Data Management Plan

Project title:

Advanced characterization of nanoparticles and nanoclusters via capillary electrophoresis and related microtechniques

Czech Science Foundation

25-17563S

HISTORY		
Version	**Date**	**Changes**
0.1	01.04.2024	First draft
0.2	18.04.2024	Second draft, after the open science course at UPOL

According to the template available at:
https://doi.org/10.48813/sstg-4g21

1. Data Summary / Shrnutí dat

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.

Použijete znovu některá stávající data? K čemu je znovu použijete? Uveďte důvody, pokud bylo opakované použití stávajících dat zvažováno, ale zamítnuto.

The project will generate new data. We did not suppose reusing of any existing data.

What types and formats of data will the project generate or re-use?
Jaké typy a formáty dat budou v rámci projektu generovány nebo znovu použity?

Data from chemical experiments will be generated including textual data (.txt, .rtf, .docx, .pdf), spreadsheets (.csv, .xlsx), statistical data (.R), and images (.jpeg, .tiff). Also experimental data (.D, .M, .LOG, .S) will be generated via Agilent instrumentation. These data are readable only by an appropriate software; OpenLabs CDS ChemStation or 3DCE ChemStation is necessary for Agilent-based data. The OpenLabs software is available only with license or 60 days free access (1).

(1) https://www.agilent.com/en/product/software-informatics/analytical-software-suite/chromatography-data-systems/openlab-cds

---

What is the purpose of the data generation or re-use and its relation to the objectives of the project?
Za jakým účelem se budou data generovat nebo opětovně používat ve vztahu k cílům projektu?

The data will be generated as the results of experiments on the analytical instrumentation, especially capillary electrophoresis, and mass spectrometry. These data will be used to confirm (or not) hypotheses given in the project.

What is the expected size of the data that you intend to generate or re-use?
Jaká je předpokládaná velikost dat, které hodláte generovat nebo znovu použít?

We expect the size of the data will less than 100 GB/year.

What is the origin/provenance of the data, either generated or re-used?
Jaký je původ dat, ať už vytvořených nebo znovu použitých?

The origin of the data is from the NPs based experiments on analytical instrumentations. Then the data will be analyzed, different types of parameters will be calculated, and graphs/figures/tables will be constructed. The types if original data are described in the previous parts.

To whom might your data be useful ('data utility'), outside your project?
Pro koho mohou být vaše data užitečná ("užitečnost dat") kromě vašeho projektu?

The data will be related to the project. We did not suppose the data itself will be useful for any other purpose. Maybe, some of the data/methodology, respectively, can be useful for other research projects dealing with the instrumentation used.

2.   FAIR data / FAIR data

Making data findable, including provisions for metadata / Zajištění dohledatelnosti dat (finability), včetně ustanovení pro metadata

Will data be identified by a persistent identifier?
Budou data opatřena perzistentním identifikátorem?

The data will be identified via the project acronym, researcher who produced the data and the date of the collection/acquisition.

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.
Budou poskytnuta bohatá metadata umožňující vyhledávání? Jaká metadata budou vytvořena? Jaké disciplinární /oborové nebo obecné standardy budou dodržovány? V případě, že ve vašem oboru neexistují metadatové standardy, uveďte, jaký typ metadat bude vytvořen a jakým způsobem.

We will generate experimental data from the analytical instrumentation. They will be catalogized based on the project WPs, instrumentation and methodologies used, and on the outputs (publications). The standards do not exist in the discipline, we will organize DM trainings to gain the knowledge and modify the DMP accordingly.

Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?
Budou v metadatech uvedena klíčová slova pro vyhledávání, aby se zlepšila možnost vyhledávání a následného potenciálního opakovaného použití?

Keywords will be not provided. The cataloging will be based on the WPs, instrumentation, methodologies, and publications which seems to be enough for the potential re-use.

Will metadata be offered in such a way that it can be harvested and indexed?
Budou metadata nabízena takovým způsobem, aby je bylo možné automaticky zpracovávat (harverstovat) a indexovat?

The data will be open; we will provide as much information as possible to allow any other re-use of the data.

Making data accessible / Zpřístupnění dat (accessibility)

Repository / Repozitáře (úložiště)

Will the data be deposited in a trusted repository?
Budou data uložena v důvěryhodném repozitáři?

All the data will be stored at the internal data backup system already developed at the Department of Analytical Chemistry, Palacký University Olomouc (GMP practices implemented; tracked access, three identical copies stored at geologically different location, etc.). Open repositories (Zenodo and the National Data Infrastructure (when starting, expected in 2025)) will be used for the long-term data storage and sharing.

