From Requirements to Models

DMP Common Standards WG

Tomasz Miksa, Paul Walk, Peter Neish

Agenda

- Part 1 – Introduction for newcomers
- Part 2 – Status update - open consultations
- Part 3 – Lightning talks from DMP tool owners
- Part 4 – Discussion
- Part 5 – Wrap-up and next steps

Collaborative note taking & attendance

https://goo.gl/crvjh1

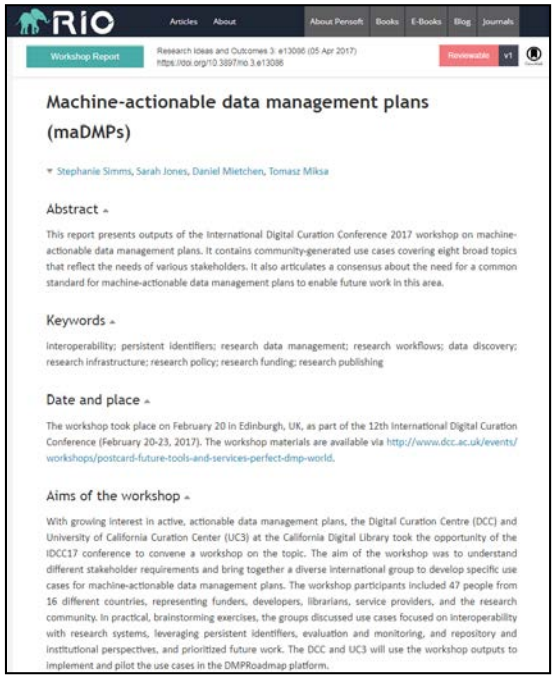
https://guidebook.com/g/rdaplenary11

Part 1

Introduction for newcomers

Challenge – 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 v1

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

Part 2

Open consultations

User story consultation

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

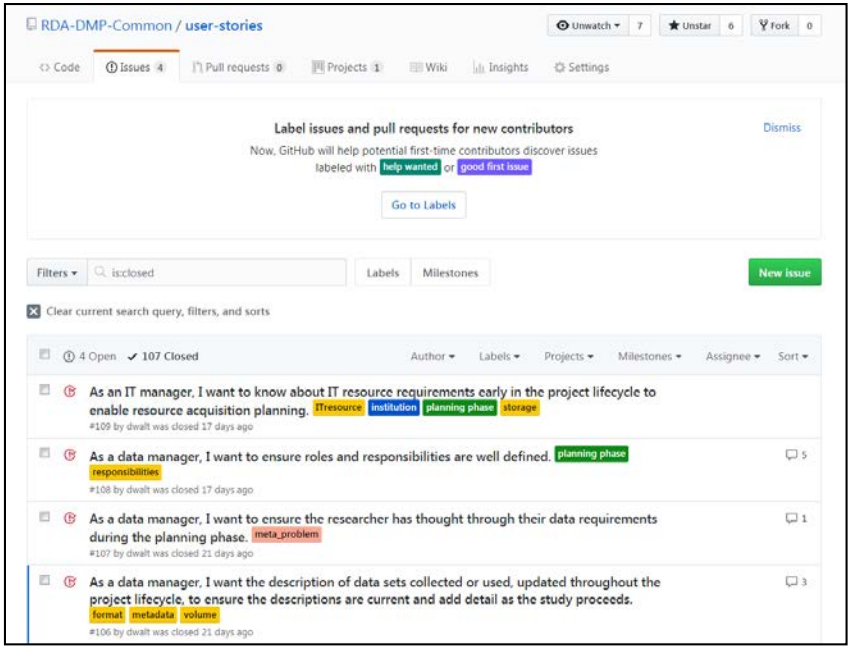
As a <stakeholder>, I want <goal> so that <reason >.

*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/

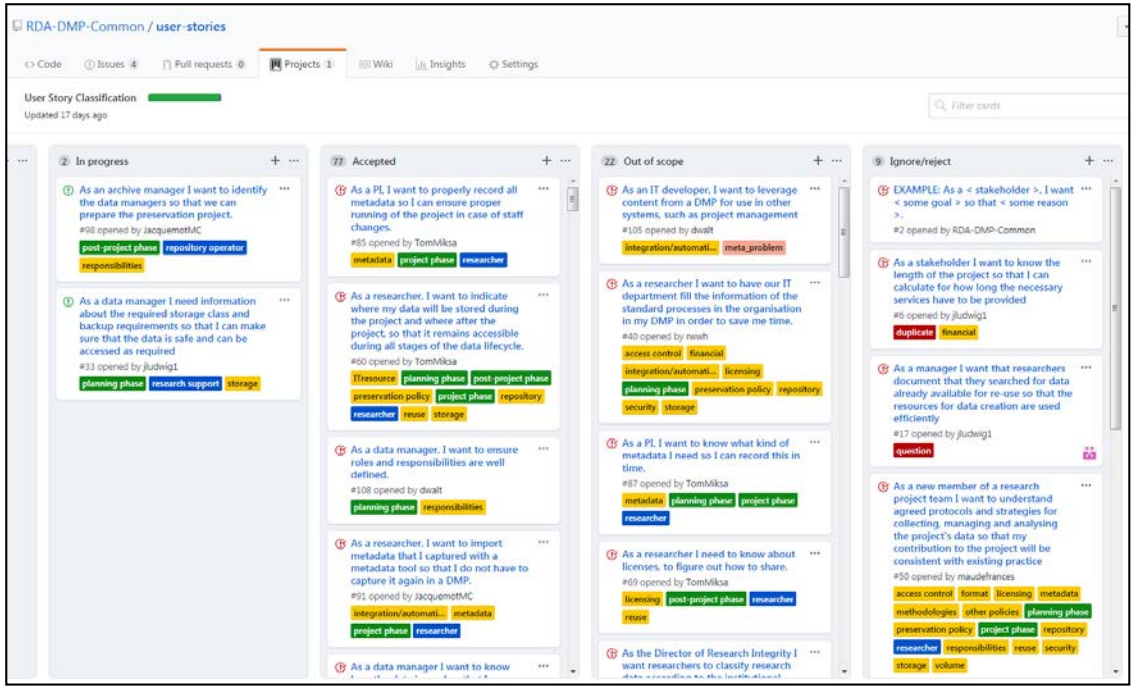
User story consultation

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



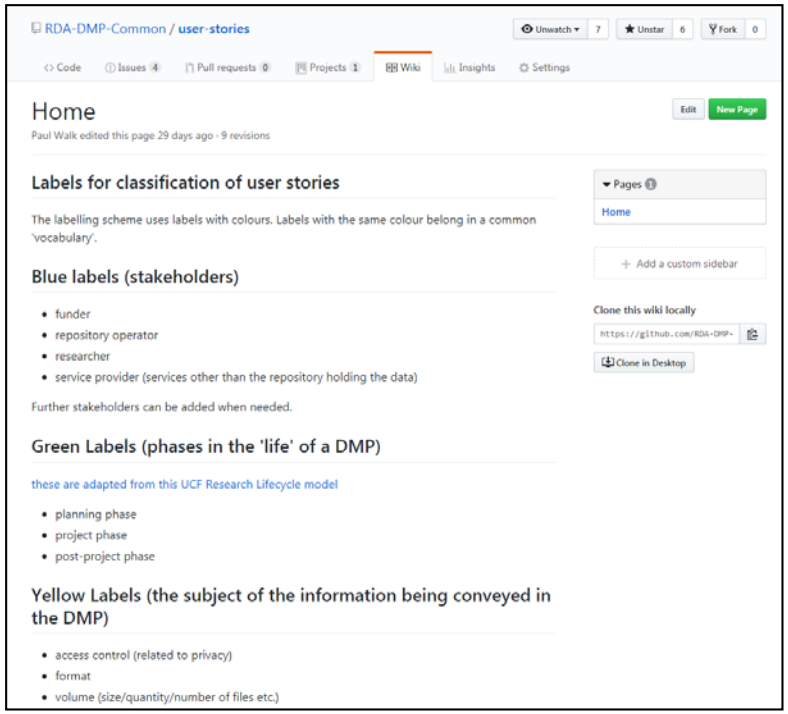
User story labelling

- https://github.com/RDA-DMP-Common/user-stories/projects/2
- Reviewed by chairs and authors
  - classified
    - in scope - useful for model definition
    - out of scope – often referring to the ecosystem, practices – important but not directly for the common data model
  - labelled



