CONTRACT N°: 101130125
ACRONYM: FLUFET
TITLE: FLow detection of virUses by graphene Field Effect Transistor microarrays

Instrument: HORIZON-EIC-2023-PATHFINDEROPEN-01

DATA MANAGEMENT PLAN

Project Starting Date: 1 March 2024
Project End Date: 31 August 2027
Duration: 42 months

PROJECT CO-ORDINATOR: UDC
Coordinator's Organization Name: Universidade da Coruña

REPORTING PERIOD: **n.1** from 1 to 12

Reference Work packages:
WP6 Project Coordination & Management

Project funded by the European Commission within HORIZON-EIC-2023-PATHFINDEROPEN-01		
Dissemination Level		
SE	Sensitive	X

Project Officer: LUCIA FARINA

DATA MANAGEMENT PLAN

PROJECT	
Project number:	[101130125]
Project acronym:	[FLUFET]
Project name:	[FLow detection of virUses by graphene Field Effect Transistor microarrays]

DATA MANAGEMENT PLAN	
Date:	[24/04/2024]
Version:	[DMP Version 1.0]

1. Data Summary

The FLUFET project main objectives are the detection of viruses, in particular the ones related to zoonosis in farms, using graphene field effect transistor (g-FET) and their functionalisation to make them more specific. The project will amass various data sets including protocols and measurements concerning graphene's characterization and functionalization, electrochemical readings, information on human receptors, viruses, and relevant proteins for detection purposes, experimental data for simulations and computational models, as well as designs for sensors and microfluidics. Additionally, certain data not initially considered may prove beneficial for achieving specific project milestones, hence it's advised not to disregard any data obtained during the project.

The data to be generated by FLUFET project collaborators will primarily consist of text, numerical formats and plots. These data will stem from measurements of the chemical, structural, and electronic attributes of materials and devices. Multimedia files like photos, videos, and visualizations may also be produced within the project. Furthermore, data used for modelling and simulating various physical or chemical processes in materials and devices can be generated, alongside data aimed at estimating broader trends, social perceptions, and economic aspects related to virus detection and gFET sensors.

Data analysis software such as Origin or Excel can be employed for storing and analysing the generated data, hence it may be stored in formats compatible with the chosen software. All data produced within the project can be reused if deemed necessary and appropriate.

In this data management plan, we outline strategies to ensure that our data adheres to the FAIR principles: Findable, Accessible, Interoperable, and Reusable. By following these principles, we aim to maximize the value and impact of our research data.
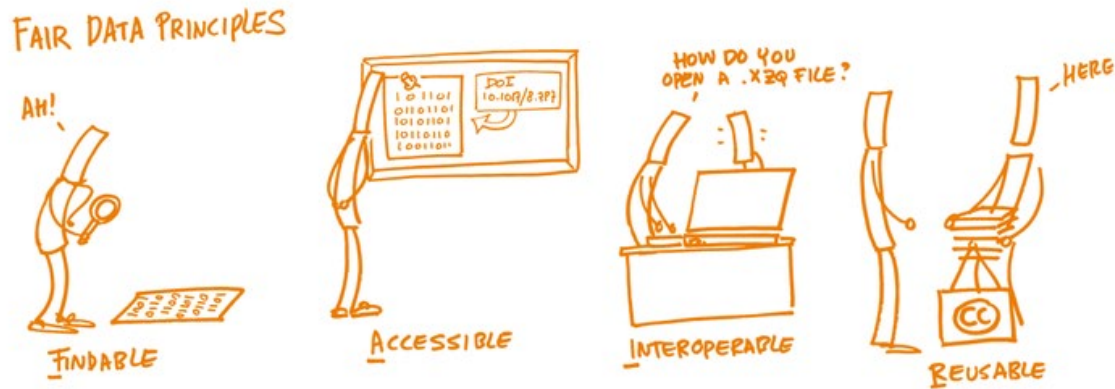
Figure 1 Image by Patrick Hochstenbach, CC0 1.0

2.   FAIR data

2.1.   Making data findable, including provisions for metadata

All non-confidential research data produced within the FLUFET project and disseminated in an open-access format will be assigned a persistent identifier, such as a digital object identifier (DOI) or Uniform Resource Identifier (URI), accompanied by descriptive metadata. These identifiers will ensure that our data remains discoverable even if its location or metadata changes.