Have you explored appropriate arrangements with the identified repository where your data will be deposited?
Prověřili jste vhodná řešení s daným repozitářem, kde budou vaše data uložena?

The internal data backup at the Department is already developed and used (it spread out from the GMP-based lab). The Zenodo was recommended by the Open science team at the university.

Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?
Zajišťuje repozitář, přidělení identifikátoru k datům? Umožní repozitář přiřadit identifikátor k digitálnímu objektu?

Yes, the repository allows to assign an identifier to the data. Moreover, the data are stored according to the preset structure, each instrument, operator, study (from method development through method validation to routine use) has own data path to tract all measurements/data generated within a project.

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.
Budou všechna data volně dostupná? Pokud určité soubory dat nelze sdílet (nebo je třeba je sdílet za podmínek omezeného přístupu), vysvětlete proč, a jasně oddělte právní a smluvní důvody pro záměrné

omezení. Mějte na paměti, že v projektech s více příjemci je také možné, aby si konkrétní příjemci ponechali data uzavřená, pokud je jejich zpřístupnění dat v rozporu s jejich oprávněnými zájmy nebo jinými omezeními podle grantové dohody.

The data will be open as most as possible. All the data will be stored at the internal backup with the limited access, however, open for the university authorities and appropriate users. The joint section will be also given for the project to share the information. Since, the basic research can generate also results that can be used for future applied research (leading to commercialization etc.), all the data will be not open for two years from their acquiring. The data which will lead to publication outputs, we will be moved to open repositories (Zenodo or National Data Infrastructure) and referred directly into the manuscripts. We will not work with any personal data within this project.

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.
Pokud je uplatňováno embargo, které má poskytnout čas na publikování nebo na ochranu duševního vlastnictví (např. patentů), uveďte, proč a jak dlouho bude časové embargo platit, s ohledem na to, že významná data by měla být zpřístupněna co nejdříve.

The embargo regarding the intellectual property was set for two years. This embargo is not valid for the data related to any publication within the frame of this project (there, the data used will be open with the submission). The setting of two years is given in the view of institutional policies (the same embargo can be applied for the theses, etc.). It should be enough to finish the experiments and protect the intellectual property.

Will the data be accessible through a free and standardized access protocol?
Budou data přístupná prostřednictvím bezplatného a standardizovaného přístupového protokolu?

The long-term data storage will be done via the open repositories (Zenodo and National Data Infrastructure) where the access is free by standardized protocols. The internal backup has the restricted access since the work is "life". However, the heads of the Dept./university have the possibility to access the data for the quality, validity, integrity, consistency, redundancy, completeness control. The joint section of the internal backup will be also established for the project to share the information within the whole team.

If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?
Pokud existuje omezení k užití dat, jak bude zajištěn přístup k datům? Jak bude nastacen přístup v průběhu projektu, a jak po jeho skončení?

The entrance to the internal backup is available only for the Department of Analytical Chemistry users with the university ID code and appropriate password. The parts of the backup are set to be accessible only for the defined IDs. This setup is already applied, and it is related to the employment (or the study) at the university. It is not related to the project duration. In the case of ending the employment (or the study), the access will be lost for the certain person. However, the organization heads (Dept./Faculty/University) have the access to the internal backup, so the data can be managed also in that case.

How will the identity of the person accessing the data be ascertained?
Jak bude ověřena totožnost osoby, která dostane k datům přístup?

The entrance to the internal backup is available only for the users with the university ID code and appropriate password (this is used for the whole information systems at the university). The entrance is recorded by the system.

Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?
Je potřeba zřídit komisi pro přístup k datům (např. pro hodnocení/schvalování žádostí o přístup k osobním/citlivým údajům)?

No, we will not deal with any personal or sensitive data.

Metadata

Will metadata be made openly available and licenced under a public domain dedication CC-0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?
Budou metadata otevřeně dostupná a licencovaná pod veřejnou licencí CC-0, jak je uvedeno v grantové dohodě? Pokud ne, objasněte, prosím, proč. Budou metadata obsahovat informace, které uživateli umožní přístup k datům?

All the data, including metadata, will be stored at the internal backup with the limited/controlled access, however, open for the university authorities and verified users. The data which will lead to publication outputs, we will be moved to open repositories (Zenodo or National Data Infrastructure) and referred directly into the manuscripts, as previously described.

How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?
Jak dlouho budou data dostupná a dohledatelná? Bude zaručeno, že metadata zůstanou dostupná i poté, co data už nebudou dostupná?