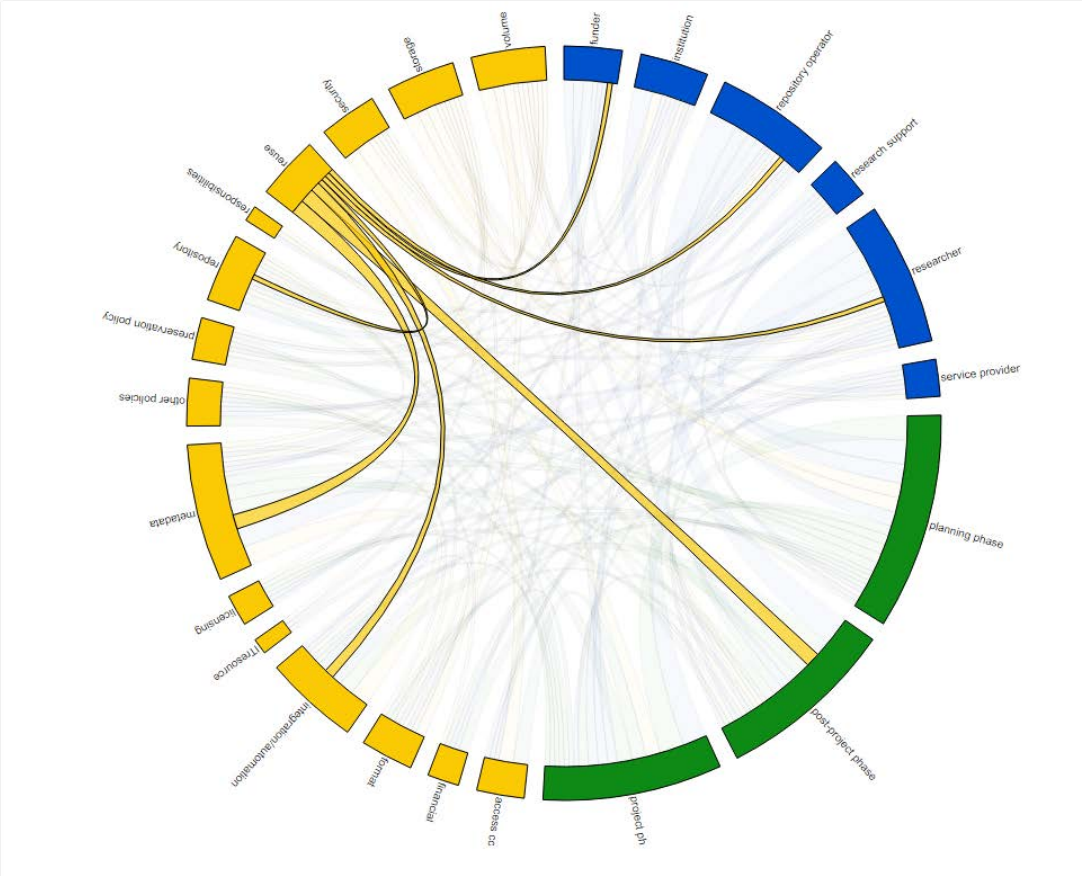
User story labelling

- 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



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]

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

'yellow' label used to classify user stories

- **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]

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

short summary of what user stories are about – more specific requirements

- 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]

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

IDs of user stories (to keep connection to the GitHub consultation)

- 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**]

Requirements grouping

- Similar requirements exist under different labels
- Example
  - information on the author of the DMP is relevant for
    - administrative activities
    - reuse
- We split requirements and grouped them using five categories
  - **Administrative, Roles and Responsibilities**
  - **Data**
  - **Infrastructure**
  - **Security, Privacy and Access Control**
  - **Policies, legal and ethical aspects**

Requirements grouping example (Data)

- DATA
  - Format
    - Format [80, 12, 99, 62, 67, 54, 80]
  - Volume
    - Data size estimate [5, 77, 80, 100]
      - For specific type of data [62]
    - Data size real [54]
  - Provenance [54]
  - Metadata
    - taxonomy/classification [14,11]
    - Links to metadata of the real data [89, 39]
    - Link publications to data [55]
    - Authorship [88]
    - Multilingual metadata [65]
    - Include raw metadata directly in the model [91, 85]
  - Reuse
    - Links to (meta-)data location [89, 90, 56, 39, 60]
  - Repository [42]
    - Persistent identifier for data [92]
    - Link publications to data [55, 88]
    - Link to License/Contract allowing data usage/storing [56]

Note: we did not move all requirements falling under a specific label, but only a subset that is relevant in this context – in the given example, relevant for data description. Other requirements for Reuse were put into other categories.

Next steps

- 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 will go deep
  - how do we model specific requirements
    - which specific fields are needed?
    - which models exist?

Consultation 2 – ‘going deep’

- https://goo.gl/DRieP4

- 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**

Consultation 2 – ‘going deep’

- Goal: reach out to experts in each category to learn
  - which concrete information (specific fields)
  - in what form they expect
- Next
  - compare collected models and fields
  - select best fitting
  - design the architecture of the model (core model with extensions vs flat model, serializations, etc.)
- Timeline:
  - end of May 2018

Part 3

Lightning talks

Slot #	DMP software	Mode
#1	Data Stewardship Wizard	in-person
#2	DMP Service (OpenAIRE)	in-person
#3	RDMOrganiser	in-person
#4	DMPRoadmap (covers DMPonline and DMPtool from DCC and UC3)	in-person
#5	ReDBox (QCIF)	remote
#6	UQ Research Data Manager (UQRDM)	remote

