

# Machine-actionable DMPs - take them and use them!

DMP Common Standards WG

Tomasz Miksa, Paul Walk, Peter Neish





## Agenda

- Part 1 Introduction for newcomers, status update, and meeting objectives
- Part 2 Common standard for machine-actionable DMPs
  - Standard overview and refinements introduced since P13
- Part 3 Ongoing and future pilot implementations
  - > Simon Oblasser: integrating maDMPs with university infrastructure
  - Tom Renner: maDMPs and Haplo repository
  - João Cardoso: using ontologies and maDMPs
  - Tomasz Miksa:
    - > maDMPs and RDM infrastructure
    - > maDMPs and repositories
    - maDMPs and funder templates
    - maDMPs and existing DMP tools
    - > maDMP schema validator
- Part 4 Discussion
- Part 5 Wrap up and next steps





## Introduction for newcomers

Part 1





## DMPs currently

- Shortcomings of existing DMPs
  - > manually completed, vague, not updated, considered bureaucracy, completed last minute, ...

	Data Officer	Who is responsible for the data management and the DMP of the project (name/email address)?			
1	Data Characteristics				
1.1	Description of the data	What kinds of data/source code will be generated or reused (type, format, volume)? How will the research data be generated and which methods will be used? How will you structure the data and handle versioning? Who is the target audience?			
Ш	Documentation and Metadata				
II.1	Metadata standards	What metadata standards (if any) will be in use and why? (see Digital Curation Centre)			
II.2	Documentation of data	What information is needed for the data to be findable, accessible, interoperable and re-usable ( <u>FAIR</u> ) in the future? Is the data machine-readable? How are you planning to document this information?			
II.3	Data quality control	What quality assurance processes will you adopt?  How will the consistency and quality of data collection be controlled and documented? (This may include processes such as repeat samples or measurements, standardised data capture, peer review of data or representation with controlled vocabularies.)			
Ш	Data Availability and Storage				
III.1	Data sharing strategy	How and when will the data be shared and made accessible? What repository will you be using? What persistent identifier will be used?			
III.2	Data storage strategy	What data are to be preserved for the long-term, and what data will not be stored? How and where will the data be stored and backed up during the research? How and where will the data be stored after the project ends? For how long will the data be stored? Are there any costs that need to be covered for storage? At what point during or after the project will the data be stored? Are there any technical barriers to making the research data fully or partially accessible?			





#### 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!



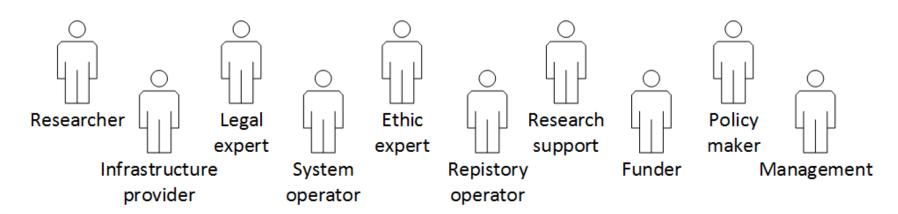






## Research data lifecycle

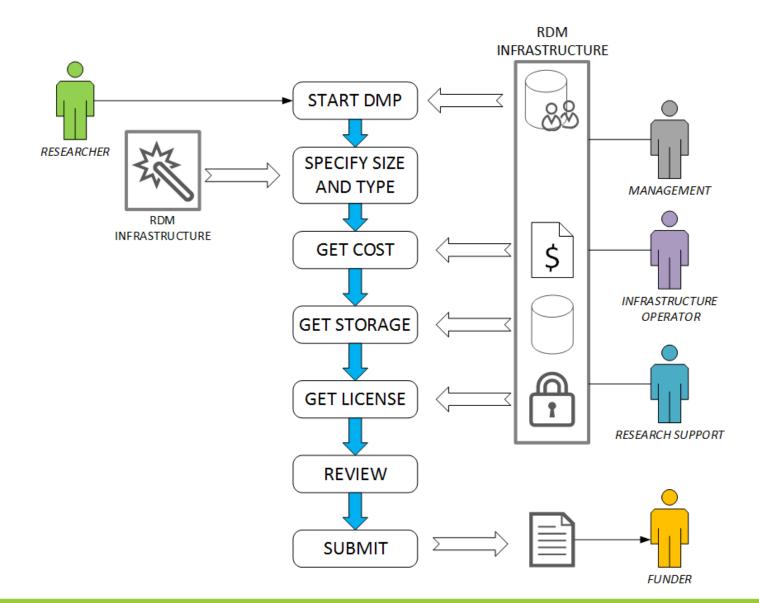
- Stakeholders involved in research data management
  - > require information at certain stages
  - > can provide information if requested at a proper stage
- Many problems can be avoided when
  - > timing is right
  - information flow is ensured







