DMP Common Standards WG at the 12th Plenary meeting in Gaborone

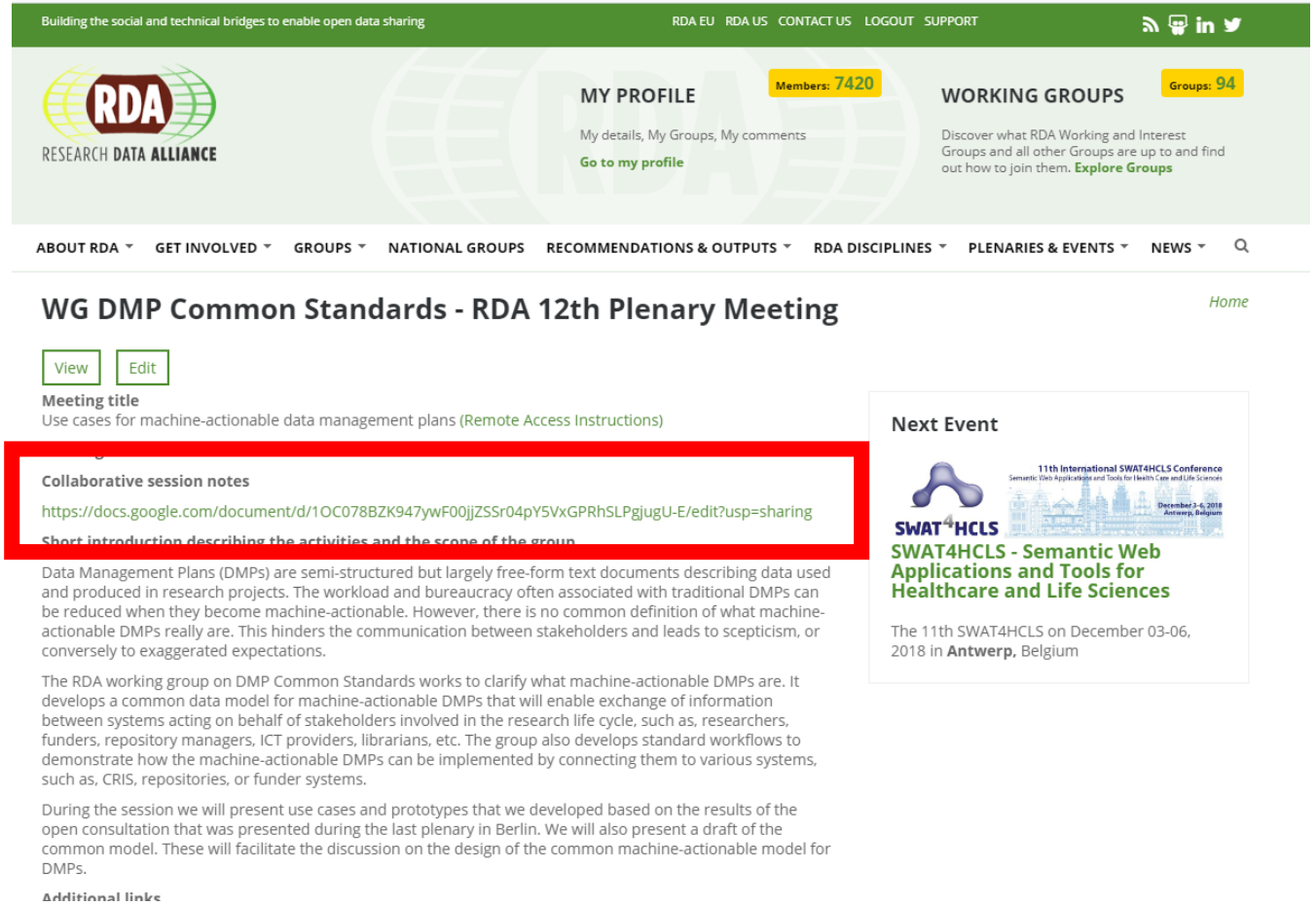
Tomasz Miksa, Paul Walk, Peter Neish

Agenda

- Part 1 – Introduction for newcomers
- Part 2 – Status update
  - Overview of two consultations performed
- Part 3 - Tools for machine-actionable DMPs
  - Examples of what can be automated
- Part 4 – Use cases and processes to include all stakeholders
  - How to define landscape of maDMPs
- Part 5 – Towards Common Data Model
  - Architecture and examples
- Part 6 – Wrap-up and next steps
- Exposing Data Management Plans WG

Collaborative notes

- https://www.rd-alliance.org/wg-dmp-common-standards-rda-12th-plenary-meeting



Introduction for newcomers

Part 1

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?

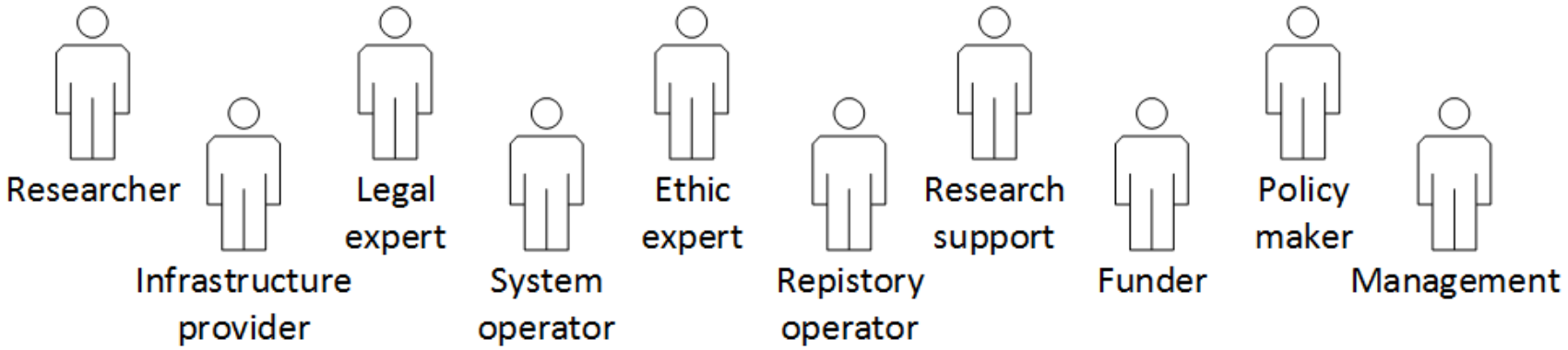
Why haven't they consulted us before?

Who is going to pay for this?

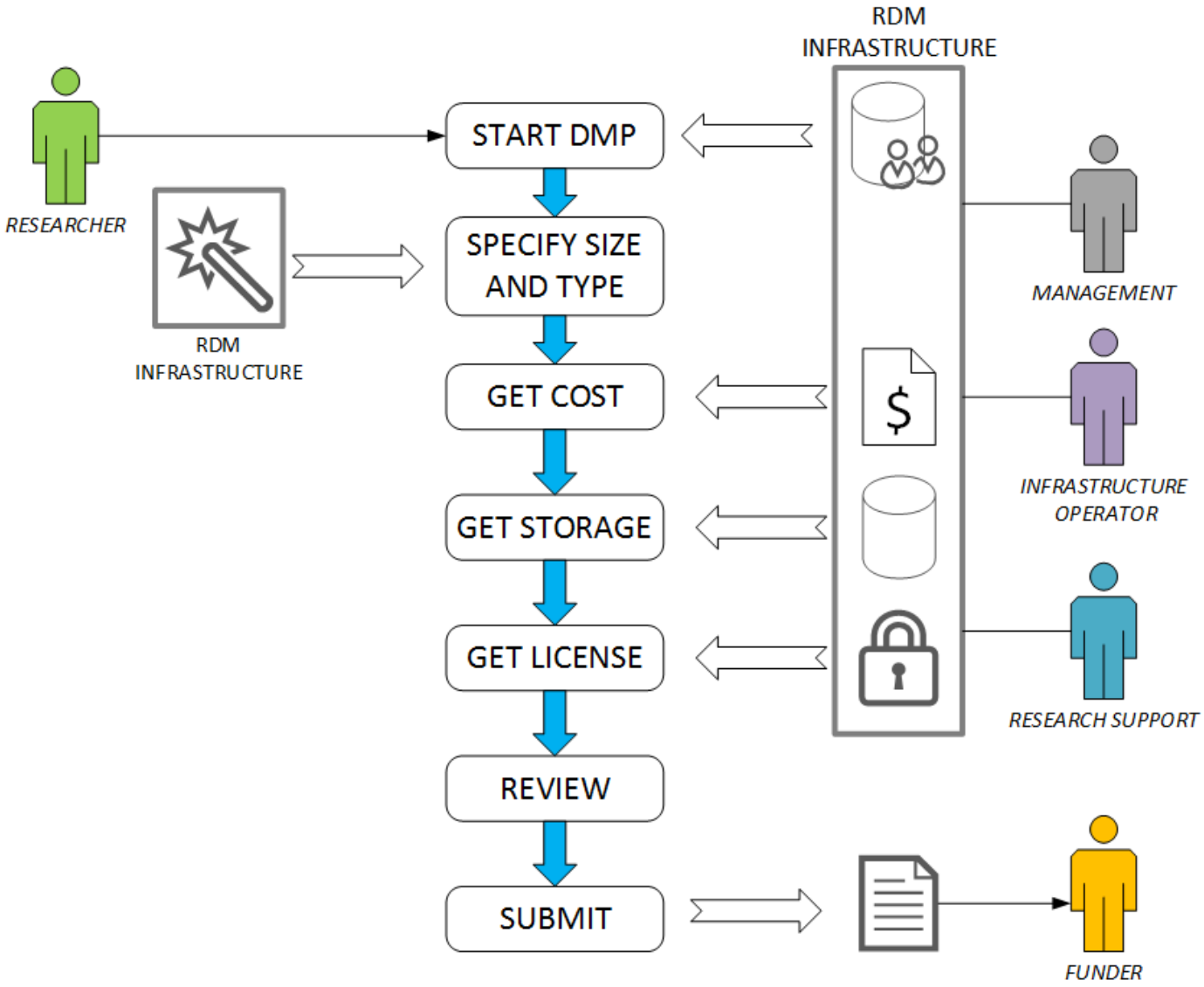
We don't have enough people for that!

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

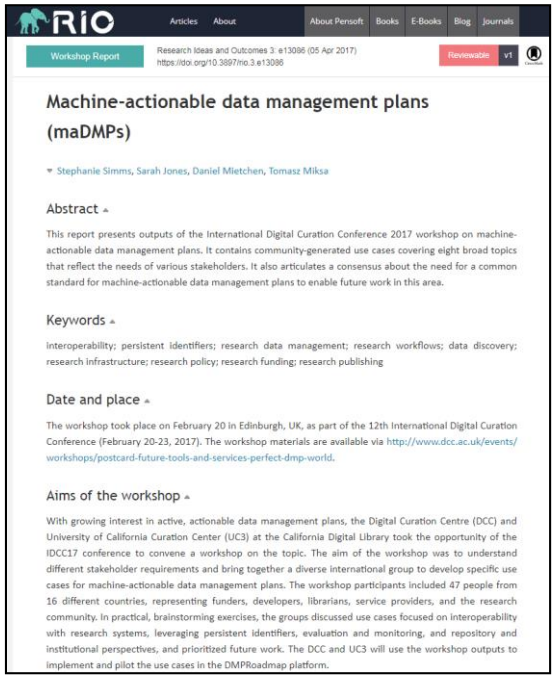


Automated Data Management Workflow



Why do we need this WG?

- Shortcomings of existing DMPs
  - manually completed, vague, not updated, …
- 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



RIO Articles About About Pensoft Books E-Books Blog Journals

Workshop Report

Research Ideas and Outcomes 3: e13086 (05 Apr 2017)
https://doi.org/10.3897/rio.3.e13086

Reviewable

Machine-actionable data management plans (maDMPs)

Stephanie Simms, Sarah Jones, Daniel Mietchen, Tomasz Miksa

Abstract

This report presents outputs of the International Digital Curation Conference 2017 workshop on machine-actionable data management plans. It contains community-generated use cases covering eight broad topics that reflect the needs of various stakeholders. It also articulates a consensus about the need for a common standard for machine-actionable data management plans to enable future work in this area.

Keywords

interoperability; persistent identifiers; research data management; research workflows; data discovery; research infrastructure; research policy; research funding; research publishing

Date and place

The workshop took place on February 20 in Edinburgh, UK, as part of the 12th International Digital Curation Conference (February 20-23, 2017). The workshop materials are available via http://www.dcc.ac.uk/events/workshops/postcard-future-tools-and-services-perfect-dmp-world.

Aims of the workshop

With growing interest in active, actionable data management plans, the Digital Curation Centre (DCC) and University of California Curation Center (UC3) at the California Digital Library took the opportunity of the IDCC17 conference to convene a workshop on the topic. The aim of the workshop was to understand different stakeholder requirements and bring together a diverse international group to develop specific use cases for machine-actionable data management plans. The workshop participants included 47 people from 16 different countries, representing funders, developers, librarians, service providers, and the research community. In practical, brainstorming exercises, the groups discussed use cases focused on interoperability with research systems, leveraging persistent identifiers, evaluation and monitoring, and repository and institutional perspectives, and prioritized future work. The DCC and UC3 will use the workshop outputs to implement and pilot the use cases in the DMPRoadmap platform.

