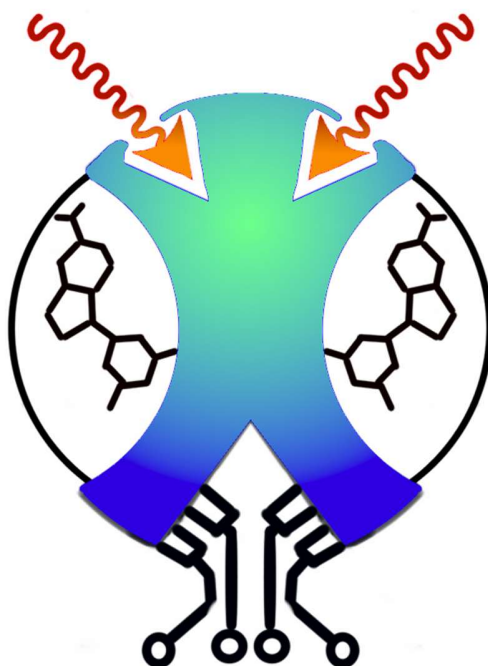


Project 101046790
InsectNeuroNano

Deliverable D5.5
Dissemination level: PU

Deliverable D 5.5 – Data Management Plan Update 2



Project acronym	InsectNeuroNano
Project Number	101046790
Project title	Insect-Brain inspired Neuromorphic Nanophotonics

Document Properties	
Nature of Document	DEC
Work Package	WP5 - Communication/Dissemination/Exploitation
Task Leader	1. ULUND
Authors	Gerda Rentschler, Anders Mikkelsen, and the InsectNeuroNano consortium
Due Date of deliverable	M30
Actual delivery date	M30
Dissemination Level	PU
Version	DMP Update 2

Disclaimer: Funded by the European Union. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Innovation Council. Neither the European Union nor the European Innovation Council can be held responsible for them.

European
Innovation
Council



Funded by the EU
Horizon Europe Programme
Project 101046790 - InsectNeuroNano



Contents

1. Data Summary	4
2. FAIR data	6
2.1. Making data findable, including provisions for metadata	6
2.2. Making data accessible	7
2.3. Making data interoperable.....	9
2.4. Increase data re-use.....	10
3. Other research outputs	11
4. Allocation of resources	11
5. Data security.....	12
6. Ethics	13
7. Other issues.....	13

1. Data Summary

Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.

What types and formats of data will the project generate or re-use?

What is the purpose of the data generation or re-use and its relation to the objectives of the project?

What is the expected size of the data that you intend to generate or re-use?

What is the origin/provenance of the data, either generated or re-used?

To whom might your data be useful ('data utility'), outside your project?

Towards the end of the second reporting period, the DMP was circulated among the principal investigators of InsectNeuroNano and the consortium members confirmed that the DMP is still valid.

The goal of the project is to develop nanophotonic on-chip devices for integrated sensing and neural computation, inspired by the insect brain. This will uniquely combine four lines of research:

1. Progress in understanding insect neurobiology that provides proven circuit designs to solve significant problems such as autonomous navigation;
2. Advanced III-V semiconductor nanowire/nanopillar and waveguide interconnects technology that exploits light to obtain a large number of interconnects with extremely low power consumption;
3. Optically efficient stable molecular dyes that can be used for novel memory components;
4. Circuit technology developed for quantum computing.

The data generated during the project will fall into one of the following categories:

- Numerical data, which can be experimental data or a result of theoretical work, such as numerical simulations. Numerical data will be stored either as tabulated values, or high density binary datafiles.
- Imagery, which allows for visual representation of device properties, including morphology, composition, crystalline quality. Standard image file formats will be used to ensure reusability.
- Computer code, which can be part of the experimental data processing chain or theoretical studies.
- Insect neurophysiology data: Anatomical data will comprise image stacks from confocal microscopy and block-face electron microscopy (both standard formats), 3D reconstructions of single-neuron morphologies (resulting from confocal microscopy and from electron microscopy; swc files), and overall brain morphologies (3D surface files, e.g. obj-files). Physiological data will comprise raw recording traces, processed data (all stored as matlab files), as well as any software code used for analysis and recording (matlab and python). Physiological data and 3D reconstructions are relatively small files (mat-files and obj files). Confocal images are large datasets between 200 and 1000MB per intracellularly recorded neuron (tiff), while EM volumes are very large datasets of several TB (tiff and jpeg image pyramids for fast access).

