# CELSA project - MaCro: Mathematics for post-quantum cryptanalysis

*A Data Management Plan created using DMPonline.be*

**Creator:** Frederik Vercauteren

**Affiliation:** KU Leuven (KUL)

**Funder:** KU Leuven (KUL)

**Template:** KU Leuven BOF-IOF

**Grant number / URL:** CELSA/23/025

**ID:** 203324

**Start date:** 01-10-2023

**End date:** 30-09-2025

**Project abstract:**

Public-key cryptography is the enabling technology for secure connections over the internet: it is used countless times per second and trillion-dollar industries such as e-commerce crucially rely on it. Nearly all currently deployed public-key systems build on the hardness of two computational problems: factoring large integers (RSA) and computing discrete logarithms (ECC). Although these systems are secure against attacks by classical computers, in 1994 Shor showed that these can be broken in polynomial time using a large-scale universal quantum computer.

Realizing such large-scale quantum computer continues to pose a stubborn engineering problem, but cryptographers are taking the threat very seriously. In 2016, the National Institute for Standards and Technology (NIST) announced an effort to standardize quantum-resistant replacements of RSA and ECC. After years of cryptanalysis, NIST announced a shortlist of candidates which they deemed secure. The problem however is that none of these candidates are provably hard, so we must rely on the failure of many experts having spent sufficient effort on trying to break them. Not spending sufficient effort on cryptanalysis is extremely dangerous and the risk of over-rushing the standardization process manifested itself in 2022: totally devastating classical, i.e. non-quantum, attacks were found against Rainbow and SIKE, both on the final shortlist of NIST.

The shortage of time and effort can only be countered by putting more emphasis on cryptanalytic efforts and on existing mathematical expertise. Moreover, the gap between cryptographers and mathematicians seems to be growing: whereas most mathematicians are well-aware of RSA and ECC, the situation is very different for the new post-quantum candidates.

The first goal of this project is to study complex mathematical structures related to post-quantum cryptography with the aim of cryptanalyzing post-quantum systems. The second goal is to bridge the gap between the community of mathematicians and cryptographers. In order to ensure that these newer systems are secure we have to take an interdisciplinary approach and bring together different communities with complementing expertise.

**Last modified:** 05-12-2023

# CELSA project - MaCro: Mathematics for post-quantum cryptanalysis

## Research Data Summary

**List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.**

| Dataset name / ID | Description | New or reuse | Digital or Physical data | Data Type | File format | Data volume | Physical volume |
|---|---|---|---|---|---|---|---|
| | | *Indicate:*<br>***N****(ew data) or*<br>***E****(xisting data)* | Indicate: **D**(igital) or **P**(hysical) | Indicate:<br>**A**udiovisual<br>**I**mages<br>**S**ound<br>**N**umerical<br>**T**extual<br>**M**odel<br>**SO**ftware<br>Other<br>(specify) | | Indicate:<br><1GB<br><100GB<br><1TB<br><5TB<br>>5TB<br>NA | |
| Code | Various programs will be implemented to execute the cryptanalytical attacks developed during the project | N | D | SO | text files | <1GB | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |

**If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:**

During the project we may re-use existing software libraries, typically Magma or Sage scripts, in accordance with the licenses under which they are published.

**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, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.**

- No

**Will you process personal data? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).**

- No

**Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.**

- No

**Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material or Data transfer agreements, Research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.**

- No

**Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.**

- Yes

The software libraries that this code would use may be published under certain restrictive licenses. Any software in this project that builds upon these libraries will be therefore also be subject to licenses that take into account the restrictions of the licenses therein.

## Documentation and Metadata

**Clearly describe what approach will be followed to capture the accompanying information necessary to keepdata understandable and usable, for yourself and others, now and in the future (e.g. in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, codebook.tsv etc. where this information is recorded).**

Any software code or script produced during this project will be documented in-file, according to the practices of the relevant language. Also, readme files will be included alongside, describing how to compile and/or operate the software and how to interpret the output produced.

**Will a metadata standard be used to make it easier tofind and reuse the data?**
**If so, please specify which metadata standard will be used.**

**If not, please specify which metadata will be created to make the data easier to find and reuse.**

- No

Any software code made publicly available will be stored on the github.com website, which is an industry reference for open-source software, on the COSIC research group's github.com page (https://github.com/KULeuven-COSIC). It will be named according to the research publications that refer to it, and links to the source code's page will also be included in such publications.

