



Stakeholder-Dialog on the topic of
“Swiss health metadata repository: prototype evaluation”
16.11-07.12.2020

Eliane Maalouf, Alessio De Santo, Paul Cotofrei

Keywords

Metadata repository, health researchers, prototype evaluation, interview protocol, Dublin Core

Authors

Eliane Maalouf, MSc, PhD assistant – Information management institute, University of Neuchâtel
Alessio De Santo, MSc, PhD assistant – Information management institute, University of Neuchâtel
Paul Cotofrei, PhD, Senior Lecturer – Information management institute, University of Neuchâtel

Address for correspondence

Information management institute

University of Neuchâtel

A.L. Breguet 2

2000 Neuchâtel

Switzerland

E-Mail: eliane.maalouf@unine.ch, alessio.desanto@unine.ch, paul.cotofrei@unine.ch

Suggested citation

The text of this stakeholder dialogue summary may be freely quoted and printed, provided proper acknowledgement is given.

Maalouf, E., De Santo, A. Cotofrei, P. (2021). Summary of the Stakeholder Dialogue on: *Swiss health metadata repository: prototype evaluation*. Swiss Learning Health System.

Summary

Following the first stakeholders dialog performed on 19.02.2020 and as discussed in our second policy brief document “Swiss health metadata repository: implementation recommendations; Requirements, technical recommendations and prototype implementation”, we established that the health researchers’ group is the users’ group most in need of a metadata repository for health research data.

Given this specific focus, we performed our dialog in the form of semi-structured individual interviews with seven researchers (two professors and five PhD students) recruited from the SLHS consortium. The interviews were conducted in November and December 2020. Each interview lasted for one hour and was held online. It was decomposed in two parts: a first part, to better learn about the interviewees current use of data and metadata resources and a second part, to evaluate an initial prototype that showcases the functionalities of the envisioned metadata repository. The first part took the form of a discussion held through Webex. During the second part, the interviewees interacted with the functional prototype through a web browser. Their interactions were observed by the interviewers through screen sharing and “think aloud” protocol. Two interviewers were present at each interview, one responsible for going through the predefined topics and probing the interviewees, and the second one responsible for note taking. All interviews were recorded through Webex, following signed consent by the interviewees. After each interview, the interviewers aggregated their notes and made updates to the protocol where needed. The complete interview protocol is detailed in Appendix I.

Key observations

Most researchers learn about the existence of relevant datasets for their research mainly from other researchers, by reading their published work or through personal network. They access those datasets also through those researchers, their advisors/managers or through the regular channels of requesting access (e.g. forms, contracts, etc.). Researchers with longer experience find it easier to identify the most relevant sources for their work.

When online databases are consulted (e.g., systematic and meta reviews), the search is often very well structured with a specific aim. The search strategy and the keywords are pre-defined as well as the coding method for the expected results. Researchers often rely on codebooks to understand the datasets that they are using.

The consulted researchers rarely needed to collect new data for their work and they relied mostly on existing sources. The normal expectation is that, when new data is collected, the owner (usually) referred to standardized vocabularies and nomenclatures to define and identify the variables. Nevertheless, in practice, this rigorous approach does not seem to be common. The difficulties in applying this rigorous work stems from the difficulty to identify the right level of granularity in those standards relative to the data at hand, the exclusive vs exhaustive character of a given standard and the difficulty to link and convert between different classifications. Classifying and annotating with the right terminology requires time to spend for understanding the structure of the standard classification/vocabulary of interest.

Researchers expect that a metadata repository would help them to:

- discover new data resources / sources that they were unaware of;
- discover links to other potentially relevant datasets from known datasets;
- trace the history of a dataset and its updates through time;
- learn about the survey/study that led to the collection of the dataset and its content before actually requesting access to it;
- learn about the procedure or legal requirements to access the data.

Furthermore, the interviewees present a higher interest with the metadata repository if they know what they can expect to get from it and if their research domain is adequately covered.

Users appreciate a visual way to navigate through the data sources. However, they are afraid that it can become overwhelming at times and potentially irrelevant if the useful information is not presented first and if interaction with the visualization tool is limited.

In discovering the resources in a metadata repository, users expect to collect as much information as possible about the related dataset as well as the study that led to the collection of the dataset, before having to actually access the dataset. Furthermore, they prefer that redundant information (e.g., repeated collections of the same dataset) be hidden by default and shown on request.

Given the diversity of researchers' background, their approach for the search process is not uniform. Nevertheless, they all expect a flexible keyword search functionality that adapts to their style (i.e. ranging

from free text with autocorrect, to specific domain search terms). Some expected to browse through a list of available data categories, or a similar structure, to learn what they can expect, before actually looking for specific items.

Some users appreciate more sharing between researchers, mainly, information about the quality of datasets that were used by other researchers and the eventual program codes used to treat and clean a common/public dataset.