- Synthesis of molecular dyes will generate synthetic procedures (text files) and NMR and MS characterization as x,y data in txt files.
- Electronic or optical material characterization techniques: profilometry, which measures the thickness of a sample and variations records a data graph in csv-format. EQE (External quantum efficiency) produces data graphs and txt format-files. Photoluminescence spectra are stored as txt-files which can be displayed as data graphs. X-ray diffraction (XRD) diffraction pattern are recorded in file formats raw, txt, xy or dql.
- Processing steps that do not create electronic files are noted in a lab journal. In fact, quite a few among the researchers are still using paper lab journals. Also, excel-files are very popular to record parameters and other descriptive information about experiment series.

The scientific progress of the consortium will almost exclusively rely on the data generated/collected throughout the project, reuse of earlier data is expected to be very limited. The only exception to this rule are two datasets of volume electron microscopy (central complex of *Megalopta genalis* and *Eciton hamatum*) that were previously acquired and which will provide a foundation for analysis relevant to this project.

The consortium foresees that the generated data will amount to less than or in order of 100GB throughout the project duration for all lines of research except insect brain connectomics, therefore no special provisions have to be taken to ensure storage for those data. In case, we find that the currently foreseen storage resources are not sufficient, the researcher will secure storage at a cost for LUSEC data storage of 50SEK/month per TB. LUSEC is the High-secure data management platform of Lund University. To date, extra storage has not been required.

For insect brain connectomics, we follow a data storage plan established for other EU projects that generate data using this method in the Heinze group. We anticipate data volumes between 4-10TB to be generated by this project. Raw images (tiff), as well as processed (aligned, filtered) images (jpg image pyramids) will both be stored at Lund Data Center (LDC), together with the annotation database and results of all automated image segmentation. The same infrastructure will be used to centrally store and backup all other data from insect neurobiology as well. Data is backed up automatically on a daily basis. The database which houses all results of annotations of connectomics data (3D traces, connectivity data), while also hosted at LDC as prime storage and backup site, is additionally manually backed up on hard drives not connected to the internet once a week. Moreover, local mirrors housing the processed image data are located at different sites to ensure fast access to image data. These are continuously synced with the main image server at LDC and thus provide an additional backup layer. Finally, automated image segmentation is performed on a dedicated GPU cluster, physically located at the department of Biology, providing an additional backup for all data generated in this process.

Storing all annotation and image data on a server with managed user rights allows not only for multiple users to work on these datasets, but to also grant access to the data to third parties after peer reviewed publications have been generated using these datasets. These datasets will be of high usefulness to the neuroethology field and can be annotated by the qualified members of the community to complete aspects of only peripheral interest to this project.

The major utility is for the creating researcher to test a research hypothesis. Representative examples of raw data will always be included in research publications (incl supplementary

material) to demonstrate the verity of the findings and to enable others to repeat an experiment. Most of the research is based on a physical sample and therefore, without the sample, the data are of limited use for third parties. Still, data might be useful as training data sets for other researchers to develop alternative reconstruction or simulation algorithms e.g. with deep learning approaches. Also, researchers in similar research fields (e.g. materials science) could make use of code deposited in GitHub to analyze their own measurement data.

2. FAIR data

2.1. Making data findable, including provisions for metadata

Will data be identified by a persistent identifier?

Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?

Will metadata be offered in such a way that it can be harvested and indexed?

Scientific documents will be provided with keywords, a DOI and online searchability where applicable, as facilitated by the chosen institutional or subject repository.

