Data Management Plan in FWO

1. General Information

- **Name applicant:** Jan-Pieter D'Anvers

- **FWO Project Number & Title:** 1238822N - Practical application of post-quantum cryptography

- **Affiliation:** imec-COSIC KU Leuven

2. Data description

- **Will you generate/collect new data and/or make use of existing data?**

Generate new data/ use existing data

- **Describe the origin, type and format of the data (per dataset) and its (estimated) volume, ideally per objective or WP of the project.**

The data types generated are the same for each work package.

Type of data	Format	Volume
Scientific articles (open access)	Latex, pdf	Up to 1GB
Computer Code, will be made available as open source	in several programming languages (C, Python, …)	Up to 1 GB
Measurements results (for published articles)	Plaintext logs, JSON, CSV, PCAP…	10 GB
Project deliverables	Latex, MS Word format, PDF	5-10GB
Presentations	Latex, PPT, PDF	1-10 GB

3. Legal & ethical issues

- **Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to the file in KU Leuven's Record of Processing Activities. Be aware that registering the fact that you process personal data is a legal obligation.**

No, in my research project, I will not collect or observe personal data.

We will make use of

- randomly generated synthetic data that has certain characteristics, but it is not linked to a natural person.

- Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, add the reference to the formal approval by the relevant ethical review committee(s)

- Yes

No ethical issues regarding experiments on humans or animals, no issues regarding human rights are in place.

Regarding dual use and misuse; whenever applicable I will submit an application to the KU Leuven Ethics Committee on Dual Use, Military use & Misuse of Research (EC DMM). The EC DMM will evaluate my research project proposal and will advise on this matter.

- Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?

- Possibly

The outcome of my research project will consist of algorithms, protocols and scientific research articles which improve the current state of the art. The final results of the project will be made available via publications according to academic standards and will be available in Open Access repositories. It is possible that some developed algorithms will be patented, which might incur a waiting period in making the results available.

- Do existing 3rd party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?

- No, there are no restrictions in place.

4. Documentation & metadata

- What documentation will be provided to enable reuse of the data collected/generated in this project?

The algorithms and protocols are self-explanatory. In the computer code comments will be integrated to improve readability. The source code will be accompanied by README files that explain the structure, the compilation order and the overall architecture.

Our research group has a detailed roadmap on how to store research data so that it can be easily re-used. The roadmap is described in detail below.

- Will a metadata standard be used? If so, describe in detail which standard will be used. If no, state in detail which metadata will be created to make the data easy/easier to find and reuse.

- No, there will be no metadata standard used.

Our research group has a detailed roadmap on how to store research data.

Data used to produce project results will be stored in a gitlab repository with access control. The gitlab repository is maintained by a DMP-officer who keeps the structure in the repository in place and manages the access control. Research data will be structured and stored per publication in a separate directory. Every directory on this gitlab repository has the same structure, so that data can be easily traced and re-used.

For manuscripts/articles: all data, sources and plotting scripts that are needed to produce the PDF are included in the directory of the publication concerned.

For software and hardware: all the necessary scripts and code are included in the directory of the publication concerned.

A selection of the relevant research data is pushed to a publicly available github repository regularly. This happens after publication of project papers. All publications are available from open access repositories.

5. Data storage & back up during the FWO project

- Where will the data be stored?

As explained above, data used to produce project results will be stored in a gitlab repository with access control. The gitlab repository is maintained by a DMP-officer who keeps the structure in the repository in place and manages the access control. In a later phase, and after the publication of the research paper concerned, the relevant data will be made available in a publicly available github repository. The gitlab repository is hosted by a server at our research department.

- How is back up of the data provided?

We make sure to replicate the data sets among at least two department computers.

The gitlab repository is hosted by a server at our research department and gets a back-up daily.

GitHub is a well-established, commercial repository service provider, which we assume to have industry-standard data availability measures in place (such that repositories will not be lost unless deleted by authorized GitHub users).

- Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely. If no or insufficient storage or backup capacities are available then explain how this will be taken care of.

- Yes.

There is sufficient storage & backup capacity during the project. We have many Terabytes of storage space available on the department machines we use to replicate our data sets. The department's Gitlab instance and GitHub are offered services and are thus ensured to have enough spare capacity.

- What are the expected costs for data storage and back up during the project? How will these costs be covered?

The costs of the gitlab maintenance are covered by our research group. These costs are integrated in the general IT-costs. The use of github is free of charge.

- Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

Access to our department machines is managed by access to the university network and standard-practice user permissions on the machine (only users with the right set of permissions can access the data, we control the management of these users).

The gitlab repository has access control and is maintained by a DMP-officer who keeps the structure in the repository in place and manages the access control. Unauthorized GitHub modifications are prevented through their industry-standard user management.

6. Data preservation after the FWO project

- Which data will be retained for the expected 5 year period after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues, ...).

All the research data will be stored on the gitlab repository for at least 10 years after the end of the project. A selection of the research data used to produce project results will be pushed to github regularly. This will happen after the publication of a project paper. The github will remain active for at least 10 years after the end of the project as well.

- Where will the data be archived (= stored for the longer term)?

To the extent feasible in terms of the volume of collected data, the data will be stored in the Gitlab and GitHub repositories. If we consider it necessary to purchase archival storage capacity for the entirety of the research project's artifacts (e.g., due to nonreproducibility of some data sets and limitations on the Gitlab and GitHub services), we will do so appropriately and in time before the project concludes. Options include the storage services offered by KU Leuven or dedicated storage hardware to be kept long-term at our department.

- What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?

The costs of the gitlab maintenance are covered by our research group. These costs are integrated in the general IT-costs of our group. The github is free of charge.

7. Data sharing and reuse

   - Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

- No

There are no restrictions in place. the data is randomly generated. There are no confidentiality concerns related to this data.

   - Which data will be made available after the end of the project?

Relevant data used to produce project results will be made available. Relevant data can consist of software or hardware scripts and code, algorithms, protocols, manuscripts, figures or others. Publications will be made available in pdf.

   - Where/how will the data be made available for reuse?

- Other: specify

To the extent feasible due to volume of the collected data sets, data used for published articles will be stored in Gitlab repositories with access control and publicly accessible GitHub repositories. Publications will be made available in open access repositories. All source code, protocols, and algorithms used for published articles will be made available as open-source software on GitHub.

   - When will the data be made available?

- Upon publication of the research results

To the extent feasible due to volume of the collected data sets, we will make our data sets publicly available on GitHub upon publication of the corresponding article. In case a data set can be reproduced, we will also publish clear instructions/scripts for how to do so (so that a large data set may also be recreated instead of downloaded).

   - Who will be able to access the data and under what conditions?

All the research data used in the project will remain available in the Gitlab with access control. The DMP-officer is in charge of the access control. A selection of the relevant data used to produce project results will be made available on the github after the end of the project and is publicly available.

   - What are the expected costs for data sharing? How will the costs be covered?

No costs are to be expected.

8. Responsibilities

- Who will be responsible for data documentation & metadata?

The PhD student/Postdoctoral researcher, Jan-Pieter D'Anvers, funded by this grant, is responsible for data documentation and metadata.

- Who will be responsible for data storage & back up during the project?

In the first place, the PhD student /Postdoctoral researcher, Jan-Pieter D'Anvers, funded by this grant, is responsible for data storage and backup. The DMP-officer maintaining the Gitlab repositories will share in the responsibility for data storage and backup.

- Who will be responsible for ensuring data preservation and reuse ?

In the first place, the PhD student /Postdoctoral researcher, Jan-Pieter D'Anvers, funded by this grant, is responsible for data preservation and reuse.

- Who bears the end responsibility for updating & implementing this DMP?

he PhD student /Postdoctoral researcher, Jan-Pieter D'Anvers, funded by this grant, is responsible for updating and implementing this DMP.