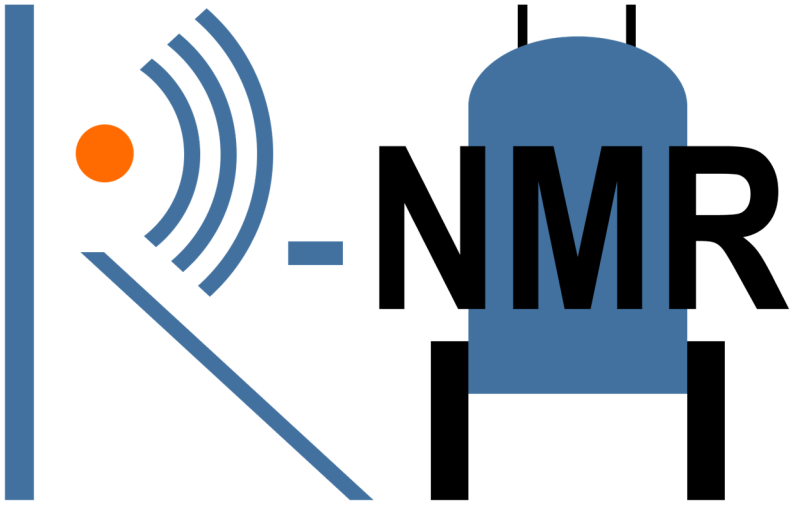

Deliverable 1.2

Data Management Plan

of Remote NMR (R-NMR):

Moving NMR infrastructures to remote access capabilities

Authors: Julia Wirmer-Bartoschek (BMRZ), Francesca Morelli (CIRMMP), Harald Schwalbe (BMRZ), Antonio Rosato (CIRMMP)

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement N. 101058595

TECHNICAL REFERENCES

Project acronym:	R-NMR
Project Title:	Remote NMR: Moving NMR infrastructures to remote access capabilities
Grant Agreement number:	10105859
Project coordinator:	Prof. Dr. Harald Schwalbe
Organization:	J.W. Goethe Universität, Frankfurt
E-mail:	Schwalbe@nmr.uni-frankfurt.de
Project website address:	http://www.r-nmr.eu/

Deliverable No.:	D1.2
Lead Beneficiary:	BMRZ
Type and dissemination level:	Report - Public
Due Date:	M5 (30 November, 2022)
Delivery Date:	28 November, 2022

History of Changes		
Version	Publication date	Changes
1.0	28.11.2022	First DMP of R-NMR project

Contents

1. Executive Summary

Nuclear magnetic resonance spectroscopy (NMR) is one of the major analytical methods applied in all chemical, physical, biological, and medical sciences. NMR's leading role stems from its analytical power in terms of molecular resolution, quantification, reproducibility, and broad application envelope. It requires sophisticated and expensive equipment, operated by scientists with diverse background ranging from service-oriented researchers to highly trained experts. The NMR community maintains excellent networking between sites and serves a broad community within the focal points of European research interests. Within the Remote-NMR (R-NMR) project, we aim to establish remote access for all NMR users throughout Europe and establish an inclusive network of NMR-infrastructures throughout Europe. We survey if and how remote access can be made possible according to the needs of the community, and implementing GDPR at facilities and sample shipment procedures. Routines for remote NMR-usage will be established, including dissemination of research and teaching protocols, archiving of data, and sample shipment.

The data management plan (DMP), aims to describe the handling of data within the project in line with the FAIR princip. During the project the following types of data will be collected: (i) survey data, (ii) raw and processed NMR data generated during the test phase. Further, common protocols and documents documenting them will also be generated as a result of the project.

2. Data Summary

The R-NMR project aims to develop standardized procedures for the remote access to NMR-machines to conduct NMR experiments. The NMR experiments required for the determination of 3D structures of biological macromolecules are typically long experiments, lasting for several days. Similarly, solid-state NMR experiments on microcrystalline samples or on non-biological materials are performed over multiple days. These experiments require an initial setup and parameter optimization, but then run in a computer-controlled manner. Experimental progress can be monitored during data collection.

To implement remote access fully, surveys of facility managers and NMR-machine users will be performed and as a result standardized procedures for the various aspects in the pipeline of NMR access will be developed and tested, ranging from sample shipment to execution of

experiments and distribution of the final data to the users. We therefore have to deal with the following types of data: (i) survey data, (ii) raw and processed NMR data generated during the test phase. Further, common protocols and documents documenting them will also be generated as a result of the project.

Survey data

Surveys are being conducted in order to collect information from NMR facility managers and facility users about, respectively, their NMR centers and their experiences with remote access. This will be the basis so that a common protocol for remote access for all facilities can be developed and adopted by facilities across Europe. Within R-NMR, WP2 is in charge of performing the surveys and analyzing the corresponding outcomes.
Surveys will be performed with the Jisc Online Survey software, and data will be downloaded from the online survey provider and transferred electronically to storage in password-protected Word/Excel files.
The results of the survey, which will be useful for the whole NMR community throughout Europe, will be made available to all partners in aggregated manner (e.g. histogram representation or pie charts). The analysis and interpretation of the results will be included in project deliverables, which will be made publicly available via the R-NMR website (https://r-nmr.eu/).

The size of data will be less than a GB.

NMR raw and processed data

During the R-NMR project, standardized methods for remote access to NMR-machines will be developed and tested. In this context, raw and processed NMR data will be generated during testing of the standardized experimental schemes that eventually will be adopted by the participating facilities. In addition, procedures for handling of remote NMR data according to the FAIR principles will be established (mainly in WP4, also with the support of the review of GDPR aspects from task 2.3).
The raw data generated typically will be in Bruker format because NMR machines are commonly equipped with Bruker consoles, so it is convenient that the data are directly stored

in the common Bruker architecture (Name/ExpNo/Pdata/ProcNo.). The same applies to processed NMR data. Acquisition parameters, pulse program, lists of variable parameters are text files, whereas raw data (ser or fid) and processed data (1r 1i, 2rr 2ir 2ri 2ii, 3rrr) files are binary coded files. Software scripts for the export of the primary NMR data in any format are established within the NMR community, and outside of the scope of the R-NMR project.

The size of the data are several GBs and the data will mainly be used by the facility staff. By using the standardized experiments on standard samples, the project will generate a series of reference spectra, of interest to the technical staff of facilities as well as to developers of new NMR schemes. To maximize the availability of these data also beyond the partnership we will leverage the BMRB public repository (https://bmrb.io/).

3. FAIR

Making data findable, including provisions for metadata

The survey data will be stored on a server of the University of Oxford, access to the raw data is only allowed to the PI of the Oxford partner (Professor Christina Redfield). The documents analyzing such data, will be made available also via Zenodo or the Open Research Europe platform, so that they will receive a PID.

For all NMR data, regardless of whether they are linked to a scientific publication, the following will apply:

- For the metadata related to the sample, common procedures for naming and identifiability will be discussed in Tasks 4.2 and 4.3 (Data & Metadata standards; Interface for data access). Some facilities already use data repository systems (e.g. LOGS, https://www.logs-repository.com/, or repositories developed in-house), providing a unique identifier and storing metadata about samples such as composition, preparation, or selected chemico-physical properties.
- The metadata related to the data acquisition will be automatically generated by the Bruker software and stored in the acquisition parameter file, from which they can be extracted using suitable text parsers available in the community.

- The raw and processed data generated in the project for its standardization purposes, will be deposited in BMRB, as mentioned in the previous section, making them publicly available with a DOI issued by the repository. The NMR data generated by the participating facilities for their users during visits (remote or in-person) fall within the remit of the respective specific access schemes (national or international) and thus will be handled in agreement with the corresponding DMPs.

Making data accessible

All the documents intended for public release produced in the context of R-NMR will be published both on the website of the project (https://r-nmr.eu/), and Zenodo or the Open Research Europe platform, as already mentioned for the analyses of surveys.

All NMR data generated in the project for its standardization purposes, will be deposited in BMRB, which is the main public repository for NMR data world-wide, making them publicly available with a DOI issued by the repository. This information can be linked to the sample metadata, via the permanent identifiers issued by LOGS, for those facilities that adopt this management system, or via the identifiers of other tools in use. Several of the partners have long-lasting collaborations with BMRB (for example, BMRZ is a curator of BMRB screening data, whereas CERM/CIRMMP infrastructure hosts a BMRB mirror).

We do not foresee any restriction on the use of the data generated by the project activities nor the need to track the identity of people within and outside the partnership accessing the data after their deposition.

Making data interoperable

The raw and processed data will be in Bruker format, which is the most common format for NMR data.
Regarding sample preparation and characteristics, there is no community standard for these metadata. We will implement during the project a common vocabulary for this purpose.

Increase data re-use

A creative commons licence will be systematically attached to the deposited data. In case of publications, the data will be made openly available at the time of the publication of the results.

The data will be stored on Zenodo, which ensures that they will be accessible after the end of the project (for at least 10 years). No paying fees up to 50 GB.

4. Other research outputs

The Open Science practices enable broad adoption and allow people worldwide to re-use the procedures and best practices developed by R-NMR. The project's results that are software code produced by the partners will be open source and made available e.g. via Github under the MIT license or similar. For outputs that are improvements to existing software, the improvement will be freely assigned to the owners of the background IP for incorporation therein. The Zenodo repository will be used to make documents available under a CC license or similar. If results are published as scientific papers, the consortium will commit to Open Access publications, including Gold and Green Open Access (or self-archiving). The partners will rely on dedicated funding from their research projects and/or institutions and store originals or pre-prints of their publications into their institution's repository or, in absence of such repositories, into arXiv

5. Allocation of resources

All documents will be stored on Zenodo, which ensures that they will be accessible after the end of the project (for at least 10 years). No fees are due up to 50 GB.

Similarly, there is no cost for data deposition in BMRB.

The coordination team will be responsible for data management in the R-NMR project, with the support of University of Patras, who is in charge of dissemination, for the management of the information on the website.

6. Data security

Will the data be safely stored in trusted repositories for long term preservation and curation?

As mentioned, all data will be stored in trusted repositories (Zenodo or the BMRB). Prior to deposition, the test data generated in the project will be managed by each individual facility, which will take care of data backup. The project will not deal with sensitive data or data of high economic value.

Data and metadata transfers will be encrypted according to standard practices.

7. Ethics

Participants (NMR facility managers) will be asked to complete an online survey that contains questions about their NMR facility and their experience with remote access to NMR spectrometer facilities; Participants will be asked if they would be willing to be interviewed online via Teams if more detailed information/feedback is required. In this case, they are asked to provide their name and email-address.

Names will only accessible to the PI of the University of Oxford, who is coordinating WP2 activities, and will not be used in the report of the surveys.

8. Other issues

Several national or other European initiatives in NMR are already working on metadata. We plan to get close to them and already are collaborating with the Panacea network.

Further we are in close contact with the eRImote network also for the discussion on FAIR as well as GDPR aspects.