#### Automated Data Management Workflow







## Why do we need this WG?

- DMPs currently
  - > 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





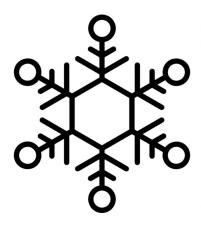
#### DMP Common Standards - Outputs

#### Common standard 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, RDF, etc.





## 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"
} ],
```



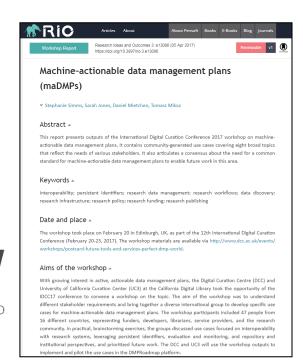


#### DMP Common Standards WG

- Launched in October 2017
- Result of a consultation made by Active DMPs IG
- Focus on machine-actionable DMPs
- > 100+ members from all continents
- >DMP tool owners are part of it







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





## 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
- >Model development

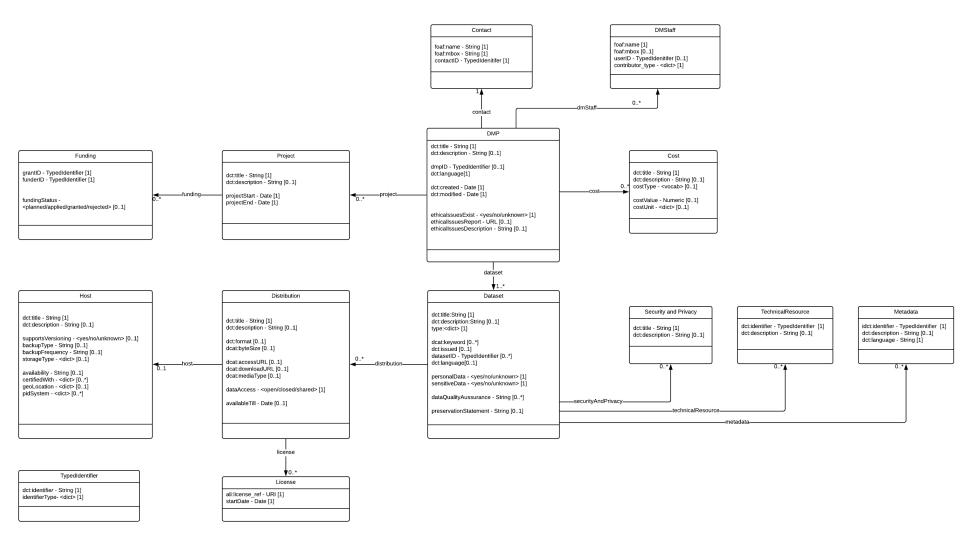




Part 2



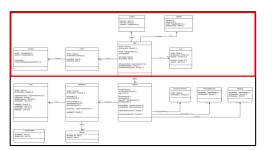


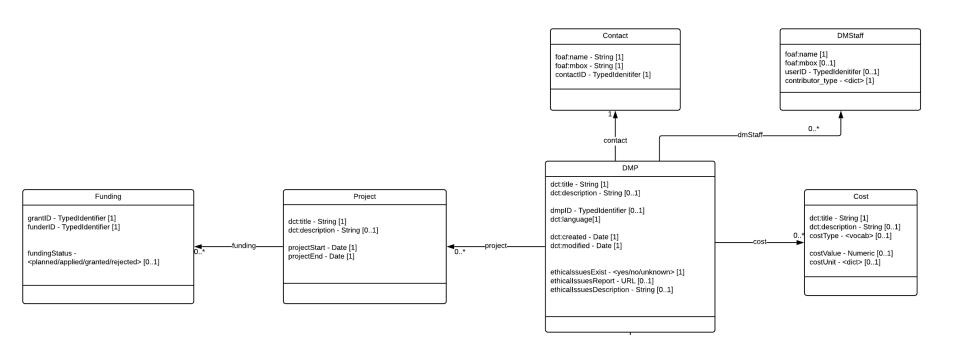


https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard/blob/master/docs/diagrams/RDA-DMP-Common-Model-Diagram-190325.pdf





