Implementing maDMPS in DMPonline

Sam Rust - Systems Developer, DCC
31 July 2019
DMPRoadmap Data Model

Targeting a minimal DMP

Required objects are a DMP, with at least one Contact and Dataset

```json
{  
  "dmp": {  
    "title": "Minimal DMP",
    "language": "en",
    "created": "2018-07-23T10:10:23.6",
    "modified": "2019-02-06T10:10:23.6",
    "ethical_issues_exist": "unknown",
    "contact": {  
      "mail": "cc@example.com",
      "name": "Charlie Chaplin",
      "contact_id": {  
        "contact_id": "http://orcid.org/0000-0000-0000-0000",
        "contact_id_type": "HTTP-ORCID"
      }
    },
    "dataset": {  
      "title": "Placeholder dataset",
      "type": "(value from dictionary)",
      "personal_data": "unknown",
      "sensitive_data": "unknown"
    }
  }
}
```

RDA-DMP-Common-Standard - Minimal DMP:

What We Already Have From DMPRoadmap

➔ DMP
  ◆ Title from Plan’s title
  ◆ Description from Plan’s description
  ◆ Creation/modification dates from Plan
  ◆ Language from plan owner’s language

➔ Contact
  ◆ Plan’s owner’s name, email, and ORCID identifier

```json
{
  "dmp": {
    "title": "Example DMP",
    "description": "An in-depth description",
    "language": "en",
    "created": "2018-07-23T10:10:23.6",
    "modified": "2019-02-06T10:10:23.6",
    "contact": {
      "mail": "cc@example.com",
      "name": "Charlie Chaplin",
      "contact_id": {
        "contact_id": "http://orcid.org/0000-0000-0000-0000",
        "contact_id_type": "HTTP-ORCID"
      }
    }
  }
}
```
Making a Few Assumptions to get a Minimal DMP

➔ Datasets
◆ DMProadmap currently does not have the concept of a Dataset
◆ Let’s assume that all answers correspond to a single dataset
  ● Title from Plan’s title
  ● Type from the most generic at
  ● Personal and sensitive data collection can safely be set as ‘unknown’

➔ DMP
◆ Ethical issues can safely be set to ‘unknown’

```json
{
  "dmp": {
    "title": "Example DMP",
    "description": "An in-depth description",
    "ethical_issues_exist": "unknown",
    "language": "en",
    "created": "2018-07-23T10:10:23.6",
    "modified": "2019-02-06T10:10:23.6",
    "contact": {
      "mail": "cc@example.com",
      "name": "Charlie Chaplin",
      "contact_id": {
        "contact_id": "http://orcid.org/0000-0000-0000-0000",
        "contact_id_type": "HTTP-ORCID"
      }
    }
  }
}
{
  "dataset": {
    "title": "Example DMP Dataset",
    "type": "dataset",
    "personal_data": "unknown",
    "sensitive_data": "unknown"
  }
}
```
Comparing to a Minimal maDMP

```json

{
    "dmp": {
        "title": "Minimal DMP",
        "language": "en",
        "created": "2018-07-23T10:23.6",
        "modified": "2019-02-06T10:23.6",
        "ethical_issues_exist": "unknown",
        "contact": {
            "mail": "cc@example.com",
            "name": "Charlie Chaplin",
            "contact_id": {
                "contact_id": "http://orcid.org/0000-0000-0000-0000",
                "contact_id_type": "HTTP-ORCID"
            }
        },
        "dataset": {
            "title": "Placeholder dataset",
            "type": "(value from dictionary)",
            "personal_data": "unknown",
            "sensitive_data": "unknown"
        }
    }
}
```

RDA-DMP-Common-Standard - Minimal DMP:

```json

```
Themes

- https://github.com/DMPRoadmap/roadmap/wiki/Themes
- Currently => A Tagging system between questions and guidance
- But, some map quite well to “string” fields on the maDMPS
  - Quickly a grey-area
- Data Description => Dataset description
- Ethics & Privacy => DMP ethicalIssuesDescription
  - ? Ethics & Privacy => Security and Privacy object ?
- Metadata & Documentation => Metadata object
- Budget => Cost object
- Roles & responsibilities => DMStaff object
- Preservation => Dataset preservationStatement
- ? Storage & Security => Security and Privacy object
Can We Do Better With The Existing Data?  --  Themes

→ Approach
   ◆ Collect answers to questions with specific themes
   ◆ Concatenate (with separators) if cardinality [0..1]
   ◆ Add separately if cardinality [0..*]

→ Remaining Questions
   ◆ Just add the answer, or the question text as well ...?
      ● Answers lose context without the questions.
   ◆ Ambiguous Themes

- "Data description",
- "Data collection",
- "Metadata & documentation",
- "Storage & security",
- "Preservation",
- "Data sharing",
- "Related policies",
- "Data format",
- "Data volume",
- "Ethics & privacy",
- "Intellectual Property Rights",
- "Data repository",
- "Roles & responsibilities",
- "Budget"
Can We Improve The Data?  --  Themes

➔ **Approach**
- Adjust the DMPRoadmap themes to more-closely align with maDMP fields and objects
- Ex: Ethics & Privacy split into two themes
- Ex: Data Quality theme added => dataset dataQualityAssurance