Source code will be searchable through GitHub, use semantic versioning (<http://semver.org/>) and commented Code versions corresponding to published paper results will be archived with a persistent identifier and citation metadata in CFF format, and registered with ModelDB for computational neuroscience models and nanoHUB for nanotechnology models.

Microscopy data /imagery will be provided with automatically generated metadata by the camera software (machine parameters like scale, camera brand, acquisition time, exposure time, light source, light intensity, and image mode completed by manually generated metadata (Creator, Title, Subject, Description) wherever possible.

Insect neurophysiology data will be organized in an online database (www.insectbraindb.org) which is an accessible platform for sharing, managing, displaying and depositing any data related to insect brain neurophysiology and anatomy (Heinze et al, 2021, eLife). Data are stored at the level of cell types and experiments. Each newly discovered cell type receives a unique internal identifier, the NIN-number (neuron identification number), as well as a persistent handle. Experiment entries can be generated at the level of cell types (experiments associated with a neuron type), brain region, or species. In each case, each experiment entry receives a persistent ID (handle), allowing to locate each individual physiological and anatomical experiment performed during the project. Each entry will be automatically embedded in its anatomical context, but will also be manually tagged with meta-information (e.g. creator, experimenter, group leader, date, etc.). Once datasets are made publicly available, they become locked and cannot be altered or deleted. If a dataset needs to be updated after initial upload, the dataset needs to be republished and will obtain a new persistent handle. As all previous handles remain resolvable, all changes to datasets are tracked and previous states of

the site can be displayed at any time. The database has highly developed search functions that allow easy and intuitive exploration of all deposited data.

For all other published datasets, metadata creation will be in accordance with DataCite, the following six fields being mandatory: Identifier, Creator, Title, Publisher, PublicationYear, and ResourceType. We will encourage the InsectNeuroNano researchers to also add the recommended fields: Subject, Contributor, Date, RelatedIdentifier, and Description. For data originating from electronic, and optical (absorption, excitation and emission) characterization techniques, other metadata fields will also be required, including the automatically generated metadata. Before uploading microscopy and (NMR) spectroscopy data to repositories, the researchers will curate the metadata manually in order to make the files discoverable and reusable.

To facilitate discoverability, adding subject keywords (corresponding to DataCite field Subject) to data and tools will be encouraged. As InsectNeuroNano comprises research in the four different areas of research listed in Section 1, there is not one controlled vocabulary that exactly matches the scope of all projects. Thesauri that might cover parts of the subject scope of the project include IEEE and INSPEC thesauri for the physics aspects, and the ACM Computing Classification System for the computing aspects. Data related to molecular dyes will be made searchable by compound names, empirical formula (Hill Notation), and unique CAS Number.

2.2. Making data accessible

Repository:

Will the data be deposited in a trusted repository?

Have you explored appropriate arrangements with the identified repository where your data will be deposited?

Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?

Data:

Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.

If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

Will the data be accessible through a free and standardized access protocol?

If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?

How will the identity of the person accessing the data be ascertained?

Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?

Metadata:

Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?

How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?

Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?

The consortium will use open data repositories that provide a persistent identifier for the research data.

Data related to insect brain neurophysiology and anatomy will be deposited in the Insect Brain Database (<https://insectbraindb.org/app/>). Each dataset receives a persistent handle once public. All data will be accessible for viewing to all internet users after the data has been published in a peer-reviewed scientific publication, or immediately after generation if the specific dataset is only of peripheral interest to the project (but might be important to other researchers; e.g. neurons recorded outside the target area). All data will be uploaded immediately after generation of the data but kept in 'private mode'. This mode of the database allows to restrict access of selected data to the owner of the dataset and the database administrator, but also allows selective sharing with other database users, facilitating collaboration within the consortium. Data can be made available to journal reviewers by generating unique links to selected datasets (prepublication) before being searchable on the database. Once published, the datasets will be made available to all internet users for viewing. Registered users of the insectbrain-database will be able to download all files.