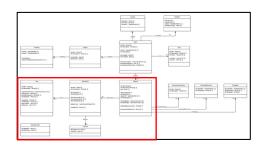


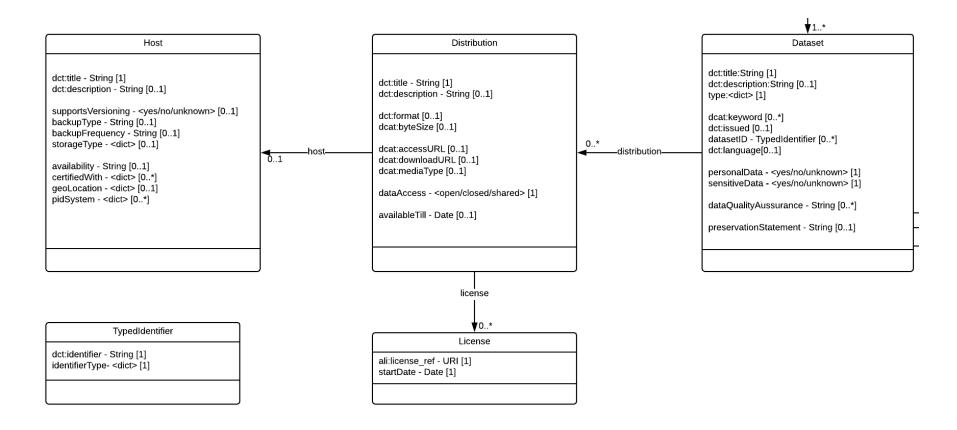


https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard/blob/master/docs/diagrams/RDA-DMP-Common-Model-Diagram-190325.pdf





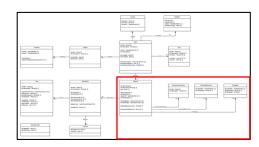


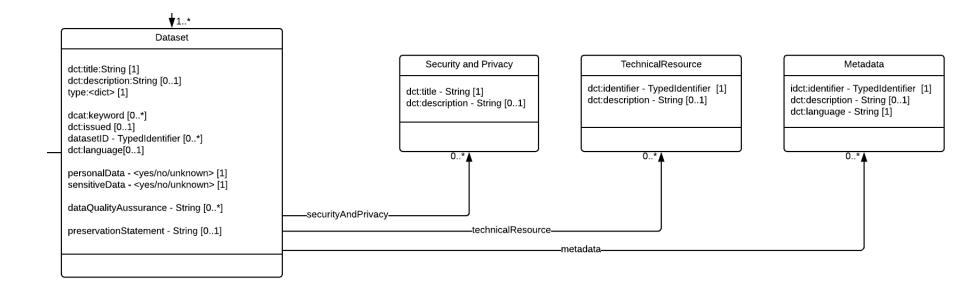


 $\frac{https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard/blob/master/docs/diagrams/RDA-DMP-Common-Model-Diagram-190325.pdf$ 











#### Model - documentation

#### Properties in 'contact'

Name	Description	Data Type	Cardinality	Example Value
contact_id	Identifier for a contact person	String	Exactly One	http://orcid.org/0000-0000-0000-0000
mail	E-mail address	String	Exactly One	cc@example.com
name	Name of the contact person	String	Exactly One	Charlie Chaplin

#### Properties in 'cost'

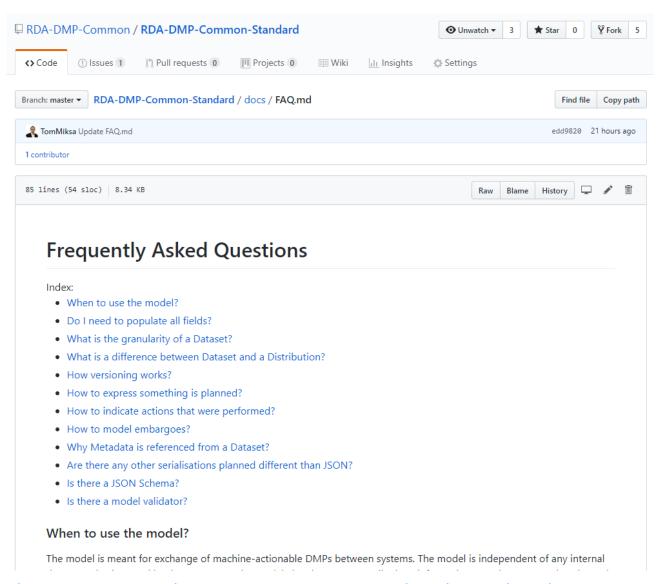
Name	Description	Data Type	Cardinality	Example Value