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

Example

- 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

- 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

- 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

- 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

- 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"
 } ],
```

Develop own concepts and vocabularies only when needed

Status update

Part 2

Summary of actions till now

- 1st consultation (user stories) went broad
  - to define scope of maDMPs
- 2nd consultation went deep
  - to identify models for specific requirements
- Proof of concept tools
  - to demonstrate how model can be used to automate tasks
- BPMN processes
  - to identify systems and stakeholders involved

1st consultation – user stories

- Goals
  - identify stakeholders at each lifecycle stage
    - define which information they **provide**
    - define which information they **expect**

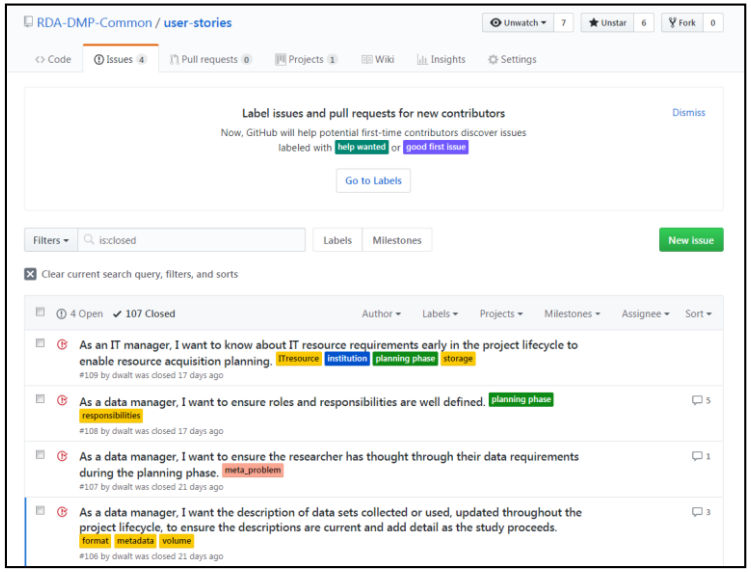
As a <stakeholder>, I want <goal> so that <reason >.

*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**.*

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

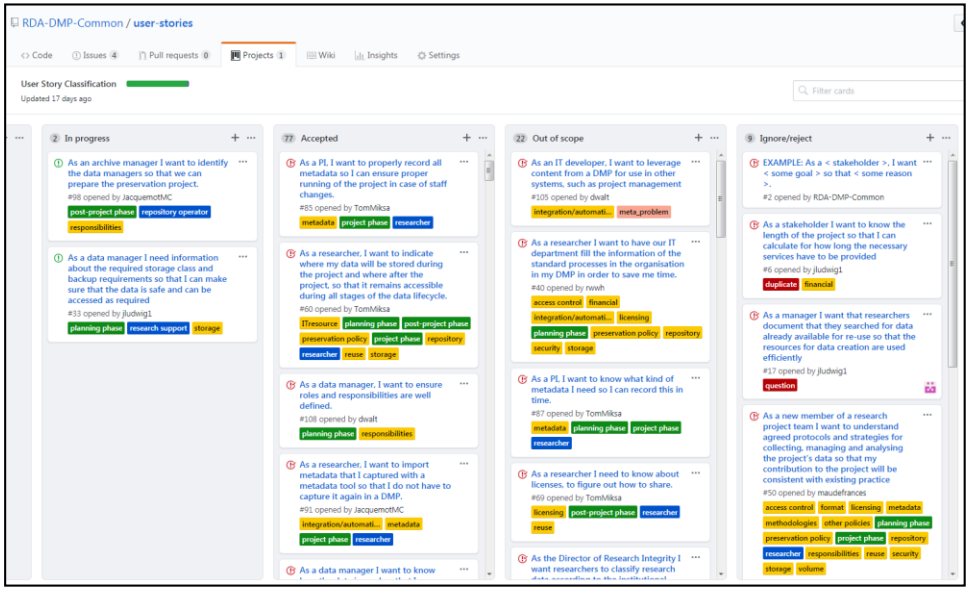
User story consultation

- https://github.com/RDA-DMP-Common/user-stories/
- 100+ issues defined
- inputs from Europe and Australia
- inputs from individuals and workshops

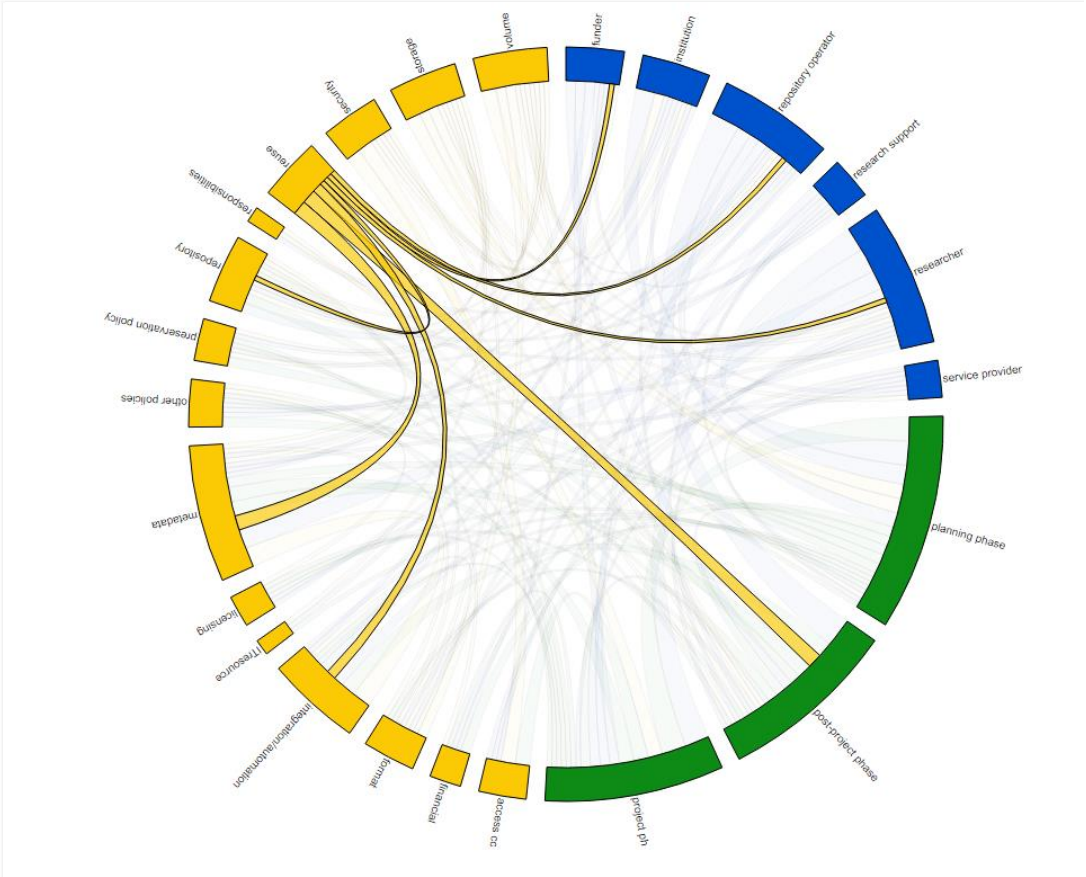


User story labelling

- https://github.com/RDA-DMP-Common/user-stories/projects/2
- https://github.com/RDA-DMP-Common/user-stories/wiki
- 3 major categories (colours)
  - stakeholders involved
  - project phase
  - subject of information conveyed
    - access control
    - volume
    - financial
    - licensing
    - metadata
    - repository
    - security
    - storage
    - etc.



User story visualisation



- https://goo.gl/znBL3F
- interactive visualisation - changes on GitHub are visible immediately
- shows relations between stakeholders, phases and information

Defining requirements for machine-actionable Data Management Plans

- Defines machine-actionability
- Describes results of user story consultation
- http://ifs.tuwien.ac.at/~miksa/papers/2018-iPres-maDMPs.pdf

From user stories to requirements

- https://docs.google.com/document/d/1sWVy0Rqj9fGsjs6GyFnBd3fH6XF2088zjK8U-1wLq4c/edit?usp=sharing
- Refactoring of user stories
- Goal: finding overlaps, gaps, duplicates
- Example below

- Metadata
  - taxonomy/classification [14,11]
  - Links to metadata of the real data [89, 39]
  - Funder information [7]
  - Link publications to data [55]
  - Authorship [88]
  - Multilingual metadata [65]
  - Include raw metadata directly in the model [91, 85]

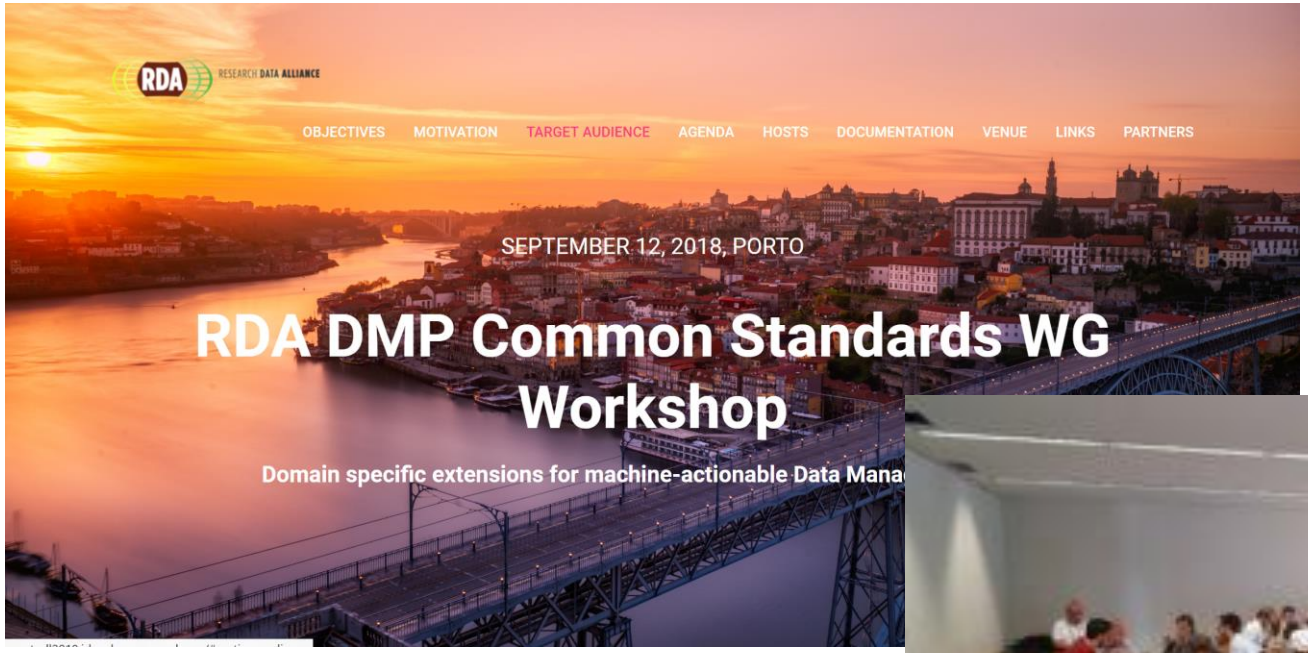
2nd consultation: From user stories to requirements

- https://docs.google.com/document/d/1mMJqmvqEAkbEWbdV7rtFU9hiQMOuH0ESn4Up_TDn1Es/edit?usp=sharing

