



Agenda for M4M Vocab Workshops

M4M workshop team: Erik Schultes, Barbara Magagna, Jeroen Belien, Nikola Vasiljevic, John Graybeal, Kristina Hettne, Mijke Jetten

Part 1: Consolidation of lists and terms describing the project assets

Feb 9: 13:00-17:00

Zoomlink: <https://zoom.us/j/94311459983>

Part 2: Building the COVID Program Vocabulary

Feb 17: Workshop 13:00-17:00

Zoomlink: <https://zoom.us/j/95165149367>

Part 1: Consolidation of lists and terms describing the project assets

Feb 9: 13:00-17:00

General links for today:

- Zoomlink: <https://zoom.us/j/94311459983>
- Registration: <https://www.go-fair.org/registration-to-zonmw-vocab-workshops/>
- Google sheet [Participant list](#)
- This document is [link](#)
- On the OSF: <https://osf.io/wj6de/>
- CEDAR Templates
 - [Admin template](#)
 - [Content template](#)
 - [Dataset template](#)
- [Selecting lists for templates - some considerations](#) (see slides 17-19)
- 6 Breakout [teams](#)

Agenda

1. 13:00-13:35: Introduction [35 min]
 - a. 5 min (Erik/Margreet): [M4M Overview](#)
 - b. 10 min (Barbara): [Vocabulary Workshops Overview](#)
 - i. 4 steps
 - c. 10 min (Jeroen): [Project content template](#)
 - d. 10 min (Barbara): [Selecting lists for templates - some considerations](#)
2. 13:35-14:15: Term selection [Plenary: lead by Jeroen (5 lists x 8 min=40 min)]
 - a. [where-when-whom](#)
 - i. The group needs to decide on "Surveillance"
 - ii. Comments:
 1. Martijn van Rooijen: With Geographic, do you also mean country names etc?
 2. Kacana Khadjavi: Maybe study region
 3. Clara: what about geographical scope of the study?
 4. Jenni: COVID-19 domains and phases?
 5. Pieter: I like's Clara's suggestion.
 6. Bruna: care /disease pathway phases vs outbreak control / pandemic phases. good to discuss further
 7. Bruna: agree COVID-19 age groups /classes
 8. Margreet: i am sorry, i missed the COVID age class reasoning. what are these classes based on? why is 65+ not included?
 9. Jessica: @Margreet if I understand correctly we aren't really concerned about the terms right now. Those are just examples and the specifics terms will be discussed later on.
 10. Clara: @Margreet if I understand correctly we aren't really concerned about the terms right now. Those are just examples and the specifics terms will be discussed later on
 11. Jessica: Discussion point for the age class terms; in terms of my research our population is anyone >16 yrs because that's the age that teens can give on their own behalf. So the only

thing I worry about with the terms is that it leads to confusion as to whether both teens and adults would be picked and then elsewhere extra text explains >16 yrs or do we create another term that accounts for age of consent >16 yrs?

12. Renate: geographical should be optional

b. [data/audio-visual](#)

i. Should we keep “audio-visual material” or combine all into “data” or “dataset”

ii. Comments:

1. Clara: I was going to say the same as Erik. You can distinguish between raw or processed data. But it is not the same as discussed here. The audio-visual or MRI are data, you could process these and it can become some derived data but the primary material is all data sets from my point of view

c. [biomaterial](#)

i.

d. [services](#)

i. repositories → qualifiers data/software, online, in house, digital repository

ii. Clara: In my opinion “repository” should be required. For cases where none is used (yet) there can be an option but it should be clear where one is used

iii. Comments:

1. Jessica: just regarding the terms that need to be discussed at a later date, I don’t think this section (and probably the standards section as well) can have pre-defined vocabulary because there are so many options. There can be suggestions, but I don’t think it will be possible to capture the majority of cases with pre-defined terms

e. [standards](#)

i. Comments:

1. Bruna: is still a bit confusing that part. you can have a semantic metadata model, and a semantic data model, don’t know why we separated those. whereas FHIR is a exchange format, which also helps to achieve semantic interoperability. Nice to keep in the notes for further discussion.