currency_code	Allowed values defined by ISO 4217.	Term from Controlled Vocabulary	Zero or One	EUR
description	Description	String	Zero or One	Costs for maintaining
title	Title	String	Exactly One	Storage and backup
type	Туре	Term from Controlled Vocabulary	Zero or One	
value	Value	Number	Zero or One	1000

https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard/blob/master/docs/index.md





#### Model – FAQ

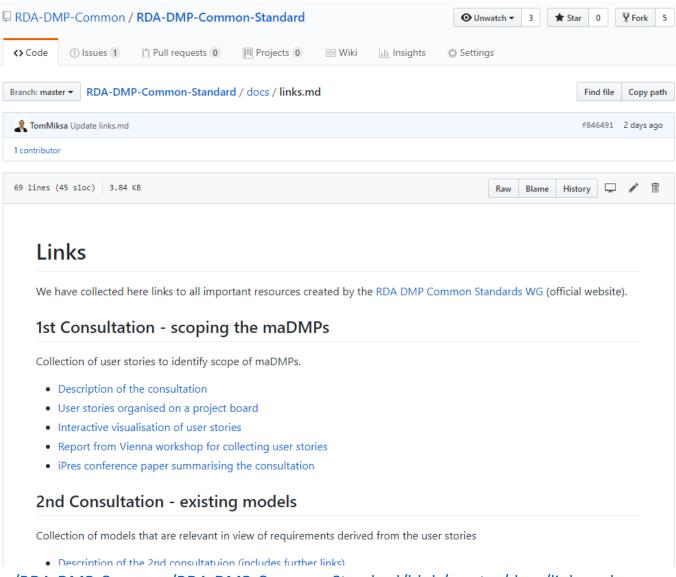


https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard/blob/master/docs/FAQ.md





#### Model – useful links

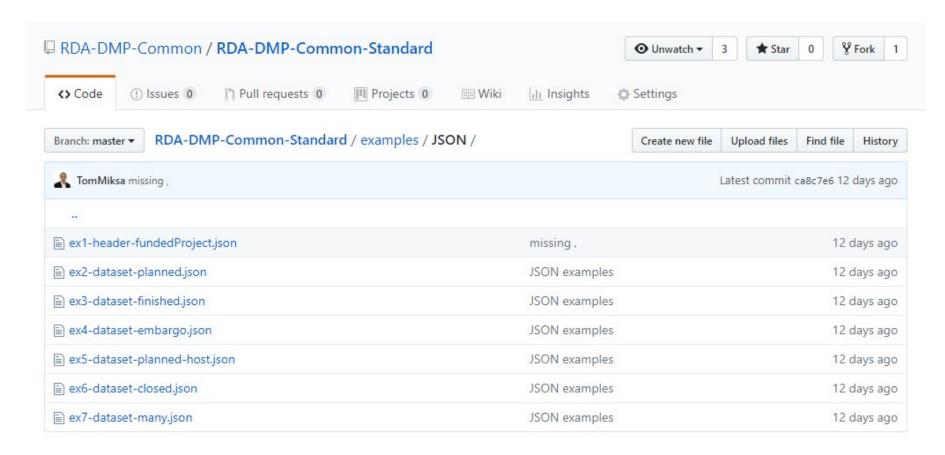


https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard/blob/master/docs/links.md





## Model – JSON examples



https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard/tree/master/examples/JSON





## Example

#### Example

> Source code will be issued on 2019-06-30 (planned) in 'some-repo'. There will be an embargo period till 2019-12-31. Later on the source code will be available on a CC-BY license.