In order to improve the accuracy of categorization and increase the visibility of both generated and published data, we will develop detailed metadata for each dataset. This metadata will consist of descriptions, keywords, and pertinent contextual details. To ensure consistency and enhance search capabilities, we will adhere to standardized schemas agreed upon by the community, such as Dublin Core. Typically, the metadata provided will include:

- Information about the author(s)
- Title of the published results
- Publication date
- Journal or platform name where the data can be accessed
- Volume and issue details (if applicable)
- Page numbers (if applicable)
- Keywords

However, the specific content of the metadata and adherence to metadata standards will be determined by the publishing platform or journal hosting the data.

In order to increase the chances of discovering data and to aid in the effective dissemination of results, keywords will be integrated into the metadata. Employing a controlled vocabulary, we will carefully choose descriptive keywords that precisely depict the content and breadth of our data. This approach will streamline search and discovery processes, allowing users to locate pertinent datasets using widely recognized terminology.

Indexing:

For indexing purposes, we will leverage both standardized indexing systems provided by data repositories and search engines, as well as custom indexing solutions tailored to the specific needs of our datasets. This will involve tagging datasets with relevant metadata, assigning appropriate keywords, and optimizing search engine visibility through techniques such as search engine optimization (SEO) strategies.

By implementing the strategies outlined in this data management plan, we are committed to fostering FAIR data practices that promote transparency, reproducibility, and collaboration in research. Through findable, accessible, interoperable, and reusable data, we aim to maximize the impact of our research and contribute to the advancement of knowledge in our field.

2.2. Making data accessible

To ensure accessibility, whenever possible, we will make our data openly accessible to the public without restrictions, either through data repositories or institutional platforms.

The non-confidential published research data of the FLUFET project will be deposited in a trusted and openly accessible repository such as Zenodo. In this scenario, a DOI will be assigned to the published data according to the Horizon Europe Funding scheme regulations. Each partner involved in the FLUFET project is tasked with uploading the data to the chosen repository and to link the data DOI to the persistent link of FLUFET project on the Horizon Europe website.

Partners will utilize the independent and openly accessible platform Zenodo, or a comparable open data repository offered by their institution, university, or company, where pertinent metadata will also be provided, along with a DOI and a link to the FLUFET project's persistent URL.

To that end we have created an open community on Zenodo: FLUFET EIC Pathfinder Community (zenodo.org).

For the data generated in the computational chemistry and simulation tasks, GitHub will be used. GitHub is a web-based platform used for version control and collaborative software development. It leverages Git, a distributed version control system, to help developers track changes in their code, collaborate with others, and manage various versions of a project. GitHub provides features like repositories, pull requests, issues, and continuous integration, making it easier to work on projects individually or in teams. It supports both public and private repositories and integrates with various development tools, enhancing productivity and facilitating seamless collaboration.

All non-confidential generated data published by FLUFET partners will be openly available under Creative Commons licences CC0, CC-BY or CC-BY-SA, in line with the "Open Definition" for content and data established by "The Open Knowledge Foundation".

In instance where the data is deemed highly promising by one or more partners of FLUFET and requires intellectual property protection, while a patent is being filed, an embargo period of 6 months can be applied. However, no embargo will be applied for non-confidential data.

The data is expected to be accessible via a free and standardized access protocol. As all non-confidential published data will be publicly accessible through the mentioned repositories, there is no requirement for a data access committee to review or approve data access requests.

Metadata will also be openly available under Creative Commons licenses as described earlier, in accordance with the GA, and will remain accessible and discoverable as long as the repositories housing the data remain accessible. In addition to metadata, we will provide detailed documentation, including data dictionaries, user guides, and README files, to aid users in understanding and utilizing the data effectively.

