Nanoscale Systems for Optical Quantum Technologies

Grant Agreement No: 712721
Start Date: 1st October 2016 - Duration: 36 months

D4.3 Data Management Plan

Deliverable:	D4.3
Work package:	WP4 Management, dissemination and exploitation
Task:	4.2 Dissemination, communication and exploitation of knowledge
Lead beneficiary:	CNRS
Type:	Open Research Data Pilot
Dissemination level:	Public
Due date:	31 March 2016
Actual submission date:	27 March 2016
Author(s):	P. Goldner, D. Serrano (CNRS-CP)

Version history

Version	Date	Author(s)	Description
V1	20/02/2017	P. Goldner, D. Serrano (CNRS-CP)	First draft
V2	09/03/2017	P. Goldner, D. Serrano, (CNRS-CP) and all partners	Revised draft
V3	16/03/2017	P. Goldner, D. Serrano, (CNRS-CP) and all partners	Complete version
V4	27/03/2017	P. Goldner, D. Serrano, (CNRS-CP)	Final version

Partners:
CNRS, Karlsruhe Institute of Technology, ICFO, Lund University, Aarhus University, Keysight Technologies.

Copyright Notice

Disclaimer

Table of Contents

Deliverable Description

This deliverable establishes the *Data Management Plan* for data collected, generated and handled by the NanOQTech consortium, during and after the project.

Data summary

Purpose of the data collection/generation and origin of the data

NanOQTech's goal is to build novel quantum devices by taking advantage of the unique coherence properties of the rare-earth ions. This pursuit starts by the demonstration of rare-earth (RE) nanostructures with long optical and spin coherence, followed by the coupling of these rare-earth nanostructures to distinct systems (high-Q optical micro-cavities, mechanical nano-resonators, graphene layers...) and development of quantum devices such as single-photon sources or quantum memories. NanOQTech is expected to yield a large amount of experimental and theoretical data resulting from:

- Multiple characterizations of bottom-up produced RE nanostructures,
- Characterizations of opto-mechanical systems (optical micro-cavities, nano-resonators...),
- RE coupled systems experimental demonstrations,
- Theoretical modelling,
- Experimental control and instrumental development (cavity stabilization control, laser stabilization...).

Experimental data in NanOQTech primarily originate from the measurement of an analogic signal, typically an electromagnetic radiation or an electric signal, by a physical instrument (detector). The detected signal is then digitalized and subsequently stored by means of a data acquisition software. On the other hand, theoretical investigations will yield data from computing physical models.

The large majority of NanOQTech data will be directly collected and generated by the partners as a result of NanOQTech's investigations. Some data may be re-used, mainly for the sake of comparison between NanOQTech's findings and already published research results. On the other hand, already existing programming code is likely to be re-used and further implemented. Re-used data and code will mostly belong to NanOQTech partners from previous investigations. Exceptionally, published data from research groups out of the consortium might be used but in such cases the original data source will be properly cited and referred to.

Data types, formats and sizes

Three major categories of data will be handled in NanOQTech. The main category consists of datasets in the form of spreadsheets or tabulated text files. Furthermore, image files and programming code will be generated. In terms of sizes, datasets and code files are expected to be lower than hundred kilobytes (KB), while image files might take a few megabytes (MB) of storage space. We estimate the total amount of data generated during the project between 100 and 300 GB.

A summary of the data formats used by NanOQTech partners is listed in Table 1:

File type	Formats
Datasets	".dat" , ".txt", ".xls", ".cvs", ".tsv", ".json", ".h5",
Images	Raster formats: ".png", ".tif/tiff", ".jpg/jpeg" Vector formats: ".eps", ".cgm" , ".amf"
Code	".m" , ".vi" , ".nb", ".pyc", ".pyd", ".pyo", ".c", ".cpp", ".h", ".hpp", ".xml", ".i", ".t", ".dox"
Other	".pdf", ".zip", ".rar"

Table 1: Summary of data formats used in NanOQTech.

Details about formats can be found in Annex 1.

Data utility

Data collection, generation and exploitation in NanOQTech is in first instance aimed to be useful to NanOQTech partners, and on a larger scale, to any scientific institution or private entity following and eventually taking advantage of NanOQTech's findings and developments.

FAIR Data

Making data openly accessible

In principle, datasets or images related to published articles or, more generally, data considered of remarkable interest for the scientific community will be openly shared. Data involved in patents or that would jeopardize further publications if made immediately open, may be kept closed or embargoed until the patents/publications are secured/published.

The data will be made available through Zenodo.org website. It is a catch-all repository for EC funded research provided by CERN, an OpenAIRE partner.

Material deposited on Zenodo.org is completely free to access and does not require an account, login or password.

Most data generated by NanOQTech do not need specialized software to access them. In the same way, images will be in standard formats and can be viewed on essentially any computer.

On the other hand, source code may need specialized software to be executed, such as Matlab[1], Mathematica[2], Python[3] or C/C++ environments[4]. Python and C/C++ codes can be edited/compiled with free, open access software. Matlab is a proprietary software but Matlab code can be often executed in open source software like Octave[5] or SciLab[6]; Mathematica is also proprietary but corresponding codes may run on the open source software Mathics[7]. Information about the relevant open source software will be given on the website.

Making data findable

Openly shared NanOQTech data will be identified according to the following convention:

- Data related to a published research article will be named as follows:

  "First author name _ Article reference_ Figure number (letter)_version.extension"

  in which the article's reference will be stated as:

  "Standard Journal Abbreviation + issue number(year)page"

  and the *version* will be abbreviated as "v" and numbered in increasing order from less to more recent, starting by 1 with increments of 1 unit, e.g. "v1, v2, v3…".

  For figures, tables or calculations involving or yielding an ensemble of data files, those will be shared within a compressed folder following the above stated naming convention, in which, the 'extension' will correspond to a standard compressed folder extension (e.g. .zip). In such cases, the name of the individual files included in the compressed folder will be freely chosen by the data owner, making sure that name repetition does never occur and that the content of each individual file is clearly described in a supporting text file. Furthermore, in cases where one or more individual files would undergo updates, a whole new version of the compressed folder will be uploaded, in which the old version of the updated files will be replaced by the new one.

- Data from other sources, e.g. public presentations, posters, summary reports, etc. will be shared within a single compressed folder named as:

  "First author _ DOI Zenodo _version.compression extension"

  in which, the DOI given by Zenodo.org will be used to link the data to its source (e.g. conference poster). As in the previous case, the name of the individual files included within the compressed folder will be freely chosen by the data owner, making sure

---

[1] https://www.mathworks.com/
[2] https://www.wolfram.com/mathematica/
[3] https://www.python.org/
[4] http://www.codeblocks.org/
[5] http://www.scilab.org/fr/
[6] https://www.gnu.org/software/octave/
[7] http://mathics.github.io/

that name repetition does never occur and that the content of each individual file is clearly described in a supporting text file. Furthermore, in cases where one or more individual files would undergo updates, a whole new version of the compressed folder will be uploaded, in which, the old version of the updated files will be replaced by the new one.

Examples:

Example 1: Data from a published research article

Data corresponding to Figure 1 in "*Optical Line Width Broadening Mechanisms at the 10 kHz Level in Eu3+:Y2O3 Nanoparticles*" by Bartholomew et al., Nano Lett. 17 (2), 778 (2017)

➔ http://pubs.acs.org/doi/abs/10.1021/acs.nanolett.6b03949

Two data files are associated to this figure, which should be named as:

- Data: *"Bartholomew_NanoLett17(2017)778_Figure1_v1.dat"*
- Image: *"Bartholomew_NanoLett17(2017)778_Figure1inset_v1.jpg"*

Example 2: Data from other sources

Data from the poster "*Towards bulk crystal coherence times in Eu3+:Y2O3 nanocrystals*" by Bartholomew et al.,

➔ https://zenodo.org/record/168412#.WKnTuH8TA8A

"Bartholomew_DOI10.5281/zenodo.168412.svg_v1.zip"

For efficient identification and search, data sets will be tagged with mandatory keywords that are general enough to cover the whole project. Specific keywords will also be added to describe the data. The table below gives the mandatory keywords and examples of keywords that could be used for each scientific WP.

	Keywords
Mandatory	Rare earth, quantum technologies, nanoqtech
Specific (examples)	
WP1	nanostructures, nanoparticles, thin films, micro-cavity, real-time control...
WP2	micro-cavity, real-time control, Purcell enhancement, qubit, single photon, telecom, quantum gate, decoherence, ...
WP3	plasmon, electrical control, nano-oscillators, strain, hybrid quantum systems, conditional dynamics...

Table 2: Keywords for NanOQTech data.

For each data set or image, the following metadata will be generated through the corresponding fields that are available on Zenodo.org when data sets/images are uploaded:

- **Digital Object Identifier**
- **Publication date**
- **Title**
- **Authors and affiliations**
- **Description**
- **Keywords**
- **Access right**
- **License**
- **Community** ('NanOQTech H2020 Project' gathers all NanOQTech related material)
- **Grants** (reference to NanOQTech is automatically generated by Zenodo.org)

Making data interoperable

The following measures will ensure that data collected in NanOQTech will be as easy to re-use as possible.
Raw and processed data (e.g. data points in an article figure) will be provided. Data and metadata will be given in standard US English. Each data set will be accompanied by a description (field available on Zenodo.org) to identify contents and experimental conditions. References to published articles or other sources may also be given for additional information. For data arranged as text files, parameters corresponding to the columns will be given too. Units from the International System of Units, or units normally

used in a field (e.g. impurity spectroscopy), will be used. For files gathered in a compressed folder, a text file will give the necessary details.

Standard vocabulary will normally be used, together with technical vocabulary that can be easily found in different sources (e.g. on-line encyclopedias). Uncommon acronyms will be avoided.

Increase data re-use

The Creative Commons Attribution 4.0[8] will be used for all data deposited on Zenodo.org.

Under this license, users are free to:
- Share — copy and redistribute the material in any medium or format,
- Adapt — remix, transform, and build upon the material, for any purpose, even commercially.

This is possible under the following terms:
- Attribution — The user must give appropriate credit, provide a link to the license, and indicate if changes were made. The user may do so in any reasonable manner, but not in any way that suggests the licensor endorses him / her or his / her use.

The licensor cannot revoke these freedoms, as long as the user follows the license terms.

Data will generally be made available for re-use immediately. In the case the data are used in a publication, they will be deposited on Zenodo.org within one month of the publication date. Access will follow the same embargo (6 months at most) as the publication deposited on Zenodo.org.

Data quality is the main responsibility of the partner(s) producing them. They follow the standard practices in physics and chemistry: reproducibility, noise/error bars evaluation, reliable statistical analysis, consistency with other experimental results and theory.

Data generated by NanOQTech will be accessible in the long term, as long as CERN is operating, which is foreseen to be at least 20 year[9]. Moreover, even if the Zenodo.org server would close, migration to suitable repositories is guaranteed[10]. Long term preservation is important since NanOQTech is dealing with highly prospective research, which may undergo future developments completely unknown at the moment.

Data formats could endure significant changes in the future, making nowadays files not readable/usable. This seems to be unlikely for the data in the form of text files or images, but could happen for software. In the latter case, it is difficult to estimate the time during which a file will be usable. As an example, for Matlab code, no official policy about version compatibility is provided by the company.

---

[8] https://creativecommons.org/licenses/by/4.0/legalcode
[9] https://zenodo.org/policies
[10] https://www.zenodo.org/faq

Allocation for resources

Zenodo.org repository is free of charge. It is developed by [CERN](#) under the EU FP7 project [OpenAIREplus](#) (grant agreement no. 283595) and OpenAIRE2020. Funding by other projects and CERN itself is actively considered by Zenodo.org managing team[11].

The extra work involved in preparing and uploading the data sets is part of the scientific work on the project and is covered by the funding for personal costs. We estimate it at 0.2 to 0.4 person-month per year and partner. The cost of creating/supervising/updating the data management is estimated at 1 pm for the project. It is covered by the personal costs affected to WP4.

The project manager, Diana Serrano, and the coordinator, Philippe Goldner, are responsible for the dissemination of the DMP within the NanOQTech consortium and for supervising its global implementation. However, partners are responsible for implementing the DMP at their level: data production, quality assessment, uploading, providing metadata, etc.

Data security

The different partners in NanOQTech, according to their institution policies, are responsible of the secure storage and data recovery of their own generated data.

Data uploaded in Zenodo.org are stored in CERN Data Center. Both data files and metadata are kept in multiple online replicas and independent replicas. CERN has considerable knowledge and experience in building and operating large scale digital repositories and a commitment to maintain this data center to collect and store 100s of petabytes of the Large Hadron Collider (LHC) data as it will grow over the next 20 years.

Transfer of data to Zenodo.org website is done through an encrypted link and a password protected login. The coordinator reviews each upload for integration in the 'NanOQTech H2020 project' community.

Ethical Aspects

No ethical or legal issues will have an impact on data sharing, as NanOQTech does not raise any of these points, as stated in the Grant Agreement.

Conclusion

This data management plan describes in details the features of data generated in NanOQTech. We also define clear procedures to maximize open access and re-use of the data, in which the open access Zenodo.org website, managed by CERN, an Open AIRE partner, plays a key role. This will ensure the long-term exploitation of NanOQTech data.

---

[11] https://www.zenodo.org/faq

Annex I

Data file extensions.

File ext.	description	Viewing /editing software
dat	Generic data file	Text editors (e.g. Notepad)
txt	Text file	Text editors (e.g. Notepad)
doc	Text file	MS Word
xls	Spreadsheet	MS Excel
tsv	Spreadsheet	Spreadsheet software (e.g. MS Excel)
cvs	Spreadsheet	Spreadsheet software (e.g. MS Excel)
ogw	Spreadsheet	Microcal Origin
json	Data interchange format	jq
h5	Hierarchical data format	Matlab, Python editors
jpg	Raster image	Standard image viewers
png	Raster image	Standard image viewers
tif	Raster image	Standard image viewers
m	Code	Matlab
nb	Code	Wolfram Mathematica
pyc	Code	Python editors
pyd	Code	Python editors
pyo	Code	Python editors
c / cpp	Code	C / C++ editors
h / hpp	Code	C / C++ editors
xml	Code	XML editors
i	Code	C / C++ editors
t	Code	Holt Software Turing
dox	Code	C / C++ editors
vi	Graphical code	LabVIEW
pdf	Portable document format	Standard PDF viewers
zip	Archive file format	Standard file archivers/unarchivers
rar	Archive file format	Standard file archivers/unarchivers