The long-term data storage, including metadata, will be done via the open repositories (Zenodo and National Data Infrastructure), as written above.

Will documentation or reference about any software be needed to access or read the data be included?
Bude zahrnuta dokumentace nebo odkaz na jakýkoliv software, který bude potřebný pro přístup k datům nebo jejich prohlížení?

Yes, we will cover the links for the software allowing to read the data. The Agilent-based data cannot be open in any different freely available software. There is 60-days free license from Agilent only. The funds

of the project do not allow us to buy a free software to open the data. However, the raw data used for the publications will be converted into commonly readable format (.csv) to keep them as open as possible. The raw data will be stored, too.

Will it be possible to include the relevant software (e.g. in open source code)?
Bude možné zahrnout příslušný software (např. v otevřeném zdrojovém kódu)?

No, we will share the links for the software.

Making data interoperable / Interoperabilita dat

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?
Jakými slovníky, standardy, formáty nebo metodikami dat a metadat se budete řídit, aby vaše data byla interoperabilní a umožňovala šíření a opakované použití dat v rámci oborů i mezi nimi? Budete dodržovat osvědčené postupy pro interoperabilitu, které jsou podporovány (vědeckou) komunitou? Které to budou?

We will follow the standards of the community as well as the standards used in the GMP labs. Since we are not working with personal and sensitive data, the practices are not so difficult. We will follow the recommendations for GMP-based analytical laboratories.

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?
V případě, že je nevyhnutelné používat neobvyklé ontologie nebo generovat ontologie a slovníky specifické pro projekt, poskytnete mapování běžněji používaným ontologiím? Budete otevřeně publikovat generované ontologie nebo slovníky, aby se dali znovu použít, upřesnit nebo rozšířit?

We will not use uncommon ontologies or vocabularies.

Will your data include qualified references to other data (e.g. other data from your project, or datasets from previous research)?
Budou vaše data obsahovat odkazy na jiná data (např. jiná data z vašeho projektu nebo soubory dat z předchozího výzkumu)?

Yes, we will include references to other data from this project. There will be not any references to datasets from previous research.

Increase data re-use / Zvyšení opakovaného použití dat

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.)?
Jak poskytnete dokumentaci potřebnou k ověření analýzy dat a usnadnění opětovného použití dat (např. soubory readme s informacemi o metodice, workflow, čištění dat, analýze dat, definicích proměnných, měrných jednotkách atd.)?

We will provide the documentation needed for validation of data including setups, methodology, codebooks, etc., as appropriate laboratory notebooks. The laboratory notebooks (.docx, .pdf) as well as the other necessary documentation will be stored in the backup, too. Then, all the information will be shared and open as the experimental data.

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?
Budou vaše data volně dostupná, aby bylo umožněno jejich co nejširší opakované použití? Budou vaše data licencována pomocí standardních licencí pro opětovné použití v souladu s povinnostmi stanovenými v grantové dohodě?

Yes, the data will be available as open as possible. Licenses will be given according to the university policies: https://openscience.upol.cz/.

Will the data produced in the project be useable by third parties, in particular after the end of the project?
Bude umožněno použití dat třetí straně během, a hlavně po skončení projektu?

Yes, the data will be available as open as possible. Licenses will be given according to the university policies: https://openscience.upol.cz/.

Will the provenance of the data be thoroughly documented using the appropriate standards?
Bude původ dat důkladně zdokumentován pomocí příslušných standardů?

The provenance of the data will be documented via the information from the instrumentation ("signatures" of the instruments in the data). The data will be linked to the information in the laboratory notebooks. All the data will be signed also by the project acronym and the operator. The date of the measurement is also automatically given from the instrumental systems.

Describe all relevant data quality assurance processes.
Popište všechny relevantní procesy zajištění kvality dat.

The data will be automatically backed up and checked for completeness by the internal backup software on daily basis. The documentation will be checked by the responsible person for the WPs, heads of the labs, and the project PI. The checks will be recorded to the appropriate documentation. The common

checks will be at least once a month. There will be a general check guided by the nominated supervisor from the open-data management team at least once a year (we will suppose once a half year at the beginning of the implementation of DPM and open policies).

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.

Kromě zásad FAIR by se DMP měly rovněž zabývat jinými výstupy výzkumu než daty a měly by pečlivě zvážit aspekty související s přidělováním zdrojů, bezpečností dat a etickými aspekty.

3. Other research outputs / Další výstupy výzkumu

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.).