Recommendations

Given the previous observations and the interactions that the interviewees had with the prototype, we recommend the following implementations:

- Enrich the Dublin Core set of metadata fields with additional information about the content of the dataset being described, mainly: the items (i.e., the variables), the type of study (e.g., cross-sectional, longitudinal, etc.), the number of observations and the collection frequency (i.e., one time collection or periodic), a link to the scientific publication.
- Improve the onboarding experience of the users by clarifying the purpose of the repository and how to go about searching for information. It was suggested to make an introductory video and tutorials.
- Provide contextual help in the interface to clarify how metadata fields need to be filled (e.g., with examples). Use a neutral, consistent, terminology or provide clear explanation of the used terms (e.g., users' guide).
- Improve the search functionality by: allowing to search for datasets by categories, add advanced search functionalities to combine filters, allow search by keys (e.g., OR, AND, "", etc.), autocorrect misspelled search keywords and allow search to look for words with the same root as well as synonyms.
- Allow users to maintain a thread throughout the navigation they perform in the repository, from the initial search, in order to trace back their path and the links they went through.
- Attach the codebook, when available, to the metadata resource.
- Distinguish descriptive information, mainly the descriptive tags on a resource, into two levels: a macro level, a broad descriptive level; and a micro level, a dataset specific level.
- Reduce the quantity of information in the graph visualization by allowing users to click on the graph to show or hide the information, keep the default graph size to a medium number of connections (e.g., 2-3 degrees of connections), show a dataset only once in the graph and related versions as extra information on another layer. When available, show all datasets that are part of the same collection study (e.g., all datasets related to a cohort study).
- Allow users to select how they want to sort the outcome of a search (e.g., by relevance, from newest to oldest, etc.). Order by relevance by default.
- Hide irrelevant information from the user, typically information related to the metadata repository management (e.g., dataset identifiers, search relevance scores).

Appendix I

Prototype evaluation interview protocol

In the following we present the detailed structure and content of the interview protocol. The interview starts with a brief explanation of the context of the metadata repository in SLHS, introduction of the interviewers, and clarifying the interview objectives as:

- 1) assess how the interviewee currently searches for data resources and how he/she interacts with metadata in general;
- 2) evaluate a preliminary prototype of a metadata repository for SLHS.

The process of the interview is then explained, according to the following parts:

- 1) general information about the interviewee
- 2) clarifying the current situation
- 3) evaluating the prototype and
- 4) concluding remarks.

The interview is reminded of his/her consent for audio-video recording and recording starts.

Part 1 - General questions about the interviewee

- What is your current position and in which organization?
- How do you evaluate the level of your acquaintance with information tools? Do you often require help or do you manage most tasks related to interacting with information systems (e.g., your computer, accessing and searching in web databases, performing analysis using informatics tools, etc.)?

Part 2 - Current situation

The goal of this series of questions is to identify how potential users are currently performing some tasks that might be integrated in the repository. We aim to identify pain areas (but also satisfying areas, if any) and missing functionalities (either in the current user situation or in our initial design).

Searching for datasets/documents/files

- Briefly inform us about the types of datasets/documents/files that you often use in your activity and for what purposes. (Ideally gives 2-3 examples)
- How do you currently procure these datasets/documents/files? Briefly describe the process of **identifying** the necessary resource, **searching** for the resource, **requesting** (if necessary), **accessing** it, **storing** it, **disposing** of it after the end of activity. (Ideally 1-2 examples of a recent event expliciting such a process)

- How much time does this process take you on average? (Especially when a new resource is needed for the activity).
- Tell us about your perceived challenges during this process. Have you faced difficulties at any of the stages of the process of acquiring and using the dataset/document/file (not related to the analysis phase)? If yes, where and why do you perceive them as difficulties?
- What in your opinion needs to be improved in order for you to be satisfied with this process? (Clarifies aspirations) How in your opinion these improvements could be made? (Clarifies missing functionalities)
- What are you currently satisfied about in this process?

Creating/Editing metadata

- In your activity, do you need to store and/or provide descriptive information (metadata) about resources (datasets/documents/files) that you produce or procure from external sources?
- If yes, how do you currently perform this task? Briefly describe the **context** (environment where this insertion is done, e.g. hospital, office, on the field), the **tools** used (currently used systems) and the **tasks** involved (steps performed, ideally the interviewee shows the steps if performed on a computer system).
- How do you procure or generate the metadata to introduce?
- Is there any standard that you follow to collect this metadata? (If yes, what are their names and acronyms) (Are these standards directly integrated in the systems/tools that you currently use?)
- Do you need to annotate the metadata following a given standardized nomenclature / ontology / controlled vocabulary? (if yes, what are their names and acronyms). Are these standards directly integrated in the systems / tools? (if yes, how? (e.g., autocomplete functionality, automatic suggestions, etc.))
- How often do you need to Edit existing metadata? How is it currently done? (Identify if the same tools and systems are used, in the same context, and involve the same tasks as a new metadata introduction).
- Do you need to create and edit metadata on behalf of someone else (or some entity), or only for your own needs? If on behalf of someone else, what is your relation to that person/entity? The goal of this question is to clarify what could motivate the user to create/edit metadata.
- Tell us about your perceived challenges during this process. Have you faced difficulties in the process of acquiring, introducing and editing metadata? If yes, where and why do you perceive them as difficulties?
- What in your opinion needs to be improved in order for you to be satisfied with this process? (Clarifies aspirations) How in your opinion these improvements could be made? (Clarifies missing functionalities)