- 5 documents to collect requirements, models, specific fields, etc.
  - **Administrative, Roles and Responsibilities**
  - **Data**
  - **Infrastructure**
  - **Security, Privacy and Access Control**
  - **Policies, legal and ethical aspects**

RDA DMP Common Standards WG Workshop

- http://rda-ws-tpdl2018.idsswh.sysresearch.org/#



Consultations summarised

- 1st consultation (user stories) went broad
  - helped us defined the scope of the maDMPs
    - what information should a maDMP contain?
    - who provides and uses this information?
- 2nd consultation goes deep (ongoing)
  - how do we model specific requirements
    - which specific fields are needed?
    - which models exist?

Tools for maDMPs

Part 3 - Examples of what we can automate with maDMPs

Horizon 2020 DMP survey report

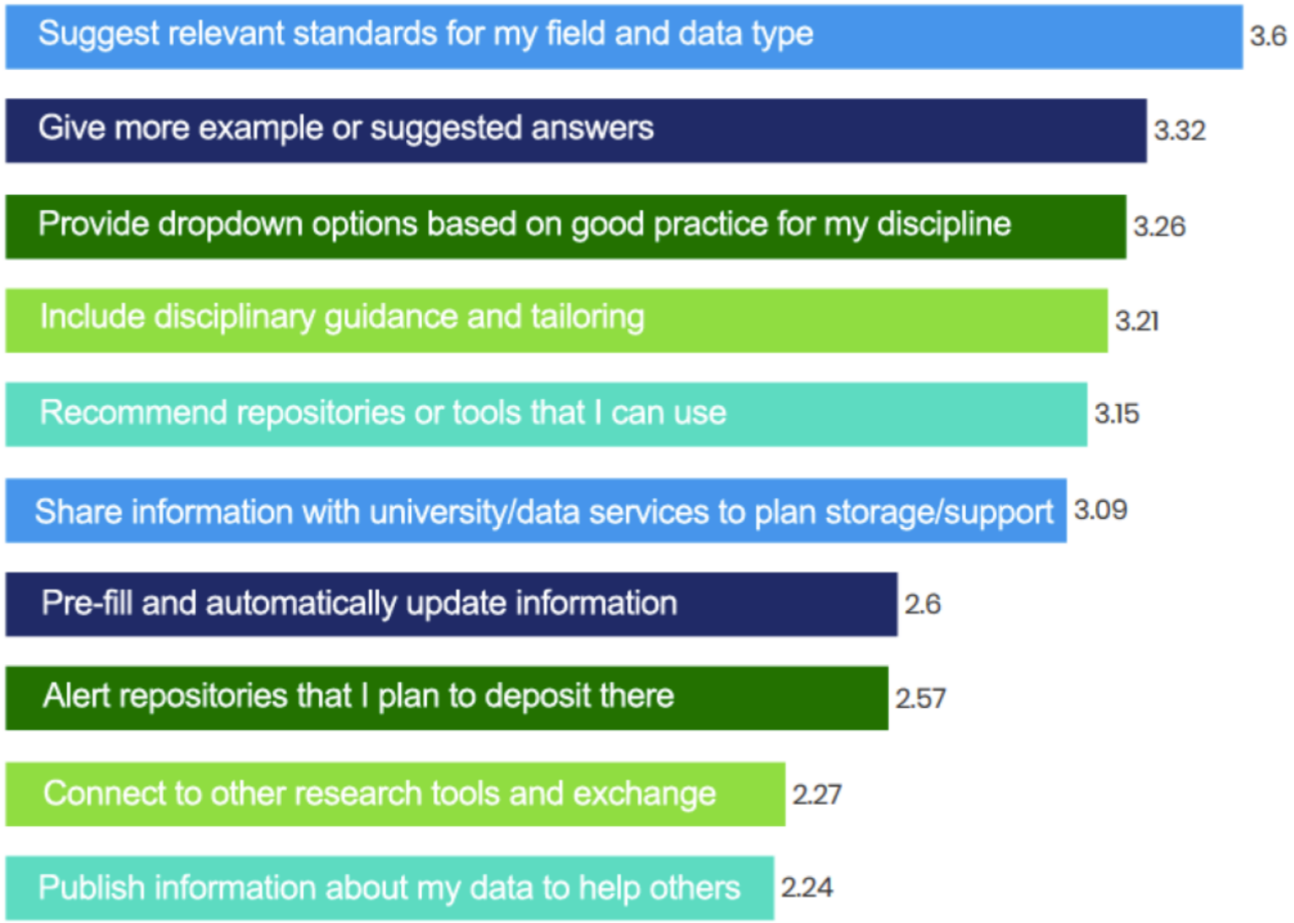


Figure 001 - Q11: priorities for a DMP template or tool

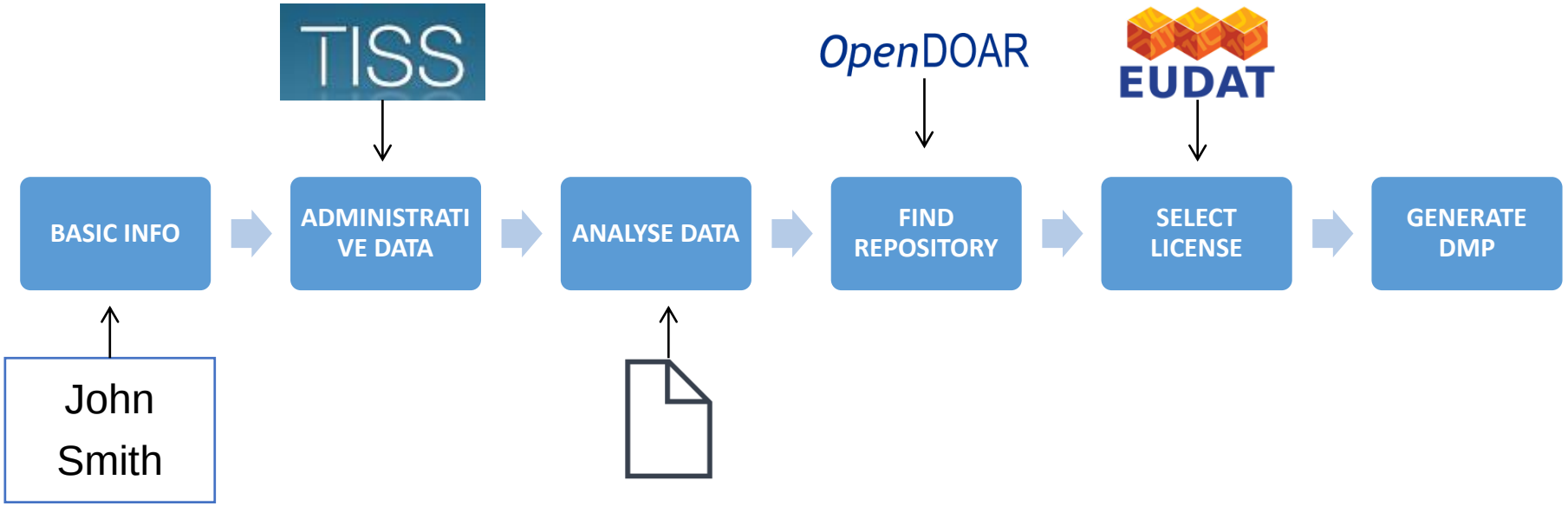
Marjan Grootveld, Ellen Leenarts, Sarah Jones, Emilie Hermans, & Eliane Fankhauser. (2018). OpenAIRE and FAIR Data Expert Group survey about Horizon 2020 template for Data Management Plans (Version 1.0.0) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.1120245

Prototypes developed by TU Wien students

- Requirements
  - Provide minimum input
  - Import as much as possible from existing systems to help in creating maDMPs
- Tools available as Docker containers on GitHub
  - https://github.com/TomMiksa/DMPGenerator
  - https://github.com/TomMiksa/digital_preservation_ex_1_2
  - https://github.com/TomMiksa/tu-dpue-lab2-ss18
  - https://github.com/TomMiksa/DigitalPreservation_2
  - https://github.com/TomMiksa/digitalpreservation-dmp-generator
  - https://github.com/TomMiksa/DMPlanner
- Example of a landing page for maDMPs
  - https://oblassers.github.io/fair-data-science/
  - https://github.com/oblassers/fair-data-science

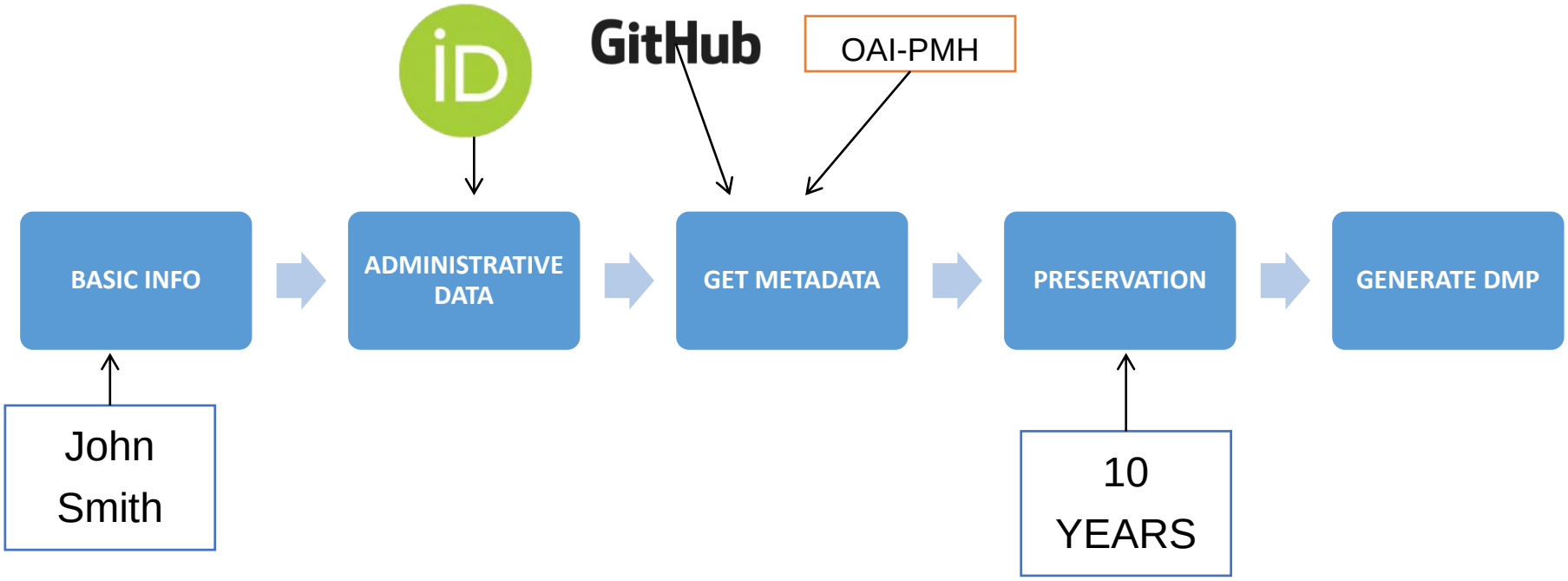
Planning phase

- Goal: get **estimations** and **recommendations** (which are feasible to implement later)



Project and Post-project phases

- Goal: **update** DMP with **real** information by **re-using** (linking) information provided elsewhere



Planning phase - demo

https://github.com/IrinaAvram/DMPGenerator

RESEARCH DATA

1. BASIC INFORMATION | 2. DATA ANALYSIS | 3. REPOSITORY AND LICENSING

Basic Information

Project Name*

TUW Report

Projektass.
Dr.techn.
Mag.

First Name

Tomasz

Last Name*

Miksa

Contact

Email: tomasz.miksa@tuwien.ac.at

Position

Projektassistent at Forschungsbereich Information und Software Engineering

NEXT

RESEARCH DATA ALLIANCE

1. BASIC INFORMATION | 2. DATA ANALYSIS | 3. REPOSITORY AND LICENSING

Upload Sample files

Choose Files 2018-TenRulesMADMPS.pdf

Name	Mime Type	Size	Input/Output	Nr. of Files	
Fig1.tif	image/tiff	1009010 Bytes	input	100	⊗
2018-TenRulesMADMPS.pdf	application/pdf	178186 Bytes	output	500	⊗