```
"DMP": {
  "modified": "2019-02-22T13:20:15.5"
  "dataset": {
           "title": "Source Code",
           "issued": "2019-06-30",
           "distribution": {
                       "accessURL": "http://some-repo...",
                       "license": {
                                   "license_ref": "https://creativecommons.org/licenses/by/4.0/",
                                   "startDate": "2019-12-31
```



#### DMP Common Standard

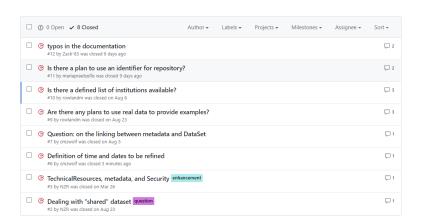
- Models information from standard DMPs
  - NOT a template
  - NOT a questionnaire
- Applicable in different settings
  - relaxed constraints within the model
    - > e.g. DMP <u>can</u> relate to a Project
  - constraints introduced at the 'business level'
    - > e.g. DMP <u>must</u> relate to a Project
- Used to exchange information between systems
  - no impact on tools/systems internal implementation





#### Updates from P13

- Only minor changes based on feedback from pilots
  - Date format changed to ISO 8601
  - > Host
    - Added identifier for a repository
  - Distribution
    - can contain many formats, not one only
  - Format
    - Mime-types recommended, but not required
  - Typed Identifiers not nested







# Ongoing and future pilot implementations

Part 3





# Simon Oblasser





# Tom Renner





# Joao Cardoso





# Tomasz Miksa

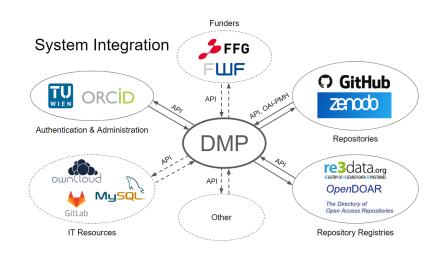




#### RDM Infrastructure @ TU Wien

- DMPs are not for funders only
- DMPs must also create benefits for researchers
  - Less work
  - Automation of tasks
  - Reuse of information
- > DMPs are the 'glue' between different systems
  - Automate
    - > getting data in
    - > getting data out
- > Pilot project





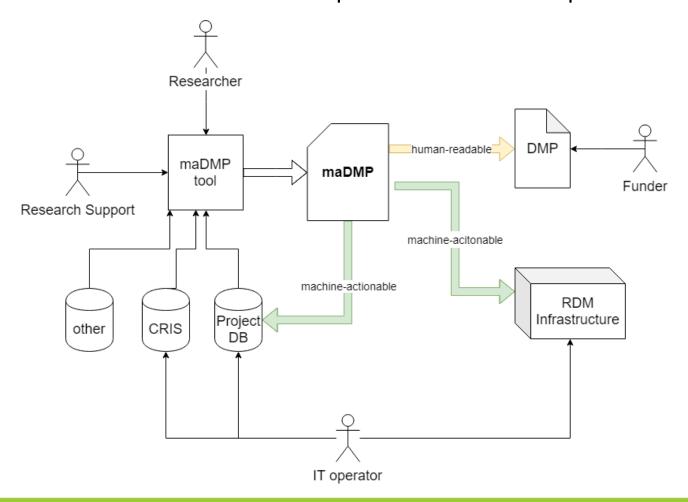




#### maDMPs and RDM infrastructure

#### **> once-only** principle

do not ask researchers same questions in different places





#### maDMPs and repositories

- Simplify data upload
  - Pre-fill fields based on maDMP
    - E.g. license, embargo, contact person, etc.



