



DMP Common Standards WG

Tomasz Miksa, Paul Walk, Peter Neish

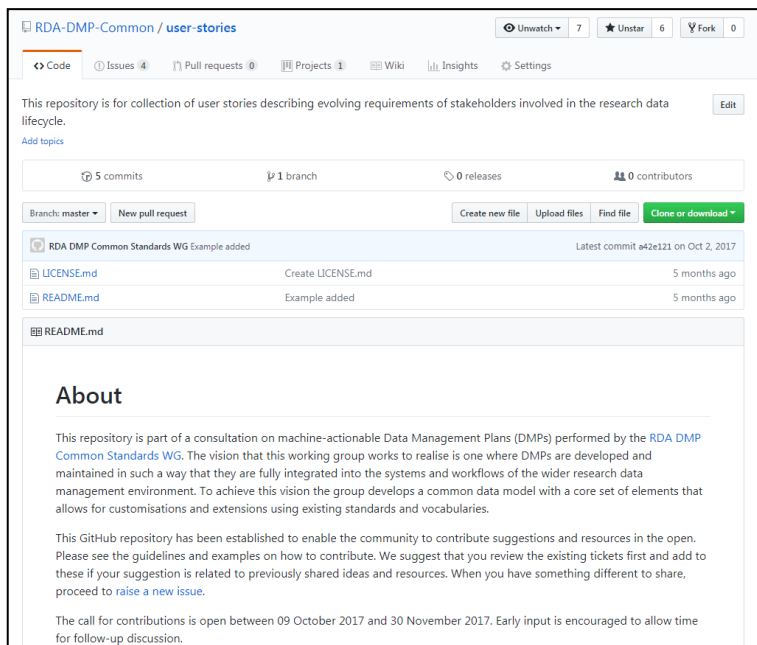
tmiksa@sba-research.org

Agenda

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

User story consultation

- <https://github.com/RDA-DMP-Common/user-stories/>
- 100+ issues defined
- inputs from Europe and Australia
- inputs from individuals and [workshops](#)



RDA-DMP-Common / user-stories

Unwatch 7 ★ Unstar 6 Fork 0

Code Issues 4 Pull requests 0 Projects 1 Wiki Insights Settings

This repository is for collection of user stories describing evolving requirements of stakeholders involved in the research data lifecycle.

Add topics

5 commits 1 branch 0 releases 0 contributors

Branch: master New pull request Create new file Upload files Find file Clone or download

RDA DMP Common Standards WG Example added Latest commit a42e121 on Oct 2, 2017

LICENSE.md Create LICENSE.md 5 months ago

README.md Example added 5 months ago

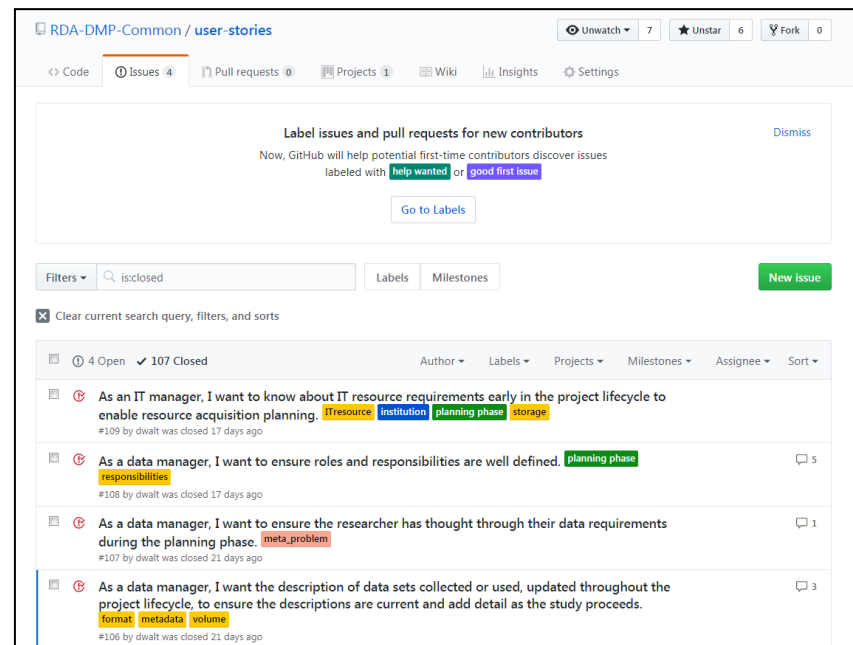
README.md

About

This repository is part of a consultation on machine-actionable Data Management Plans (DMPs) performed by the [RDA DMP Common Standards WG](#). The vision that this working group works to realise is one where DMPs are developed and maintained in such a way that they are fully integrated into the systems and workflows of the wider research data management environment. To achieve this vision the group develops a common data model with a core set of elements that allows for customisations and extensions using existing standards and vocabularies.

