Case description from the project "FAIRify Humanities" of the National Data management Forum 2019.

Title: Data generated by teachers tagging lesson plans with competencies in a learning platform

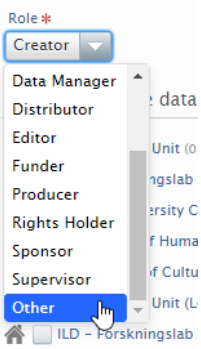
Name and affiliation: *Benjamin Brink Allsopp, AAU, Anders Tamborg (former AAU, now KU), Karsten Kryger Hansen, AAU*

Research area: Competencies and learning platforms: the use of competencies to characterize and measure learning in learning platforms

Data set(s): The dataset focused on in this short description and during the duration of the project is quantative data from an external party, and a small focus on self-generated qualitative data from interviews.

What efforts was carried out to make your data more FAIR?

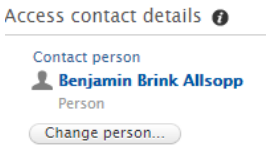
***In general*:** The group has looked into specific datasets that originates from the public school sector in Denmark, describing the fulfilling of learning objectives in specific courses. Secondary, the group has looked into the FAIR-principles, with an intention of understanding the FAIR principles, and put these into specific practice. The group have been in contact with the data supplier, in order to facilitate a talk on how these data could become more FAIR. Due to time shortcomings there was not an ability to look further into these issues, and a high level of uncertainty of the ability to publish this kind of data, especially due to IPR and GDPR related issues. The other data type that is often part of these kind of studies are qualitative data based on interviews.



***More "Findable"*:** As part of the FAIR principles, data should become searchable in an indexed resource. As the data cannot be made public available, the findability (and possible option for assigning a DOI), will also determine on the granularity of the research data. E.g. metadata could describe a specific dataset from a given year, or there could be a record about the holdings of these data in general. The group has looked into the options of making research datasets findable through the Pure portal at AAU; VBN. An important factor to consider here, is that the dataset is made available to the research group, is not produced by the research group. The person relation can therefore be misleading in issuing metadata making it a quotable data set. The most obvious possibility is to use the role "Other". So in making other organizations data "viewable" by a research group, where the dataset does not have an official metadata record, is challenged by the metadata formats – at least in Pure. The republication issue is much aligned with what is described here; https://www.nature.com/news/legal-confusion-threatens-to-slow-data-science-1.20359

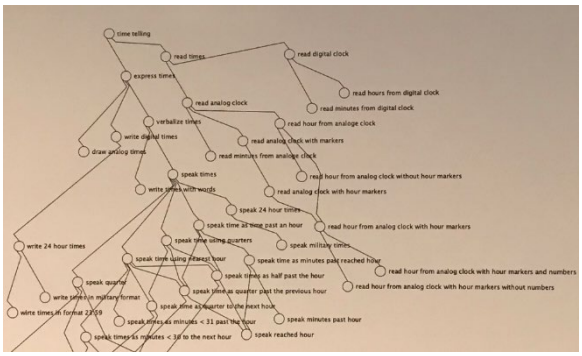
***More "Accessible"*:**

As part of the investigation the group had conversations with the rights holder for the quantitative dataset. As the access to the data set is given by request and agreement with the rights holder, the ability to make research data like these are quite limited. There is a lot of discussion going on in the community, on how much of these data should be accessible to a larger audience. As for both the quantitative and qualitative data, there are ethical and legal considerations that does not allow for open access. However, metadata records (as investigated in Pure), holds good options for having contact person details on data set. It is important that administrators of these



systems pay close attention to keeping access contact details updated as part of the off-boarding procedures at the university, in order not to have “orphan” data, and pointers to people who are no longer able to respond.

More “Interoperable”:



For the quantitative data, there has been initiatives, as part of previous work by the research group to make common terminologies for learning objectives. This would make a FAIR - as the taxonomy should be shared – resource that could be used in (meta)data, making it easier to map learning objectives in a Danish context, but perhaps also in an international context. The original data set include text based string indicating learning objectives. Examples are; “*Du kan løse problemer, hvor du skal bruge flere regningsarter.*” and ”*Du kan opstille udtryk med variable, der beskriver omkreds og areal af enkle polygoner.*”.
However, there is a need for such taxonomies to have support both by schools, teachers, ministries etc, to have the interoperability over time. And of cause support in the systems collecting the learning objectives. It is a time consuming process to build and incorporate these kind of frameworks and put them into data, that is currently captured in learning platforms. More about this approach can be found here; https://vbn.aau.dk/da/publications/ability-maps-in-the-context-of-curriculum-research. Given the amount of learning objectives, these taxonomies also become huge and complex.

More “Re-usable”:
There is a limited focus on making data in this context reusable, and it has not been possible during this project to find a proper way to e.g. license these kind of data for public use. This come down to ethical considerations (including GDPR), but also Intellectual Property Rights (IPR). The scope of the research area is also currently also in a lot of considerations on the protection of pupils, also when it comes to minimizing a potential “misuse”. As part of the project we initiated a talk with the data provider, but due to time shortcoming we could not reach a consensus on how even a subset of the data might be licensed for further reuse.
For the interview data and the ethnographic observation notes, these are often embedded as part of the article, and mostly being supportive for the researcher during the writing process. It would take a lot of detailed descriptions that is hard to make machine readable, to provide sufficient provenance around these kind of data, and being sure the context is well understood for an “outsider” finding the data.

What was the biggest challenges to make your data more FAIR?
IPR and GDPR. And lacking standardization of terminology.