Internal Funds

Data Management Plan: Syntax and Implementation Principles for Multimodal Dependent Type Theory

1 Plan Overview

Creator Andreas Nuyts
Affiliation KU Leuven
Template LaTeX calque of KU Leuven BOF-IOF template

2 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.

We will generate:

- Mathematical proofs (source files typically .tex),
- Software code and proof code, likely in Agda (.agda), Lean (.lean), Haskell (.hs), perhaps Rust (.rs) or Coq (.v).

This data will be digital and textual in nature, and its volume will be small: on the order of MBs or GBs at most. It is not meaningful at this point to create a concrete subdivision in "datasets" (or rather, code projects and technical reports).

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:

N/A

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.

Not directly.

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.

N/A

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.

See above.

3 Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data 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).

- The mathematical proofs should be readable by themselves to experts.
- Software and proof code should be documented as is standard practice for such materials, and will typically be associated to a publication which further explains the purpose and design. Low level documentation will be found within the code files, as is customary. The source files for higher level documentation will be part of the same repository that contains the code.

Will a metadata standard be used to make it easier to find 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.

At a technical level: Many programming languages have a standard format for documentation, which allows more human-readable documentation to be automatically generated.

At a higher level: Adherence to any metadata standards is not customary in our community. Rather, readme files (which display e.g. on Github as renderd web pages), documentation pages and personal websites are used to connect information.

Andreas Nuyts has a habit of connecting publications, associated slides, videos and code on his personal webpage `https://anuyts.github.io`, which in itself is based on a git repository in standard formats and is therefore easy to backup and move to a different hosting service even in the event where Github Pages would be discontinued.

4 Data Storage & Back-up during the Research Project

Where will the data be stored? How will the data be backed up?

Software and proof code and mathematical proofs, as well as scientific writing: written on our personal work laptops and several times daily pushed to at least one online git repository; such as github.com, bitbucket.com, codeberg.org or gitlab.kuleuven.be.

Considering reports regarding information removals contrary to the scientific strife for accumulation of human knowledge, by or under the command of the US government, both in the years 2017-2021[1] and since January 2025[2], we favour European solutions, even though our data are not particularly politically sensitive.

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?

[1] `https://en.wikipedia.org/wiki/Political_interference_with_science_agencies_by_the_first_Trump_administration`

[2] `https://en.wikipedia.org/wiki/2025_United_States_government_online_resource_removals`

- The researchers' personal devices are password-protected,
- Access to code repositories is protected by password (and 2FA) or ssh-key,
- Data is not particularly sensitive as the project does not involve personal data,
- We are in favour of open access and even creative commons publications and the field has a sufficiently cooperative spirit that even publicly accessible WIP is feasible, so *access* by unauthorized persons is not really an issue to begin with.

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

Given the small total size of the data, this cost is negligible. In fact, the aforementioned git repository hosting services provide the service for free (or funded from overhead costs).

5 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. . .).

We intend to retain all the data relevant to understanding, (re)using or reproducing[3] our research, i.e. all data mentioned in section 2.

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

Software and proof code will be stored on an online git repository (of the same kind we also use for backup, see above) and, possibly, additionally submitted to a publisher as an artefact officially associated to a published conference or journal paper. While paper artefacts are typically targeted at *reusability*, protecting them against 'bit rot', online git repositories typically pose a lower barrier towards researchers who want to (re)use or build on one's work.

Mathematical proofs will be stored on a permanent open archive such as ArXiv or HAL and/or become an integral part of scientific publications, in which case they will be deposited in Lirias.

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

Negligible to zero.

6 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.

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

N/A.

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.

We have no non-disclosure agreements or similar, and there is no reason or expectation why such agreements would be made. We are in favour of free and open source software (FOSS) as well as open

[3] Insofar as this concept is relevant for our type of research.

access, and even creative commons publications. Some creative commons and open source licenses, stipulate restrictions, e.g. a ban on commercial use or a 'share alike'-clause. When contributing to an existing FOSS project such as the Agda proof assistant, we will be expected to subscribe to the open source license already in effect.

Where will the data be made available? If already known, please provide a repository per dataset or data type.

See above.

When will the data be made available?

Either the data will be public even during its creation (this is quite customary in particular for code), or it will be published when a state of maturity has been reached, and at the latest at publication time of the associated paper.

Which data usage licenses are you going to provide? If none, please explain why. Question not answered.

We favour FOSS and creative commons licenses.

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.

No.

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

Negligible to zero.

7 Responsibilities

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

The researcher creating the data.

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

The researcher creating the data.

Who will manage data preservation and sharing?

Andreas Nuyts and Dominique Devriese.

Who will update and implement this DMP?

Andreas Nuyts and Dominique Devriese.