

Anna Salmi & Mari Elisa Kuusniemi

Project MILDRED: Charting Ground for Research Data Management Services at University of Helsinki

Anna Salmi, University of Helsinki, annamerisalmi@gmail.com

Mari Elisa Kuusniemi, University of Helsinki, mari.elisa.kuusniemi@helsinki.fi

This paper describes a topical case study conducted at the University of Helsinki. Current states of research data management (RDM) practices within the academic community have been under close scrutiny during summer 2016 in Project MILDRED, Development Project of Research Data Infrastructure at University of Helsinki (UH). As relatively little is still known about the broad picture of the researchers' current research data depositing and preserving practices, the project undertook a three-stage charting of the situation in the UH context. An inventory of 250 peer-reviewed, UH-authored scientific journal articles published between 2015 and 2016 was conducted, revealing a selection of data repositories representing different domains that house UH data, plus a variety of RDM statements by the authors. (Salmi 2016.)

The inventory laid the basis for a research data repository e-survey sent to UH researchers in July 2016. The survey gathered 258 answers, providing a corpus of information about 1) what existing repositories are in use; 2) what domains the repositories cover; 3) what kinds of data type the repositories support; 4) reasons for not

having deposited data; and 5) what kind of alternative storage and preservation services and devices are being utilized. The survey also recorded wishes and criticism concerning topical issues around RDM at University of Helsinki. 62 % of the answers represented life sciences, 21 % humanities and social sciences, and 17 % natural sciences. (Salmi, Ojanen & Kuusniemi 2016.)

According to the survey results, the respondents' lack of specific knowledge about data depositing possibilities is the main reason for not making use of repositories (28 % of the respondents stated this). Data sensitivity, irrelevance to the research field, small amounts of data generated, and general lack of need to deposit were the next most common explanations. 11 % of the respondents explicitly named sensitivity issues, another 11 % general irrelevance. Need for guidance was also called for. On the other hand, personal hard drives, UH network hard drives, external hard drives, and USB memory sticks were chosen most often for storage. (Salmi, Ojanen & Kuusniemi 2016.)

As a result of the inventory and the survey together, a listing of 48 repositories included in

the Re3data.org data repository register was created. As the registry databases provide API features to promote data system interoperability, information about e.g. data types, data access type, data licenses, software, citation guidelines, quality management, and metadata standards for UH data could be harvested. Repository specific metadata and access to it were here the focus of interest. (Salmi & Pitkänen 2016.)

This final stage of the research revealed that most of the repositories housing UH data are mainly data type specific, with only 19 % of the sample featuring organization as a specific metadata field, and none currently including ORCID metadata. All repository search engines were browsed by a test query: researcher's name and/or by "Helsinki". Name of researcher turned out to be a valid search term in 21 of the cases, organization name only in nine. 36 of 48 repositories yielded no information when searched by organization, and in a couple of cases the engine was down. Repositories where organization could be identified included e.g. FSD; Gene Expression Omnibus; Inspire-HEP; NCBI Database of Genotypes and Phenotypes; The Finnish Language Bank; and Zenodo. (Salmi & Pitkänen 2016.)

To sum up, there now exists a preliminary map of repositories storing and/or preserving UH research data as well as new knowledge about individual researchers' depositing needs, preferences, and concerns. A large number (44 %) of the e-survey respondents used international data repositories or databases; as many as 21 % had data in two or more repositories and 10 % in three or more. On the other hand, a striking number stated that they had not enough knowledge about best depositing practices. Information about the variety of reasons for lack of need to deposit is also interesting. (Salmi, Ojanen & Kuusniemi 2016.) The results of the metadata browsing in particular encourage certain realism in ongoing institutional RDM service development, since institutional data/metadata can be harvested from only a fraction of the identified repositories (Salmi & Pitkänen 2016). All in all, growing knowledge about RDM practices and preferences helps

orientate towards new possibilities of promoting the principles of producing and curating findable, accessible, interoperable, and re-usable (FAIR) research data in an institutional setting (European Commission, Directorate-General for Research & Innovation 2016, 3–4, 6.; van den Eynden et al. 2011, 31–33).

Sources

European Commission, Directorate-General for Research & Innovation (2016). H2020 Programme. Guidelines on FAIR Data Management in Horizon 2020. Version 3.0. 26 July, 2016, s. 3–4, 6. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.

Salmi, Anna (2016). Project MILDRED data inventory [unpublished dataset]. University of Helsinki.

Salmi, Anna, Ojanen, Mikko & Kuusniemi, Mari Elisa (2016). Project MILDRED Research Data Repository Survey, University of Helsinki. University of Helsinki. Figshare. DOI: https://figshare.com/articles/Project_MILDRED_Research_Data_Survey/3806394

Salmi, Anna & Pitkänen, Timo (2016). Project MILDRED metadata inventory [unpublished dataset]. University of Helsinki.

van den Eynden, Veerle, Corti, Louise, Woollard, Matthew, Bishop, Libby & Horton, Laurence (2011/2009). *Managing and Sharing Data. Best Practice for Researchers*, s. 31–33. Essex: UK Data Archive, University of Essex.