Required Storage space

Input: 100901000 Bytes
Output: 89093000 Bytes
Total:189994000 Bytes

COMPUTE REQUIREMENTS

Repositories

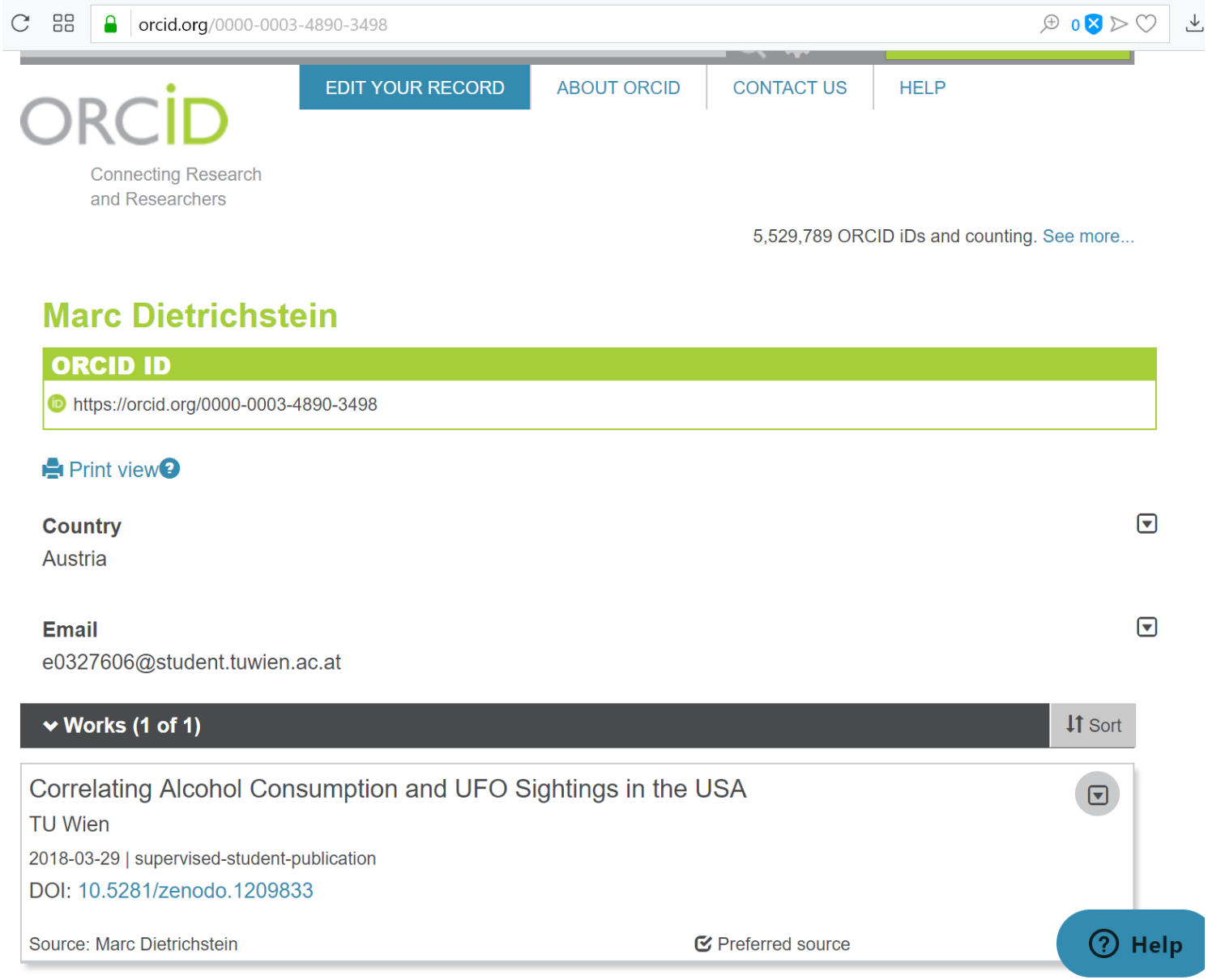
Architektur-Informatik http://architektur-informatik.scix.net/cgi-bin/works/Home

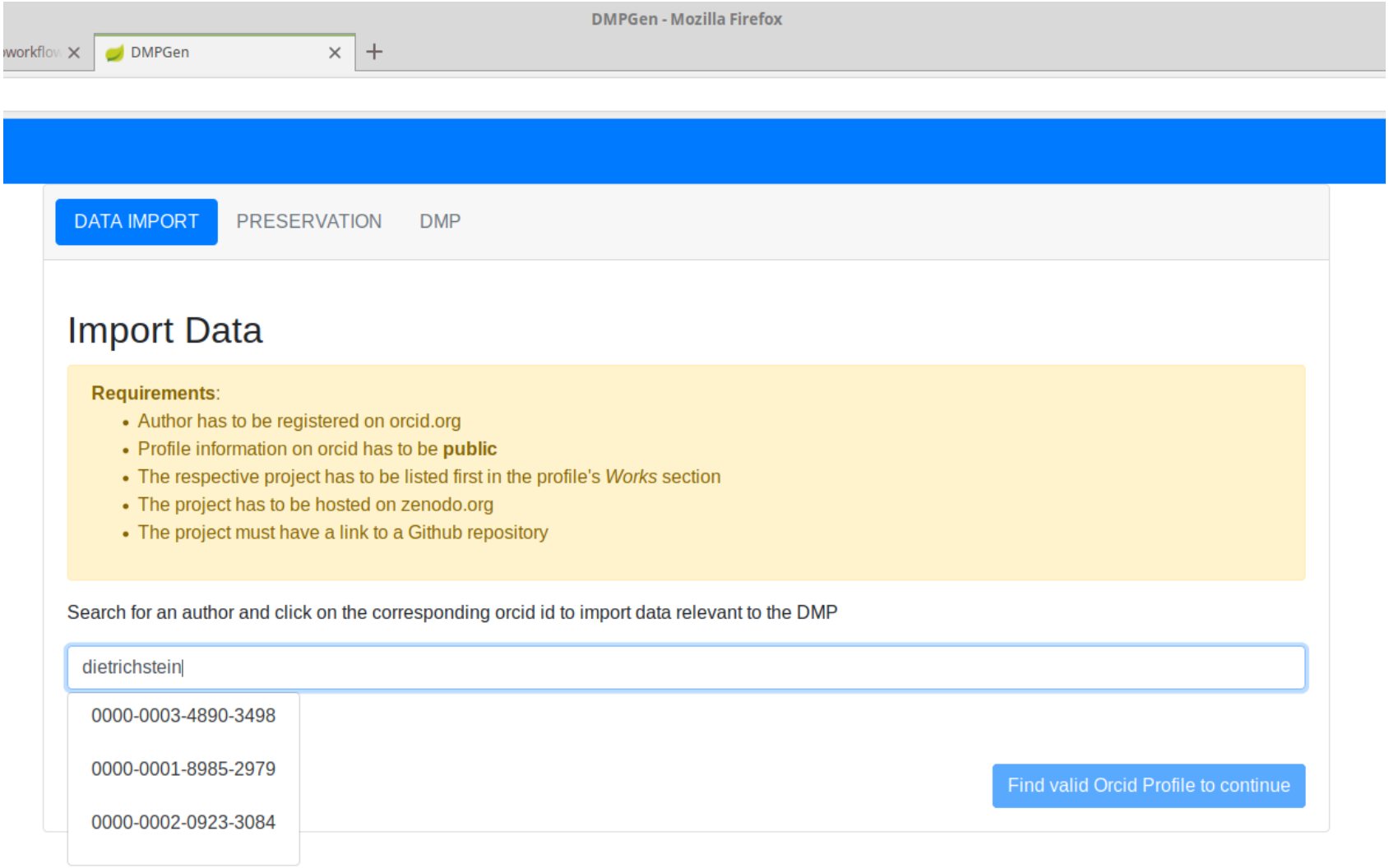
CumInCAD Digital Archive http://cumincad.architexturez.net/

Digitale Landesbibliothek Oberösterreichische http://digi.landesbibliothek.at/

Project and post-project phase - demo 1

https://github.com/mdietrichstein/digitalpreservation-dmp


orcid.org/0000-0003-4890-3498
EDIT YOUR RECORD
ABOUT ORCID
CONTACT US
HELP
Connecting Research and Researchers
5,529,789 ORCID iDs and counting. See more...
Marc Dietrichstein
ORCID ID
https://orcid.org/0000-0003-4890-3498
Print view
Country
Austria
Email
e0327606@student.tuwien.ac.at
Works (1 of 1)
Sort
Correlating Alcohol Consumption and UFO Sightings in the USA
TU Wien
2018-03-29 | supervised-student-publication
DOI: 10.5281/zenodo.1209833
Source: Marc Dietrichstein
Preferred source
Help


DMPGen - Mozilla Firefox
DMPGen
DATA IMPORT
PRESERVATION
DMP
Import Data
Requirements:
Author has to be registered on orcid.org
Profile information on orcid has to be public
The respective project has to be listed first in the profile's Works section
The project has to be hosted on zenodo.org
The project must have a link to a Github repository
Search for an author and click on the corresponding orcid id to import data relevant to the DMP
dietrichstein
0000-0003-4890-3498
0000-0001-8985-2979
0000-0002-0923-3084
Find valid Orcid Profile to continue

Search for an author and click on the corresponding orcid id to import data relevant to the DMP

Search for author on orcid

ORCID Author Info

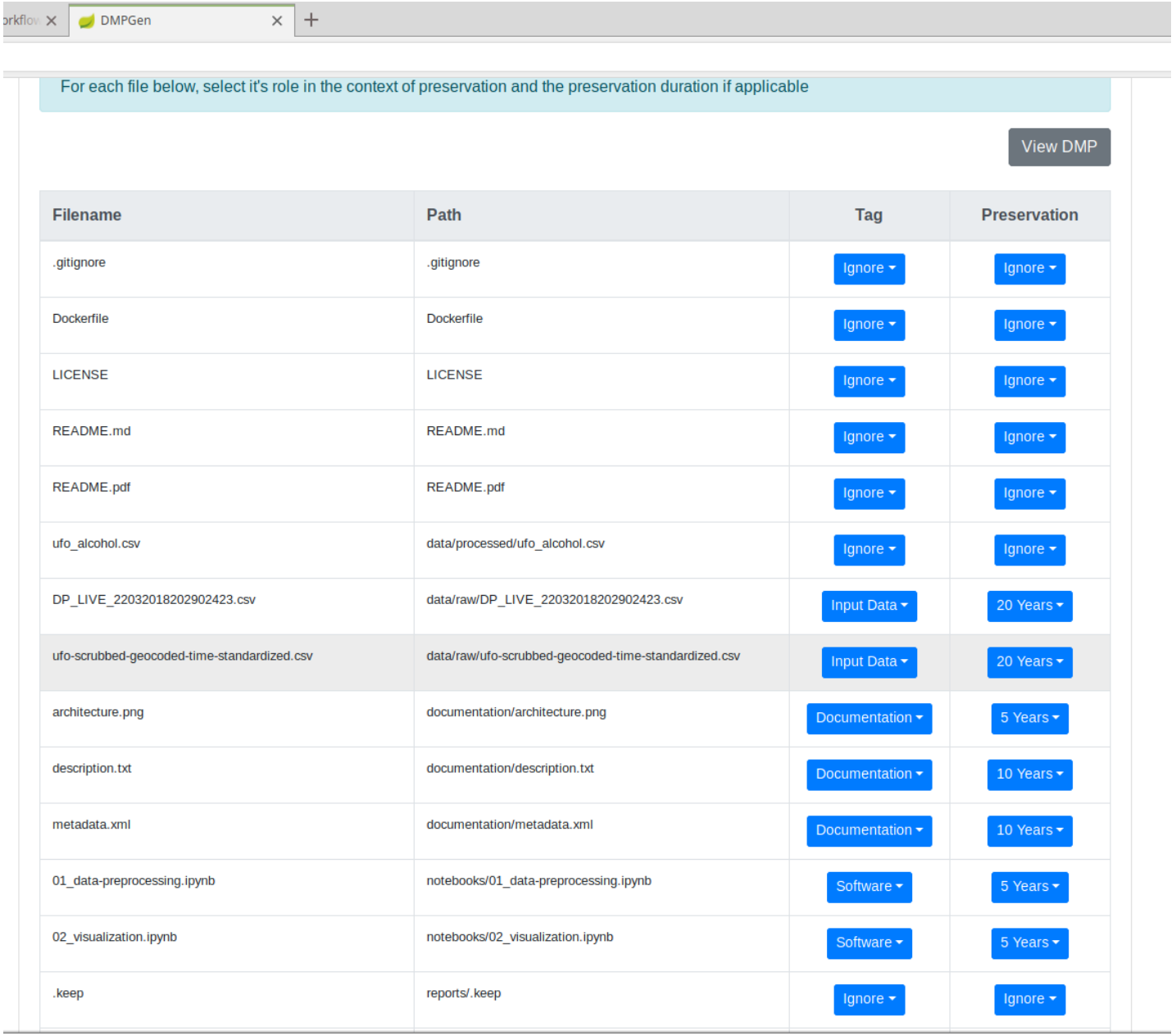
Orcid Id	0000-0003-4890-3498
Given Name	Marc
Family Name	Dietrichstein
E-mail	e0327606@student.tuwien.ac.at
Project Title	Correlating Alcohol Consumption and UFO Sightings in the USA
Project DOI	https://doi.org/10.5281/zenodo.1209833
Publication Date	2018/03/29

ORCID Author Info

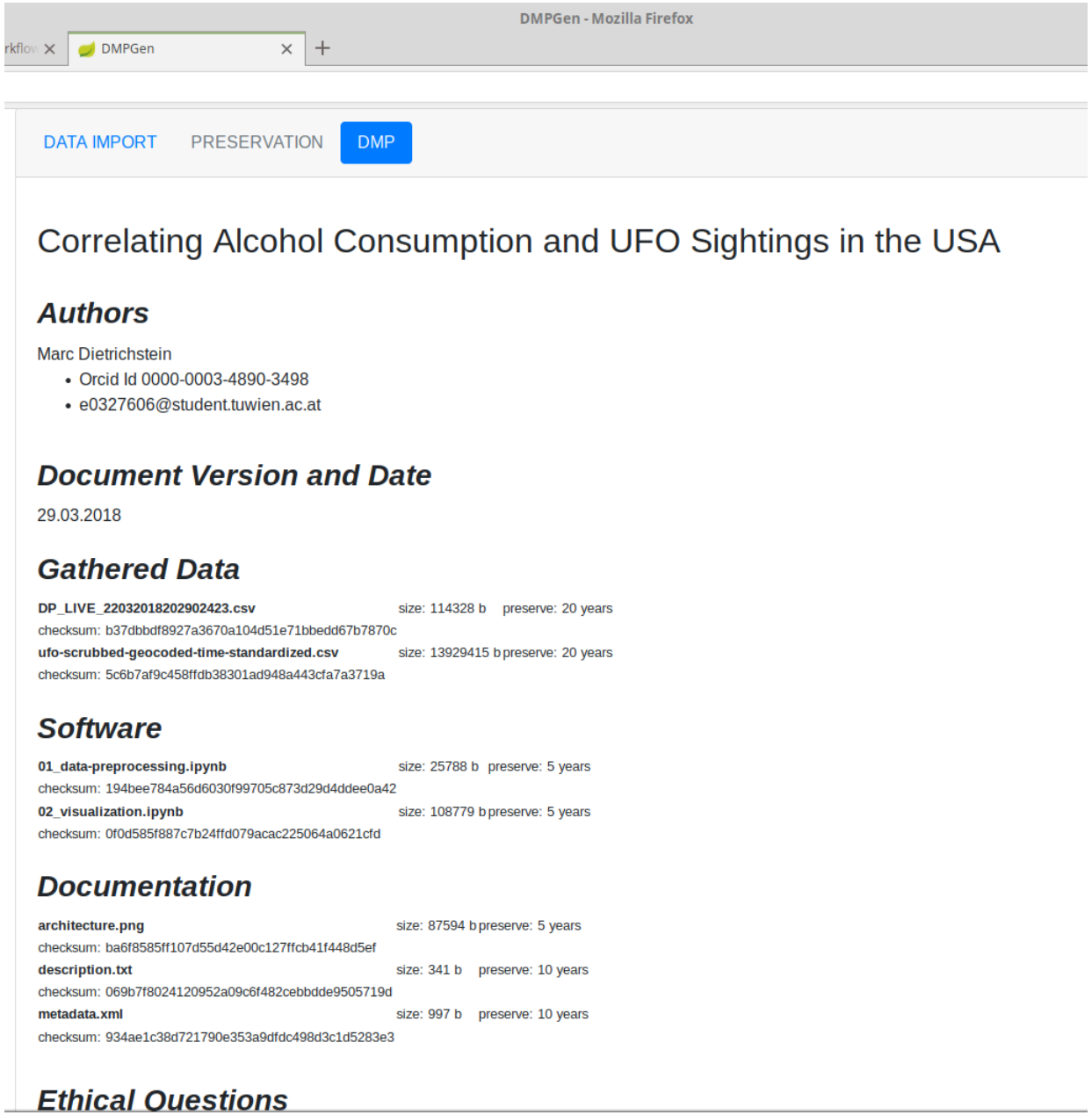
Orcid Id	0000-0003-4890-3498
Given Name	Marc
Family Name	Dietrichstein
E-mail	e0327606@student.tuwien.ac.at
Project Title	Correlating Alcohol Consumption and UFO Sightings in the USA
Project DOI	https://doi.org/10.5281/zenodo.1209833
Publication Date	2018/03/29

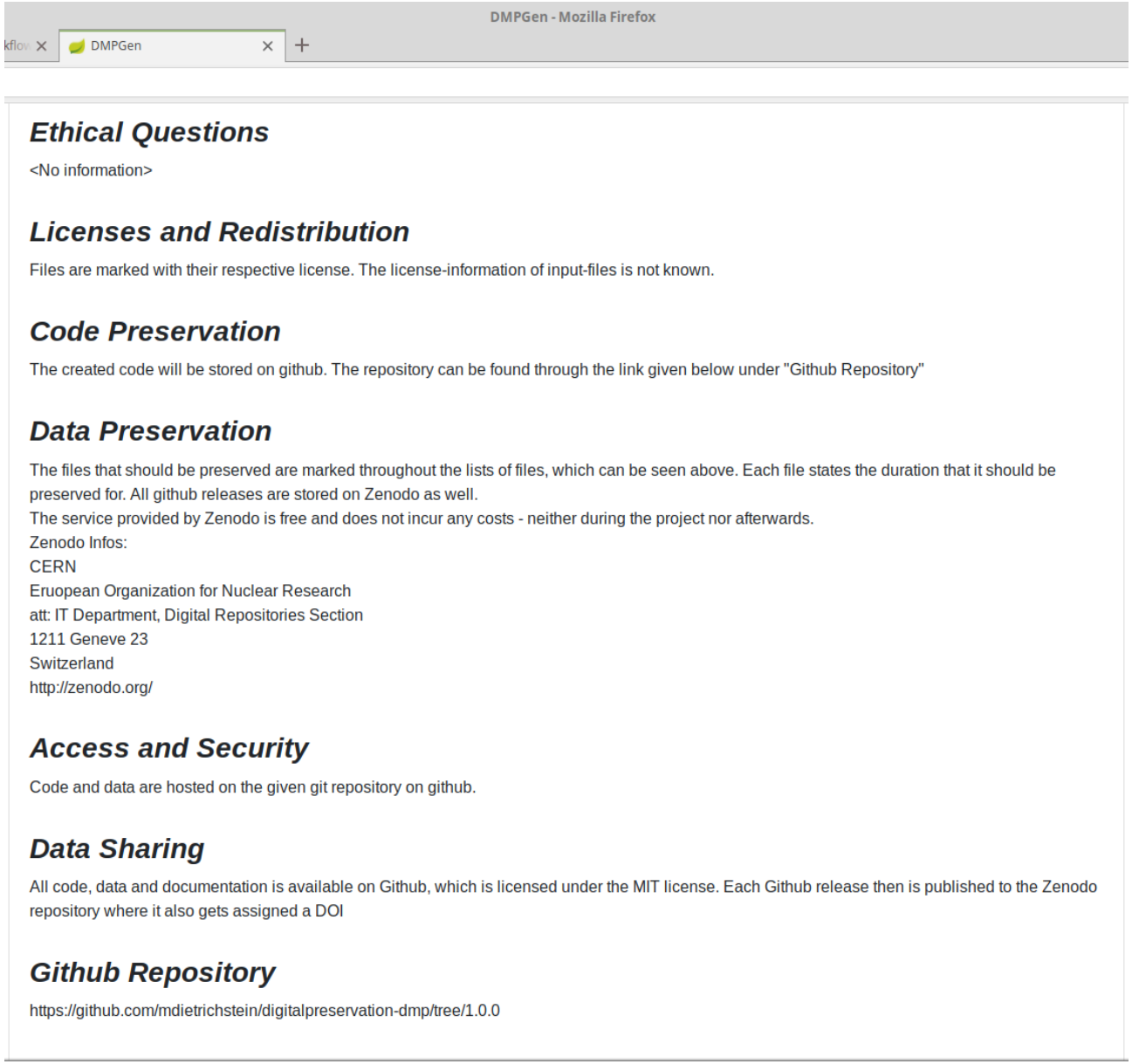
Zenodo Project Info

Creators	Marc Dietrichstein, sorx
Rights	openAccess
Type	software
Github Url	https://github.com/mdietrichstein/digitalpreservation-dmp/tree/1.0.0



Filename	Path	Tag	Preservation
.gitignore	.gitignore	Ignore	Ignore
Dockerfile	Dockerfile	Ignore	Ignore
LICENSE	LICENSE	Ignore	Ignore
README.md	README.md	Ignore	Ignore
README.pdf	README.pdf	Ignore	Ignore
ufo_alcohol.csv	data/processed/ufo_alcohol.csv	Ignore	Ignore