DTL

DUTCH TECHCENTRE FOR LIFE SCIENCES

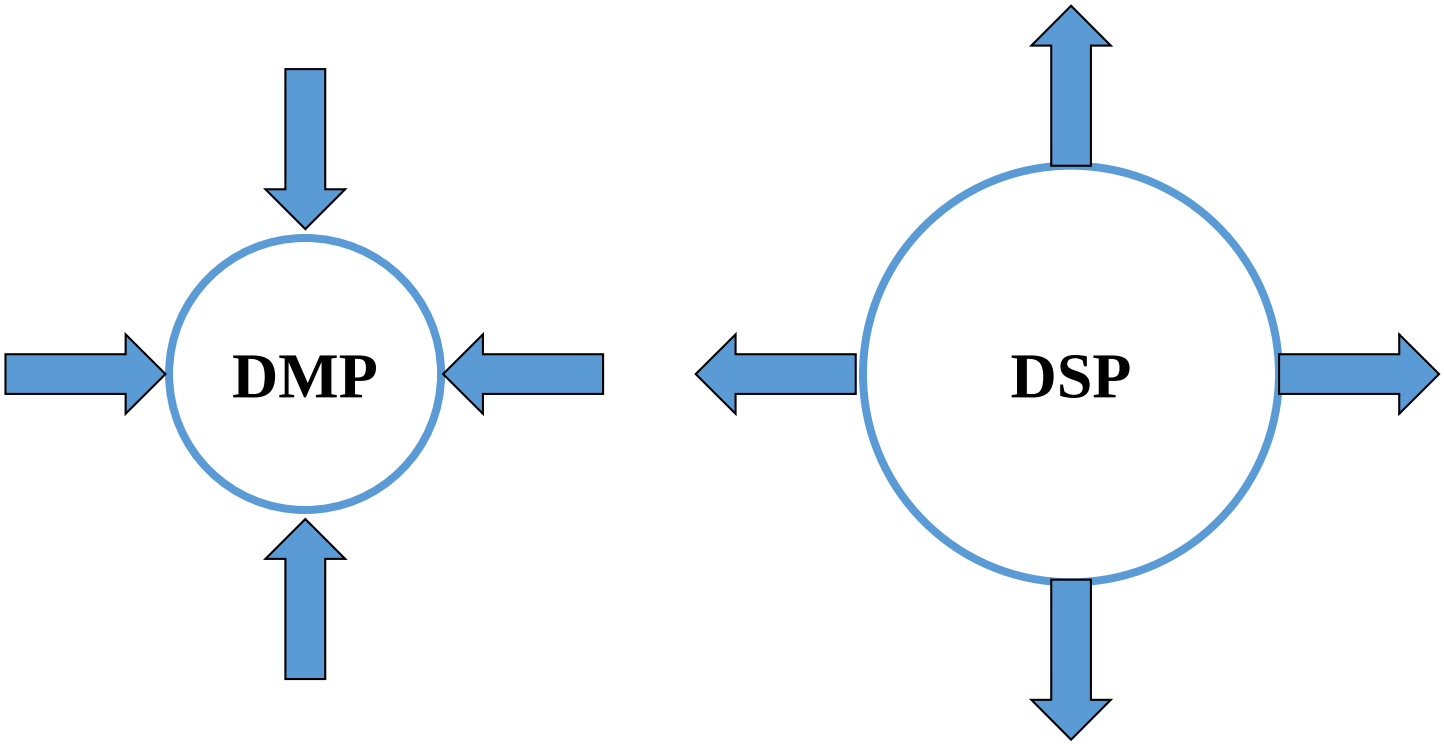
elixir NETHERLANDS

DATA MODEL OF THE ELIXIR DATA STEWARDSHIP WIZARD

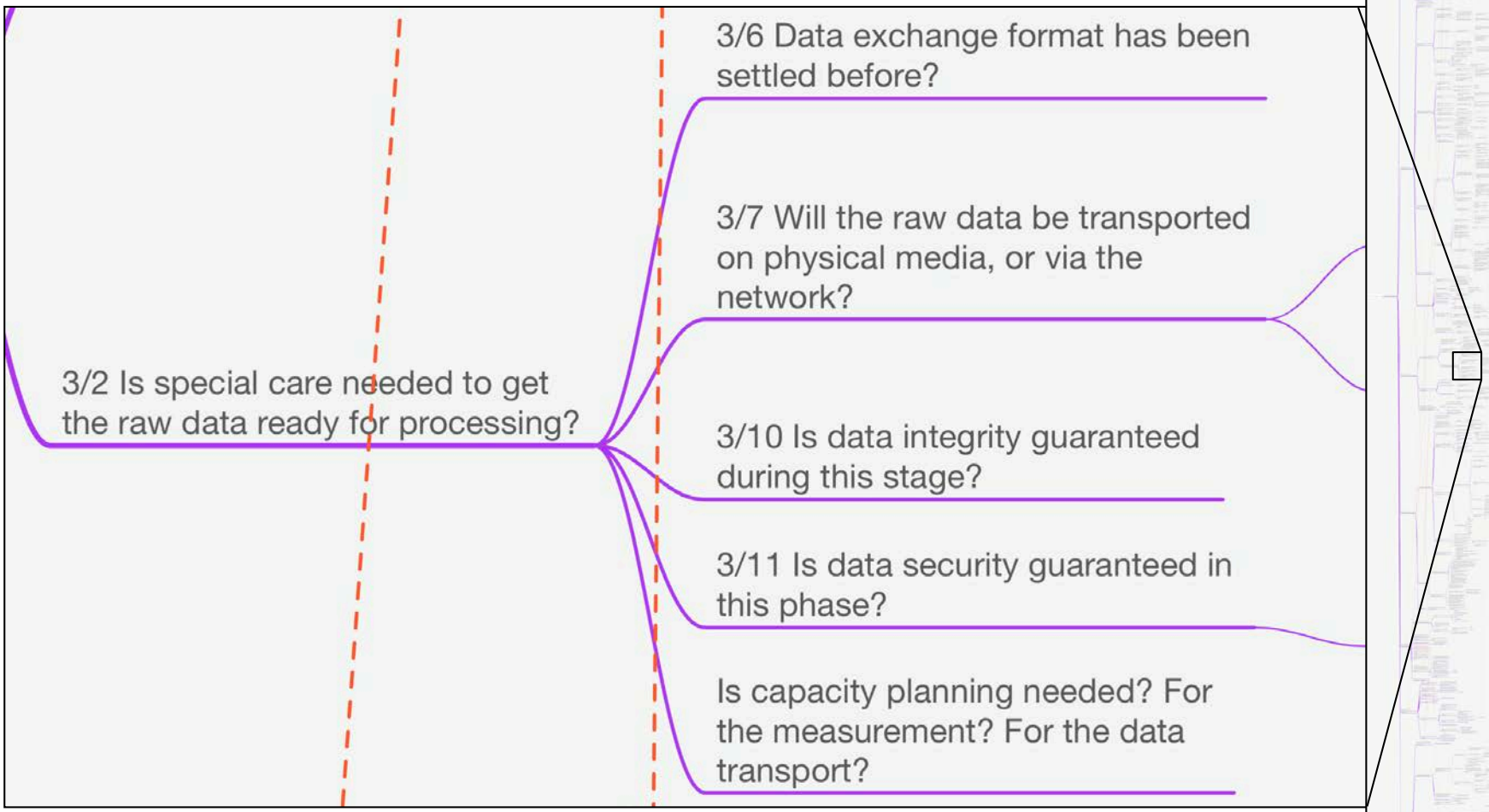
ROB HOOFT

March 2018

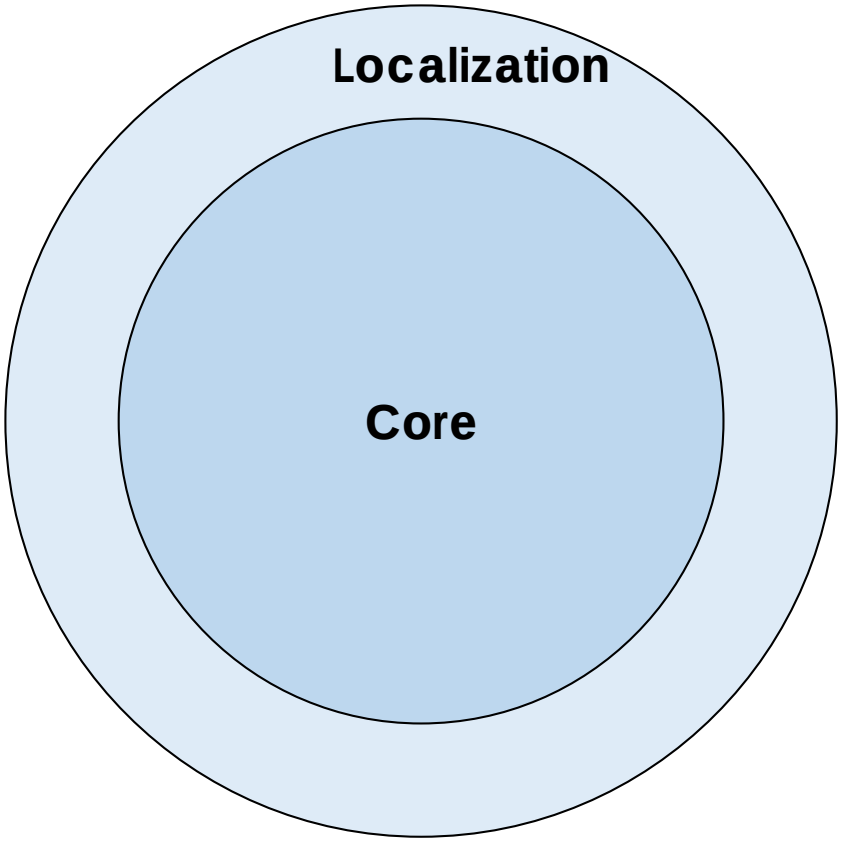
The incredibly shrinking DMP?



Hierarchical



EXTENSIBLE



JSON

```
{
  "questionid": 30, "type": "option",
  "precondition": { "questionid": 26, "answerid": 0 },
  "title": "When is the raw data archived?",
  "answers":
     [
       { "id": 0, "label":
         "As soon as it comes in, in chunks" },
       { "id": 1, "label":
         "As soon as it has all arrived, in one session"
},
       { "id": 2, "label":
         "All at once with the results at project end" }
     ]
}
```

Slot #	DMP software	Mode
#1	Data Stewardship Wizard	in-person
#2	DMP Service (OpenAIRE)	in-person
#3	RDMOrganiser	in-person
#4	DMPRoadmap (covers DMPonline and DMPtool from DCC and UC3)	in-person
#5	ReDBox (QCIF)	remote
#6	UQ Research Data Manager (UQRDM)	remote

Research Data Management Organiser

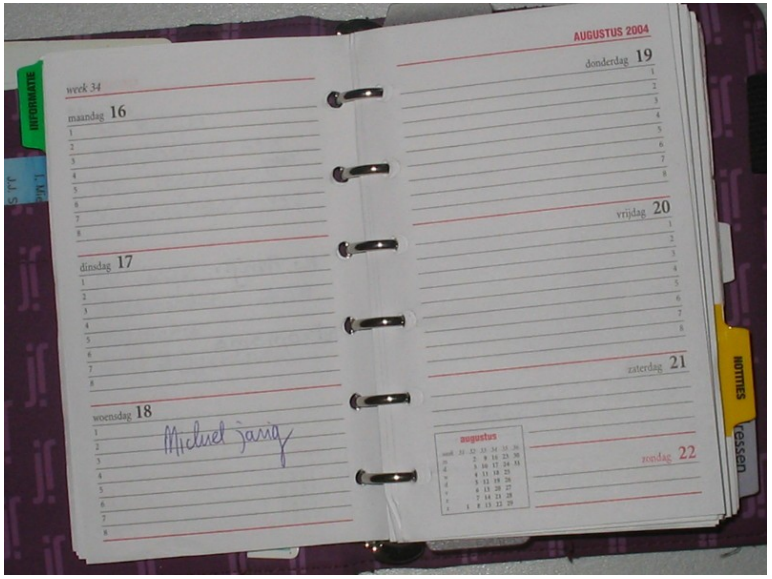
rdmorganiser.github.io

@rdmorganiser

Why RDMO?

