Revolutionizing Spatial Biology with a cutting-edge Multi-Scale Imaging platform

Grant Agreement Nº 101136680

WP 6

D6.1: Data Management Plan

Lead Beneficiary: ABBELIGHT (ABBELIGHT)

Due date of deliverable: 31.03.2024

Type and dissemination level: Report, Confidential

Funded by the European Union

nanoSCAN project is funded by the EU's HORIZON EUROPE programme under grant agreement number 101136680.

Table of Contents

Disclaimer

This document is part of the deliverables from the project nanoSCAN, which has received funding from the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the granting authority can be held responsible for them.

About nanoSCAN

The nanoSCAN project aims to transform tissue analysis with a novel 3D spatial biology platform that provides crucial insights into cellular and tissue functions. Spatial biology visualizes the interaction of molecules with their 3D environment, which is essential for cell and tissue screening. However, most spatial biology imaging technologies, based on wide-field microscopy, have limited spatial resolution and insufficient molecular profiling. A major obstacle to quantitative tissue imaging progress is the lack of a single instrument that can cover various complementary scales from tissue to molecule with high speed, high throughput, and high accuracy.

To address these limitations, we propose to develop a new imaging platform, the SAFe-nSCAN, which combines multi-scale optical microscopy solutions, from structured illumination microscopy [Calvarese 2022] for rapid cell and tissue inspection and classification to single-molecule localization microscopy techniques [Mau 2021, Jouchet 2021, Friedl 2023] for deeper and higher nanoscopic 3D information over preselected regions.

The consortium consists of academic partners who will develop the technology, a non-profit association that will facilitate beta testing and promote the technology, and an SME that will collaborate with a new start-up company to manufacture chips and bring molecular resolution spatial biology to the market.

Executive summary

The first release of the Data Management Plan of the project nanoSCAN is presented here. The document is drafted following the EU guidelines and public template for DMP, initially foreseen for Horizon 2020 and in force in Horizon Europe. The key issues on description of the dataset to be generated, compliance with the FAIR principles, general aspects related to conservation and protection of data, ethical issues are reported. This document can be updated to reflect new evolution in the DMP during the course of the project.

Abbreviations

Abbreviation	Definition
DMP	Data Management Plan
FAIR	Findable, Accessible, Interoperable, and Reusable.
ASTER	Adaptable Scanning for Tunable Excitation Regions
SIM	Structured Illumination Microscopy
SMLM	Single Molecule Localisation Microscopy
ModLoc	Modulated Localisation Microscopy

1 Data Summary

The nanoSCAN project will generate two major categories of data: documents related to the management and communication of the project (WP 5 and WP6), and scientific data (WP1/4) as detailed below, all governed by the FAIR principles. NanoSCAN will generate new data both in the development process and in the validation and first application phase. More specifically, the main categories of data that will be generated and specific datasets are:

- **Chip performances**
  Images to evaluate modulation contrast and modulation frequency
  Waveguide coupling and propagation losses at different wavelengths
  Pattern generator insertion losses
  Polarization behaviour in waveguides
- **SAFe-nSCAN components characterisation**
  Laser power output
  Scanning unit for ASTER performances, achievable field of view
  Repositioning precision for mapping
- **SAFe-nSCAN validation**
  optical resolution and localisation precision of different integrated imaging techniques: widefield, SIM and ModLoc
- **Beta-testing validation**
  Report with improvement and solution-problem fit description.
- **Simulations**
  Synthetic data of optical propagation in the microscope under structured illumination
  Mode propagation in the waveguides
  Thermal response and switching velocity

The formats of the data generated during the project will be:

- **Microscopy Images**: microscopy acquisition files are saved in TIFF and bigTIFF format.
- **SMLM measurements**: processing of SMLM data are tables of coordinates saved as CSV files.
- **Componenent performances**: text files.
- **Metadata**: optical microscopy images will be annotated with metadata, in the form of structured description (json file), following the QUAREP initiative, promoting quality control and reproducibility.

- **Reporting**: text files and PPT presentation files.

The size of the data is highly variable. SMLM data (TIFF movies) are prone to generate big data set, easily leading to TB level for whole measurement sessions (with a predicted flow of 2GB/s data size is expected to be on the order of 1.8B per acquisition). The issues of data conservation, together with the principle of limitation of long-term huge data storage will be considered. Either reduction of the raw datasets to synthetic forms (once TIFF movies are validated for SMLM acquisition datasets only CSV files will be stored) or lossless compression of SMLM images to reduce data volumes by at least 20% of the initial size.

The data generated during NanoSCAN project will be of use to the very same partner of the consortium also after the conclusion of NanoSCAN for further development on the components and system, for alternative analysis of the acquired data, or to devise new approaches. External users from the scientific community could benefit from the available dataset to proper exploit the combined use of Widefield, SIM and SMLM microscopy to scan biological tissue.

2 FAIR Data

2.1 Making data findable

All published data will be assigned unique permanent identifiers (DOI) through publication in public domain or institutional repositories. The creation of metadata depends on the used data storage system but a common set of metadata for the project has been agreed upon by the partners. Institutional repositories provide usually the environment to describe data with rich metadata.

