Methods in Research on Research
MiRoR

D2.13 - Data Management Plan

Project GA	**676207**
Project acronym	**MiRoR**
Project website	**www.miror-ejd.eu**
EC Project Officer	**Audrey Arfi**
Deliverable version	**Final**
Deliverable nature	**Other**
Dissemination level	**PU**
Contractual Date of Delivery	**31/08/2016**
Actual Date of Delivery	**28/10/2016**
File name	**MiRoR_D2.13.pdf**
Author	**UPD**
Contributors	**All consortium members**

Quality Control Assessment - History of changes			
Issue	**Date**	**Comment**	**Author / Contributor(s)**
V0.1	13/07/2016	Document creation	Laura De Nale (UPD)
V0.2	19/07/2016	First draft	Laura De Nale (UPD)
V0.3	25/07/2016	Updated draft internally reviewed	Laura De Nale (UPD) / Isabelle Boutron, Raphael Porcher (UPD)
V0.4	16/09/2016	Updated draft after peer review from partners	Patrick Bossuyt (AMC), Els Goetghebeur (UG), Bridget Young(UL), Aurélie Névéol (CNRS), Isabelle Boutron (UPD)
V0.5	28/10/2016	Final version approved by all consortium partners	Isabelle Boutron (UPD), Laura De Nale (UPD)
V1.0	28/10/2016	Submission to the EC	Laura De Nale (UPD)

Disclaimer and acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No 676207. This document reflects the views of the author(s) and does not necessarily reflect the views or policy of the European Commission. The REA cannot be held responsible for any use that may be made of the information this document contains.

Project abstract

Our aim is to create, in Europe, an innovative and ambitious multidisciplinary intersectoral joint doctoral training programme, dedicated to Methods in Research on Research (MIROR) in the field of clinical research. “Research on Research”, is an emerging new scientific discipline that aims to reduce waste in research and increase research value. Waste in research represents tens of billions of Euros spent each year on studies that are redundant, flawed in their design, never published or poorly reported. The public is the main victim of this waste and reducing waste and increasing value of research represents a major societal challenge.

Our proposal involving 15 early-stage researchers, aims to:

1. Prepare students for envisioning the future challenges in clinical research and find innovative solutions to face them,
2. Train students to go well beyond the state-of-the-art in their research,
3. Help students think differently, taking advantage of the multidisciplinary expertise and intercultural diversity of the network,
4. Teach students how to move from research to action and convert knowledge and idea into a product,
5. Help students develop skills to match the public and private sector needs and create new professional opportunities.

MIROR brings together 7 world-class research teams in various disciplines (computer sciences, appliedmathematics, biostatistics, bioinformatics, clinical epidemiology, psychology, social sciences and translational medicine) from 6 different European countries; 6 non-academic partners involved in diverse sectors, and 4 major academic partners. We will tackle several steps of a clinical research project (planning, conduct, reporting and the peer-review); various study designs (observational studies, randomised trials, systematic reviews); various study questions (therapeutic, diagnostic, and prognostic evaluation) using various methods (meta-epidemiologic studies, qualitative studies, experimental studies, simulations etc).

Table of contents

Introduction

The present document is a deliverable of the MiRoR project (Grant Agreement #676207), funded by the European Commission Research Executive Agency (REA), under the Innovative Training Networks Programme of the Marie Sklodowska Curie Actions (H2020-MSCA-ITN-2015).

The purpose of this document is to provide the general approach to the project data management and it will serve as a framework for all Data Management Plans (DMP) to be prepared by MiRoR research fellows before the beginning of each project.

Carefully managing research data is an essential part of good research practice and starts with adequate planning. A DMP is a document specifying how data will be handled both during and after a particular research project. The DMP will reflect on data collection, data storage, data security and data retrieval.

This document was discussed and approved during the MiRoR network meeting in Ghent (Oct. 17-18, 2016) and it will be updatedwithin six months from the recruitment of MiRoR Research Fellows (M13 - March 2017), before the mid-term review meeting (M18 - August 2017), at the end of the first reporting period (M24 - February 2018) and whenever new datasets arise or new projects start.

1 – Research data and open access

Research data refers to information, in particular facts or numbers, collected to be examined and considered and as a basis for reasoning, discussion, or calculation.

In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, and interview recordings. The focus is on research data that is available in digital form.[1]

MiRoR participates on a voluntary basis in the Open Research Data Pilot (see article 29.3 of the grant agreement). This new feature of Horizon 2020 is designed to improve and maximise access to and reuse of research data generated by projects.

This implies thatforthe digital research data generated in the action ('data'), the beneficiaries must:

> *(a) deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the following:*
>
> *(i) the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible;*
>
> *(ii) other data, including associated metadata, as specified and within the deadlines laid down in the data management plan;*
>
> *(b) provide information — via the repository — about tools and instruments at the disposal of the beneficiaries and necessary for validating the results (and — where possible - provide the tools and instrument themselves).*

However, not all data can be open. Projects can therefore opt out at any stage and so free themselves retroactively from the obligations associated with the conditions – if:

- Participation is incompatible with the obligation to protect results that can reasonably be expected to be commercially or industrially exploited,
- Participation is incompatible with the need for confidentiality in connection with security issues,
- Participation is incompatible with rules on protecting personal data,
- Participation would mean that the project's main aim might not be achieved,
- The project will not generate / collect any research data or
- There are other legitimate reasons.