- Simplify reporting
  - e.g. update a DMP when data was uploaded to a repository
- > Prototype
  - > maDMP used to automate upload of data into Dataverse
    - https://hido1994.github.io/madmp/



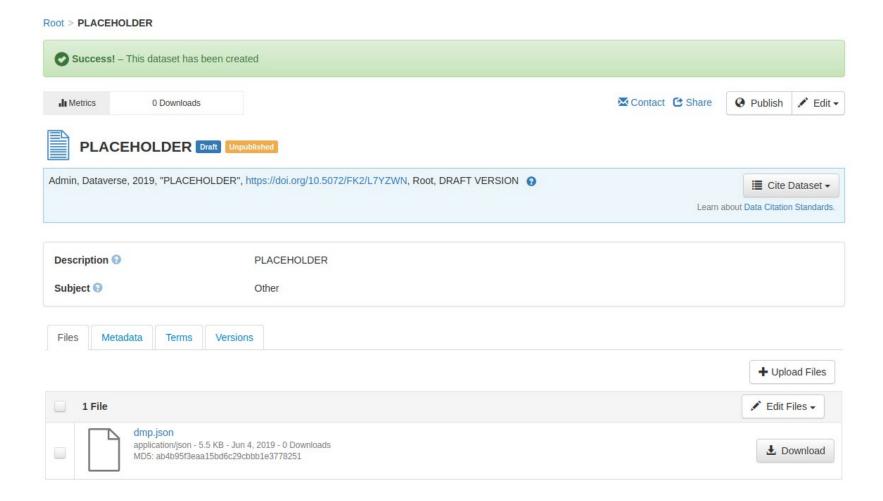








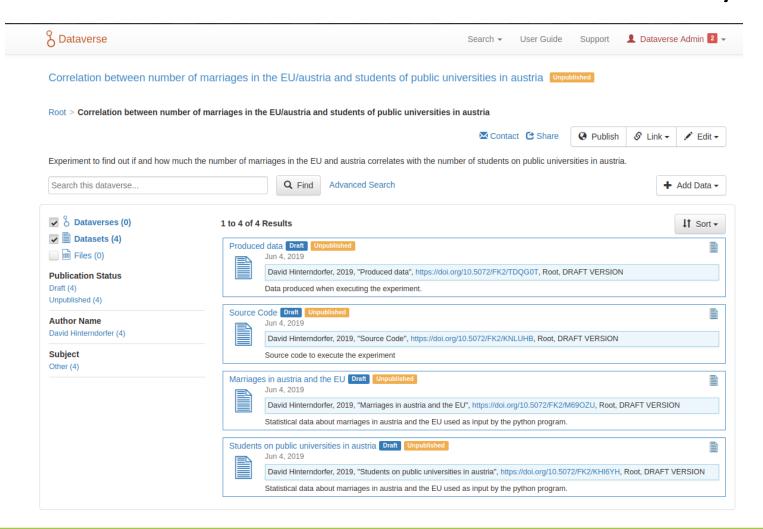
#### Upload maDMP into Dataverse







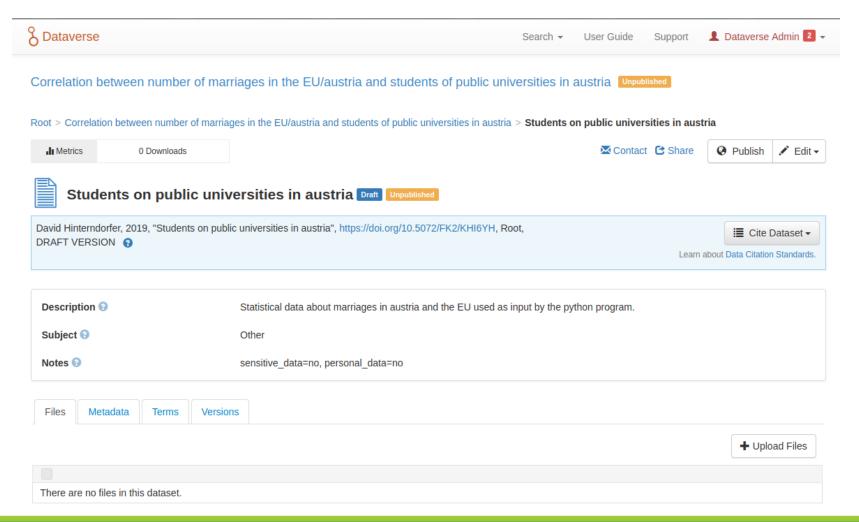
Datasets created and metadata set automatically



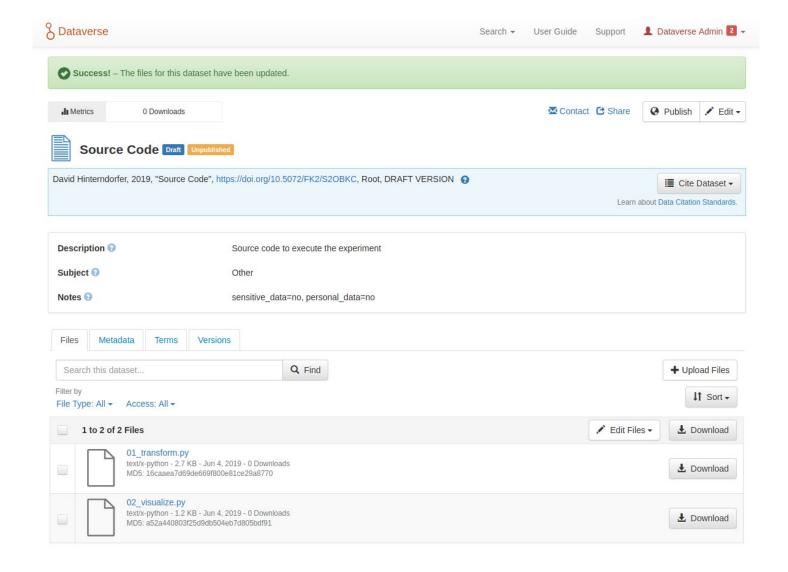




#### User must upload data only





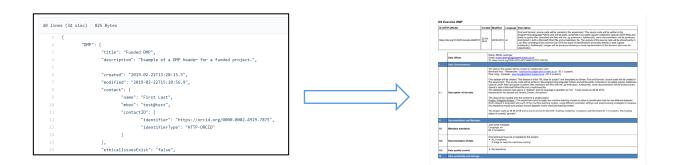






# maDMPs and funder templates

- > Funders still require traditional DMPs
- > Generate human-readable DMPs from maDMPs
  - Edit and review
    - > No blank page syndrome
  - > Pre-filled with information
  - Tailored for a specific funder
    - Template, questions asked, etc.







## maDMPs and funder templates

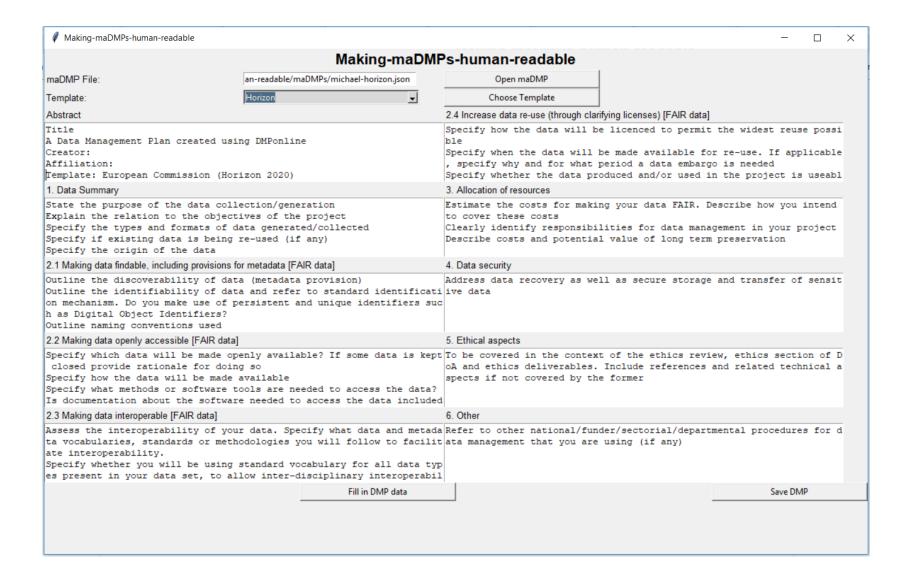
#### > Prototypes

- Developed by Data Stewardship students at TU Wien
- Include mapping
  - > Horizon 2020
  - FWF (Science Europe compliant)
- > Input
  - maDMP as JSON
- Output
  - > PDF
- Code and (ma)DMP examples available on GitHub
  - https://github.com/datastewardstuw/DataStewards2
  - https://github.com/MBAigner/Making-maDMPs-human-readable
  - https://hasankhatib.github.io/maDMPTranslator/





## Making maDMPs human readable

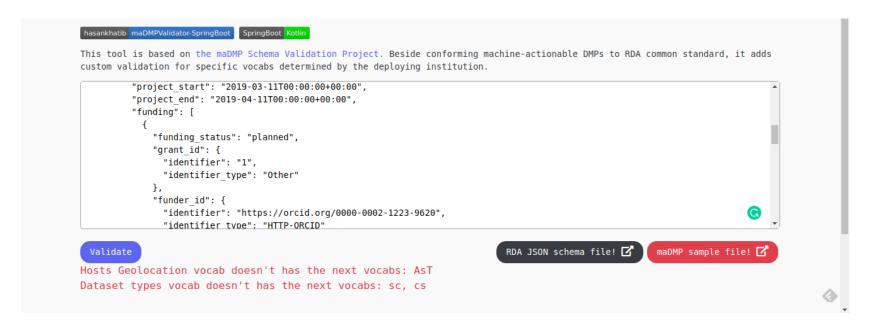






#### maDMPs schema validation

- Validates JSON files against JSON Schema
- Additionally checks
  - Values in controlled vocabularies
  - Links accessibility
- Configurable schema and dictionaries



https://github.com/HasanKhatib/maDMPValidator-SpringBoot





#### DMP Online and maDMPs

- > Prototype
  - by Sam Rust (DCC DMP Online)
  - > exports a minimal maDMP
- See webinar and visit joint session this afternoon

# Implementing maDMPS in DMPonline ... Sam Rust - Systems Developer, DCC 31 July 2019

https://www.rdalliance.org/sites/default/files/Implementing%20m aDMPS%20in%20DMPonline.pdf

