Canada Excellence Research Chair in Nano-Optical Biosensing and Molecular Diagnostics

A Data Management Plan created using DMP Assistant

Creator: Niko Hildebrandt

Principal Investigator: Niko Hildebrandt

Affiliation: McMaster University

Funder: NSERC

Template: McMaster - NSERC Pilot DMP

ORCID iD: 0000-0001-8767-9623

Project abstract:

A smart combination of nanotechnology, optical sensing, and biomolecular analysis into advanced hybrid systems has an enormous potential to establish the next generation of products and applications for clinical research and diagnostics as well as environmental and food sensing. Niko Hildebrandt is an internationally renowned expert in multiplexed biosensing and Förster resonance energy transfer (FRET) technologies using luminescent molecules and nanomaterials. As the Canada Excellence Research Chair (CERC) in Nano-Optical Biosensing and Molecular Diagnostics at McMaster University, he will build an ambitious research and technology centre to develop high-performance methods for biosensing and bioimaging. The interdisciplinary and translational research approach will range from fundamental studies and computational modelling of mechanisms related to luminescence, energy transfer, biomolecular interactions, and nanomaterials, over the development and engineering of advanced methods for versatile and multiplexed biosensing and bioimaging, to the application of energy transfer based sensors for in-vitro diagnostics, biomolecular analysis, food safety, and environmental sensing. Niko Hildebrandt and his multidisciplinary team will make McMaster University a world-leading player in the nanobio sciences, apply and commercialize break- through biosensing technologies, and translate and mobilize knowledge about luminescence and FRET biosensing in Canada and beyond. Through developing, understanding, characterising, and applying novel luminescence and energy transfer concepts, the CERC in Nano-Optical Biosensing and Molecular Diagnostics will break the current limits of sensitivity, selectivity, simplicity, and spatial and temporal resolution. The research results will significantly advance both the technology and the performance of state-of-the-art biological analysis, environmental sensing, and disease diagnosis to promote excellent research for a technologically advanced Canada and for enhancing the health and wellness of Canadians.

Identifier: CERC-2022-00072

Start date: 01-09-2023

End date: 31-08-2031

Last modified: 19-03-2024

Grant number / **URL:** https://www.cerc.gc.ca/chairholders-titulaires/index-eng.aspx?filter=Hildebrandt

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Canada Excellence Research Chair in Nano-Optical Biosensing and Molecular Diagnostics

Data Collection

What data will you use in your research (types and file formats)? How it will be collected, created, linked to, acquired, and/or recorded? How will you structure, name, and version-control files and folders?

Data generated within the CERC program will consist of experimental data, such as absorption and luminescence spectra and intensities, fluorescence lifetimes, or microscopy images from cells, theroretical data, such as modeling of nanobiohybrid nanostructures and their properties, or information about the investigated samples (e.g., temperature, concentration, volume etc.). All data of this CERC project will be generated by project staff (from the CERC group and/or collaborators and partners) in experiments or modeling related to the research program. The data will be collected manually (and noted in lab books or note books) or automatically (by the equipment). Most of the collected data will be analyzed and processed by the project staff. Data formats will be variable and can be handwritten text and schemes, photos and images, videos and sound recordings, data tables, equations, codes, graphs, or diagrams. The size of the data can vary from relatively small (10 kB) ASCII files or Excel spreadsheets to very large (500 MB) datasets, Origin Pro files, Powerpoint presentations, or microscopy image files. As far as possible, all data will be saved in accessible formats (e.g., ASCII).

There will be a dedicated CERC samples-and-experiments database to which all data will be integrated. If existing data sets will be used, they will be implemented in the CERC data base or an appropriate link to those data sets will be provided in the CERC database. Whereever possible, data formats that offer broad usability (e.g., txt, ASCII etc.) will be used.

All files will start with the date the data were acquired (YYYYMMDD) followed by the labbook code and page (e.g., NH2024-01-P17) and a very brief data description (e.g., Immunoassays). Additional files related to the data (e.g., with description and analysis) can be added using the same format and adding the additional description (e.g., rawdata, info, analysis01). Example for an Origin Pro file: 20241121_NH2024-02-P59_QDfluorescence_Spectra01.opju

To assure findabable, accessible, interoperable, reusable data, folders will be systematically structured as follows:

- username
---- subproject
-------- type of experiment
---------- info data (e.g., readme.txt)
---------- raw data (e.g., ASCII file)
---------- processed data (e.g., Origin file)
---------- analyzed data (e.g., pptx file with graphs, schemes, and tables)
-------- report data (e.g., presentation, manuscript draft, report draft)
- finalization
---- YYYY
-------- reports (e.g., word files)
-------- papers (e.g., word files)
-------- presentations (e.g., pptx files)
- published
---- YYYY
-------- reports (e.g., word and pdf files)
-------- papers (e.g., word and pdf files)
-------- presentations (e.g., pptx and pdf files)

Documentation and Metadata

How will data be documented to read and interpret it in the future? How will you make sure you update documentation? If you have any discipline-specific standards or variables, include it here.