3. **14:15-14:20 BREAK (5 min)**

4. 14:20-15:20 [Term selections](#) consolidation and template completion [Plenary, in [groups](#), check also [participants list](#), lead by Rita/Erik (60 min)]

a. clean/consolidate terms (5 breakout groups à 2 persons)

b. finalize content template (Barbara with a second person)

5. 15:20-15:55: Open discussion [Plenary lead by Kristina Hettne [35 min]

a. decide on critical terms

6. **15:55-16:00 BREAK (5 min)**

7. 16:00-16:30: Recommendations for .. [Plenary, Barbara (30 min)]

a. the usage of vocabs in CEDAR

b. [building FAIR vocabularies](#)

8. 16:30-17:00: Wrap up [Plenary (30 min)]

- a. 5 min, Erik: review the day / Q&A
- b. Jeroen: explain homework and show current portal with local data behind password
 - i. finalize checking lists and terms so we can take next step February 17th

Common Notes

- **General question: Sarah:** Is there more information available on the FAIR data points? We really need more information on this so that it can be decided if we need an institutional one or a UAS one that falls under the DCC. Regardless, its realization requires quite a bit of work on our part.
 - **Erik:** this question is coming up more often. It is another process than what we are focusing on today. At a future moment we will come back with more information about that. If you need information quickly, please contact me personally. Some links:
 - <https://www.fairdatapoint.org>
 - Deployment: <https://fairdatapoint.readthedocs.io/en/latest/deployment/production-deployment.html>
 - Download <https://hub.docker.com/r/fairdata/fairdatapoint>
 - **Rita:** please also join our Q&A sessions on Monday afternoons (15:00-16:00) <https://zoom.us/j/97415399025>
- **Mijke:** With regard to a Train the Trainer: make sure to contact Erik or me and we'll get a group together so we can facilitate/help you to get all the M4M efforts implemented in local organisations
- Discussion in the background: We will need at some point to decide on a governance structure for the COVID Program vocabs and metadata. We'll come back to this later. There are also governance issues around setting up and running FAIR Data Points.
- **1.c question: Jan:** So if entries are fixed, do we use drop down list instead of free text? OK, has just been answered, thx.
 - You have the possibility to select or do type ahead and select the narrowed choices
- **1.d question: Petra:** can you also select terms from ICD, orphanet and Snomed?
 - Barbara: Yes, this is possible if you know the url of the vocabulary. It can then be included in the template.
- **1.d question: Martijn:** Why is it that we do not want to register also the size of the collection?
 - Barbara: we will come to this in the next 30 min.
- **1.d Decisions to be taken:**
 - In which CEDAR template does the "size of collection" go (Content or Dataset)?
 - DECISION TAKEN: Move to Dataset
 - "Functional Space" (Where/When/Whom)
 - DECISION TAKEN: use as Subject in Dataset
 - "Gender" is too simple to be a vocab
 - DECISION TAKEN: convert to multiple choice
 - Discussion on which to use as label:

- [sex](#) (NCIT: definition -> The assemblage of physical properties or qualities by which male is distinguished from female; the physical difference between male and female; the distinguishing peculiarity of male or female.)
 - [gender](#) (NCIT: definition-> The assemblage of properties that distinguish people on the basis of the societal roles expected for the two sexes.)
 - Data
 - DECISION TAKEN: Data Type → embed in variable of Dataset
 - DECISION TAKEN: delete from Content and move to Dataset
 - Biomaterial
 - DECISION TAKEN: Keep Substance, Specimen, Derivative in lists for Content level
 - DECISION TAKEN: Species → Organism (Variable in Dataset)
 - Services
 - DECISION TAKEN: delete “desktop applications”
 - DECISION TAKEN: move Devices to Dataset
- did we write somewhere the preferred ontologies to use? Do we have a preference? Or a "no-go" list because of licenses outside NL (e.g. SNOMED CT)
- Margreet: in the breakout about Biomaterials, we also looked into <https://swissbiobanking.ch/dataset-bacteria/> >> we plan to use this for AMR related biobanks. Would it help to look into this for the Biomaterial template? in the same link, you can also find datasets for tissue and liquid. Or would the SBP dataset be an example of a data model (together with MIABIS, etc)

Renate: don't we need to archive data using sustainable data formats? we need to decide on a list of formats.

Barbara: This is not something to decide in the vocabulary workshop but when writing the DMP.

Jeroen: It should be at the distribution level.