# HIGHLIGHTING SOLUTIONS PROPOSED BY RDA ACTIVE DMPS, EXPOSING DMPS AND DMP COMMON STANDARDS WORKING GROUPS



Daniel Bangert Göttingen State and University Library RDA Secretariat



Peter Neish University of Melbourne RDA DMP Common Standards WG



Sam Rust Digital Curation Centre



Kathryn Unsworth
Commonwealth Scientific and
Industrial Research Organisation
RDA Exposing DMPs WG



Tomasz Miksa SBA Research RDA DMP Common Standards WG





https://www.rd-alliance.org/rda-working-groups-solutions-dmp-recording-and-slides-webinar-now-available-0





#### DMP Roadmap export to maDMPs

- > Prototypes
  - developed by Data Stewardship students at TU Wien
    - https://github.com/martinpichler/dmp-roadmap-parser
    - https://github.com/lukasinschlag/maDMP\_export\_tool
  - export maDMPs from DMP Roadmap
    - DMP Online, DMP Tool
- Mappings
  - > Horizon 2020
  - > FWF



#	Plan	Description	
11	Markus Test 1	Abstract -> description	<b>≜</b> Download maDMP
12	Data stewardship: Climate and energy resources	This experiment is used to show the correlation between the yearly climate values and the used energy resources, in the years 1992-2016. The idea is that the climate has an impact on the amount of energy needed. The input data for this experiments is taken for the climate values from data.gv and for the energy resources from Statistik Austria.	Download maDMP
13	Data stewardship: Climate and energy resources	This experiment is used to show the correlation between the yearly climate values and the used energy resources, in the years 1992-2016. The idea is that the climate has an impact on the amount of energy needed. The input data for this experiments is taken for the climate values from data.gv and for the energy resources from Statistik Austria.	Download maDMP





## RDM Organizer

- Developed by Data Stewardship students at TU Wien
- Export maDMPs from RDM Organizer
- Mappings
  - > Horizon 2020,
  - > FWF (Science Europe compliant)



- > Prototype
  - https://helmuthb.github.io/dmp-tools-actionable/
- Planned by the RDM Organizer team