➔ **Result**
- Easier to correspond freetext answers in DMPRoadmap to string fields in the maDMP
Can We Improve The Data?   --  Adding Fields

Lots of little ones here, with minor backend/UI changes:

Start and end dates on the Plan would enable Project object

Adding identifiers to organisations (ROR, Funder Registry), along with a funding_status on the plan enables the Funding object
# A Common Problem: Controlled Vocabularies

The maDMP format enforces the use of many controlled vocabularies, which DMPRoadmap currently has no concept of:

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost.currency_code</td>
<td>Allowed values defined by ISO 4217.</td>
</tr>
<tr>
<td>DMStaff.contributor_type</td>
<td>Contributor Type. Allowed values as defined by DataCite. See:</td>
</tr>
<tr>
<td>Dataset.type</td>
<td>Type according to:</td>
</tr>
<tr>
<td></td>
<td><a href="http://vocabularies.coar-repositories.org/pubby/resource_type.html">http://vocabularies.coar-repositories.org/pubby/resource_type.html</a></td>
</tr>
<tr>
<td>Metadata.identifier</td>
<td>Controlled Vocabulary using an Identifier</td>
</tr>
<tr>
<td>Funder.funder_id</td>
<td>Funder ID, recommended to use CrossRef Funder Registry. See:</td>
</tr>
<tr>
<td></td>
<td><a href="https://www.crossref.org/services/funder-registry/">https://www.crossref.org/services/funder-registry/</a></td>
</tr>
</tbody>
</table>
Can We Improve The Data?  --  Extending Roadmap Model

➔ New, Structured Question Formats

◆ Cost
  - currency type, numerical amount field, and comment
  - Maps onto cost object when tagged with budget theme

◆ Boolean
  - Simple Yes/No
  - in-conjunction with ethics, personal, and sensitive themes, and a potential unanswered status maps onto the corresponding data/issuesExist fields

◆ Staff
  - Link to user’s identifiers (i.e. ORCID) & list allowed contributor_type options
  - Maps onto DMStaff

◆ Metadata
  - Restrict choice to a controlled vocabulary to allow saving an identifier
Existing Work - RDA Metadata Standards Question Type

Provides an extendable way of storing the data for new question-formats. Example of pulling in controlled vocabulary from an external API.

```ruby
def answer_hash
  default = { "standards" => {}, "text" => "" }
  begin
    h = text.nil? ? default : JSON.parse(text)
    rescue JSON::ParserError => e
    h = default
  end
  end

def update_answer_hash(standards = {}, text = "")
  h = {}
  h["standards"] = standards
  h["text"] = text
  self.text = h.to_json
end
```
A Big Remaining Question -- Datasets

A large assumption we made at the beginning was that all questions were answered about one dataset. While this is needed with our current data model, to properly support the maDMP standard, we would want to be able to express multiple datasets for a DMP.

DMPOPIDoR team has piloted functionality to support answers for multiple, defined, datasets in their fork of the DMPRoadmap project: https://github.com/OPIDoR/DMPOPIDoR/tree/dmpopidor_branding

An upshot to the way this is implemented is all of the theme/question mapping logic would remain the same, and just map onto multiple dataset objects.
A Minimal DMP in DMPonline

```json
json.prettify!

json.dmp do
  json.title @plan.title
  json.description @plan.description if @plan.description.present?
  json.language @plan_info[:language]
  json.created @plan.created_at
  json.modified @plan.updated_at
  json.ethicalIssuesExist @plan_info[:ethical_issues]

  # TODO: add in other ethical issues fields
  # TODO: Map answers to question with theme 'Ethics & privacy'
  json.ethicalIssuesDescription

json.contact do
  json.name @plan_info[:contact][:name]
  json.mail @plan_info[:contact][:email]
  json.contact_id do
    json.contact_id @plan_info[:contact][:id]
    json.contact_id_type @plan_info[:contact][:id_type]
  end
end

# TODO: make this an array with a single item
json.dataset do
  json.personal_data "unknown"
  json.sensitive_date "unknown"
  json.title @plan.title # NOTE: this is a bad mapping
  json.type "dataset" # NOTE: this is a bad mapping
  json.description @plan_info[:dataset][:desc]
end
end
```

```ruby
# frozen_string_literal: true

class Api::V0::RdaController < Api::V0::BaseController

  before_action :authenticate

  def export_dmp
    @plan = Plan.find(params[:id])
    @plan_info = {}
    @plan_info[:ethical_issues] = "unknown"
    @plan_info[:language] = @plan.owner.language || Language.default
    @plan_info[:language] = @plan_info[:language].abbreviation
    @plan_info[:contact] = {}
    @plan_info[:contact][:name] = @plan.owner.name
    @plan_info[:contact][:email] = @plan.owner.email
    @plan_info[:contact][:id_type] = "HTTP-ORCID"
    @plan_info[:contact][:id] = "http://orcid.org/0000-0000-0000-0000"
    @plan_info[:staff] = []
    @plan_info[:staff] = @plan.roles.editor.not_creator.map(&:user)
    @plan_info[:dataset] = {}
    data_desc = Theme.find_by(title: "Data description")
    @plan_info[:dataset][:desc] = @plan.answers.select { |a| a.question.themes.pluck }
  end
end
```