DP_LIVE_22032018202902423.csv	data/raw/DP_LIVE_22032018202902423.csv	Input Data	20 Years
ufo-scrubbed-geocoded-time-standardized.csv	data/raw/ufo-scrubbed-geocoded-time-standardized.csv	Input Data	20 Years
architecture.png	documentation/architecture.png	Documentation	5 Years
description.txt	documentation/description.txt	Documentation	10 Years
metadata.xml	documentation/metadata.xml	Documentation	10 Years
01_data-preprocessing.ipynb	notebooks/01_data-preprocessing.ipynb	Software	5 Years
02_visualization.ipynb	notebooks/02_visualization.ipynb	Software	5 Years
.keep	reports/.keep	Ignore	Ignore


DMPGen - Mozilla Firefox
DMPGen
DATA IMPORT
PRESERVATION
DMP
Correlating Alcohol Consumption and UFO Sightings in the USA
Authors
Marc Dietrichstein
Orcid Id 0000-0003-4890-3498
e0327606@student.tuwien.ac.at
Document Version and Date
29.03.2018
Gathered Data
DP_LIVE_22032018202902423.csv size: 114328 b preserve: 20 years
checksum: b37dbbdf8927a3670a104d51e71bbedd67b7870c
ufo-scrubbed-geocoded-time-standardized.csv size: 13929415 b preserve: 20 years
checksum: 5c6b7af9c458ffdb38301ad948a443cfa7a3719a
Software
01_data-preprocessing.ipynb size: 25788 b preserve: 5 years
checksum: 194bee784a56d6030f99705c873d29d4ddee0a42
02_visualization.ipynb size: 108779 b preserve: 5 years
checksum: 0f0d585f887c7b24ffd079acac225064a0621cfd
Documentation
architecture.png size: 87594 b preserve: 5 years
checksum: ba6f8585ff107d55d42e00c127ffcb41f448d5ef
description.txt size: 341 b preserve: 10 years
checksum: 069b7f8024120952a09c6f482cebbdde9505719d
metadata.xml size: 997 b preserve: 10 years
checksum: 934ae1c38d721790e353a9dfdc498d3c1d5283e3
Ethical Questions


DMPGen - Mozilla Firefox
DMPGen
Ethical Questions
<No information>
Licenses and Redistribution
Files are marked with their respective license. The license-information of input-files is not known.
Code Preservation
The created code will be stored on github. The repository can be found through the link given below under "Github Repository"
Data Preservation
The files that should be preserved are marked throughout the lists of files, which can be seen above. Each file states the duration that it should be preserved for. All github releases are stored on Zenodo as well.
The service provided by Zenodo is free and does not incur any costs - neither during the project nor afterwards.
Zenodo Infos:
CERN
Eruopean Organization for Nuclear Research
att: IT Department, Digital Repositories Section
1211 Geneve 23
Switzerland
http://zenodo.org/
Access and Security
Code and data are hosted on the given git repository on github.
Data Sharing
All code, data and documentation is available on Github, which is licensed under the MIT license. Each Github release then is published to the Zenodo repository where it also gets assigned a DOI
Github Repository
https://github.com/mdietrichstein/digitalpreservation-dmp/tree/1.0.0

DMPGen - Mozilla Firefox

DMPGen