In cases where accessing specific data extensions requires particular software, details regarding the required software tools will be provided alongside the data.

2.3.   Making data interoperable

To promote interoperability, data and metadata vocabularies, standards, formats and methodologies will follow the general practices within the scientific community. We will use common data formats and standards within our field to ensure compatibility with existing tools and workflows. This consistency with other data generated outside of the FLUFET project is expected to make the data interoperable and flexible for exchange across disciplines.

In case it becomes unavoidable or inefficient to avoid the use of uncommon or only project specific ontologies and vocabularies, mappings and references to the more commonly used ontologies will be provided. Such uncommon ontologies will be published together with the data to allow their reuse, refinement and possible extension.

If deemed relevant, the published data could also include qualified references to other data generated within FLUFET project or already published data from previous research. Where applicable, we will adopt linked data principles to connect our datasets with related resources, enabling seamless integration and analysis across diverse datasets.

2.4.   Increase data re-use

To enhance the reusability of our data, we will provide clear terms of use through appropriate licensing, such as Creative Commons licenses, enabling others to understand how they can legally reuse and redistribute the data. Additionally, we will implement version control mechanisms to monitor changes and updates to our datasets over time, ensuring transparency and reproducibility.

Guidelines and documentation for data validation analysis will be included alongside the published data to further enhance its reusability. These materials may consist of readme files detailing methodology, codebooks, data cleaning procedures, analyses conducted, variable definitions, units of measurement, and other pertinent information. Whenever possible, data files will include header information specifying the type and units of the contained data.

Once published, the generated data will be freely accessible under Creative Commons licensing agreements, as previously described, in compliance with standard reuse licenses outlined in the General Agreement (GA). Proper attribution and citation of our data will be encouraged through citation metadata and guidelines, enabling accurate acknowledgment and referencing of our work.

Prior to publication, internal reviewers within the FLUFET consortium and external reviewers outside of the consortium will scrutinize the data for consistency and quality assurance.

3. Other research outputs

If other research outputs are generated within FLUFET project, partners will evaluate sharing these results in accordance with the FAIR principles outlined above. Such research outputs outside of the scope of generated or re-used data could be either digital outputs (e.g., protocols, models, sensorics concepts) or physical (e.g., new materials, measurement devices) output type.

In cases where the aforementioned principles may not directly apply to these research outputs, supplementary procedures for their data management will be devised and presented, ensuring consistency with the FAIR principles.

4. Allocation of resources

FLUFET uses standard tools and a free of charge research data repository. The costs of data management activities are limited to project management costs and will be covered by allocated resources in the project budget.

Long-term preservation of the public data is ensured through Zenodo. Other resources needed to support reuse of data after the project ends will be solved on a case-by-case basis.

The overall responsibility for data management lies with the project coordinator, the UDC team, supported by members of each partner, with experience in Horizon Europe projects.

5. Data security

All data generated, reused, and other research outcomes from the project will be securely stored on the internal data servers of FLUFET partners. Access will be restricted to ensure reliable data security, and backup strategies will be implemented to prevent data loss in accordance with the institution's internal IT policies. Utilizing SharePoint/OneDrive, hosted on UDC's servers, robust data security measures will be established to facilitate secure storage, recovery, archiving, and transfer of data among project partners.

Furthermore, data will be securely archived in trusted repositories belonging to each partner within the FLUFET consortium for long-term preservation and curation, wherever such repositories are available.

6. Ethics

There are no anticipated ethical or legal impediments that would affect the sharing of non-confidential data. While the research conducted within the FLUFET project is not foreseen to pose ethical dilemmas, partners will diligently monitor this aspect.

7. Other issues

Initially the use of other national/funder/sectoral/departmental procedures for data management is not envisaged, but with the rise of next-generation funds for modernization we cannot rule out applying for funds to improve data management in the course of the project.