Kromě správy dat by příjemci měli také zvážit a naplánovat správu dalších výstupů výzkumu, které mohou být generovány nebo znovu použity v rámci jejich projektů. Tyto výstupy mohou být digitální (např. software, pracovní toky, protokoly, modely atd.) Nebo fyzické (např. nové materiály, protilátky, činidla, vzorky atd.).

We will not generate other outputs than previously reported. Protocols will be covered into the laboratory notebooks. In the case of new materials (e.g., modified nanoparticles), the step-by-step protocols will be prepared, validated (for the reproduction of the processes by different people) and stored.

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.

Příjemci by měli zvážit, které z otázek týkajících se výše uvedených FAIR standardů se mohou vztahovat na řízení dalších výstupů výzkumu a měly by se snažit poskytnout dostatečné podrobnosti o tom, jak budou jejich výstupy z výzkumu řízeny a sdíleny nebo zpřístupněny k opětovnému použití v v souladu se zásadami FAIR.

We will not generate other outputs than previously reported.

4. Allocation of resources / Alokace zdrojů

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.) ?

Jaké budou náklady na vytvoření FAIR dat nebo jiných výstupů výzkumu ve vašem projektu (např. přímé

a nepřímé náklady související s úložištěm, archivací, opětovným použitím, zabezpečením atd.)?

The costs related to the data management is related mainly to the data storage hardware (approx. 20 kCZK/year). The other costs (security, archiving) will be covered by the Dept., the personal costs related to the data management will be covered partially by the project. The costs for personal trainings will be covered by the project indirect costs.

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)
Jak budou tyto náklady pokryty? Upozorňujeme, že náklady související se správou dat o výzkumu/výstupech jsou způsobilé jako součást grantu Horizon Europe (pokud jsou v souladu s podmínkami grantové dohody)

The costs related to the data management is related mainly to the data storage hardware (approx. 20 kCZK/year). The other costs (security, archiving) will be covered by the Dept., the personal costs related to the data management will be covered partially by the project. The costs for personal trainings will be covered by the project indirect costs.

Who will be responsible for data management in your project?
Kdo bude zodpovědný za správu dat ve vašem projektu?

The PI (Dr. Jan Petr) will be responsible for the data management in the project. The data management will be partially discussed also by Dr. Tomáš Pluháček, who is one of the heads of the internal backup.

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)?
Jak bude zajištěno dlouhodobé uchování? Diskutujte o nezbytných zdrojích, jak toho dosáhnout (náklady a potenciální hodnota, kdo a jak rozhoduje, jaká data budou uchovávána a jak dlouho)?

The long-term data preservation will be done in the open platforms (Zenodo or National Data Infrastructure). Here, the costs will be covered by the institution (Palacký University Olomouc) who joined the open policies according to the national as well as European legislation. The other future costs are unpredictable because they depend on the management of the university.

5.   Data security / Bezpečnost dat

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)?
Jaká opatření jsou nebo budou zavedena pro zabezpečení dat (včetně obnovy dat, zabezpečeného úložiště/archivace a přenosu citlivých dat)?

The entrance to the internal backup is available only for the authorized users with the university ID code and appropriate password. The data are stored at two servers in two separate buildings (in approx. 3 km distance). The internal backup meets the GMP criteria.

Will the data be safely stored in trusted repositories for long term preservation and curation?
Budou data bezpečně uložena v důvěryhodných úložištích pro dlouhodobé uchování a udržování?

Open repositories (Zenodo and the National Data Infrastructure) will be used for the long-term data storage and sharing. Here, those platforms deal with preservation and curation of the repositories.

6. Ethics / Etika

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).
Existují nebo mohly by existovat nějaké etické nebo právní překážky, které mohou mít dopad na sdílení dat? Lze o nich také diskutovat v kontextu etického přezkumu. Je -li to relevantní, zahrňte do popisu akce Descriptio of Action (DoA) odkazy na etické výstupy a kapitolu etiky.

No, there is not any ethics or legal issue with the data generated within the frame of this project.

Will informed consent for data sharing and long term preservation be included in questionnaires dealing with personal data?
Bude informovaný souhlas se sdílením dat a dlouhodobým uchováváním zahrnut do dotazníků zabývajících se osobními údaji?

No, we will not deal with any personal or sensitive data.

7. Other issues / Ostatní

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)?
Využíváte, nebo budete využívat jiné národní/grantové/sektorové/institucionální postupy pro správu dat? Pokud ano, jaké (uveďte prosím jejich seznam a stručně je popište).

We will use Departmental (internal) backup of the data. Then, we will use the procedures defined by the university at: https://openscience.upol.cz/en/.