Projects can also choose to keep selected datasets or even all data closed for any of the reasons above, via their Data Management Plan.[2]

[1,2] Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020, version 2.1, 15 February 2016: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

2 - MiRoR research data

It is necessary to make a distinction among different kind of data that will be used/generated within the differentresearch projects. The MiRoR project will involve the following data:

1. **Data publicly available**, such as data publicly available on bibliographic databases (e.g. CENTRAL, MEDLINE); clinical trials registries (e.g. WHO ICTRP); and other data publicly available (e.g. Global Burden of Diseases of the Institute of Health Metrics and Evaluation).

2. **Data generated by the project from readily accessible information,** by adding our own scores to existing data (e.g., published report, trial records) such asassessment of the risk of bias from published reports, identification of the type of study design, classification of form of 'spin, etc.

3. **Data generated by the project from peer-review reports and responses to peer reviewers.** Review reports for published papers are available online as they come from open peer review journals whereby all peer reviewers' comments and authors responses are made available along with the published article. For rejected manuscripts peer-review comments and author responses will be gathered prospectively and the authors of the manuscripts submitted and peer reviewers involved in the peer review will be informed that the follow up of their manuscript/peer review will be analyzed with all potentially identifying data being kept strictly confidential by researchers. They will be able to refuse without any consequences on the acceptance or not of the manuscript.

4. **Data generated by the project, involving participants:** through qualitative studies and surveys such as telephone interviews, focus groups or online surveys. This type of research will not involve any risk for the participants. Participants will be patients interviewed on their experience in participating in clinical trials or in the peer review process; undergraduate and post-graduate students; researchers; and clinicians.

5. **Data generated by RCTs randomizing:**

    a) **Manuscripts submitted to journals**: the authors of the manuscripts submitted and peer reviewers involved in the peer review will be informed that the follow up of their manuscript/peer review will be analysed in order to improve the journal quality by the comparison of different editorial processes. They will be able to refuse without any consequences on the acceptance or not of the manuscript. All potentially identifying data will be kept strictly confidential by researchers.
    b) **Sections of a manuscript/protocol to be written with or without a writing aid tool**: the participants will be students invited to participate in a training session dedicated to manuscript/protocol writing. They will be informed they are participating in a randomized controlled trial and they will sign an electronic consent form. Data recorded will be data related to the manuscript such as drafted text and time-to-completion. Personally identifying information such as names and date of births will not be collected. Participants will be given a unique identifier known only to the researchers and linked to the participant's general demographic information such as previous experience in manuscript writing, level of comfort with writing in English and level of education.

3 – MiRoR data management: general principles

a) Research data repository

The consortium agreed to deposit the data and publications generated by the project in Zenodo, unless for a specific project there is a subject specific repository that is considered more relevant. In any case, the chosen repository will be indicated in each DMP. For all data not involving participants, we will make sure it is possible for third parties to access, mine, exploit, reproduce and disseminate the data. For data involving participants, we will follow the procedure for data sharing described in c).

b) Responsibilities

It is the responsibility of each ESR and his/her main supervisor to manage data and to deposit project research data and publications in Zenodo - or in another more relevant repository.

c) Informed consent of participants in case of data sharing

For research projects involving participants, prior to the conduct of the research, they will receive full information on the research to be undertaken and the data sharing policy.

Participants will be informed that:

- The research will contribute to improve future research;
- Data will be de-identified and stored in a secure repository;
- To gain access to their data, researchers need to submit a protocol and sign a data use agreement whose details will be provided to participants;
- If it is accepted by the ethics committee to which the project is submitted, we will offer the possibility to participate in the research while opting out from data sharing.

Participants will be asked to sign a written consent form providing assurance of confidentiality of information shared and de-identification. A contact person will be named to allow participants to obtain answers to any questions about the research and the participants' right.

Templates of the information sheets and informed consent documents will be made available upon request to the REA.

d) Data protection

The personal data collected as part of the research will be very limited and informed consent of participants about the use of personal data will be required. Personal identity will be protected by the use of anonymous codes and IDs. The relation of real names and codes/IDs will only be known to the leading researcher who will keep the records in secure place. All analysis will be blinded to the ID. When required, we will obtain approvals from the competent Data Protection Office. All copies of approvals / notifications regarding the processing of personal data by the competent institutional Data Protection Officer will be made available upon request to the Research Executive Agency (REA). Personal data will be encrypted and stored securely.

e) Intellectual Property rights

The intellectual property rights related to the results achieved within the project are detailed in the Grant Agreement (page 29) and in the consortium agreement linked to the Grant Agreement n°676207. According to the MiRoR consortium agreement, version 3 of 7 April 2016, the copyright and intellectual property right (IPR) for the data generated, collected or used during each project is owned by the Party that generates them.