This GitHub repository has been established to enable the community to contribute suggestions and resources in the open. Please see the guidelines and examples on how to contribute. We suggest that you review the existing tickets first and add to these if your suggestion is related to previously shared ideas and resources. When you have something different to share, proceed to [raise a new issue](#).

The call for contributions is open between 09 October 2017 and 30 November 2017. Early input is encouraged to allow time for follow-up discussion.



RDA-DMP-Common / user-stories

Unwatch 7 ★ Unstar 6 Fork 0

Code Issues 4 Pull requests 0 Projects 1 Wiki Insights Settings

Label issues and pull requests for new contributors

Now, GitHub will help potential first-time contributors discover issues labeled with [help wanted](#) or [good first issue](#)

[Go to Labels](#)

Filters is:closed Labels Milestones [New issue](#)

Clear current search query, filters, and sorts

4 Open 107 Closed Author Labels Projects Milestones Assignee Sort

As an IT manager, I want to know about IT resource requirements early in the project lifecycle to enable resource acquisition planning. [ITresource](#) [institution](#) [planning phase](#) [storage](#) #109 by dwalt was closed 17 days ago

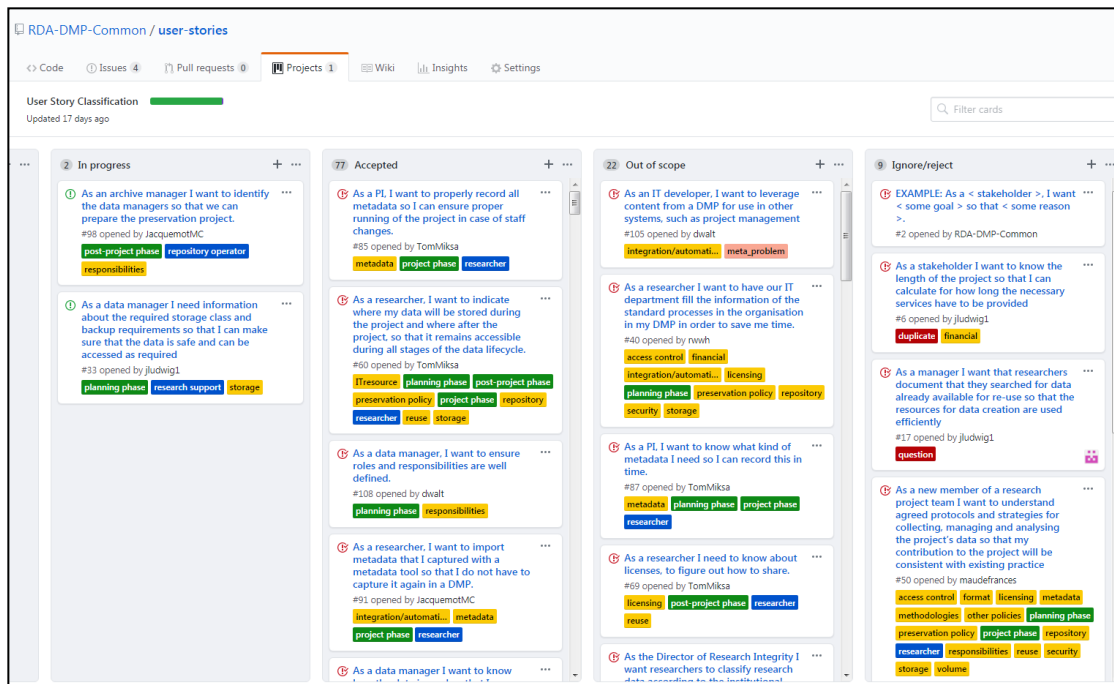
As a data manager, I want to ensure roles and responsibilities are well defined. [responsibilities](#) [planning phase](#) #108 by dwalt was closed 17 days ago 5