```
{
  "@context": {
    "dmp": "http://purl.org/madmps#",
    "foaf": "http://xmlns.com/foaf/0.1/",
    "dc": "http://purl.org/dc/elements/1.1/",
    "dcterms": "http://purl.org/dc/terms/",
    "premis": "http://www.loc.gov/premis/rdf/v1#"
  },
  "@id": "http://example.org/dmps/mydmp",
  "@type": "dmp:DataManagementPlan",
  "dcterms:title": "mdietrichstein/digitalpreservation-dmp: Submission Release",
  "dcterms:description": " 👽 Exploring the connection between alcohol consumption and the number of ufo sightings in the USA 👽 ",
  "dc:creator": [
    {
      "@id": "0000-0003-4890-3498",
      "foaf:name": "Marc Dietrichstein",
      "foaf:mbox": "e0327606@student.tuwien.ac.at"
    }
  ],
  "dc:date": "29.03.2018",
  "dmp:hasDataObject": [
    {
      "@id": "https://doi.org/10.5281/zenodo.1209833",
      "@type": "dmp:SourceCode",
      "dmp:hasIntelectualPropertyRights": {
        "dcterms:license": "https://opensource.org/licenses/MIT"
      },
      "dmp:hasDataRepository": "https://github.com/mdietrichstein/digitalpreservation-dmp/tree/1.0.0",
      "dmp:hasPreservation": "All files that need preservation, are marked with their respective preservation duration. The files the
      "dmp:hasDataSharing": "All code, data and documentation is available on Github and is licensed under the MIT license. To make t
      "dmp:hasEthicsAndPrivacy": " <No information> ",
      "dmp:hasDocumentation": "The documentation can be found in all files that are marked as type documentation. These files can be
      "dmp:hasDataCollection": "All files that are collected from external sources are marked as input-files.",
      "dmp:hasDataObject": [
        {
          "@type": "dmp:input_data",
          "dc:title": "DP_LIVE_22032018202902423.csv",
          "dmp:hasIntelectualPropertyRights": {
            "dcterms:license": "https://opensource.org/licenses/MIT"
          },
          "dmp:hasMetadata": {
            "premis:hasObjectCharacteristics": {
              "premis:fixity": {
                "premis:hasMessageDigestAlgorithm": "premis:Fixity:SHA",
```

Project and post-project phase - demo 2

https://github.com/alexschwarzresearch/DMPlanner

Name

Please provide your full name.

Search ORCiD

full_name Tomasz Miksa

orcid 0000-0002-4929-7875

current_employment_name SBA Research

Resources

Add as many Github repositories or OAI-PMH compliant DOIs as you like.*

Add Resource

Zenodo Ten Simple Rules For Machine-Actionable Data Management Plans (Preprint)

documentation

Remove

Github TomMiksa/DMPlanner

software

Remove

Preservation Time

Choose how many years the data for each group should be kept.

Software 10 years

Documentation 20 years

TUW DMP

A Data Management Plan created using DMPlanner.

Creator

Name: Tomasz Miksa
ORCiD: 0000-0002-4929-7875
Current Work: SBA Research

How will you manage copyright and Intellectual Property Rights (IPR) issues?

The software which was created in the course of the project has the license restrictions "MIT License".

Which data are of long-term value and should be retained, shared, and/or preserved?

In this project especially the documentation, as well as the software has a long-term value and should at least be as long preserved as the targeted preservation time specifies. The targeted preservation time for the documentation is 20 years. The targeted preservation time for the software is 10 years.

What is the long-term preservation plan for the dataset?

One of the main strategies of the long-term preservation plan is the use of public accessible repositories to save the components of the project. The documentation resource "Ten Simple Rules For Machine-Actionable Data Management Plans (Preprint)" is hosted on Zenodo. The software resource "DMPlanner" is hosted on Github.

How will you share the data?

The data will be primarily shared through the public repositories listed above. This way the data is openly accessible and findable, as well as searchable. The data is available at the repositories as of this moment.

Are any restrictions on data sharing required?

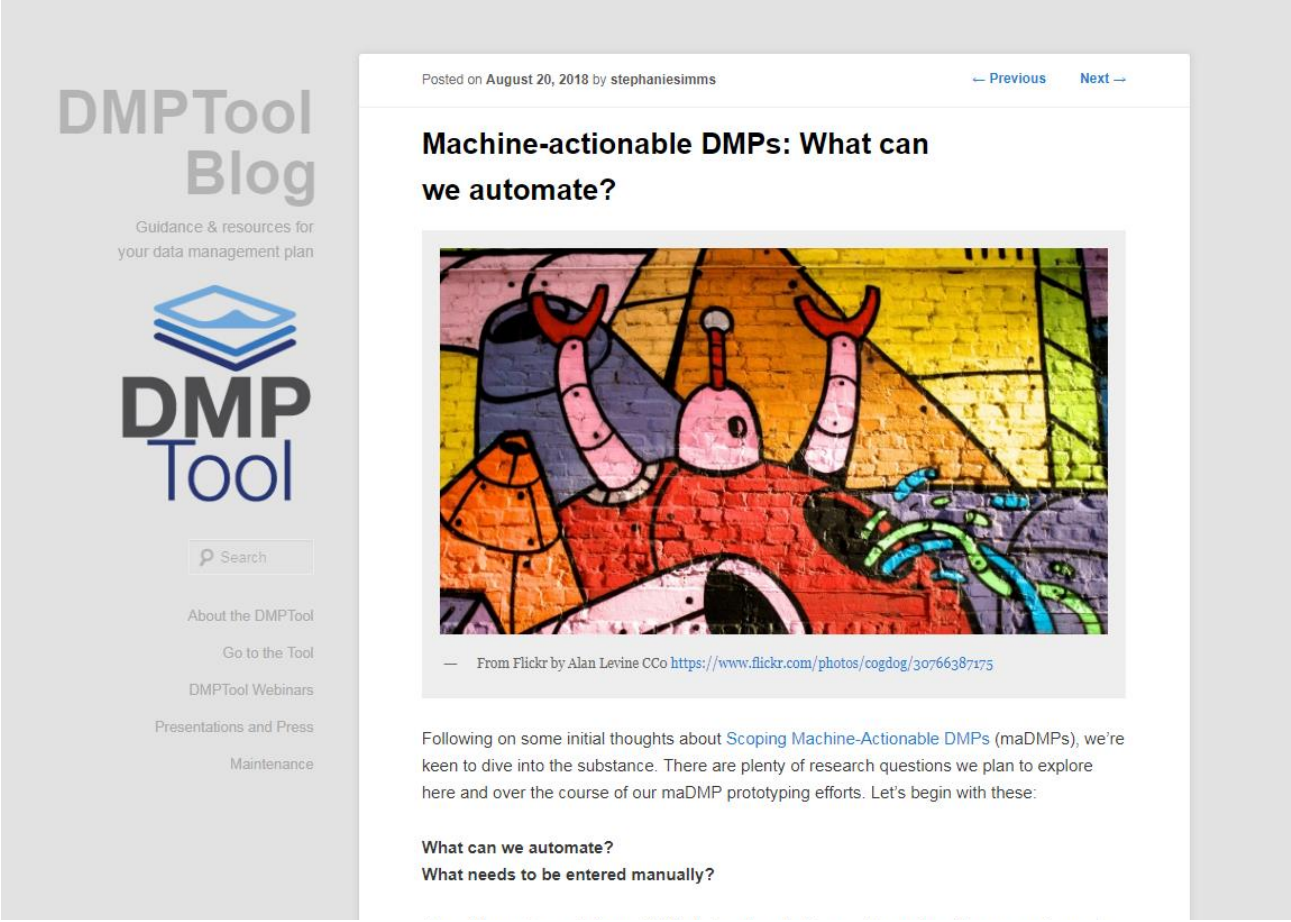
The restrictions on data sharing are composed of the used licenses together with the long-term preservation plan. With this in mind the following restrictions for the resources of the project apply. The documentation resource "Ten Simple Rules For Machine-Actionable Data Management Plans (Preprint)" will be hosted on Zenodo for at least 20 years. The software resource "DMPlanner" will be hosted on Github for at least 10 years.

Who will be responsible for data management?

The creator of this data management plan is Tomasz Miksa. Therefore Tomasz Miksa is also the reference person for possible reviews and revisions regarding this data management plan in the future. Unless amended Tomasz Miksa is additionally responsible for the adherence to the plan.

