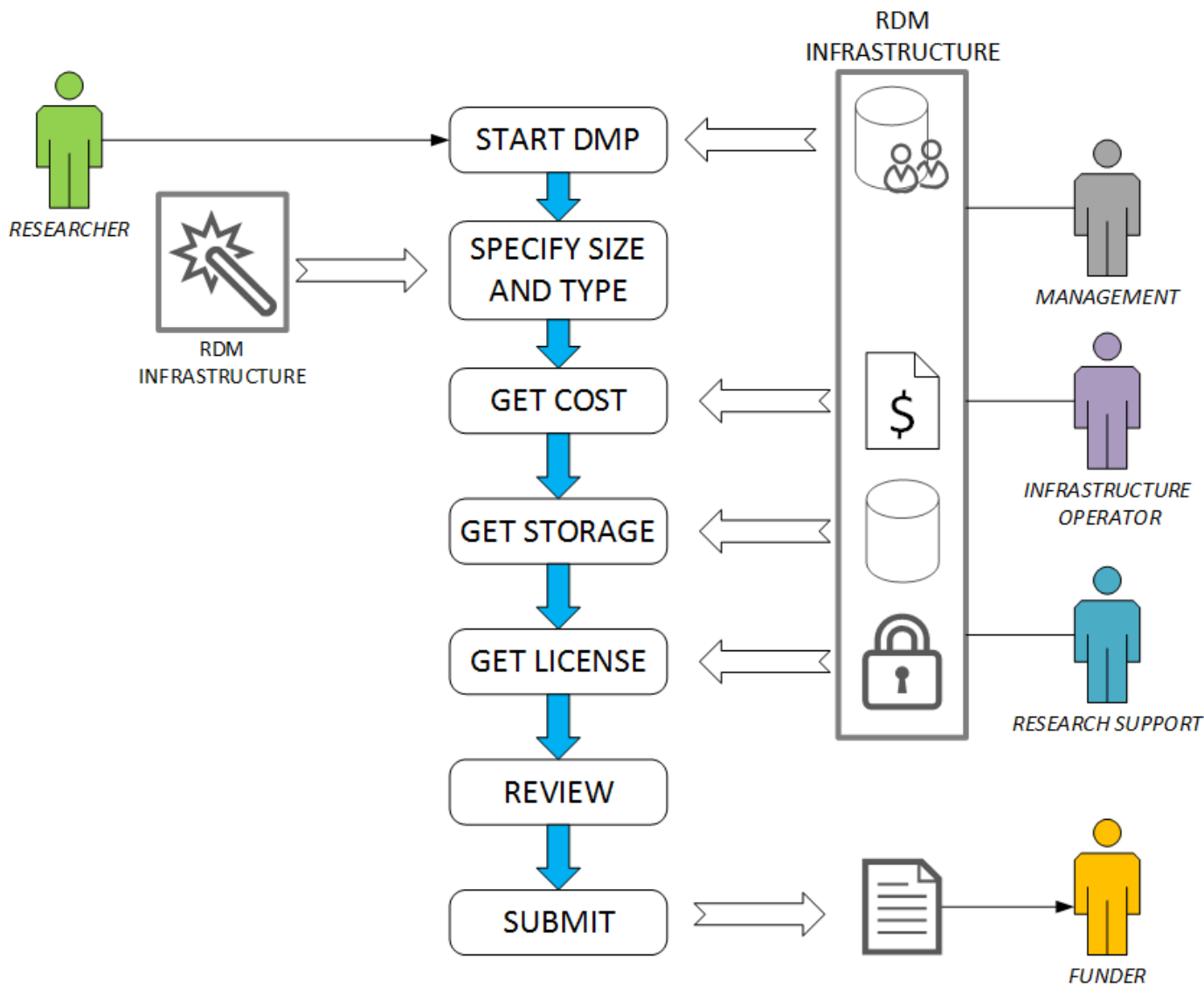


# Research data lifecycle

- › Stakeholders involved in research data management
  - › require information at certain stages
  - › can provide information if requested at a proper stage
- › Many problems can be avoided when
  - › timing is right
  - › information flow is ensured