## Data Storage & Back-up during the Research Project

**Where will the data be stored?**

- Other (specify below)

The publicly available software code and scripts will be primarily stored on the COSIC GitHub repository (https://github.com/KULeuven-COSIC). Code that is not publicly available will be stored on the COSIC research group's GitLab infrastructure, which is private and also maintained by the ESAT department of the KU Leuven.

**How will the data be backed up?**

- Other (specify below)

The publicly accessible COSIC GitHub repository (https://github.com/KULeuven-COSIC) is backed up automatically by GitHub. The COSIC research group's GitLab infrastructure, is private and also maintained by the ICTS services of the ESAT department of the KU Leuven.

**Is there currently sufficient storage & backup capacity during the project?**

**If no or insufficient storage or backup capacities are available, explain how this will be taken care of.**

- Yes

**How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?**

During the course of the project, software code and scripts will be protected by access rights maintained by the ESAT network and within the GitLab.

**What are the expected costs for data storage and backup during the research project? How will these costs be covered?**

The costs of the GitLab maintenance are covered by the COSIC research group as part of the general IT costs. The use of GitHub is currently 5USD/month and is funded from general COSIC funds.

## Data Preservation after the end of the Research Project

**Which data will be retained for 10 years (or longer, in agreement with other retention policies that are applicable) after the end of the project?**

**In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).**

- All data will be preserved for 10 years according to KU Leuven RDM policy

**Where will these data be archived (stored and curated for the long-term)?**

- KU Leuven RDR
- Other (specify below)

These data will be archived in several locations. First, the ESAT GitLab will hold the code and scripts so that it can be easily retrieved for further use by researchers; this data is backed up.  Second, a copy of the totality of the data will be stored in the KU Leuven RDR

**What are the expected costs for data preservation during the expected retention period? How will these costs be covered?**

Given the small size (< 1GB) of the data, the storage cost will either be free of charge, or already covered by COSIC's or KU Leuven's IT costs.

## Data Sharing and Reuse

**Will the data (or part of the data) be made available for reuse after/during the project?**
**Please explain per dataset or data type which data will be made available.**

- Yes, as restricted data (upon approval, or institutional access only)
- Yes, as open data

Projects that are deemed advanced enough will be made publicly available on COSIC's GitHub page, with appropriate licenses.  All projects will be stored on COSIC's GitLab so that it can be shared (conditioned on internal approval) with other researchers.

**If access is restricted, please specify who will be able to access the data and under what conditions.**

For the projects stored internally on GitLab, approval will be granted to COSIC researchers, or affiliated, under the condition of legitimate use for further research. This acts as a sanity check so that the technology is not removed from COSIC's internal control without further approval.

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

**Please explain per dataset or data type where appropriate.**

- No

Some open-source software libraries used during this project come with licenses that limit their commercial use, but no proprietary code will be used so that all projects can be made open-source without requiring further authorisation.

**Where will the data be made available?**

**If already known, please provide a repository per dataset or data type.**

- KU Leuven RDR (Research Data Repository)
- Other data repository (specify below)

The data will be made publicly available on COSIC's GitHub account because this is a widely-used repository for open-source code. Alternative access may be made available via KU Leuven RDR.

**When will the data be made available?**

- Upon publication of research results
- Other (specify below)

If software projects are in enough of a developed state, they will be made available upon publication of research results.

**Which data usage licenses are you going to provide?**

**If none, please explain why.**

- CC-BY 4.0 (data)
- MIT licence (code)

The data that is made publicly available will use appropriate licenses to permit the use of the code for further research, but prohibit commercial use, in addition to respecting any licenses of libraries used within.

**Do you intend to add a persistent identifier (PID) to your dataset(s), e.g. a DOI or accession number? If already available, please provide it here.**

- Yes, a PID will be added upon deposit in a data repository

The KU Leuven RDR will be used for this.

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

Given the small size (< 1 GB), data sharing is expected to be free of charge.

## Responsibilities

**Who will manage data documentation and metadata during the research project?**

The authors of the code are responsible.

**Who will manage data storage and backup during the research project?**

Daniel Ishimwe is the COSIC sysadmin in charge of linking the COSIC GitHub repo and internal GitLab repos with the KU Leuven RDR.

**Who will manage data preservation and sharing?**

Daniel Ishimwe is the COSIC sysadmin in charge of linking the COSIC GitHub repo and internal GitLab repos with the KU Leuven RDR.

**Who will update and implement this DMP?**

The co-supervisors of the project: Wouter Castryck & Frederik Vercauteren