Data that form software/algorithms for modelling, are stored on GitHub/GitLab. Stable versions of code will be made available directly after publishing of a peer-reviewed articles or (as is more increasingly required) made available to reviewers and archived (as above) on publication in a form that allows direct reproduction of the published results

In most peer-reviewed research publications, representative examples of raw data are included in the supplementary material which provides long-term preservation.

Research publications in peer-reviewed journals are findable through DOIs. To ensure visibility, the raw dataset have to be clearly referenced in all related scientific publications.

Publicly posted datasets in the other disciplines of InsectNeuroNano will be in repositories like NoMaD Repository, Materials Cloud, Crystallography Open Database, Zenodo or ERDA.ku.dk. All of these repositories either manually or automatically assign DOIs to the uploaded datasets. In case we want to share data via repositories with more specialized subject areas, we will check their entries in FAIRsharing and re3data, respectively, to ensure that the repositories have high quality standards.

While data concerning insect brain neurophysiology can be made publicly accessible early on, the data in nanotechnology, materials science and chemistry will be assessed for their commercial exploitation potential before publication. IP protection may delay publication by some months. Trade secrets: in case it turns out that the scale-up of the biomimetic circuit will need longer time to commercial exploitation than 5 years, the partners may decide to keep it secret.

All other datasets can be made available upon request except those where legal restrictions exist. These are datasets that are covered by an exclusive right under a patent, a protected design, or specifically in Sweden, the Act (1992:1685) on the Protection of Circuit Designs for Semiconductor Products, as well as data that form computer programs.

Researchers in the project follow the file and folder naming conventions as described by Stanford Libraries "Best practices for file naming" guide (<https://library.stanford.edu/research/data-management-services/data-best-practices/best-practices-file-naming>). As they have to adapt to the folder structure provided by the IT support at the infrastructure where instruments are used (e.g. Max IV, Lund Nano Lab, Electronic Research Data Archive at UCPH (ERDA)) the naming conventions can take slightly varying forms.

2.3. Making data interoperable

What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?

Will your data include qualified references¹ to other data (e.g. other data from your project, or datasets from previous research)?

Datasets available in portable formats (pdf, csv) allow data exchange and reuse. Also, most of the image formats. Code shared in GitHub also allows exchange and re-use, although in the case of plug-ins, ownership of a Matlab license may be a precondition for being able to use the shared code. For archived code versions the environment and dependencies for running the code will be detailed in supporting files. A software quality checklist (Software Sustainability Institute) will be used to document the status and usability of the code and included in the README file.

The entire insect brain-database is web-browser based and no additional software is needed. All 3D neuron data will be stored online as swc files, a standard format for this purpose. Additionally, if raw data of neuron skeletons differ from swc files, we will provide those files for download. For online display, uploaded swc files are automatically translated into object-files that are readable by WebGL engines. Confocal image stacks will be both provided as jpg-series for quick viewing and browsing, but also in the full-resolution native format for download (tiff). To view those files, the open source ImageJ implementation FIJI is recommended. The search engine of the database is based on a graphical representation of the insect-brain that is founded in the insect-brain nomenclature that has been established by Ito et al. 2014 and is widely accepted and used across insect species. This common denominator will make it straight-forward to locate each dataset according to the brain region containing the relevant cell-type. As no such system for naming cell-types is in place, we have developed the NIN number system to identify each dataset. Commonly used names of each cell are linked as tags to each NIN, allowing to search for those names as well. For connectomics datasets, neuron skeleton/connectivity datasets that belong to a specific publication are deposited as interactive experiment entries in the insect brain database. Image

¹ A qualified reference is a cross-reference that explains its intent. For example, X is regulator of Y is a much more qualified reference than X is associated with Y, or X see also Y. The goal therefore is to create as many meaningful links as possible between (meta)data resources to enrich the contextual knowledge about the data. (Source: <https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/>)

data due to their large size can be made available for community based annotation via our online CATMAID server. User accounts granting rights to view and/or annotate volume EM image data will be generated upon request. Tools to interact with connectome based circuit models will be developed during the project and will be made available on the insect brain database.