```
{
  "@context": {
    "dc": "https://purl.org/dc/elements/1.1/",
    "dcterms": "https://purl.org/dc/terms/",
    "dmp": "https://purl.org/madmps#",
    "foaf": "http://xmlns.com/foaf/0.1/",
    "premis": "http://id.loc.gov/ontologies/premis.html#",
    "schema": "https://schema.org/",
    "time": "https://www.w3.org/2006/time#"
  },
  "@type": "dmp:DataManagementPlan",
  "dc:creator": {
    "@id": "https://orcid.org/0000-0002-4929-7875",
    "@type": "foaf:Person",
    "foaf:Organization": "SBA Research",
    "foaf:name": "Tomasz Miksa"
  },
  "dcterms:created": "2018-09-16",
  "dcterms:title": "TUW DMP",
  "dmp:hasDataObject": [
    {
      "@id": "https://doi.org/10.5281/zenodo.1172673",
      "@type": "dmp:Documentation",
      "dmp:hasDataObject": [],
      "dmp:hasIntelectualPropertyRights": [
        {
          "dcterms:license": "Creative Commons Attribution 4.0"
        },
        {
          "dcterms:license": "Open Access"
        }
      ],
      "dmp:hasMetadata": {
        "dc:date": "2018-02-13T15:41:09Z",
        "dcterms:abstract": "
Data management plans (DMPs) are documents accompanying research proposals and project outputs. They describe data
```

Machine-actionable DMPs: What can we automate?



DMPTool Blog

Guidance & resources for your data management plan

DMP Tool

Search

About the DMPTool

Go to the Tool

DMPTool Webinars

Presentations and Press

Maintenance

Posted on **August 20, 2018** by **stephaniesimms**

← Previous Next →

Machine-actionable DMPs: What can we automate?

— From Flickr by Alan Levine CCo https://www.flickr.com/photos/cogdog/30766387175

Following on some initial thoughts about Scoping Machine-Actionable DMPs (maDMPs), we're keen to dive into the substance. There are plenty of research questions we plan to explore here and over the course of our maDMP prototyping efforts. Let's begin with these:

What can we automate?
What needs to be entered manually?

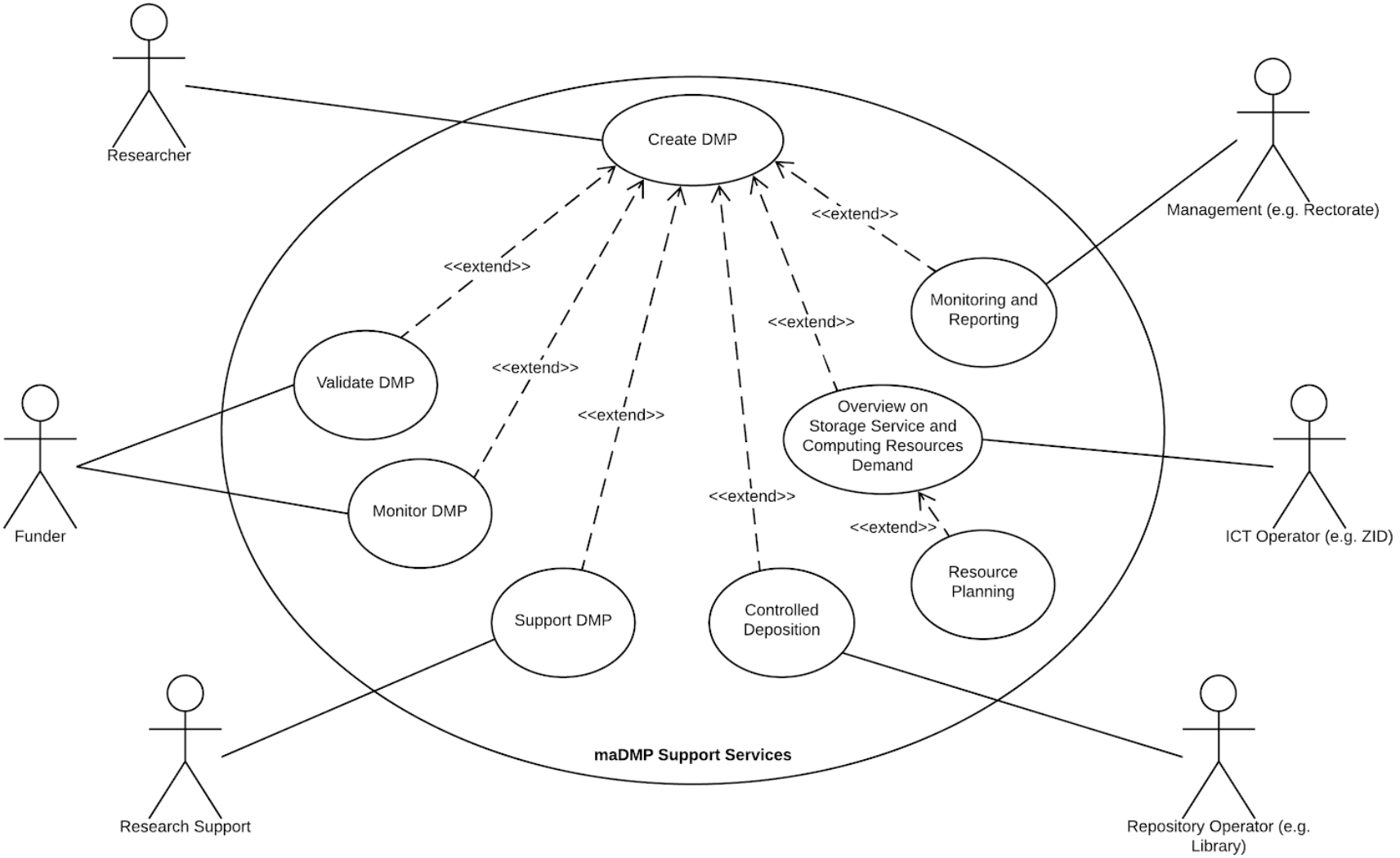
https://blog.dmptool.org/2018/08/20/machine-actionable-dmps-what-can-we-automate/

Discussion

Processes for maDMPs

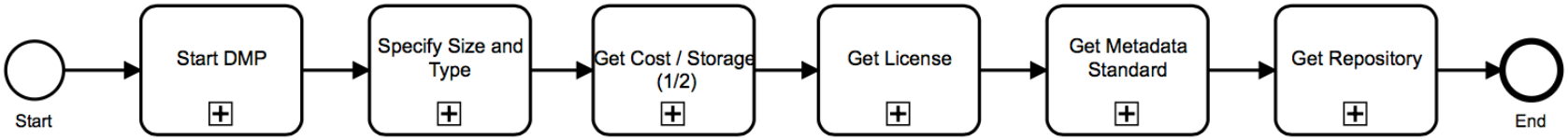
Part 4 – Identifying stakeholder interactions and services

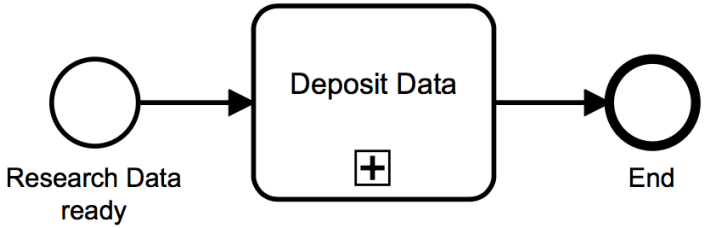
maDMPs use cases



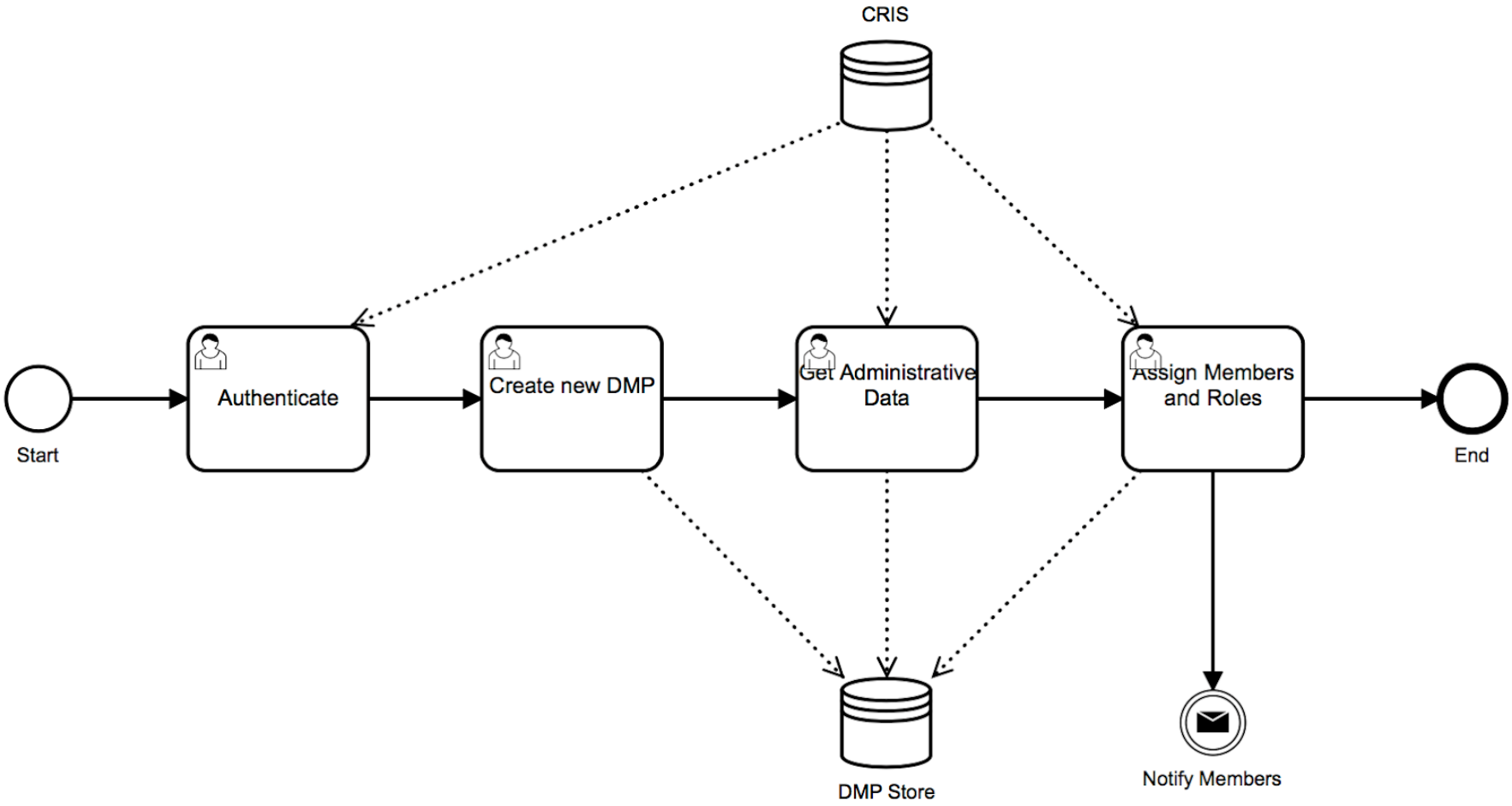
BPMN process - overview

- Business Process Modelling Notation (BPMN)
- Defined 10 workflows

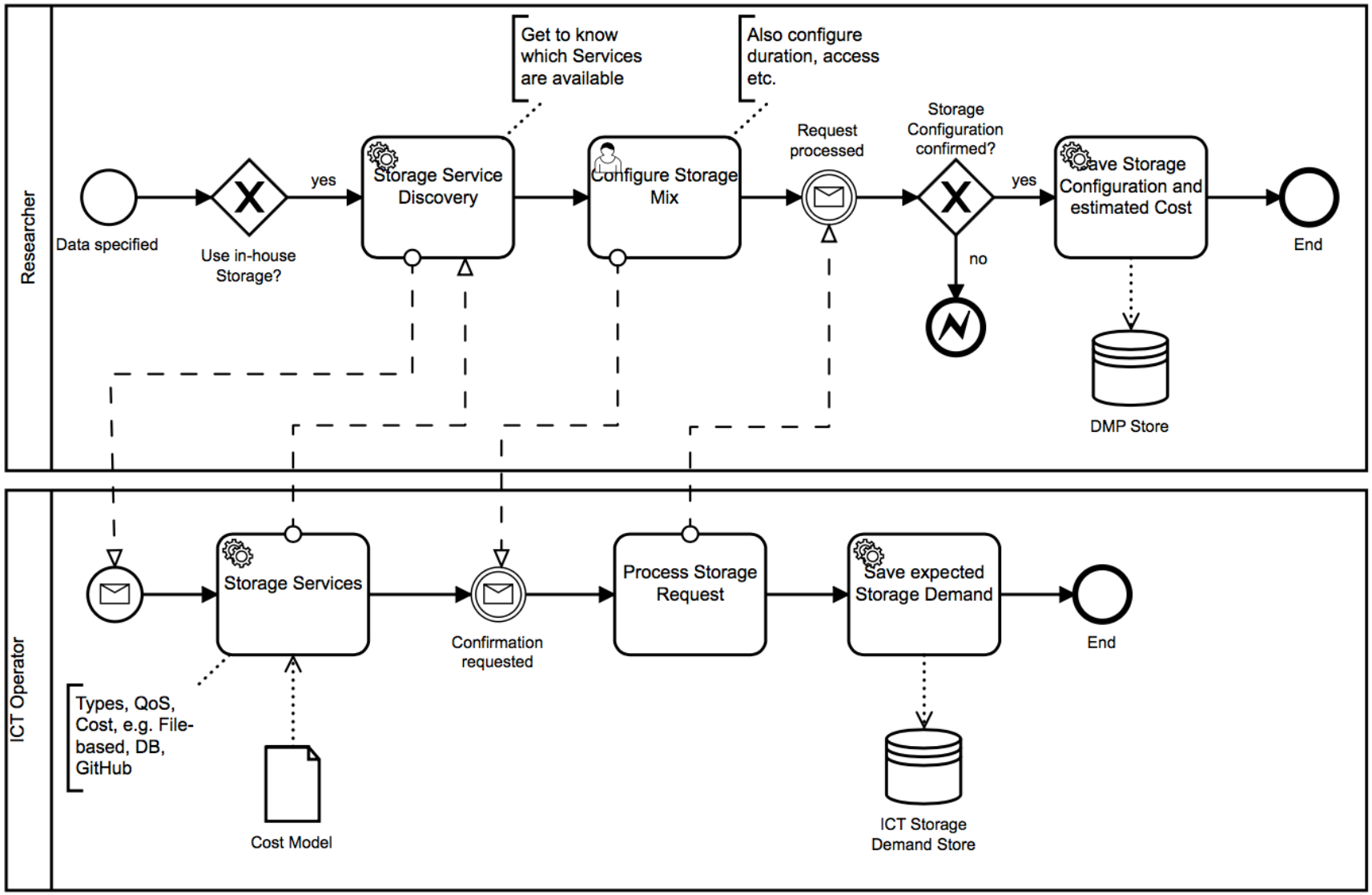


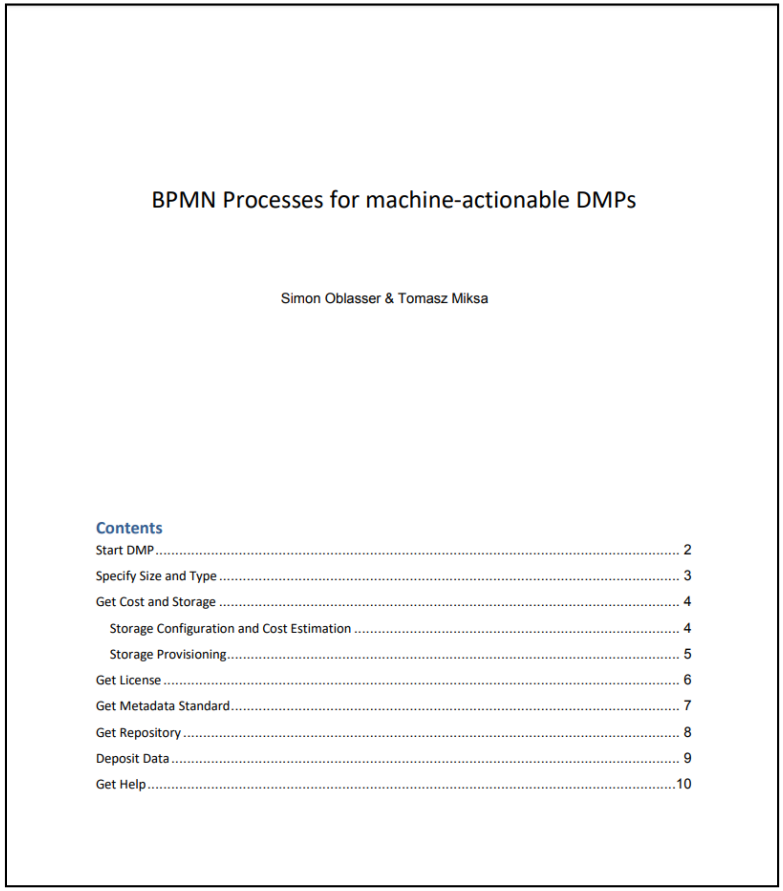


Start DMP



Get Cost / Storage





BPMN Processes for machine-actionable DMPs

Simon Oblasser & Tomasz Miksa

Contents

http://rda-ws-tpdl2018.sysresearch.org/documents/2018-TPDL-Porto-Handout-BPMN.pdf

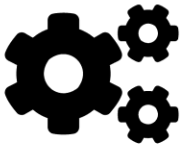
Processes - summary

- Processes help identify
  - **tasks** performed by stakeholders
    - e.g. ICT operator provide costs of storage
  - **systems** needed to be put in place
    - e.g. maDMP repository or costing service
  - **concepts** to be developed or agreed
    - e.g. cost model for storage
- Useful in deploying maDMPs
- Allow us to narrow down focus of this WG
  - common model does not contain business logic
    - e.g. cost estimation is done by a service that provides a value
  - common model is an information carrier
    - tools, services, processes make maDMPs *machine-actionable*

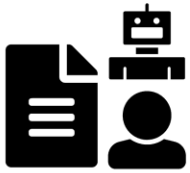
Discussion

10 principles for maDMPs

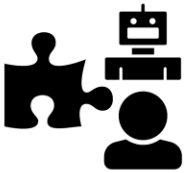
1 Integrate DMPs with the workflows of all stakeholders in the research data ecosystem



2 Allow automated systems to act on behalf of stakeholders



3 Make policies (also) for machines, not just for people

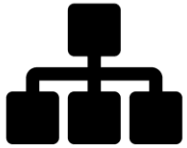


4 Describe—for both machines and humans—the components of the data management ecosystem

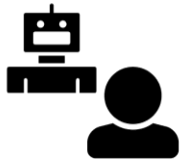
5 Use PIDs and controlled vocabularies

Miksa, Tomasz, Simms, Stephanie, Mietchen, Daniel, & Jones, Sarah. (2018). Ten simple rules for machine-actionable data management plans (preprint). http://doi.org/10.5281/zenodo.1434938

10 principles for maDMPs



6 Follow a common data model for maDMPs



7 Make DMPs available for human and machine consumption

8 Support data management evaluation and monitoring

9 Make DMPs updatable, living, versioned documents

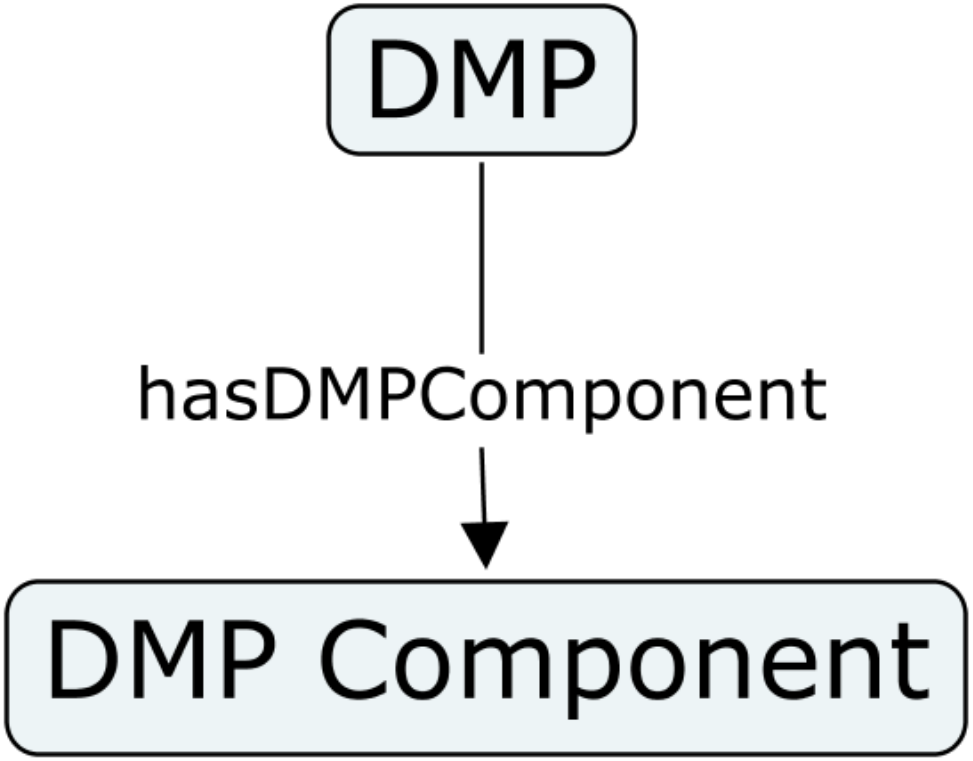
10 Make DMPs publicly available

Miksa, Tomasz, Simms, Stephanie, Mietchen, Daniel, & Jones, Sarah. (2018). Ten simple rules for machine-actionable data management plans (preprint). http://doi.org/10.5281/zenodo.1434938

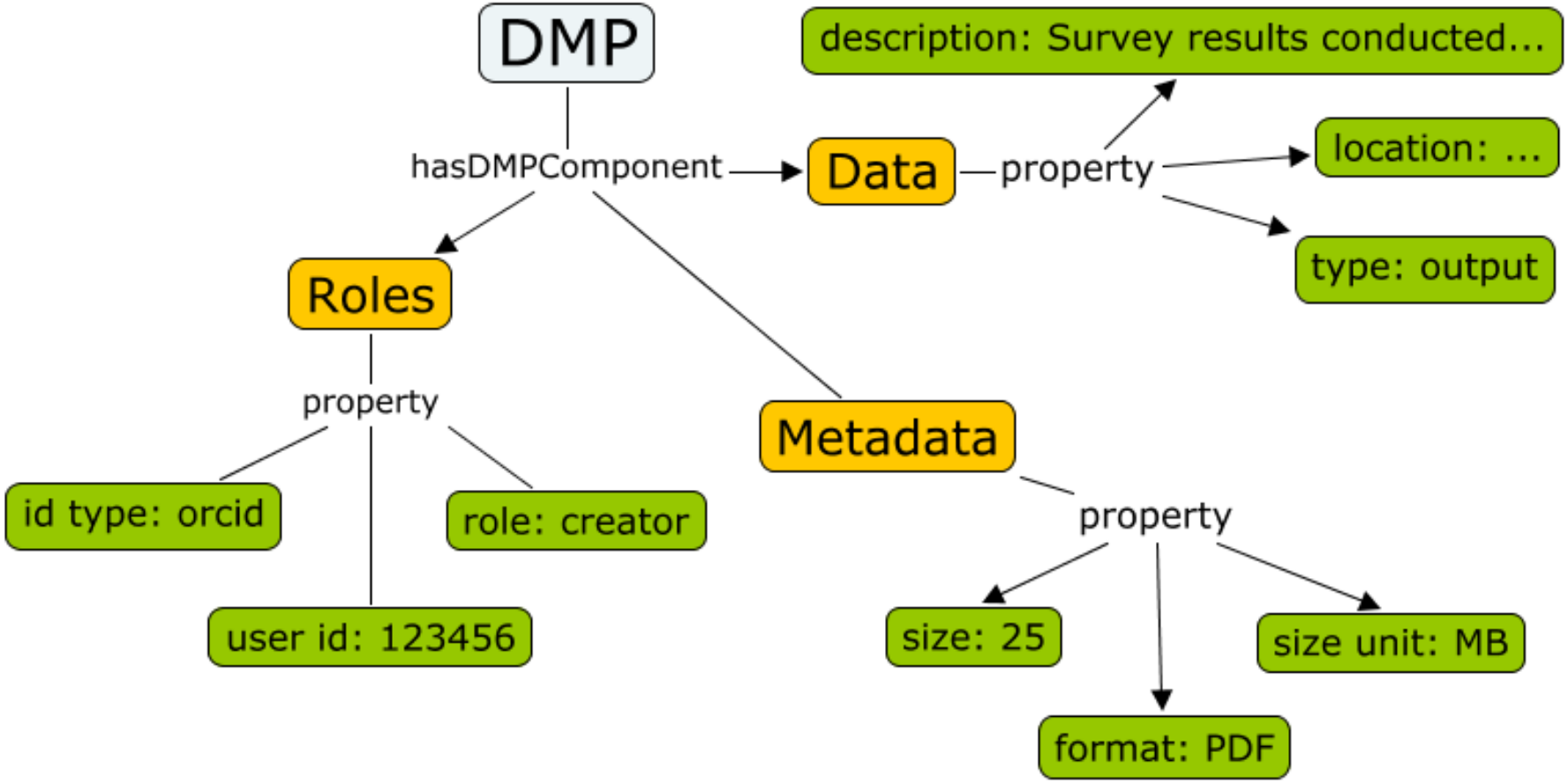
Towards Common Data Model

Part 5 – work in progress!

Core model



Core model instance - example



DMP Components

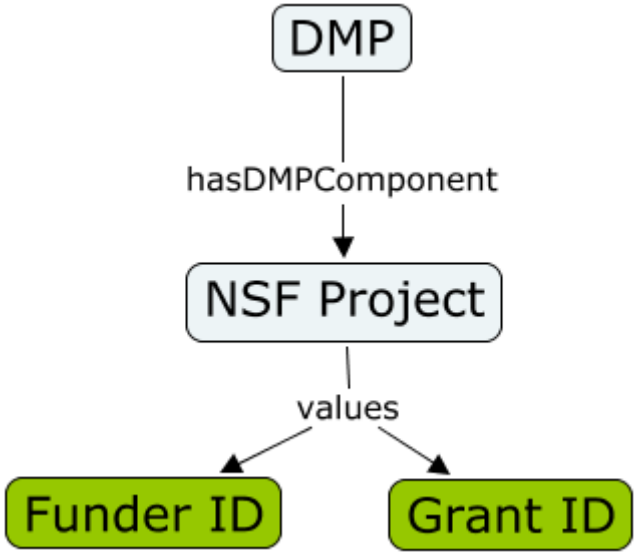
- Give flexibility to address various needs
  - currently there is no single DMP template that suits all…
  - A mix can be selected to suit institutions/funders needs
- Allow to decouple and encapsulate logical parts of the model
  - e.g. license, metadata, repository, cost, preservation, …
  - can be similar to DMP themes, but do not have to
- Allow to avoid