Organizer instead of plan

- Organize the data management over the whole lifetime of a project and beyond
- Engage all stakeholders

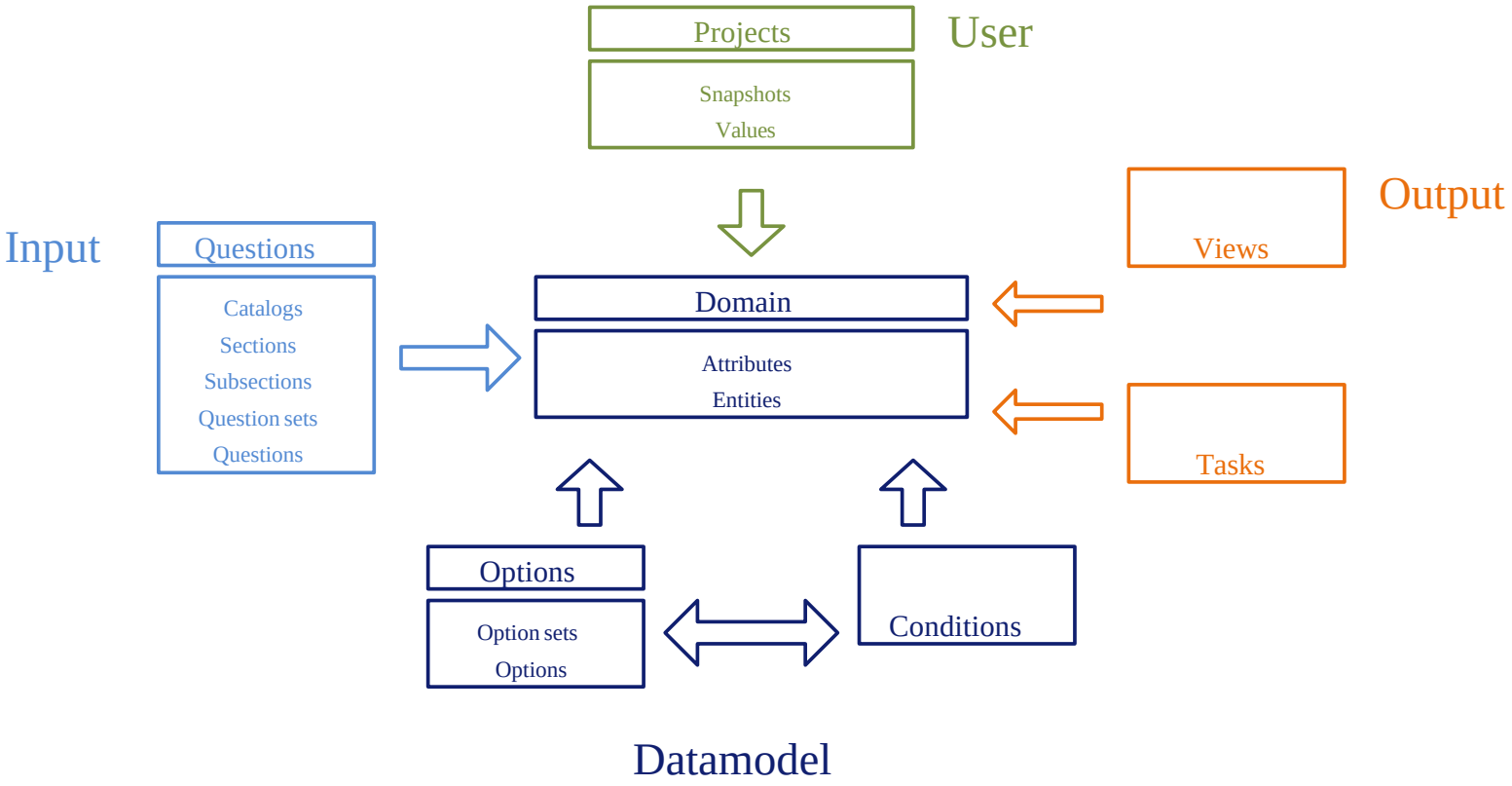


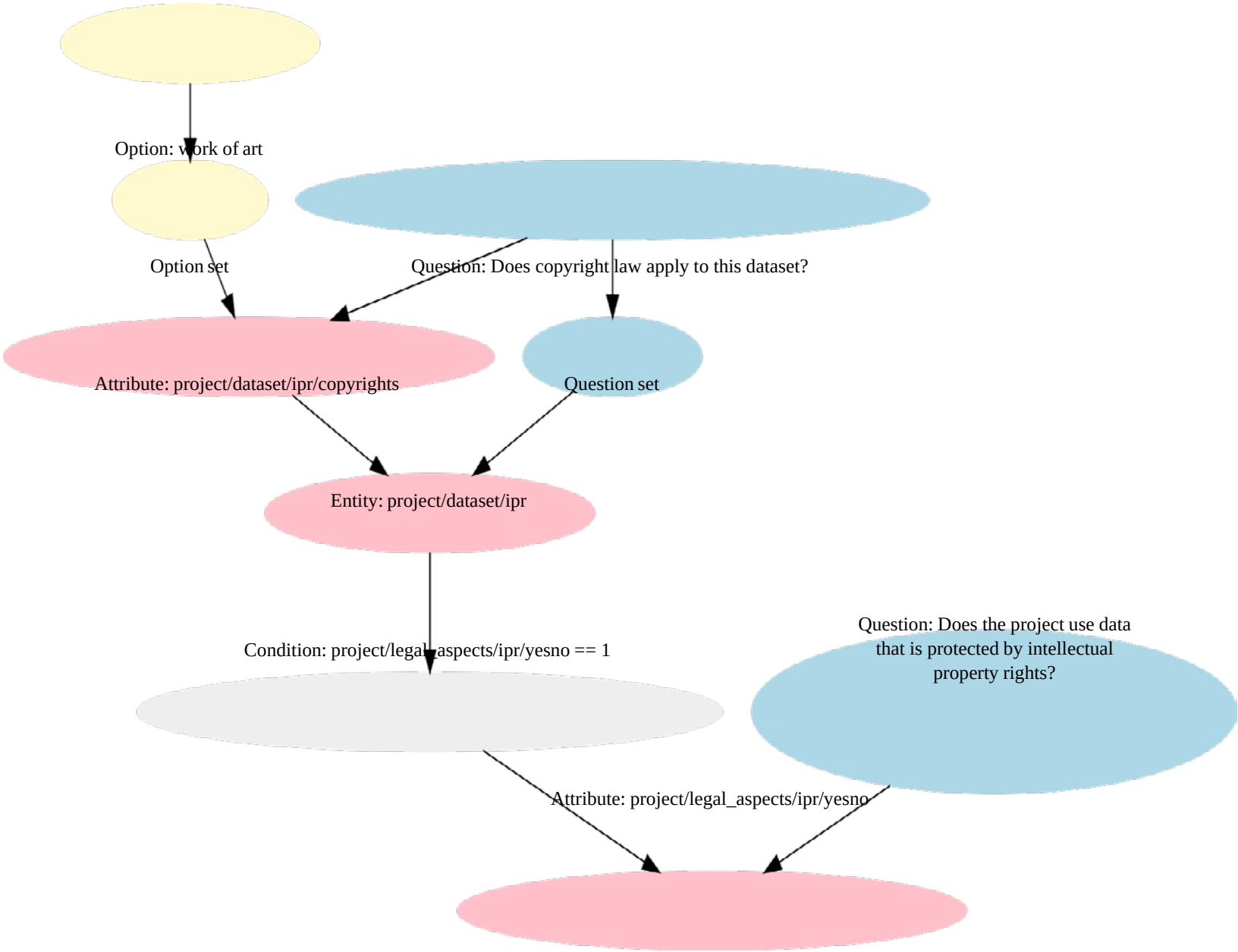
Local instead of central

- Full customization of the content
  - disciplinary context
  - institutional environment
- Easily deployable for universities or infrastructure providers
- Embed into existing infrastructure

Image credit: M.Minderhoud, Agenda, CC BY-SA 3.0

RDMO Architecture




Option: work of art
Option set
Question: Does copyright law apply to this dataset?
Attribute: project/dataset/ipr/copyrights
Question set
Entity: project/dataset/ipr
Condition: project/legal_aspects/ipr/yesno == 1
Question: Does the project use data that is protected by intellectual property rights?
Attribute: project/legal_aspects/ipr/yesno

Examples for attributes and entities

- project/coordination/name
- project/costs/
- project/dataset/
- project/dataset/description
- project/dataset/data_security/
- project/dataset/ipr/
- project/dataset/metadata/
- project/dataset/pids/
- project/dataset/preservation/
- project/dataset/sensitive_data/
- project/dataset/sharing/
- project/dataset/size/
- project/dataset/storage/
- project/dataset/versioning_tool
- project/funder/
- project/legal_aspects/
- project/partner/
- project/preservation/
- project/schedule/
- project/support/

RDMO XML Format

```
<catalog xmlns:dc="http://purl.org/dc/elements/1.1/">
  <dc:uri>https://rdmorganiser.github.io/terms/questions/rdmo</dc:uri>
  <dc:comment/>
  <order>1</order>
  <title lang="en">RDMO</title>
  <title lang="de">RDMO</title>
  <sections>
    <section>
      <dc:uri>https://rdmorganiser.github.io/terms/questions/rdmo/general</dc:uri>
      <dc:comment/>
      <order>0</order>
      <title lang="en">General</title>
      <title lang="de">Allgemein</title>
      <subsections>
        <subsection>
        <dc:uri>https://rdmorganiser.github.io/terms/questions/rdmo/general/topic</dc:uri>
        <dc:comment/>
        <order>0</order>
        ...
```

Programmable JSON API

```
curl -X GET -H 'Authorization: Token oojoh3phaighaebiNeiyeeCeiY3Peuv2eitoojoh' \
  https://rdmo.aip.de/api/v1/projects/values/?attribute__path=project/dataset/size/volume
```


```
[
  {
    "id":10061,
    "project":"https://rdmo.aip.de/api/v1/projects/projects/69/",
    "attribute":"https://rdmo.aip.de/api/v1/domain/attributes/262/",
    "set_index":0,
    "collection_index":0,
    "text":"",
    "option":null,
    "created":"2017-05-29T14:50:20.009917Z",
    "updated":"2017-05-29T14:50:20.009924Z"
  },
  ...
]
```

Resources

Website:	rdmorganiser.github.io
GitHub organisation:	github.com/rdmorganiser
RDMO source code:	github.com/rdmorganiser/rdmo
RDMO questionaire:	github.com/rdmorganiser/rdmo-catalog
Documentation:	rdmo.readthedocs.io
Demo instance:	rdmo.aip.de
Mailinglist:	rdmo@listserv.dfn.de
Twitter:	@rdmorganiser
Slack:	rdmo.slack.com
GitHub issues:	github.com/rdmorganiser/rdmo/issues
Supplemental material	cloud.aip.de/index.php/s/dmrTbnb1liqS3Es

Slot #	DMP software	Mode
#1	Data Stewardship Wizard	in-person
#2	DMP Service (OpenAIRE)	in-person
#3	RDMOrganiser	in-person
#4	DMPRoadmap (covers DMPonline and DMPtool from DCC and UC3)	in-person
#5	ReDBox (QCIF)	remote
#6	UQ Research Data Manager (UQRDM)	remote

DMPRoadmap Data Model

Jimmy Angelakos

RDA 11th Plenary – Berlin 21/03/2018

roadmap

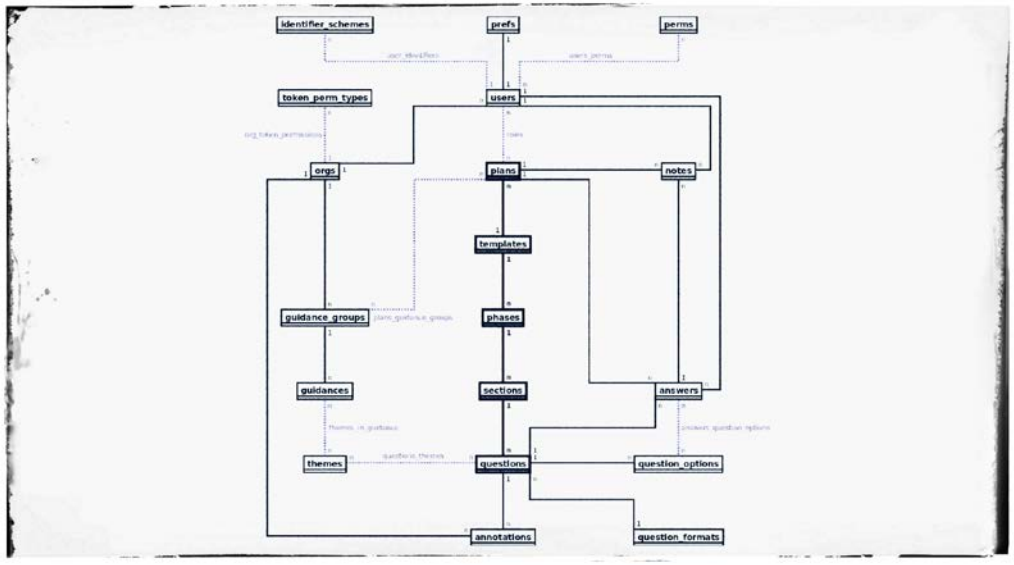
...is Open Source!

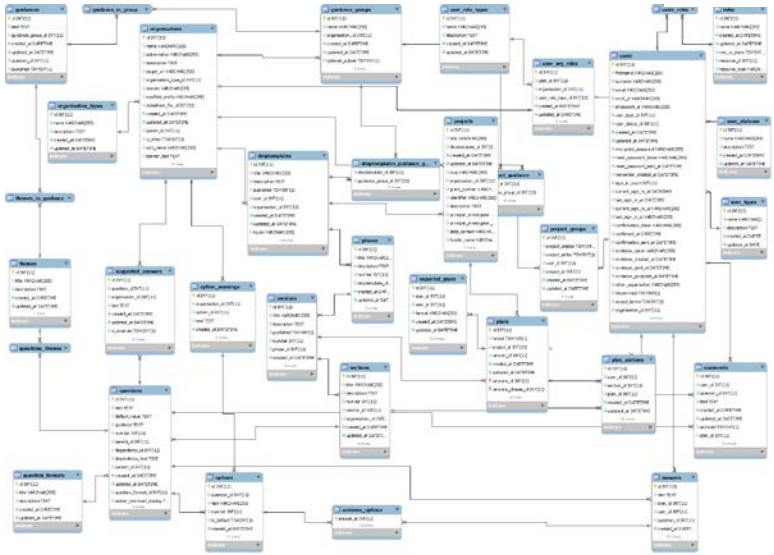
- **DMPRoadmap users:**
  - **DMPonline (DCC, UK)**
  - **DMPTool (UC3, USA)**
  - **DMP Assistant (Portage, Canada)**
  - **DMP OPIDoR (France)**
  - **DMPTuuli (Finland)**
  - **DMPMelbourne (U of Melbourne)**
  - **and more:** https://github.com/DMPRoadmap/roadmap/wiki/Local-installations-inventory

roadmap

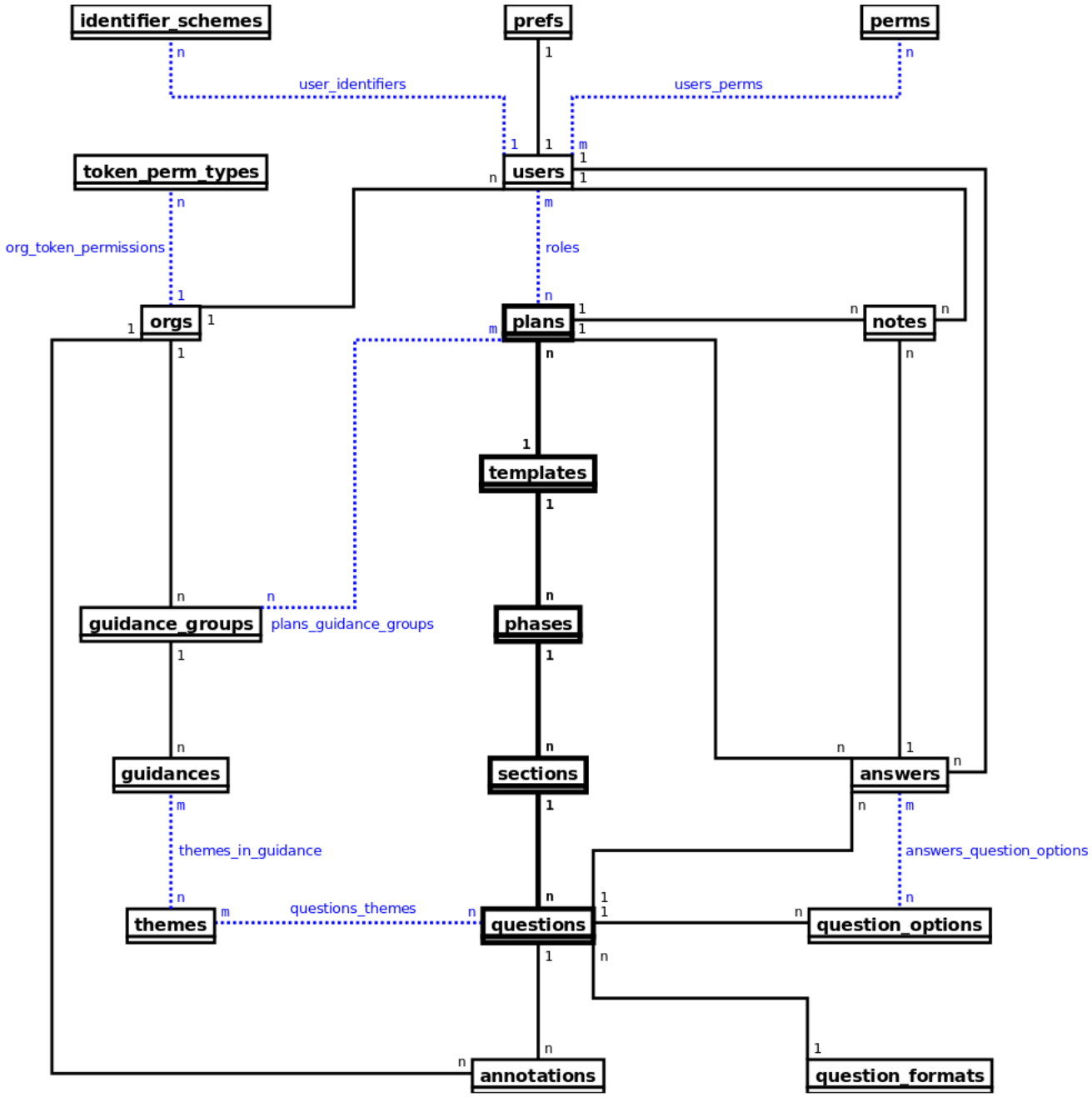
Data Model

Not as complicated as it looks!



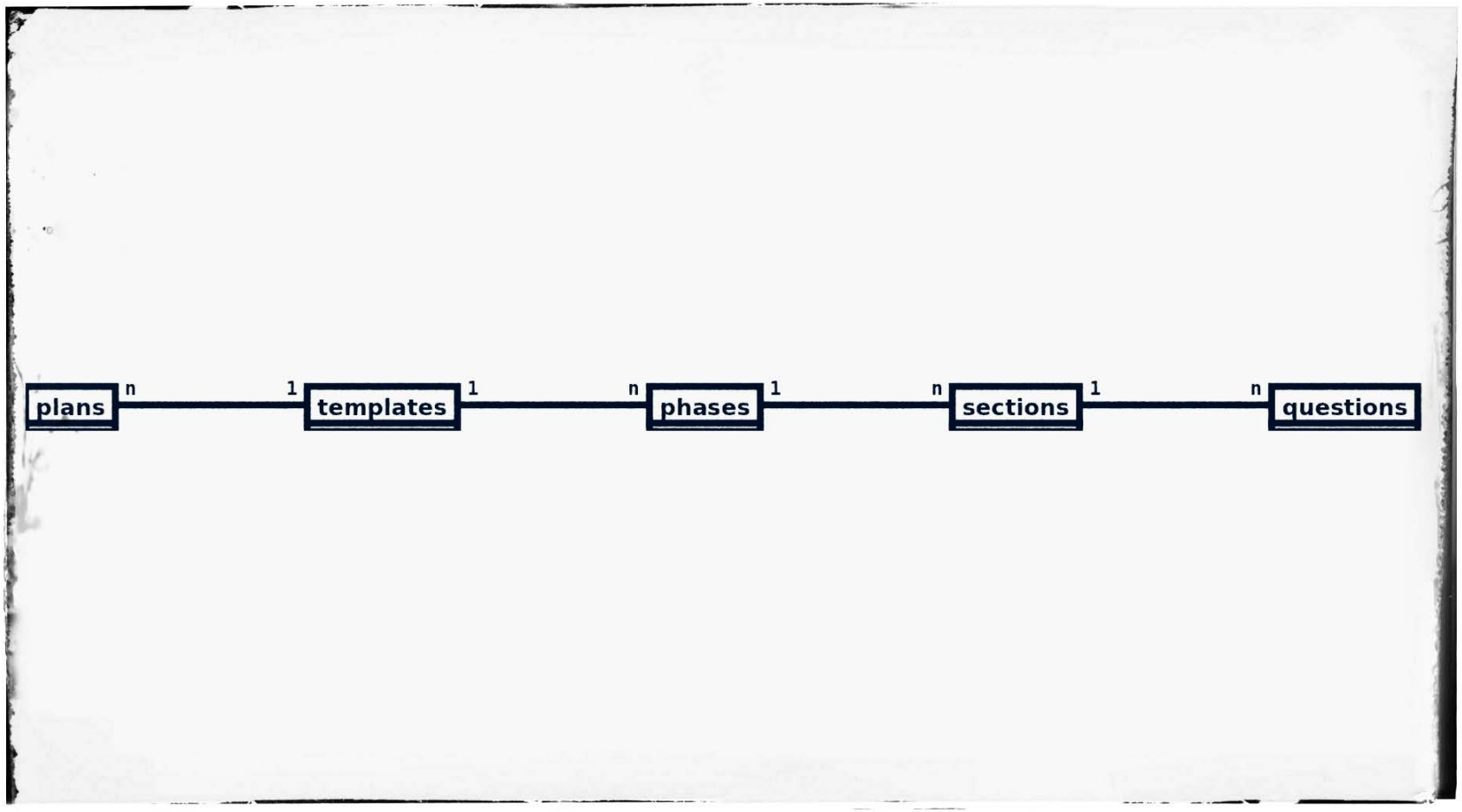


← **Old was:**


identifier_schemes
prefs
perms
user_identifiers
users_perms
token_perm_types
users
org_token_permissions
roles
orgs
plans
notes
templates
guidance_groups
plans_guidance_groups
phases
guidances
sections
answers
themes_in_guidance
answers_question_options
themes
questions_themes
questions
question_options
annotations
question_formats

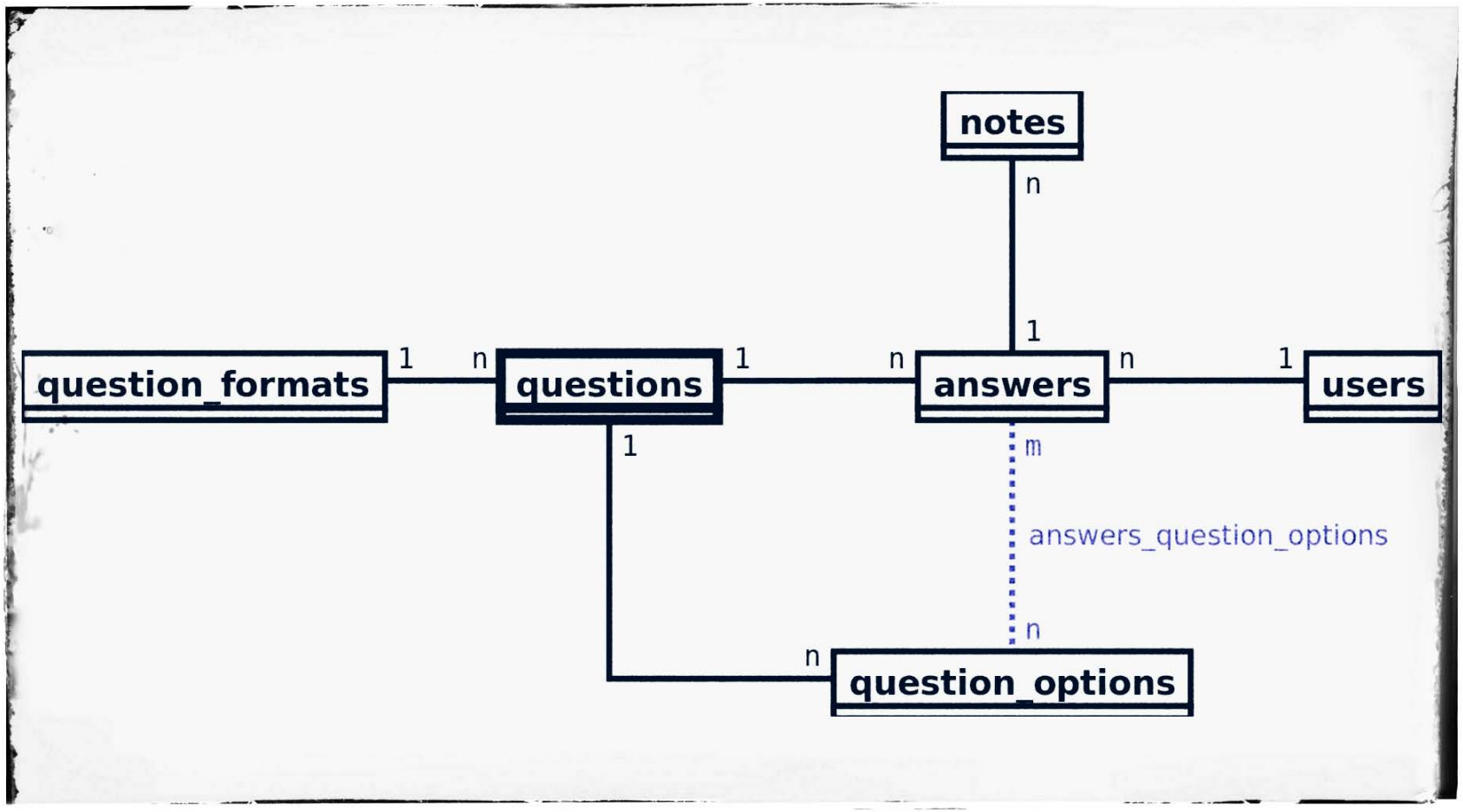
roadmap

“Main axis” closeup



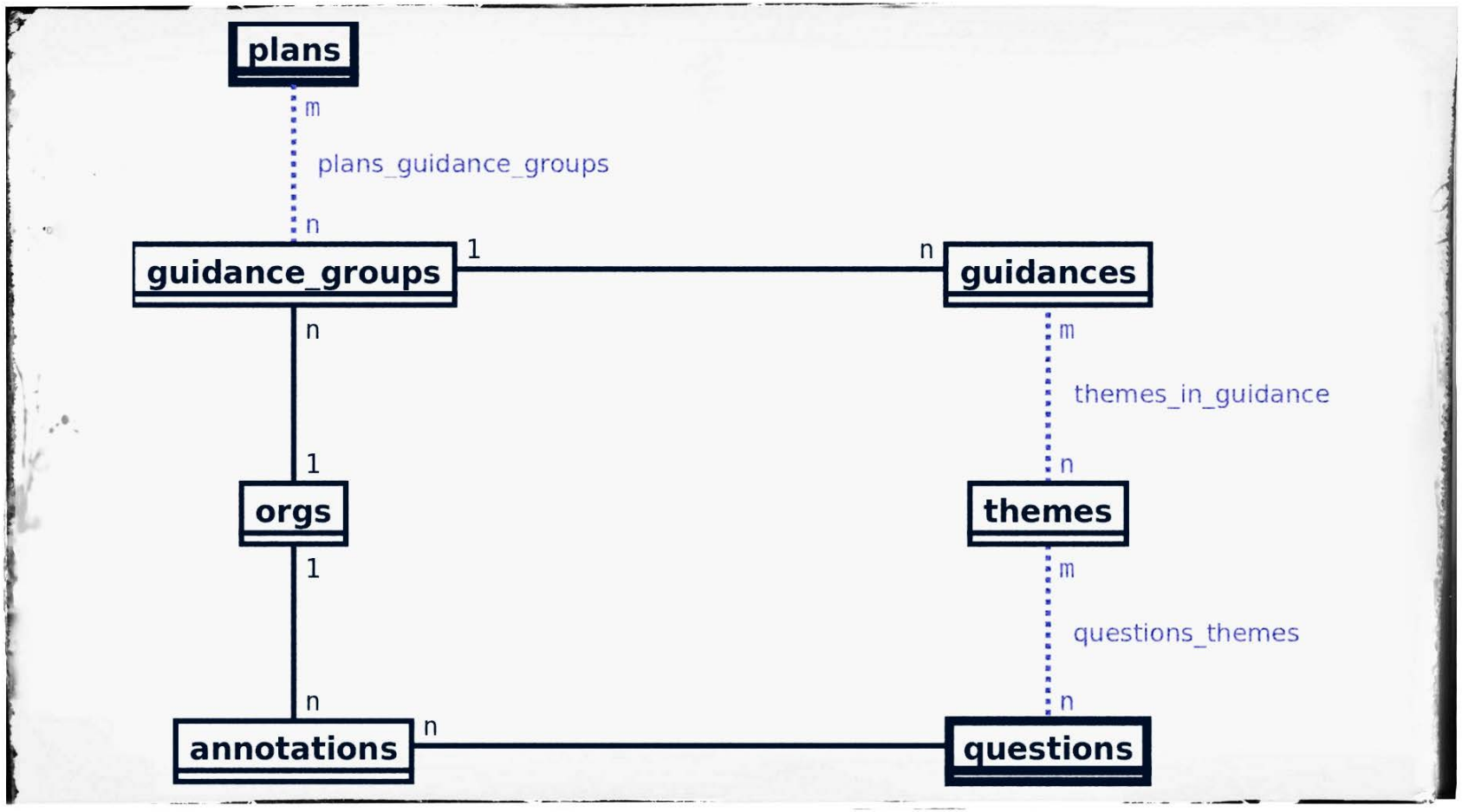
roadmap

Questions closeup



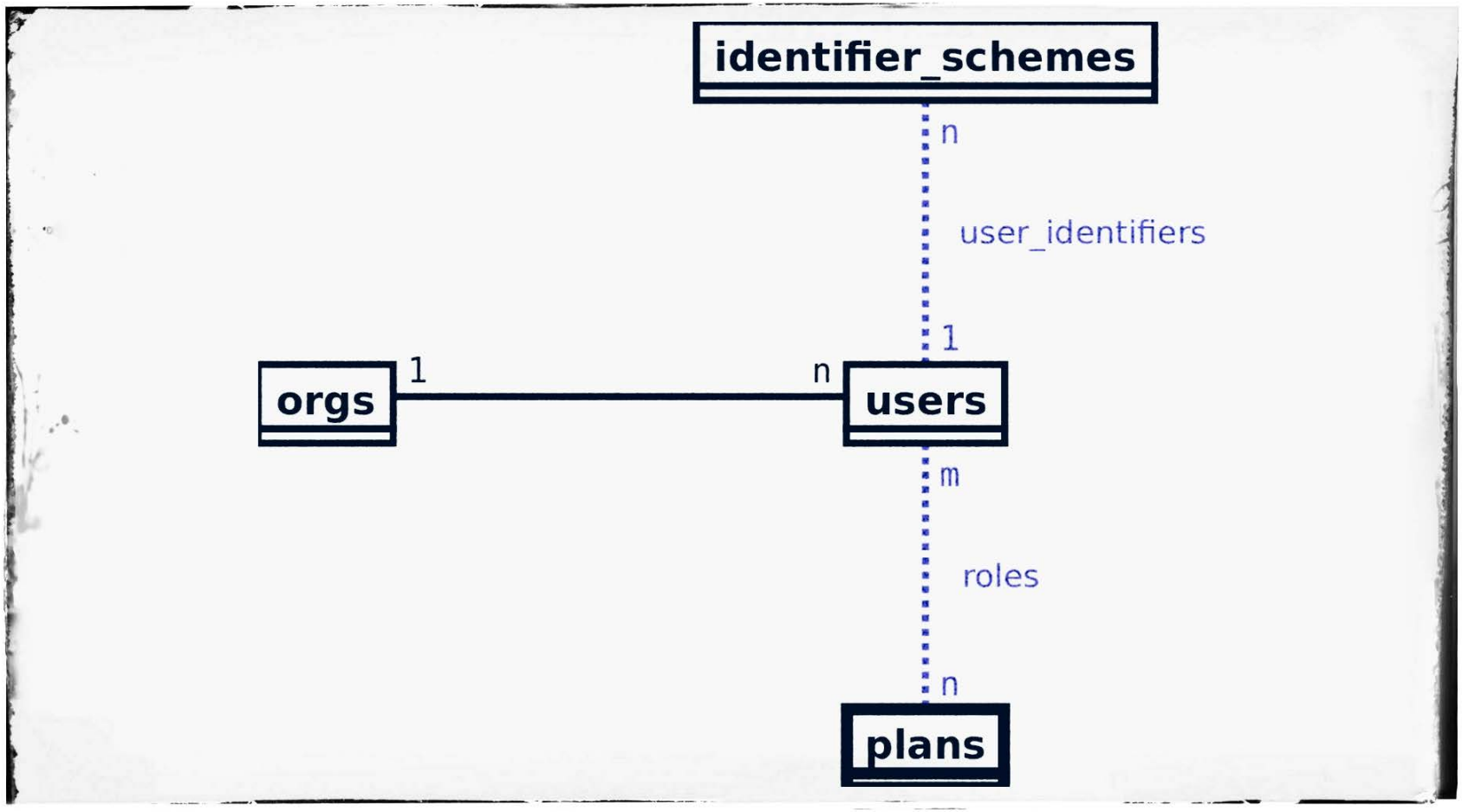
roadmap

Guidance closeup



roadmap

Users closeup



Part 4

Discussion

Part 5

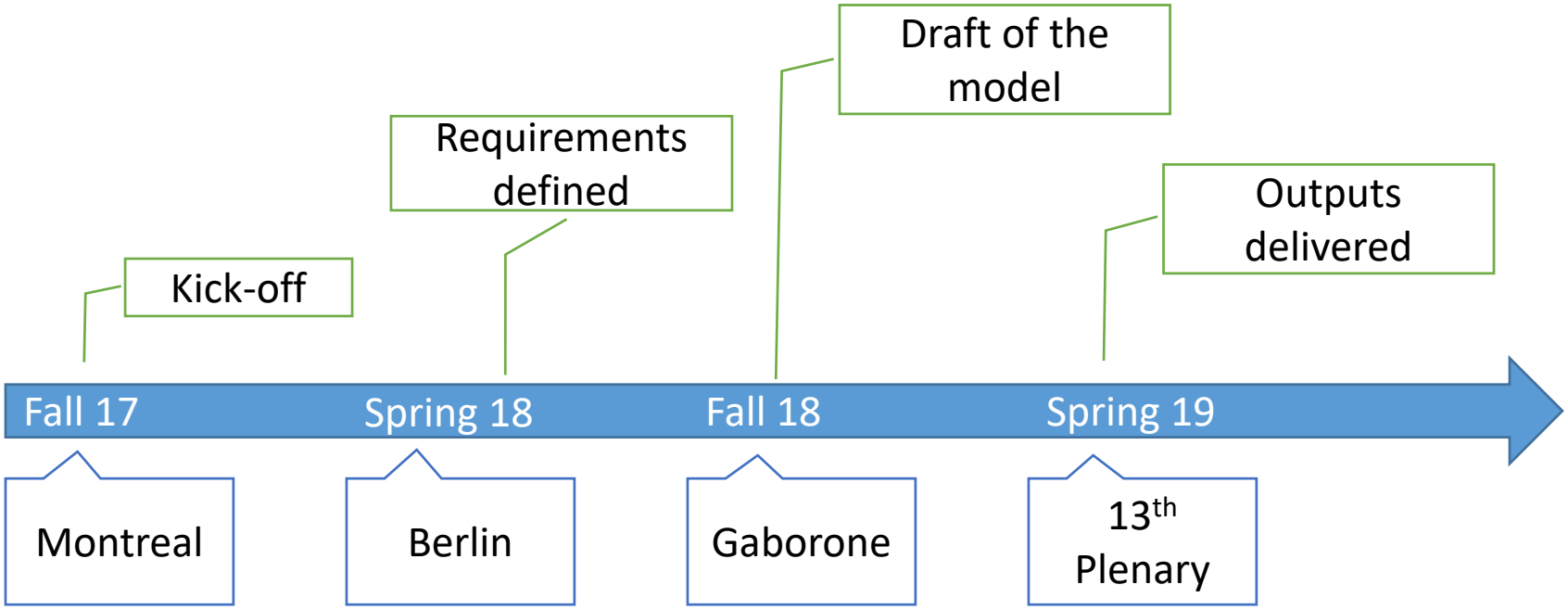
Wrap-up and next steps

Contributions needed!

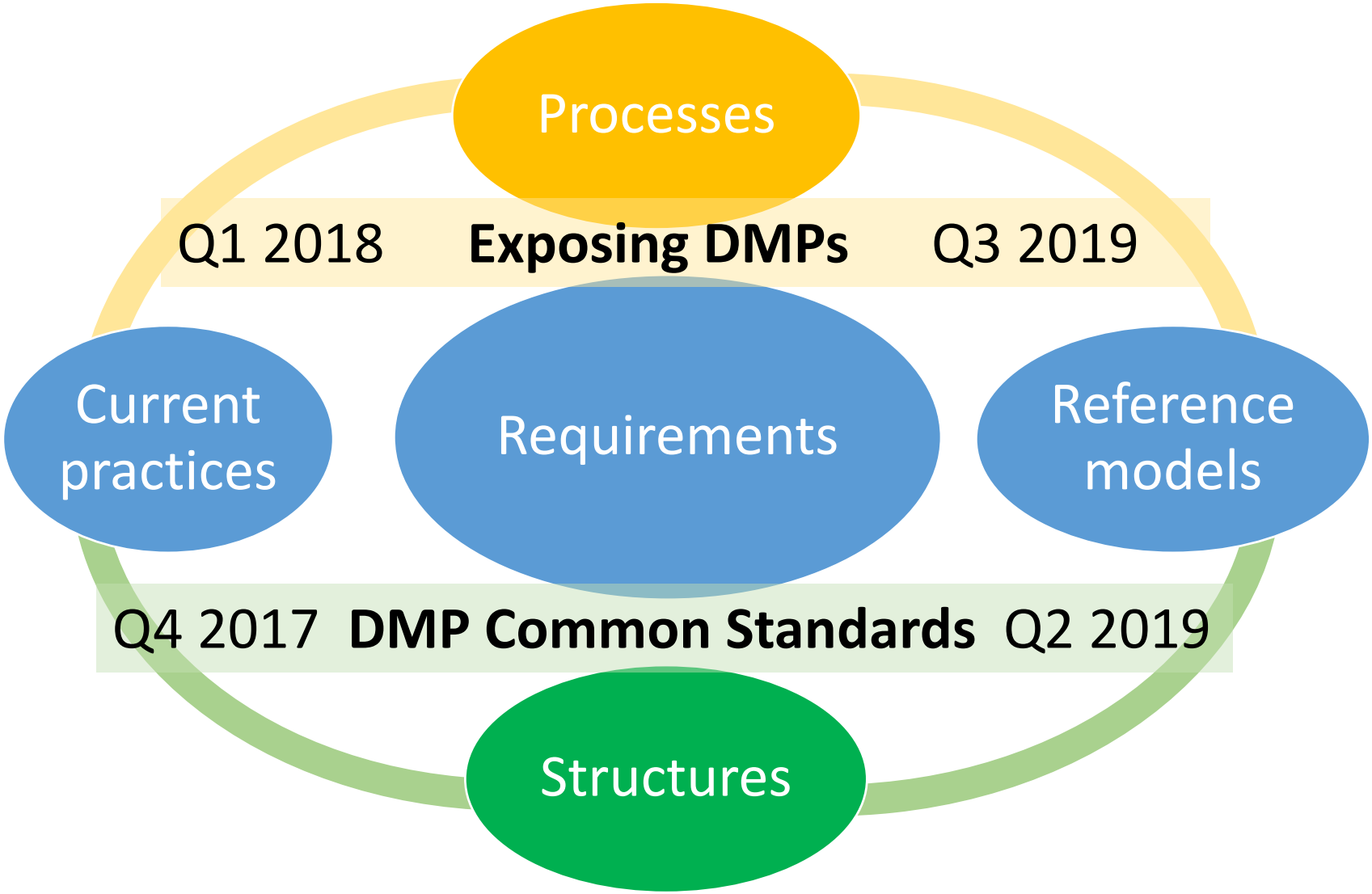
- Champions to lead requirements and model collection in each area:
  - Administrative, Roles and Responsibilities
  - Data
  - Infrastructure
  - Security, Privacy and Access Control
  - Policies, legal and ethical aspects

Timeline

- Our main focus: 12th Plenary Meeting in Botswana
  - 5-8 November 2018



Working Groups Complementary Activity



Staying in touch!

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