For the other research areas in the project, thesauri that might cover parts of the subject scope of these, include IEEE and INSPEC thesauri for the physics aspects, and the ACM Computing Classification System for the computing aspects, as well as search keywords (compound names, empirical formula (Hill Notation), and CAS Numbers). However, due to the research being interdisciplinary and frontline, no controlled vocabulary can cover the whole project, therefore, subject keywords will be chosen freely by the researchers.

2.4. Increase data re-use

How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?

Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?

Will the data produced in the project be useable by third parties, in particular after the end of the project?

Will the provenance of the data be thoroughly documented using the appropriate standards?

Describe all relevant data quality assurance processes.

Further to the FAIR principles, DMPs should also address research outputs other than data, and should carefully consider aspects related to the allocation of resources, data security and ethical aspects.

Research line 1— insect neurophysiology: Data published in the insectbrain-database will be available for reuse after it has been published in a peer-reviewed journal, or upon the discretion of the data-owner. Data will be available for the lifetime of the database, which is aimed at a minimum of 10 years. All downloads by registered users will be logged in the database, so that re-use of data can be traced. As all online data is formally published, re-use is free of charge. A CC-BY-NC4.0 license is attached to each dataset, and a CC-BY4.0 is attached to all images produced via the visualization tools of the insect brain database (using deposited data). An API is in place to allow automated queries and download of data via third party software packages (e.g. in R, or python).

Each species in the database is managed by one or several Curators. Before a dataset becomes searchable in the database, it has to be approved by a Curator (ensuring consistency, compliance with database conventions, lack of objective errors).

The openly available research data from the other three research lines including computer code will be published under the following licenses:

- Research data: Open Data Commons Attribution License, or similar
- Computer code: open source license, e.g. Apache-2.0.

Data related to scientific publications will be made openly available at the time of publication. The usage of standard dataservers will ensure long-term storage and availability, with typical

guaranteed storage times exceeding 10 years after the deposition of the data. Where appropriate, data will be deposited on the Edinburgh DataShare facility which is a CoreTrustSeal repository for long-term retention and access to research data.

3. Other research outputs

In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.).

Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.

All semiconductor samples created within the InsectNeuroNano project will be for a specific set of experiments. Molecular dyes synthesized will be used up within the project. Special synthesis of dyes and semiconductor samples is possible but would deduct from the project resources. Thus, semiconductor samples and molecular dyes cannot be shared for free.

All insect neuroscience data will be publicly shared after peer reviewed publications have been generated. This first includes all data based on electrophysiology and all anatomical data resulting from light microscopy (all deposited on the insect brain database). All registered users of the insect brain database can download and reuse all data (with CC-BY-NC4.0 license). Additionally, the datasets generated by the connectomics studies will be made freely available to the community once published by granting access to our image- and database servers. As our central servers have managed user rights, we will grant read access to anyone interested, and read/write access to anyone with appropriate training. This process will increase the impact of the collected data, increase the quality of existing annotations by community based proof reading and allows for third parties to develop own lines of inquiry based on our datasets.

4. Allocation of resources

What will the costs be for making data or other research outputs FAIR in your project (e.g. direct and indirect costs related to storage, archiving, re-use, security, etc.) ?

How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)

Who will be responsible for data management in your project?

How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?

As a basic rule, each research group is responsible for the management of the data they produce. Also, they are supported by IT administrators who provide a system architecture that permits the safe storage of the produced data. For each instrument a manager has been appointed. The instrument manager is also responsible for data management on the instrument computer, where the raw data are stored.

Research line 1 - insect neurophysiology: The insectbrain-database is housed on a commercial

server operated by Amazon. The physical location of the server is Frankfurt, Germany. The database has been developed and programmed together with a local start-up company in Lund, Sweden (KevinTedoreInteractive), with whom a maintenance contract exists. Maintenance is expected to cost ca. 5000 EUR per year, and these costs include server fees, update of software to current data-formats, etc. Connectomics data are stored and maintained at LDC (Lund Data Center), at a cost of currently 800 EUR per month for all datasets produced and annotated. A fraction of this cost is associated with the data produced in this project (ca. 100 EUR per month).