  - Tree model that can be too complex and not flexible
  - Flat model that can be too generic

DMP Component

DMP Component - example

Help addressing contradictive requirements

- USA
  - DMPs mandated by funders (NSF)
  - DMPs include funder id, grant id, etc.
- Australia
  - DMPs used as a management tool to support good practice
  - Funder and grant id are optional



Registry of DMP Components

- DMP Components
  - Projects [0..n]
    - NSF Project
    - EC Project
    - Wellcome Project
  - Data [1..n]
    - …
  - Costs
    - …
  - Security
    - …
  - …

Registry of DMP Components

- Can RDA maintain the registry?

- Possible process to manage the registry
  - core model entities are governed by some fairly strict RDA process (this group initially)
    - e.g. Projects, Data, etc.
  - extensions can be submitted and must pass a quick review process
    - e.g. NSF Project, EC Project, …

Guidelines

- Provide guidance on how to
  - customise model
  - implement model
- Draft:
  1. **Create maDMPs for data, not for projects**
  2. **Create maDMP views to address stakeholder requirements**
  3. Version maDMPs and track source of information
  4. Follow core model for maDMPs
  5. Use extension mechanism to address specific requirements
  6. Requirements for extensions
  7. Include information that is subject to change in other systems

G1: Create maDMPs for data, not for projects

- maDMPs will be living documents when they accompany data
- data has different lifecycles
  - different licenses
  - different storage
  - different repositories
  - different preservation actions
- data is not always a result of a project
  - permanent data collection of satellite images
- maDMPs NOT for each file, but for logical collection
  - e.g. input data, output data, software, documentation

G2: Create maDMP views to address stakeholder requirements

DMP 1				
Type	Size	Cost	License	Repository
Software	300	500	CC-BY	a

DMP 2				
Type	Size	Cost	License	Repository
Output data	600	300	CC-BY	B

DMP 3				
Type	Size	Cost	License	Repository
Metadata	100	100	CC-BY	C

+

Total size	Total cost	Licenses	Repositories
1000	900	CC-BY	A, B, C

Funder's view

Common model - summary

- Ongoing developments
  - Core Model
  - Registry of components
  - Guidelines
- To model DMP Components we will
  - use input from consultations (especially the 2nd one)
  - new inputs from you!
- Your contribution needed
  - Develop components that are relevant in your setting
  - For each field we still need to know who, when and why needs it!
    - e.g. NSF needs grant ID when DMP is submitted for a review
- We will provide soon
  - Wiki / google docs for collaborative work
  - Guidance for developing components

Discussion

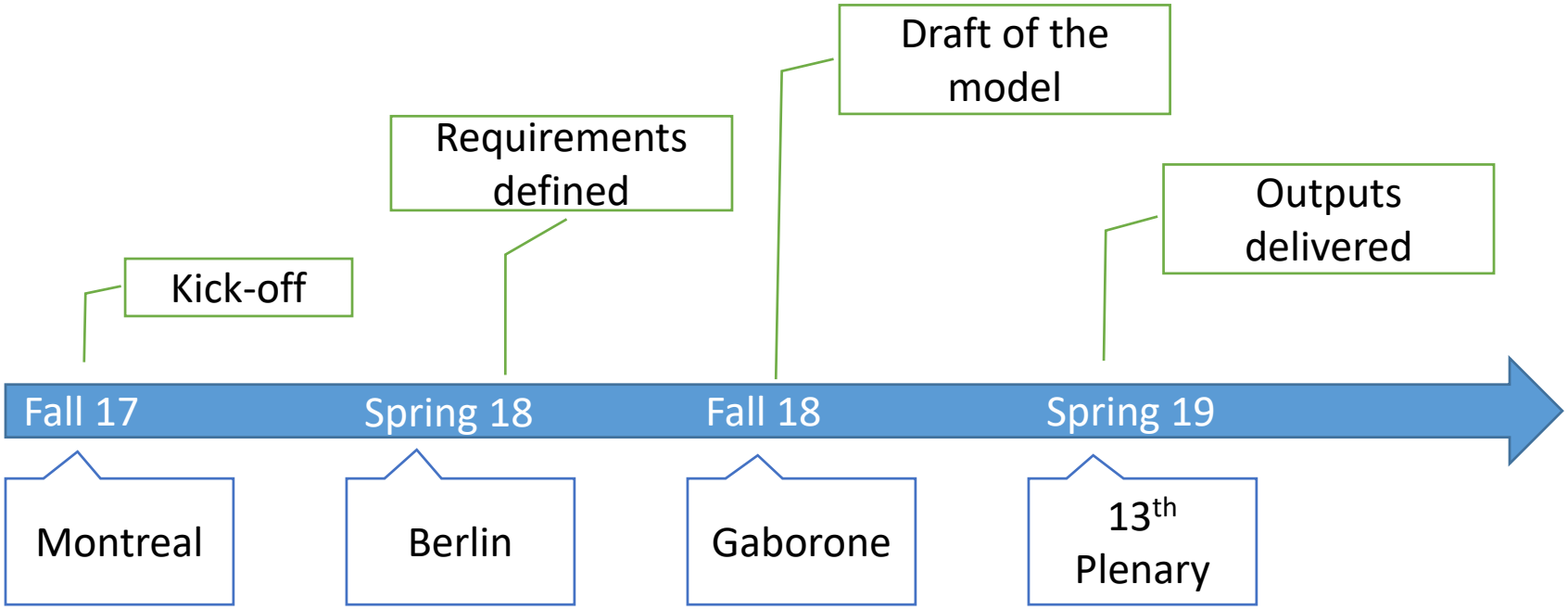
Wrap-up and next steps

Part 5

- 1st consultation (user stories) went broad
  - to define scope of maDMPs
- 2nd consultation went deep
  - to identify models for specific requirements
- Proof of concept tools
  - to demonstrate how model can be used to automate tasks
- Processes modelling
  - to identify systems and stakeholders involved
- Common data model
  - Core model
  - DMP Components
  - Guidelines

Timeline

- Slide taken from the Berlin plenary presentation
  - Seems we're on track



Staying in touch!

- Sign up to the group
  - https://www.rd-alliance.org/groups/dmp-common-standards-wg
- Participate in model development
- Contact group chairs

Tomasz Miksa

Paul Walk

Peter Neish

Exposing Data Management Plans WG Update

IDW 2018 RDA 12th Plenary
November 6, 2018
Gabarone, Botswana

Who we are:

Chair (s): Angus Whyte, Fiona Murphy, Natalie Meyers, Kathryn Unsworth, Marie-Christine Jacquemot-Perbal Secretariat Liaison: Lynn Yarmey | TAB Liaison: Francoise Genova

Complementary Aims

Related groups with complementary aims:

- DMP Common Standards WG Chair (s): Tomasz Miksa, Paul Walk, Peter Neish | Secretariat Liaison: Lynn Yarmey | TAB Liaison: Wenbo Chu
- Active Data Management Plans IG Chair (s): David Giaretta, Helen Glaves, Kevin Ashley, Stephanie Simms, Sarah Jones | TAB Liaison: Jane Wyngaard California Digital Library has received an NSF EAGER AWARD #1745675 to support active, machine-actionable data management plans (DMPs).
- Sharing Rewards and Credit (SHARC) IG Chair(s) Anne Cambon-Thomsen, Laurence Mabile, Laurent Dollé, Anne-Marie Tassé, Michele De Rosa | Secretariat Liaison: Lynn Yarmey | TAB Liaison: Paul Uhlir

What we're up to:

A needs assessment/survey to query community on their attitudes toward exposing data management plans.

Please try the Exposing DMPs Survey Pilot (https://bit.ly/2xxFZsc)

Anticipated Outputs:

Reference model including alternative strategies for exposing plans to best serve community interests in meeting FAIR principles, based on shared experience of ‘early adopters’ in test implementations.

Supported by the survey work aimed toward gauging user needs and motivations for exposing DMP, as well as perceived risks and disbenefits.

Join Us

RDA Web Page:

https://rd-alliance.org/groups/exposing-data-management-plans-wg

Email Us: exposing-plans@rda-groups.org

These Slides: