DMP Common Standards WG

Tomasz Miksa, Paul Walk, Peter Neish

tmiksa@sba-research.org

CC BY-SA 4.0

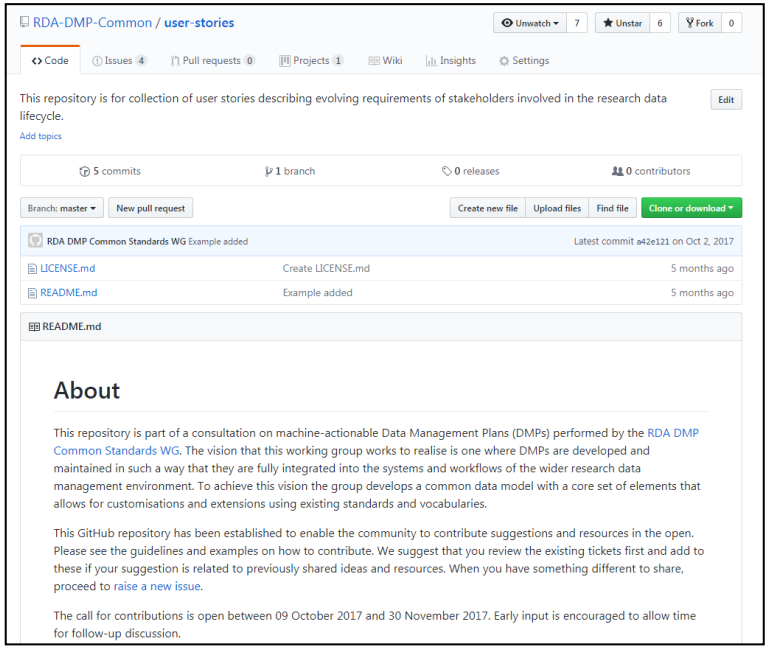
Agenda

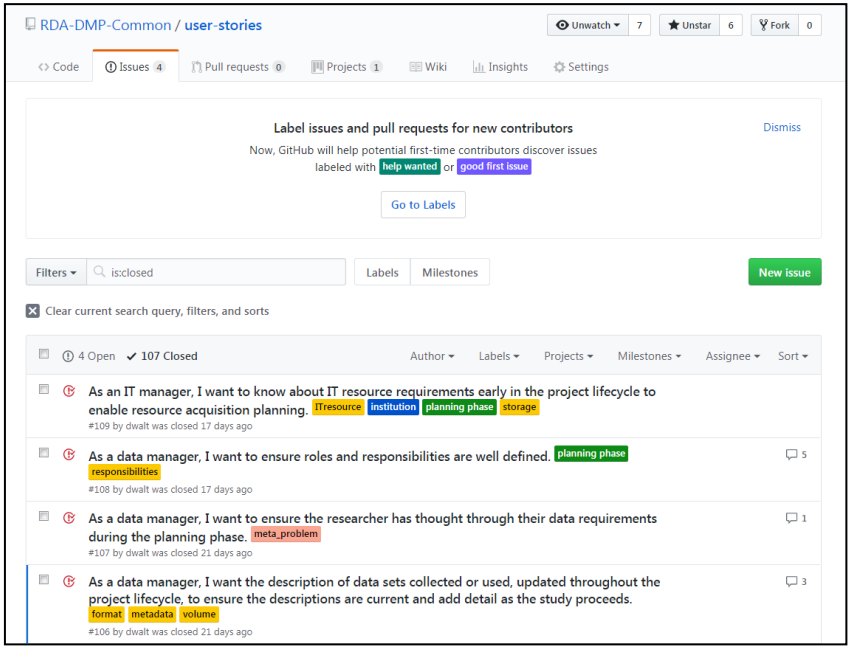
- Progress update and planning
  - results of the user story consultation
  - next activities
- Plenary in Berlin
- Organisational

CC BY-SA 4.0

User story consultation

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

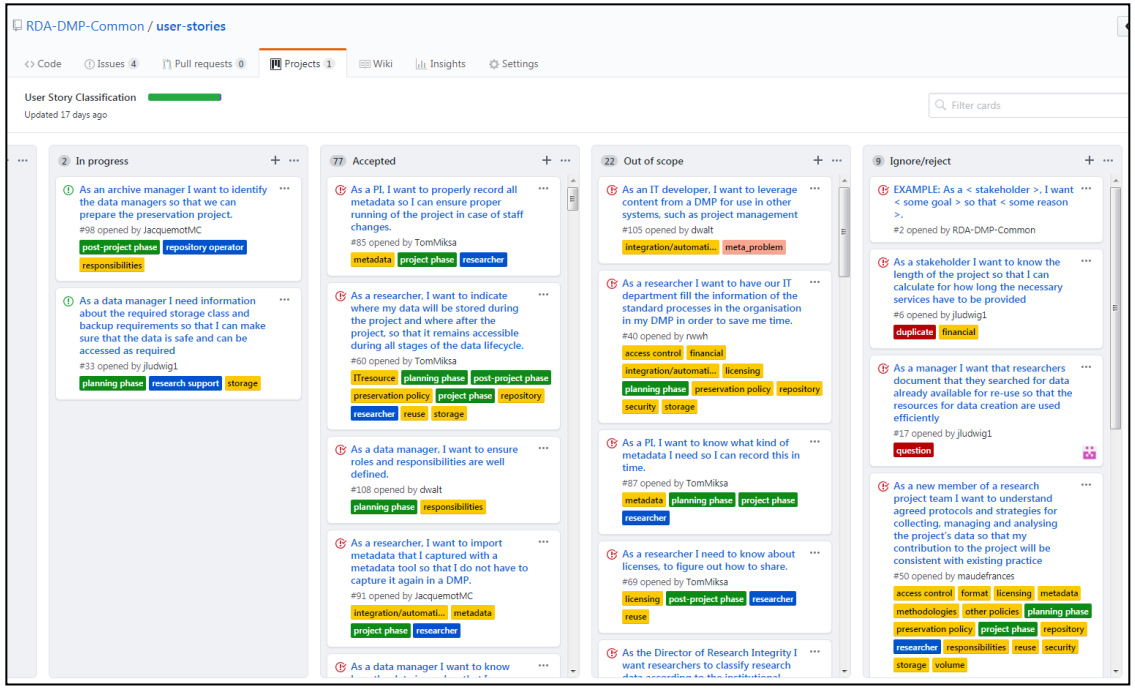




CC BY-SA 4.0

User story labelling

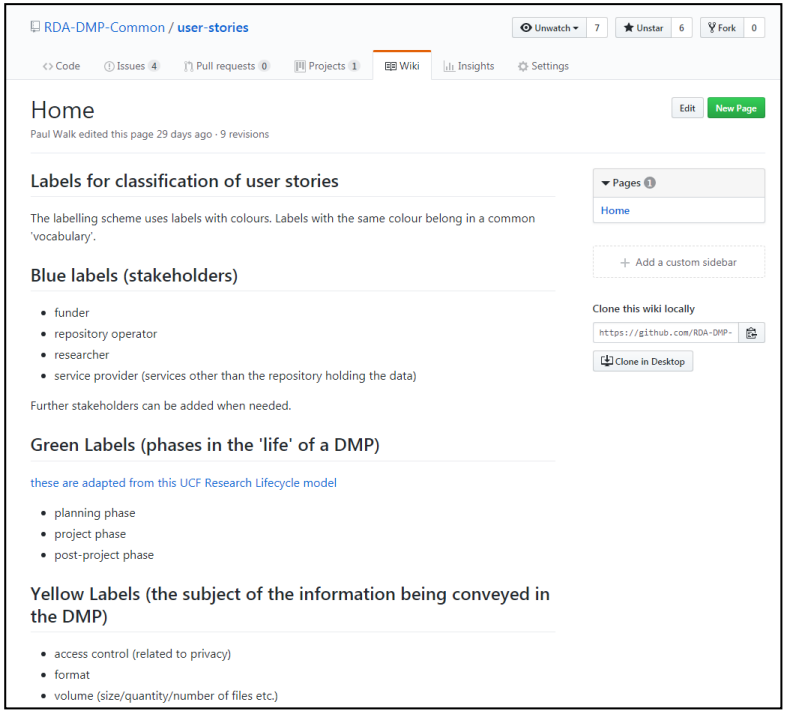
- https://github.com/RDA-DMP-Common/user-stories/projects/2
- Reviewed by chairs and authors
  - classified
    - in scope - useful for model definition
    - out of scope – often referring to the ecosystem, practices – important but not directly for the common data model
  - labelled



CC BY-SA 4.0

User story labelling

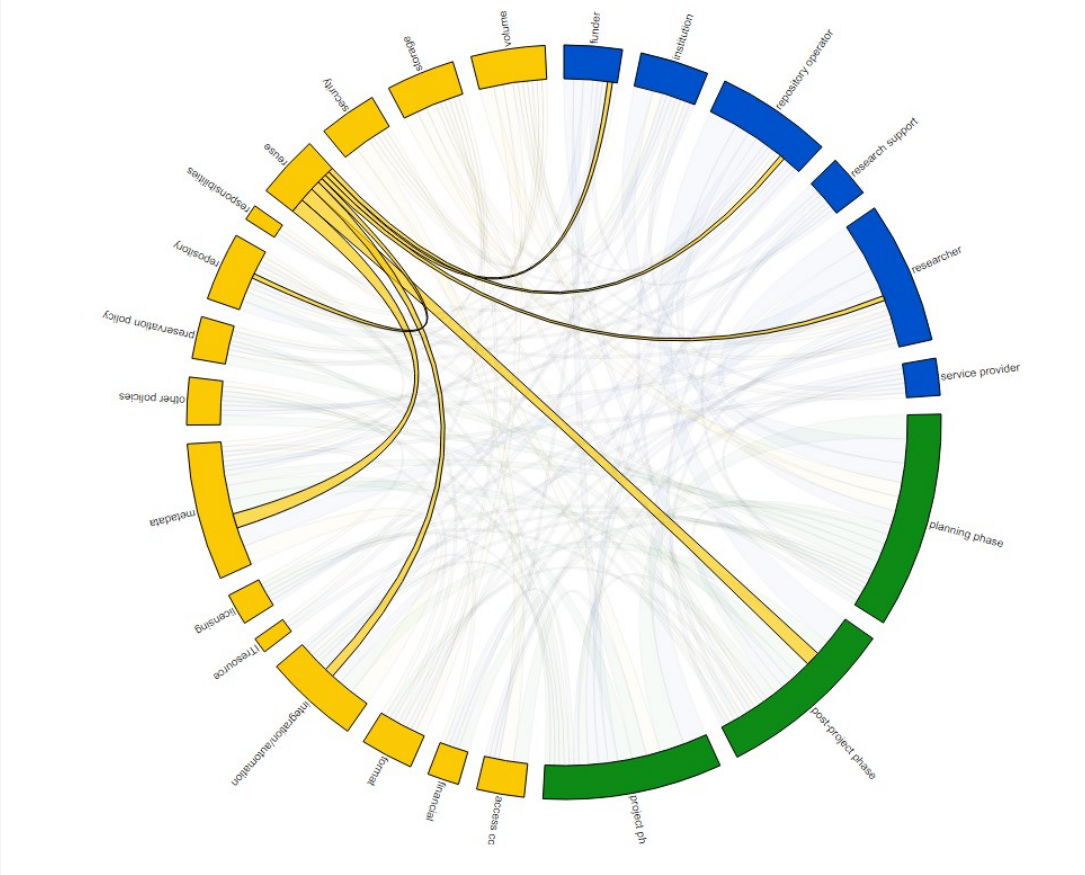
- https://github.com/RDA-DMP-Common/user-stories/wiki
- 3 major categories (colours)
  - stakeholders involved
  - project phase
  - subject of information conveyed
    - access control
    - volume
    - financial
    - licensing
    - metadata
    - repository
    - security
    - storage
    - etc.



CC BY-SA 4.0

User story visualisation

- https://bl.ocks.org/peterneish/f6dad14e46327011f0ccf15d49dd27fb
- interactive visualisation - changes on GitHub are visible immediately
- shows relations between stakeholders, phases and information



CC BY-SA 4.0

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]

CC BY-SA 4.0

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]

CC BY-SA 4.0

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]

CC BY-SA 4.0

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**]

CC BY-SA 4.0

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**

CC BY-SA 4.0

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.

CC BY-SA 4.0

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?

CC BY-SA 4.0

Consultation 2 – ‘going deep’

- 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**

CC BY-SA 4.0

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

CC BY-SA 4.0

Plenary in Berlin

CC BY-SA 4.0

Plenary in Berlin

- https://www.rd-alliance.org/dmp-common-standards-wg-rda-11th-plenary-meeting

- Part 1 – Introduction for newcomers, status update, and meeting objectives
- Part 2 – Open consultation on stakeholder requirements towards machine-actionable DMPs
  - Presentation of objectives, methodology, and results
  - Discussion
- Part 3 – Data Models of Existing DMP Tools
  - Lightning talks from tool owners describing how they organize DMP information currently (e.g. DMP Online, DMP tool, RDM Organizer, DM Wizard)
- Part 3 – Towards Common Data Model
  - Moderated discussion
- Part 4 – Wrap up and planning

CC BY-SA 4.0

Plenary in Berlin

- Template for presentations
  - no marketing talks
  - focus on data models and information modelling

- Break out groups
  - face to face meeting gives a chance to do some collaborative work
  - should we book any extra rooms (if possible)?
  - are there any topics to discuss?

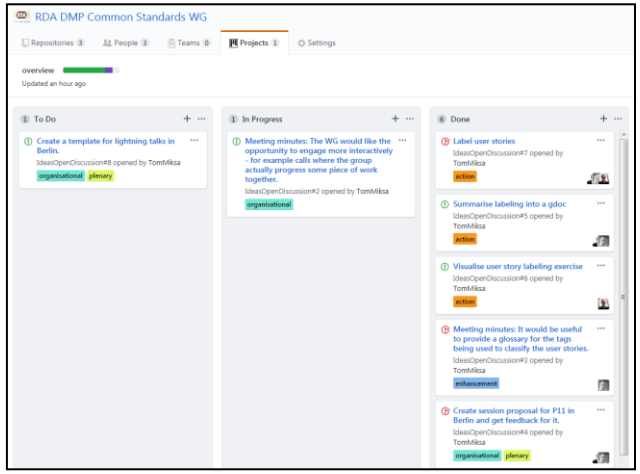
CC BY-SA 4.0

Organisational

CC BY-SA 4.0

Feedback from previous calls

- Tracking progress
  - https://github.com/orgs/RDA-DMP-Common/projects/1
- Providing new ideas and feedback not related to consultations
  - https://github.com/RDA-DMP-Common/IdeasOpenDiscussion
- Of course you can also reach any of the chairs by mail.



CC BY-SA 4.0

See you in Berlin!

CC BY-SA 4.0