# Automated Data Management Workflow



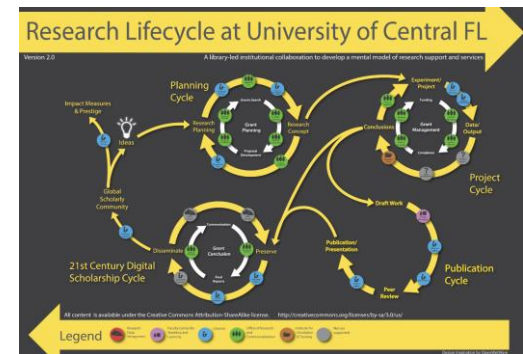
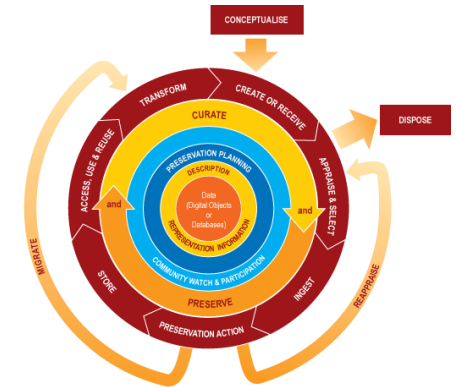
# Machine-actionable DMPs

- › living documents
- › automate data management
  - › collect information from systems
  - › trigger actions in systems
- › facilitate validation
- › This requires
  - › well-defined RDM workflows
  - › data management infrastructure
  - › common data model

```
"dc:creator":[ {  
  "foaf:name":"John Smith",  
  "@id":"orcid.org/0000-1111-2222-3333",  
  "foaf:mbox":"mailto:jsmith@tuwien.ac.at",  
  "madmp:institution":"AT-TU-Wien"  
}],
```

# Research Data Lifecycle Models

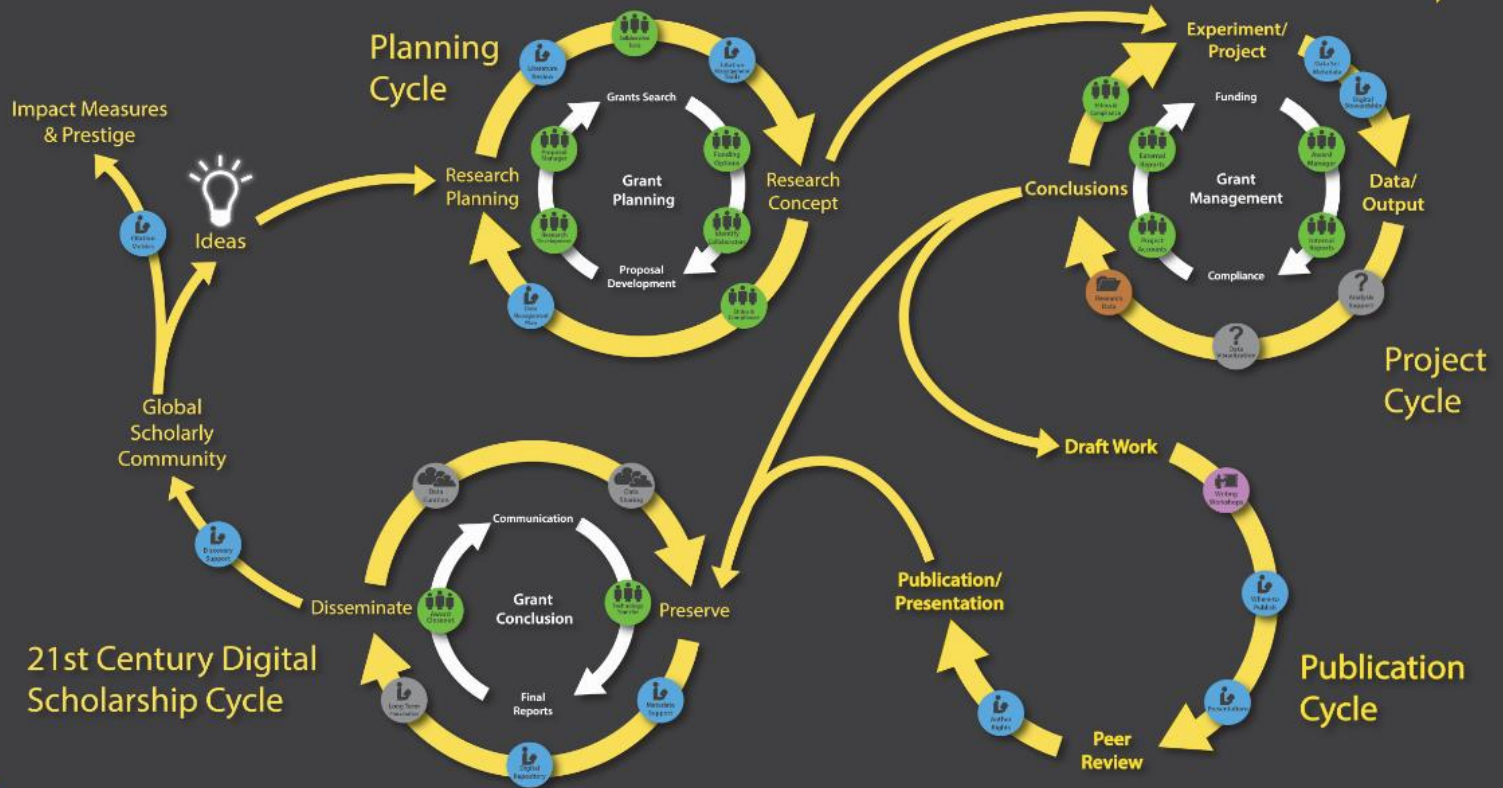
- Describe roles of stakeholders
- Help in
  - tailoring services
  - identifying responsibilities
  - defining infrastructure
- NOT to be used by researchers
- Examples
  - Digital Curation Centre
  - University of Central Florida



# Research Lifecycle at University of Central FL

Version 2.0

A library-led institutional collaboration to develop a mental model of research support and services



All content is available under the Creative Commons Attribution-ShareAlike license. <http://creativecommons.org/licenses/by-sa/3.0/us/>

## Legend



Design Inspiration by OpenWetWare



# Exercise 1

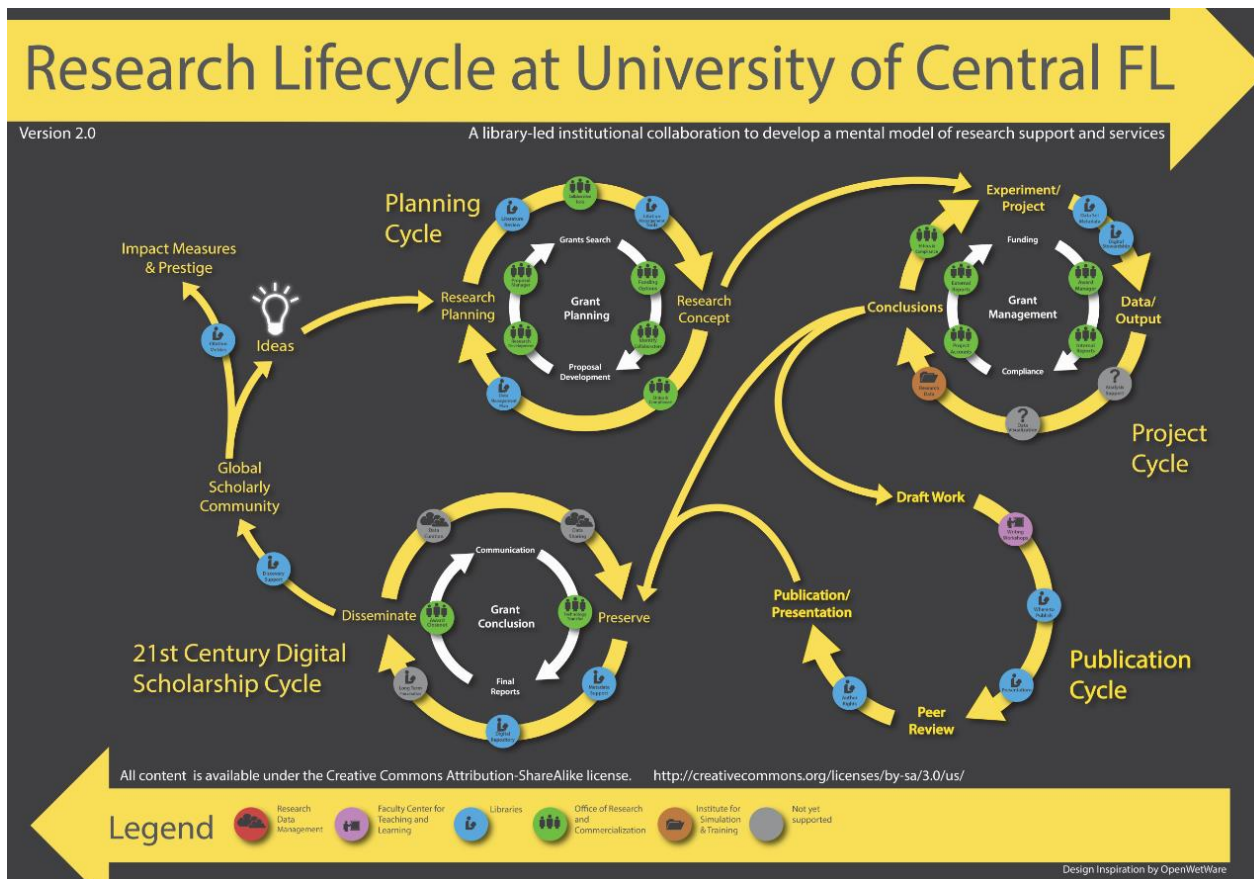
# Exercise description

- › Work in groups
- › Analyse sample data lifecycle
  - › UC Florida
  - › 4 phases
- › For each stage
  - › identify stakeholders
  - › write user stories
- › Time
  - › 4 x 10 minutes
- › Use sticky notes
- › Report afterwards

# User stories - examples

- *As a **<stakeholder>**, I want **<goal>** so that **<reason>**.*
- *As a **researcher**, I want to inform repository operator on the amount of data in the planning phase, so that they provide information on costs.*
- *As a **funder**, I want to have a list of DOIs for articles published within a research project, so that I can track how the grant money was spent.*
- *As a **repository operator**, I want to know the embargo periods for ingested data, so that I can restrict access to specific contents.*

# Planning cycle



# Reporting

# Coffee break

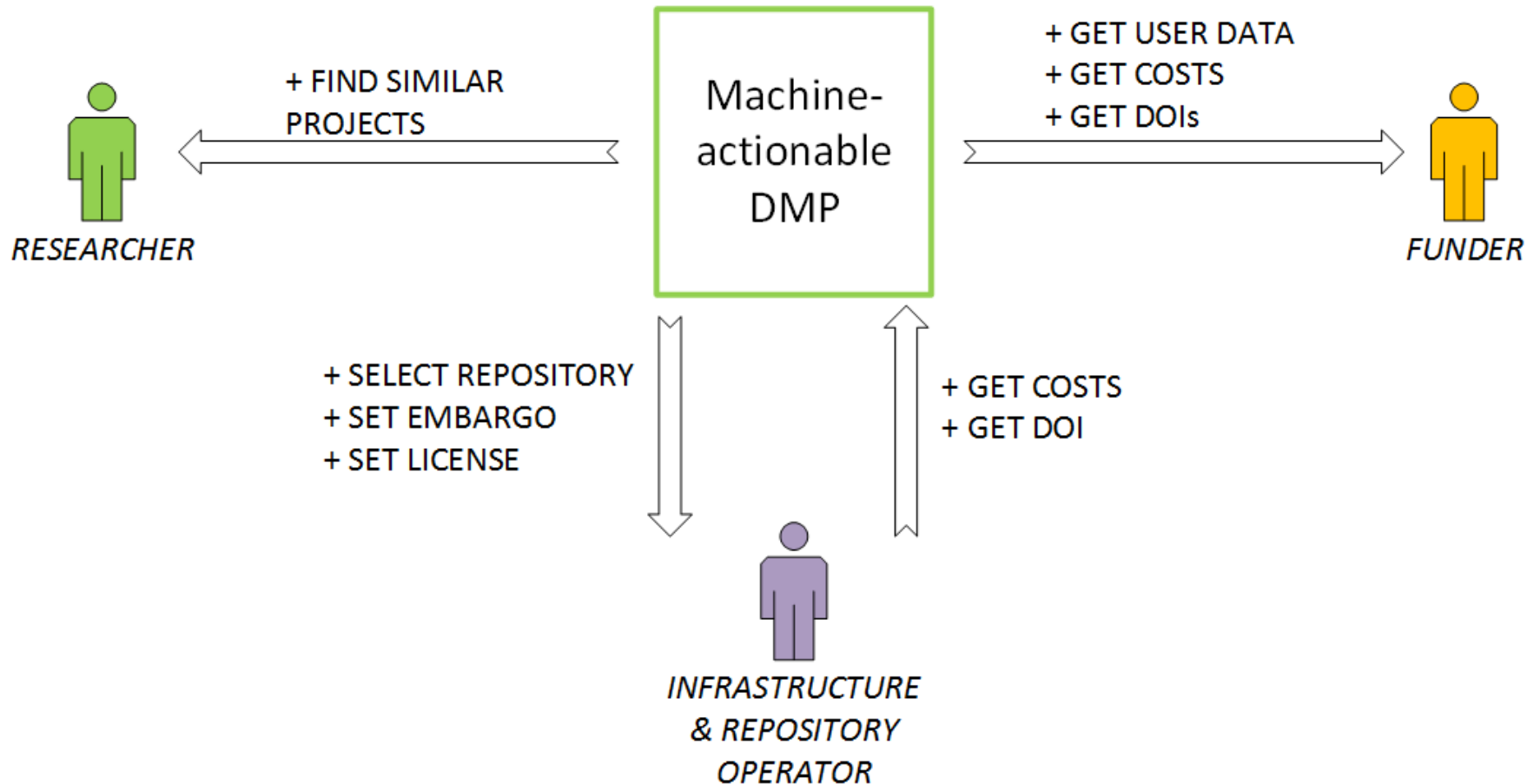
# Setting the scene

# Machine-actionable DMPs

- › living documents
- › automate data management
  - › collect information from systems
  - › trigger actions in systems
- › facilitate validation
- › This requires
  - › well-defined RDM workflows
  - › data management infrastructure
  - › common data model



# Machine-actionable DMPs



# Example (not the real model)

- Current DMPs – model questionnaires

```
<administrative_data>
```

```
  <question>Who will be the Principle Investigator?</question>
```

```
  <answer>The PI will be John Smith from our university.</answer>
```

```
</administrative_data>
```

- Machine-actionable DMPs – model information

```
"dc:creator":[ {
```

```
  "foaf:name":"John Smith",
```

```
  "@id":"orcid.org/0000-1111-2222-3333",
```

```
  "foaf:mbox":"mailto:jsmith@tuwien.ac.at",
```

```
  "madmp:institution":" AT-Vienna-University-of-Technology"
```

```
}],
```

# Example (not the real model)

- Currently available – not very useful

<administrative\_data>

<question>Who will be the Principle Investigator?</question>

<answer>The PI will be John Smith from our university.</answer>

Reuse existing standards, e.g. Dublin Core, PREMIS, etc.