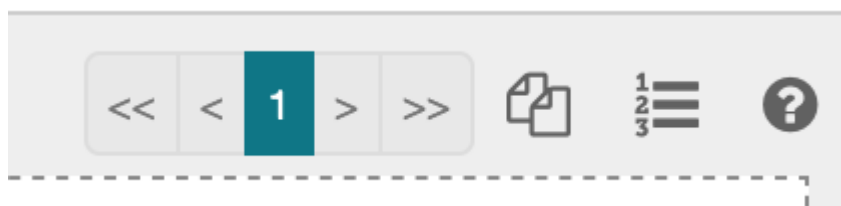
Margreet:

<https://dans.knaw.nl/en/about/services/easy/information-about-depositing-data/before-depositing/file-formats>

Petra: For file formats, we should be careful in restricting a list. Then I will have a problem with machine-specific formats that can't yet be saved in a preferred file format.

Clara: Apologies, I am a bit lost. Where do we see the vocab lists in the templates? When I hover over the ? it shows “choose the type of data” for example but I can't see any dropdown.

Jeroen: Help for Clara on CEDAR:



Clara you can click on either the green 1, the list with 1,2,3 in front. there you can either select or use type ahead. Does this help?

Jessica: Do we have a general idea of what a content metadata template should tell us or what we in general want it to contain? That's I think what makes it hard to decide what should go in content and what in dataset.

Pieter: Yes, @Jessica, I was thinking about that as well. With deleting 'qualitative' and 'quantitative', for instance, the admin template really doesn't tell so much any more about what type of research the project contains. At the same time, I understand that it only makes sense to be specific at the dataset level. My question is answered :)

Barbara: Jeroen presented this in the slides.

Jeroen: Let's keep this question in mind for the next workshop and see if it is still relevant after that one.

Bruna: did we write somewhere the preferred ontologies to use? Do we have a preference? Or a "no-go" list because of licenses outside NL (e.g. SNOMED CT)

Barbara: Hi Brunna, we are addressing semantic artefacts separately. We have prepared a long list for this.

Homework Champions !

Clara Amid to Everyone (5:11 PM)

Happy to do something as homework

Pieter Dronkers to Everyone (5:11 PM)

Perfect! No problem:)

Jessica Hruvey to Everyone (5:11 PM)

I can help - 4

Jenni to Everyone (5:11 PM)

Jenni Breel, Group 1, I will continue on our list

Petra van Overveld to Everyone (5:11 PM)

I will work on 5!

Pieter Dronkers to Everyone (5:12 PM)

I will work on 2:)

Mirjam to Everyone (5:12 PM)

I will check with Brunna, we could try to work on 6

Part 2: Building the COVID Program Vocabulary

Feb 17: Workshop 13:00-17:00

document links:

[homework](#)

6 Breakout [teams](#)

Zoomlink: <https://zoom.us/j/95165149367>

1. **13:00 - 13:45: Introduction (35 min)**
 - a. 5 min (Erik): Overview on day activities
 - b. 15 min (Jeroen): [Recap work done \(homework\)](#)
 - c. 15 min (Nikola): [FAIR vocabularies](#)
 - d. 10 min (Barbara): [Re-using semantic artefacts](#)
2. **13:45 - 14:15: Mapping part 1 (30 min)**
 - a. check mappings for different sets of terms, identify missing definitions (4 groups à 2 persons)
3. **14:15 - 14:20: BREAK (5 min)**
4. **14:20 - 15:00: Mapping part 2 (40 min)**
 - a. continue checking for different sets of terms, identify missing definitions (4 groups à 2 persons)
5. **15:00 - 15:55: Definitions (55 min)**
 - a. provide (final) definitions for undefined terms (4 groups à 2 persons)
6. **15:55 - 16:00: BREAK (5 min)**
7. **16:00 - 16:30: Quality check (30 min)**
 - a. check quality for term definitions (4 groups à 2 persons)
8. **16:30 - 16:45: Vocabulary for Dataset Template (15 min)**
 - a. Considerations around the dataset template
 - b. Considerations around FAIR variable descriptions
9. **16:45 - 17:00: Wrap up and long-term perspective (15 min)**
 - a. Metadata and vocabulary maintenance
 - b. Sustainability aspects

Notes:

Link for vocabularies, standards and repositories: 'RDA covid-19 recommendations and guidelines' document

https://zenodo.org/record/3932953#.Xx_f7Z4za70

Questions:

- Martijn: what is the maximum number of altLabels you can assign in SKOS?
- Answer Nikola: no limits

Icons created, used and presented by Nikola can be found at zenodo: the icon is named MetaManMachine:-) <https://zenodo.org/record/4471098#.YC0VLhNKIL4>

- the same sort of slides but curated for wind energy domain:
<https://zenodo.org/record/4533760#.YC0WExNKiL4>