As a data manager, I want to ensure the researcher has thought through their data requirements during the planning phase. [meta_problem](#) #107 by dwalt was closed 21 days ago 1

As a data manager, I want the description of data sets collected or used, updated throughout the project lifecycle, to ensure the descriptions are current and add detail as the study proceeds. [format](#) [metadata](#) [volume](#) #106 by dwalt was closed 21 days ago 3

User story labelling

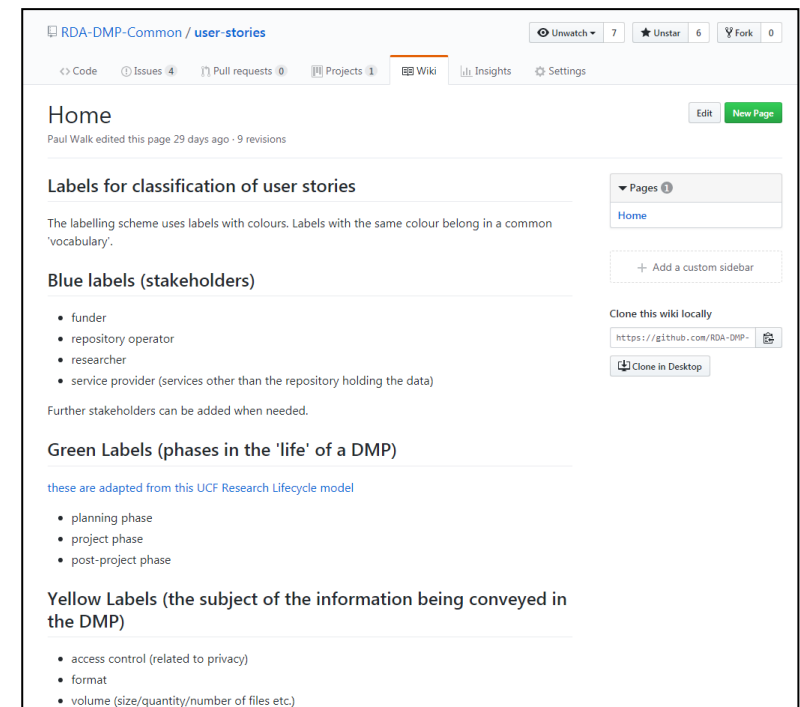
- <https://github.com/RDA-DMP-Common/user-stories/projects/2>
- Reviewed by chairs and authors
 - classified
 - in scope - useful for model definition
 - out of scope – often referring to the ecosystem, practices – important but not directly for the common data model
 - labelled



The screenshot shows a GitHub repository named 'RDA-DMP-Common / user-stories'. The main view is a Kanban board with four columns: 'In progress', 'Accepted', 'Out of scope', and 'Ignore/reject'. Each column contains several user stories, each represented by a card. The cards are labeled with their status and contain text describing the user story. Below the text, there are labels indicating the project phase and the role of the user. For example, in the 'In progress' column, the first card is labeled 'post-project phase' and 'responsibilities'. In the 'Accepted' column, the first card is labeled 'metadata', 'project phase', and 'researcher'. In the 'Out of scope' column, the first card is labeled 'integration/automation' and 'meta problem'. In the 'Ignore/reject' column, the first card is labeled 'EXAMPLE: As a < stakeholder >, I want < some goal > so that < some reason >'. The labels are color-coded: green for 'planning phase', yellow for 'post-project phase', blue for 'research support', orange for 'storage', red for 'metadata', purple for 'project phase', brown for 'researcher', light blue for 'integration/automation', pink for 'meta problem', light green for 'access control', light orange for 'financial', light purple for 'licensing', light blue for 'preservation policy', light green for 'repository', light orange for 'security', light purple for 'reuse', light blue for 'question', light green for 'format', light orange for 'metadata', light purple for 'methodologies', light blue for 'other policies', light green for 'planning phase', light orange for 'preservation policy', light purple for 'project phase', light blue for 'repository', light green for 'researcher', light orange for 'responsibilities', light purple for 'reuse', light blue for 'security', light green for 'storage', light orange for 'volume'.