2.2 Making data accessible

Project results will be published in peer-reviewed journals following Plan-S initiative for Open Access, once covered by IP laws. For publications of scientific results arising from the project itself, journals with open access and with high impact factors, or the open access publishing platform “Open Research Europe” will be considered with high priority. In the case of exceptional results meriting inclusion in very high impact publications without the option of open access, then open access (green open access) will be ensured by archival of the peer-reviewed preprints in an online repository, either in partner institutional archives (such as HAL in France) or in disciplinary archives as Zenodo (or arXiv). The same repositories will be used to archive gold open access papers. Authors will be asked to provide proper justification for choosing publication routes in hybrid open access journals.

Social media platforms will be used to disseminate results and seek community feedback.

POLIMI developed Napari-SIM-processor, a user-friendly Python plugin for reconstructing SIM data. ABBELIGHT will integrate the demodulation processing and provide image reconstruction software for end-users and the open-source community.

2.3 Making data interoperable

NanoSCAN will adopt standardised data formats, such as standard image file formats (e.g., TIFF) and open data standards (e.g., JSON, XML), to ensure compatibility with various software tools and platforms. QUAREP-LiMi is an initiative with members from both academia and industry aiming at improving the reproducibility of light microscopy experiments in Life and Material Sciences. Within nanoSCAN project we will follow QUAREP-LiMi initiative [Nelson 2021] towards standardised vocabulary and formats to describe light microscopy instrumentation and experiments which will make increase data re-use.

2.4 Making data Reusable

During the whole project, care will be taken in assuring best practices in laboratory measurements and definition of measurements. To ensure the data storage meets Data Policy and Open Access requirements, the core data of this project will be automatically backed up and kept for 25 years, in line with the research council's recommendations. All data will be kept, at least, in duplicate and centrally indexed along with the associate metadata as described in the previous section in order to facilitate re-use.

3 Other research outputs

While developing the hardware for structured illumination, POLIMI will advance and finalize the open-software (based on python and the Napari package) for reconstruction of SIM data (https://www.napari-hub.org/plugins/napari-sim-processor). The software will be capable of reconstructing the data acquired in the SIM microscopes and during the beta testing. Synthetic data will be generated using python, based on the generation of realistic Point Spread Function of the microscope.

4 Allocation of resources

Costs for initial curing, maintenance and storing of data will be covered by standard procedures at local institutions. Long term preservation and open access will be granted by the service offered free of charge by external repositories (e.g. Zenodo)

5 Data security

All data generated by NanoSCAN do not require specific data protection measures, as it would be the case if the data would involve personal information and fall under the European data protection legislation. At the partner institutions, data are systematically saved on internal and protected servers with a daily back-up of hard drives. Some partners are using open-source cloud software and local RAID storage technologies. According to local regulations, primary data used for publications sometimes has to be securely stored for a defined period of time, e.g. ten years, in a durable form in the institution of origin. Central data repositories for open access mostly are managed by institutional departments, e.g. libraries, which ensure data security - including protection against malevolent hacking – as well as reliability and sustainability. HAL (CNRS) is for instance going to be certified by CoreTrustSeal to demonstrate it is a trustworthy data infrastructure.

6 Ethics

The beneficiaries do not expect any ethical or legal issues impacting data sharing and do not deal with personal data. If according questionnaires will be created, informed consent for data sharing and long term preservation will be included

7 Other issues

This DMP will be modified as needed also in agreement to new requirements or suggestions issued by:

- national and EU legislation;
- indications produced at EU level, in particular in the framework of Open Science and EOSC;
- best practices issued at the level of the Laserlab-Europe network of laser infrastructures;
- local requirements set forth by partner institutions.

In the present stage, this field is quite dynamic with potentially new standards or practices in preparation. Therefore, we will be flexible in adopting best practices as needed.

8 References

- M. Calvarese, P. Pai., A. Candeo, G. Calisesi, F. Ceccarelli, G. Valentini, R. Osellame, H. Gong, M. A. A. Neil, F. Bragheri and A. Bassi, Opt. Express, 2022, 30, 30246–30259.
- M. Calvarese, P. Pai., F. Ceccarelli, F. Sala, A. Bassi, R. Osellame and F. Bragheri, Sci. Rep., 2022, 12, 239.
- Mau, A., Friedl, K., Leterrier, C. et al. Fast widefield scan provides tunable and uniform illumination optimizing super-resolution microscopy on large fields. Nat Commun 12, 3077 (2021). https://doi.org/10.1038/s41467-021-23405-4
- Jouchet, P., Cabriel, C., Bourg, N. et al. Nanometric axial localization of single fluorescent molecules with modulated excitation. Nat. Photonics 15, 297–304 (2021). https://doi.org/10.1038/s41566-020-00749-9
- Friedl K, Mau A, Boroni-Rueda F, Caorsi V, Bourg N, Lévêque-Fort S, Leterrier C Asessing crosstalk in simultaneous multicolor single-molecule localization microscopy Cell Reports Methods 3(9), 100571 (2023) https://doi.org/10.1016/j.crmeth.2023.100571
- Nelson, Glyn, Ulrike Boehm, Steve Bagley, Peter Bajcsy, Johanna Bischof, Claire M. Brown, Aurélien Dauphin, et al. “QUAREP-LiMi: A Community-Driven Initiative to Establish Guidelines for Quality Assessment and Reproducibility for Instruments and Images in Light Microscopy.” *Journal of Microscopy* 284, no. 1 (2021): 56–73.