- What are you currently satisfied about in this process?

Part 3 - Prototype evaluation

The goal of this series of questions (mainly observations) is to evaluate how potential users would perceive the ease of use of the prototype and the completeness of its functionalities. To avoid biasing the user, the interviewer needs not to explain how to perform the tasks to the interviewee but instead leave her to discover and observe where and how the user interacts with the interface. The interviewer will need to collect as much information as possible and probe the interviewee interactively (while avoiding bias) to explain what they are doing and why. Ideally, record by audio-video the interactions on the screen.

Note: *if the user gets stuck on a task in the interface, before helping her out to advance in the interview, try to collect information on what made her stuck. Is it an element missing from the interface? Or not at an unintuitive place? Unintuitive or incomprehensible name? Missing explanation? Let the user speak out loud on why she feels stuck or unable to achieve the task requested. Then provide a hint and see how the user continues from that point on, again making sure to collect information on further pain points for the user. Keep in mind that the goal of the interview is to understand the intentions of the user and if the interface helps them achieve those intentions, so the interviewer needs to intervene as little as possible in guiding the user around the interface.*

This part requires (1) user access to the prototype (2) user understanding of a metadata repository.

Evaluating functionalities & content

For each of the following tasks proposed to the user, collect observations about the following: where the user looks for the functionality, how they interact with the functionality and when and if they are stuck. Ask the user to comment out loud on the task they are doing, explaining what is their intention and how and if this intention is being met by the system.

Searching for dataset/document/file

- Request from the user to use the application **to search for a dataset** they might need.
- How is the user interacting with the **graph visualization**? Do they perceive it as useful/interactive enough, or would they rather prefer another way of presenting a collection of resources? If they prefer another mode, let them describe it briefly?
- Request from the user to **look up more detailed information** about a dataset. Are their metadata information the user is looking for but does not find? Let them briefly speak about what is missing in the currently available metadata and why they might need this information?
- Let the user bring a recent search experience to mind, ask the user to try to perform it in the system. Let them describe out loud what they are looking for and note how they are doing it on the interface. Are they trying to fulfill all the steps of the search process, from identification to acquisition, or just part of it? Let them briefly describe their intentions and steps.

- Is the user looking for certain types of **filters**? Are they introducing specific **keywords**? Given the current filters, does the user perceive them as useful/interactive enough? If not, let the user briefly explain what they would rather have.
- Stating that the prototype is not final, ask the user what they perceive missing in the dataset collection, in order to identify missing metadata that are perceived useful by users.
- Let the user speak about how they generally perceived performing the tasks of searching and browsing? What do they perceive as still missing? What do they perceive as unintuitive or unnecessary? What do they appreciate?

Creating/Editing metadata

- Ask the user to perform a **metadata creation** for a resource of their choice. Let them speak out loud about the process they are following. Are they trying to input a metadata information for which there is no previewed field? Let them describe what they perceive as missing.
- What is the user expecting if there is a missing metadata field they need? Do they expect to be able to add it or request for it? Let them describe how they would proceed.
- Ask the user to perform an **Edit metadata** task on a resource. Let them speak out loud about the process they are following.
- Let the user bring a recent create/edit metadata experience to mind, let the user compare that experience to their current experience with the interface. Let the user describe what they perceive missing in the prototype interface and why.
- If the user usually works with standards and standardized vocabularies, let them describe how they would rather see these vocabularies integrated in the current interface?
- Ask the user to perform the task of visualizing all the metadata resources they have created. Where do they expect to find this information?
- Let the user speak about how they generally perceived performing the tasks of creating and editing metadata information? What do they perceive as still missing? What do they perceive as unintuitive or unnecessary? What do they appreciate?

Part 4 - Conclusion, next steps

- In case a metadata repository is available, would you be inclined to use it? If yes, why? If not, why not?
- Would you like to add something about the concept of the metadata repository or about the prototype you just tested that was not covered during the interview? Are there aspects that you would recommend to take into account in the repository development process?

Finish the interview by informing the interviewee how and where the results of the analysis will be available. Inform the interviewee on how the video and audio recordings will be managed by the end of the analysis.