Research line 2 - III-V semiconductor nanowire nanowire/nanopillar and waveguide interconnects technology: Data storage will be provided by Lund University via LUSEC. All data will also be stored on local computers for laboratory equipment. The LUSEC will thus serve as long term storage of the original data. The LUSEC comes at a cost for LUSEC data storage of 50SEK/month per TB. Further, for the servers of the laboratory equipment and a cold storage for long-term archiving the major cost is the salary of the IT administrators (covered by indirect costs). Data security rules of Lund University apply. Regarding the III-V nanopillar and waveguide interconnects data of INL, the original data will be stored at internal server backup folder own by INL and supported by INL IT administrator. Beyond the duration of the project the older data will be moved to a cold storage at INL for long-term archiving of the project data.

Data derived from research lines 3 and 4 (molecular dyes and electronic circuit technology), are originally stored on servers owned by University of Copenhagen (e.g. ERDA.ku.dk), thus major resources are salaries of IT personnel (indirect costs).

Relevant data generated at the University of Edinburgh will use the fully backed-up and secure DataStore service which is provided free for all research staff and postgraduate research students. Metadata for research outputs (including both publications and datasets) is maintained in the PURE research information system.

The libraries of the InsectNeuroNano partners are equipped for long term preservation of both, printed and electronic copies of publications. Digital archiving of the underlying data is regulated at departmental level.

Costs for personnel and infrastructure are partly covered at departmental level and partly centrally. The costs are included in the indirect costs.

After 10 years, the decision whether data are deemed to be useful is made by the research group leaders.

5. Data security

What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?

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

The partners individually ensure the secure and reliable storage of the data generated by them, in accordance with their respective internal ICT policies.

The hosting contracts of commercial providers (Amazon, Box, etc) include provisions for data security including back-up and recovery.

Data stored on Lund University servers are backup up daily and recovery of lost data for up to three months is possible. Likewise, UCPH's Electronic Research Data Archive (ERDA) is a secure central storage facility that can be accessed by all researchers in the project. The local departments have backup systems linked to key experimental infrastructure for short term protection.

Data that are not made publicly available but must be stored for 10 years are conserved in a cold storage.

6. Ethics

Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

Will informed consent for data sharing and long term preservation be included in questionnaires dealing with personal data?

The InsectNeuroNano project sets a high aim for the ethical awareness and full compliance with national, EU and international law, by systematic integration of ethical thinking into all processes. This includes the courses in research ethics that are available for all researchers.

In practice for data management, this means that raw data that were included in all research need to be archived and accessible to enable investigations for suspected research misconduct. Each partner will follow their own established routines for detection and handling of scientific misconduct.

No personal or protected data will be gathered in the research.

7. Other issues

Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?

In Sweden, the national law Arkivlag (SFS 2019:866) applies to all data produced within Lund University. The law is implemented at Lund University through the Plan for handling of documents (Dokumenthanteringsplan Version 1.12, in force since 09/05/2022). Departmental procedures have been developed to comply with this law.

Secondly, since the 1st of August 2022, the law SFS 2022:818 on making public sector data available applies. SFS 2022:818 transfers the "Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information" into national law and for Universities this specifically applies to research data.

All research data must be made available in the existing format and only upon request and if it is not disproportionate effort in a format that is open, machine-readable and, where appropriate, accessible and searchable, together with the associated metadata. There are a few exceptions, those that can be relevant here is data that is covered by an exclusive right under the Patent Act

(1967:837), the Design Protection Act (1970:485), the Act (1992:1685) on the Protection of Circuit Designs for Semiconductor Products, as well as data that form computer programs.

For UCPH, the Danish Code of Conduct for Research Integrity (2014) specifies the proper management of primary materials and data.

For the UK, the data management and sharing requirements are as specified by the UKRI, which requires that publicly funded research data should be made openly available with as few restrictions as possible in a timely and responsible manner, including best practice for data preservation and discoverability.