In case of Joint ownership, each of the joint owners shall be entitled to use their jointly owned Results for non-commercial research activities on a royalty-free basis, and without requiring the prior consent of the other joint owner(s).

The PhD Candidate does not acquire, due to his assignment, any right of industrial or commercial property on the research results, the equipment, the knowledge or the expertise of the laboratory in which he is hosted, including unpublished results.

4 - MiRoR Data Management Plan

The consortium agreed on the following procedure for the draft and submission of the different versions of the DMP:

a) The MiRoR DMP template is provided to MiRoR research fellows upon their arrival in each team;
b) Each research fellow will be asked to prepare a DMP in the first months of his project, with the help of her/his supervisor, co-supervisor and mentor. Despite working on several projects, each research fellow will prepare only one DMP, that will be regularly updated;
c) Updated DMPs are submitted by the research fellow to the responsible supervisor and to the MiRoR project coordinator on a regular basis;
d) The coordinator gathers together all DMPs received in the period and submits them as a unique document to the European Commission:
   - within six months from the recruitment of MiRoR Research Fellows (M13 - March 2017),
   - before the mid-term review meeting (M18 - August 2017)
   - at the end of the first reporting period (M24 - February 2018), and
   - whenever new datasets arise or new projects start.

MiRoR DMP template

According to Annex 1 of the *Guidelines on Data Management in Horizon 2020*[3], within each research project, there is a number of information to be provided for each dataset (rows in blue below). Additional sections (grey rows) have been added by the consortium because considered important forclarifying data management within each project.

The consortium discussed and approved this template before the publication of the new

[3] Guidelines on Data Management in Horizon 2020, Version 2.1, 15 February 2016: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.
The consortium discussed and approved this template before the publication of the new *Horizon 2020 FAIR Data Management Plan template:* *http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.*

Horizon 2020 FAIR Data Management Plan template but, in line with this document, we commit to make our research data FAIR, that is findable, accessible, interoperable and reusable.

As a result, the information that will be provided by research fellows for each dataset of their research projects will be organised as follows:

Project details	ESR number:*title* Project objectives: Submission date: *dd/mm/yyyy*
Data set reference and name	Identifier for the data set to be produced such as for example the name of the project (i.e., MiRoR) followed by the ESR number and the project number e.g. MiRoR1-p1-xx
Data set description	Description of the data that will be generated or collected, its origin (in case it is collected), nature and scaleand to whom it could be useful, and whether it underpins a scientific publication. Information on the existence (or not) of similar data and the possibilities for integration and reuse.
Data formats	Description of the format used for storing each digital data type (e.g. excel, XML, SPSS, MS Word, AVI, NVivo) and explanation on why certain (proprietary) formats could be selected over and above other open formats and why this would not be possible otherwise.
Risks and potential difficulties during data collection and processing	Description of factors posing a threat for the quality of the data while collecting and processing the data. If possible, description of how these risks will be dealt with and potential difficulties.
Standards and metadata	Reference to existing suitable standards for metadata creation. If there are no standards in this discipline, describe what type of metadata will be createdand how. *Questions to consider:* - *How will the data be created?* - *What standards or methodologies will you use?* - *How will you structure and name your folders and files?* - *How will you ensure that different versions of a dataset are easily identifiable?*
Data sharing	Description of whether and how data will be shared, including access procedures, embargo periods (if any), outlines of technical mechanisms for dissemination and necessary software and other tools for enabling re-use, and definition of whether access will be widely open or restricted to specific groups. *Questions to consider:* - *Specify if the data will be stored in Zenodo or in another*

	subject specific repository - *Are there restrictions on data sharing? Describethese restrictions if any(e.g. for data involving participants researchers need to submit a protocol and sign a data use agreement)* - *What are the access procedures?* - *Is there embargo period?* - *How data will be licenced to permit the widest reuse possible?* - *Have you costed in time and effort to prepare the data for sharing?*
Archiving and preservation (including storage and backup)	Description of thedata backup procedures that will be adopted to ensure the data and metadata are robustly stored during the lifetime of the project. Description of measures taken to ensure the security of data (especially for sensitive data).Description of the procedures that will be put in place for long-term preservation of the data. Indication of how long the data should be preserved, what is its approximated end volume, what the associated costs are and how these are planned to be covered. *Questions to consider:* - *Where will your digital data be stored?* - What are the digital data backup procedures? - What are themeasures taken to ensure the security of data? - *What is the long-term preservation plan for the dataset? e.g. deposit in a data repository such as Zenodo*
Ethical issues	Description of potential ethical issues of collecting, storing, processing and archiving data. Description of procedures of ethical approval related to the project.
Responsibilities	Indication of who, within each team, will be responsible for data management, metadata production, dealing with quality issues and the final delivery of data for sharing or archiving. If several people will be responsible, specification of their roles and responsibilities. For collaborative projects explanation of the coordination of data management responsibilities across partners. Description of who will be responsible for the data once the researcher has left the laboratory.
Reported by	Name of the research fellow, supervisor, co-supervisor and mentor involved.