All file names will include a labbook code and page (e.g., NH2024-01-P17), such that the full description of the experiments can be found. In addition, the folder structure will contain an "info data" subfolder in which readme.txt files that contain information to put the data into context are stored. The readme.txt file will contain information regarding the dataset, including an overview of data files, methodological information that is not available in the individual data files, and information regarding any potential sharing or access restrictions.

Updates of documentation will be ensured by mentioning the dates of the updates in the readme.txt files (txt data format will be easiest to read by the most common systems). All data will be stored in CERC-program-specific folders on OneDrive, such that files can be optimally shared and synchronized.

Storage and Backup

How will data be stored during the research project? How much space do you need (in MB, GB, TB) and for how long? Where and how will it be backed up? How will team members access and contribute data?

All data will be stored and shared on a OneDrive folder of 1 TB, in which data is automatically encrypted. A new folder will be created every year. Backups will be created on MacDrive, a McMaster hosted cloud storage platform. In addition, every user will keep password-protected backups of their data on their computer-internal hard drives or external hard drives.

Considering that the research group is growing, it is difficult to estimate how much data will be created. Also, whereas spectroscopy data is usually not very large (ASCII files of up to 1 MB), time-resolved fluorescence curves can contain many data points (several MB). Microscopy images can be significantly larger (hundreds of MB).

Preservation

Where will data be deposited for long-term preservation and access at the end of your research project? Indicate what preservation-friendly file formats you will use and what supporting documentation you will include.

All data will reside on the in-house servers until the servers are decommissioned.

Data will be exported in non-proprietary formats, such as simple ASCII text files that contain all relevant metadata and sample information.

Datasets will be preserved by publishing them on McMaster Dataverse, which is a research data repository for McMaster faculty, students, and staff. Researchers can choose to make content available publicly, to specific individuals, or to keep it private. The Research Data Management Services team is available to assist in preparing data for sharing and preservation

Sharing and Reuse

Describe whether and how the data will be shared, and its potential to be reused. What will you share and in what form? What end-user license will you include? How will you the research community know your data exists?

The research community will know that the data exists because of published articles. In these articles and the supporting information, the analysis of the data will be available. After publication, all related data (and the articles and supporting information in non-journal-formatted versions) will be published on the McMaster Dataverse data repository. The repository will also create DOIs that can be shared on the CERC-program and/or research group website. Creative commons data licenses will be used.

High risk and sensitive data will not be publicly shared. Such data can be patient information, ethically sensitive data of community partners, or data that is under consideration for patenting. Such data will be stored and only shared with specific users who are authorized for reviewing the data.

Responsibilities and Resources

What data management task roles and responsibilities (applicant and research team, as appropriate), succession planning, and resources will be required to implement the DMP?

All HQP of the CERC program will create data under the supervision of the PI. The HQP will also process and store the data and submit it to Dataverse. The CERC program manager and PI in collaboration with the Research Data Management Services team will manage the database and ensure that it is properly maintained and functional.

All HQP of the CERC program will be trained (via the McMaster Research Data Management Services) in accessing, contributing to, and sharing the database. Should the program mamager or the PI become unavailable, the research team will have full access to the data.

The CERC program manager (with help from the Research Data Management Services team) will be responsible for succession planning.

Ethics and Legal Compliance

What ethical, legal and commercial constraints are the data subject to? How will you manage, secure, and share data accordingly?

Sensitive data will be treated using the highest ethical standards and the guidelines of the McMaster Research Ethics Board. No data from human samples will be collected before the approval of relevant ethics committee has been obtained.

Person-identifiable data will not leave the unit from which they originated, and keys to identification numbers will be held confidentially within the respective clinical units. In cases when blood or tissue samples have already been collected in cohort studies all patient information will already have been anonymised and samples will not be traceable back to individuals.

The McMaster Industry Liaison Office (MILO), which facilitates working with industry by negotiating all industry-sponsored research agreements and all research contracts performed by the university, whether from government, industry or other sources and works with researchers and industry to ensure that appropriate legal agreements are in place to enable the exchange of cells, other biological or chemical materials (MTAs) and confidential information and data (NDAs/DTAs), has extensive experience regarding legal and commercial constraints. All ownership, licensing, and intellectual property rights of data that may be subject to such constraints will be managed by MILO.