  - maDMPs to be used in the internal data model
  - update from Robert Ulrich





#### maDMPs at German Aerospace Center (DLR)

- > maDMPs as one of the key elements in RDM
- > Project funded and starting soon







# Discussion





# Wrap-up and next steps

Part 4





#### Next steps

- Standard is frozen
- Start the process for the official RDA Recommendation
- Two adoptions needed
  - What is the level of adoption at your organisation? (E.g., pilot, production, ...)
  - Do you plan to continue to use the Recommendation?
  - Did you need to modify the Recommendation for your use?
  - Can you give an estimate of how much time / effort you have spent on the adoption so far?
  - What's your overall experience? (E.g., Very Good, Good, Fair, Poor)
  - Would you do it again?
- Maintenance through this WG or Active DMPs IG tbd.





#### maDMPs - take them and use them!

- Standard specification
  - https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard
- Participate in model adoption!
- Contact group chairs
  - Questions
  - Ideas
  - Success stories



Tomasz Miksa



Paul Walk



Peter Neish





## Visit our joint session

- More on machine-actionable DMPs
  - > Today, after lunch

- Towards FAIR Data Management Plans, from principles to practice
  - Active Data Management Plans IG
  - Exposing Data Management Plans WG
  - DMP Common Standards WG





#### **Publications**

- Tomasz Miksa, Peter Neish, Paul Walk, Andreas Rauber: Defining requirements for machine-actionable Data Management Plans. iPres 2018
- Tomasz Miksa, Stephanie Simms, Daniel Mietchen, Sarah Jones: Ten principles for machine-actionable data management plans. PLOS Computational Biology (in press)
- Tomasz Miksa, João Cardoso, José Luis Borbinha: Framing the scope of the common data model for machine-actionable Data Management Plans. BigData 2018: 2733-2742
- Asztrik Bakos, Tomasz Miksa, Andreas Rauber: Research Data Preservation Using Process Engines and Machine-Actionable Data Management Plans. TPDL 2018: 69-80