able DMP

"dc:creator":[ {

"foaf:name":"John Smith",

"@id":"orcid.org/0000-1111-2222-3333",

"foaf:mbox":"mailto:jsmith@tuwien.ac.at",

"madmp:institution":"AT-Vienna-University-of-Technology"

}],

# Example (not the real model)

- Currently available – not very useful

```
<administrative_data>
```

```
  <question>Who will be the Principle Investigator?</question>
```

```
  <answer>The PI will be John Smith from our university.</answer>
```

```
</administrative_data>
```

- Machine-actionable DMP

Use PIDs whenever possible, e.g. ORCID

```
"dc:creator":[ {  
  "foaf:name":"John Smith",  
  "@id":"orcid.org/0000-1111-2222-3333",  
  "foaf:mbox":"mailto:jsmith@tuwien.ac.at",  
  "madmp:institution":"AT-Vienna-University-of-Technology"  
}],
```

# Example (not the real model)

- Currently available – not very useful

<administrative\_data>

<question>Who will be the Principle Investigator?</question>

<answer>The PI will be John Smith from **our university**.</answer>

</administrative\_data>

- Machine-actionable DMP

```
"dc:creator":[ {  
  "foaf:name":"John Smith",  
  "@id":"orcid.org/0000-1111-2222-3333",  
  "foaf:mbox":"mailto:jsmith@tuwien.ac.at",  
  "madmp:institution":"AT-Vienna-University-of-Technology"  
}],
```

Use controlled  
vocabularies

# Example (not the real model)

- Currently available – not very useful

```
<administrative_data>
```

```
  <question>Who will be the Principle Investigator?</question>
```

```
  <answer>The PI will be John Smith from our university.</answer>
```

```
</administrative_data>
```

- Machine-actionable DMP

```
"dc:creator":[ {  
  "foaf:name":"John Smith",  
  "@id":"orcid.org/0000-0001-1111-2222-3333",  
  "foaf:mbox":"mailto:jsmith@tuwien.ac.at",  
  "madmp:institution":"AT-Vienna-University-of-Technology"  
}],
```

Develop own  
concepts and  
vocabularies only  
when needed

# DMP Common Standards

- Identify existing standards, models, vocabularies, ontologies, ...
- Analyse how they can model information from existing DMPs
  - use DCC DMP themes as benchmark
- Develop our own model
- Run pilots and use it!
- We have on board
  - DMP Online
  - DMP Tool
  - DM Organizer
  - DM Wizard

# DCC DMP Themes

- › DATA DESCRIPTION
- › DATA FORMAT
- › DATA VOLUME
- › DATA COLLECTION
- › METADATA & DOCUMENTATION
- › ETHICS & PRIVACY
- › INTELLECTUAL PROPERTY RIGHTS
- › STORAGE & SECURITY
- › DATA SHARING
- › DATA REPOSITORY
- › PRESERVATION
- › ROLES & RESPONSIBILITIES
- › BUDGET
- › RELATED POLICIES



# Exercise 2

# Exercise description

- › Work in groups
- › Analyse DMP themes
  - › **who** are the stakeholders that need information on the particular theme?
  - › **when** is the **information needed** and **by whom**?
  - › **who** are the stakeholders that **can provide** expertise on this?
- › Time
  - › 30 minutes
- › Use flipcharts and pens
- › Report afterwards

# Exercise 2 - example

THEME	PROVIDES INFORMATION	WHEN PROVIDES	NEEDS INFORMATION	WHEN NEEDS	COMMENTS (SYSTEMS, MODELS, EXAMPLES)
DATA VOLUME	RESEARCHER	PLANNING	REPOSITORY OPERATOR	DATA PUBLICATION	...
...	...	...	...	...	...

# Reporting

# Wrap-up

- › What can we automate?
- › Which services will always be manually performed?
- › How broad can we integrate?
  - › university-wise?
  - › country-wise?
  - › European-wise?
- › Which services can we share?



# Thank you!

<https://www.rd-alliance.org/groups/dmp-common-standards-wg>