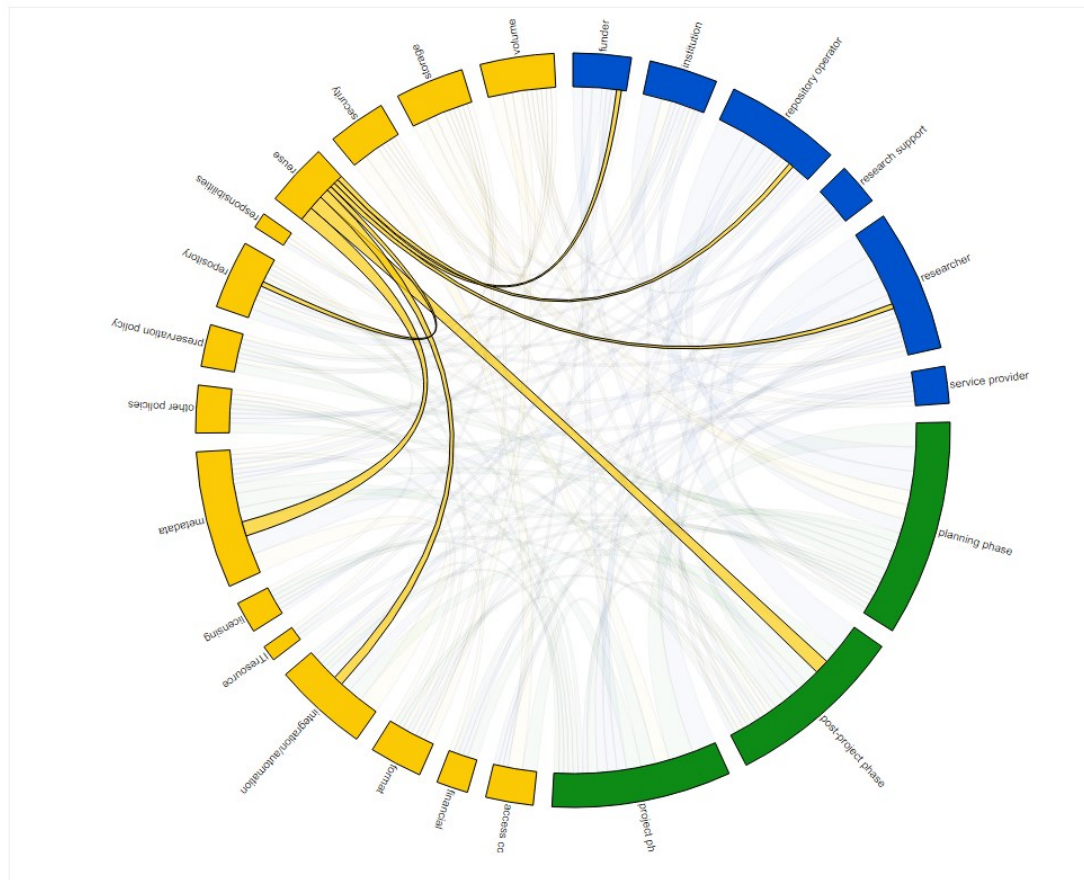
User story labelling

- <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://bl.ocks.org/peterneish/f6dad14e46327011f0ccf15d49dd27fb>
- 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

› 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**]

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

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://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](#)

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

Plenary in Berlin

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

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?

Organisational

Feedback from previous calls

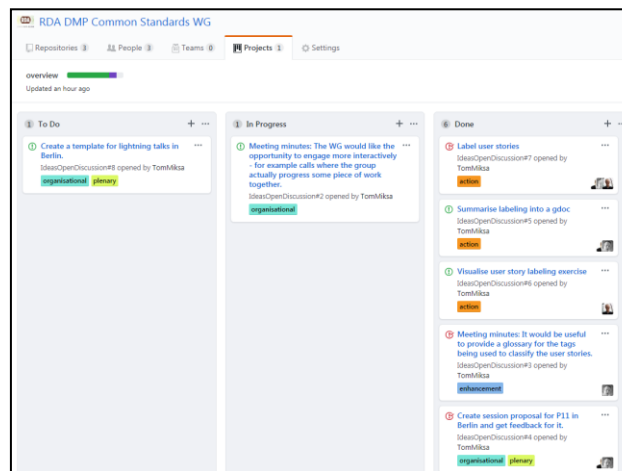
- Tracking progress

- <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.



See you in Berlin!

RDA

11 PLENARY

MEETING

21-23 MARCH 2018

Berlin, Germany