FWO DMP Template

Project supervisors (from application round 2018 onwards) and fellows (from application round 2020 onwards) will, upon being awarded their project or fellowship, be invited to develop their answers to the data management related questions into a DMP. The FWO expects a **completed DMP no later than 6 months after the official start date** of the project or fellowship. The DMP should not be submitted to FWO but to the research co-ordination office of the host institute; FWO may request the DMP in a random check.

At the end of the project, the **final version of the DMP** has to be added to the final report of the project; this should be submitted to FWO by the supervisor-spokesperson through FWO's e-portal. This DMP may of course have been updated since its first version. The DMP is an element in the final evaluation of the project by the relevant expert panel. Both the DMP submitted within the first 6 months after the start date and the final DMP may use this template.

1. General Information	
Name applicant	Katrien Antonio
FWO Project Number & Title	EOS ID 40007517 and FWO file G0I3422N
Affiliation	☒ KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☐ Other:
2. Data description	
Will you generate/collect new data and/or make use of existing data?	☒ Generate new data ☒ Reuse existing data

Describe the origin, type and format of the data (per dataset) and its (estimated) volume *If you **reuse** existing data, specify the **source** of these data.* *Distinguish data **types** (the kind of content) from data **formats** (the technical format).*	We will reuse (open and proprietary) insurance data sets. Roughly, these data sets will be up to 20GB, in .txt, .csv, SAS, R or Python format. The insurance companies collect the data and we will use them in pseudonymized form for research purposes. These insurance data sets will securely be transferred to us or they will exclusively be available on-site. We will also analyze data collected by the Belgian Cancer Registry (BCR). These datasets are exclusively available on-site, using a computer maintained by BCR. The researchers working on this specific work package will follow a training on privacy awareness, under the lead of BCR. The datatypes consist of claim numbers, claim sizes (and their time stamps), survival data, and features/covariates of policyholders, individuals or insured events. At the moment of writing we are in the process of recruiting PhD students and postdocs, and composing the scientific advisory board dedicated to the project. First steps in launching the different research lines and work packages have been initiated. However, no specific actions are launched yet regarding the collection of data sets from industry or societal partners. Therefore, no further specifics of the collected data can be listed yet.

3. Ethical and legal issues	
Will you use personal data? If so, shortly describe the kind of personal data you will use AND add the reference to your file in your host institution's privacy register. *In case your host institution does not (yet) have a privacy register, a reference is not yet required of course; please add the reference once the privacy register is in place in your host institution.*	☒ Yes ☐ No If yes: - Privacy Registry Reference: to do - Short description of the kind of personal data that will be used: see our answer in box 2 of this template. Our research will process personal information, received in pseudonymized format from insurance companies. Hereto, we will register our data sets in the KU Leuven's privacy register. We will also register the work packages within the ASTeRISK EoS project where personal data are reused, with both the Social and Societal Ethics Committee at KU Leuven and the Commission d'éthique SPLE at UCLouvain.

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 If yes: - Reference to ethical committee approval:
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?	☒ Yes ☐ No If yes, please comment: as part of the work packages we may develop simulation machines or engines, capable of generating artificial data sets, inspired by insights derived from working with original, confidential proprietary insurance data sets. No IP restrictions will be claimed; the simulation machines or engines will be open source, with dedicated documentation.
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?	☒ Yes ☐ No If yes, please comment: see our answer in box 2 of this template. Data sets received from insurance companies will be subject to a non-disclosure agreement, restricting the dissemination or exploitation of the received data.

4. Documentation and metadata

What documentation will be provided to enable understanding and reuse of the data collected/generated in this project?	All our papers aim for full reproducibility of our proposed models, estimators and algorithms. Hereto, the papers will list all technical details. For each paper, relevant programming code and detailed documentation will be retained in a version control system (git, hosted by GitHub). Moreover, the following files will be saved in the same folder as the dataset: • project documentation will be provided in a readme file. This file will include the project name, keywords, name of involved researchers and their ORCID ID, name of funder, funding code, start- and end date of the project, links to publications, creative common license, registration code from ethical committee for privacy review • the privacy review application will be saved as a PDF document and the registration code will be added to the project documentation. • the data management plan will be saved as a word file. • data preparation and the statistical analyses will be documented in an annotated analysis code file (as an R code or Python script). The version of the used software will be documented carefully. • data generation will be documented in an annotated data-generating code file. The variables in the created datasets will be described in a codebook. The statistical analyses conducted on the simulated data will be documented in an annotated analysis code file (i.e. R code or Python file). The version of the used software will be documented.
Will a metadata standard be used? If so, describe in detail which standard will be used. If not, state in detail which metadata will be created to make the data easy/easier to find and reuse.	☐ Yes ☒ No If yes, please specify:

5. Data storage & backup during the FWO project

Where will the data be stored?	During the research we will store the data on the researchers' university managed computer and on an external drive, encrypted and protected. We will also store the data in a cloud service securely accessible to both KU Leuven and UCLouvain researchers (OneDrive for Business or NextCloud). Programming code and detailed documentation will be retained in a version control system (git, hosted by GitHub). If necessary, file transfers will be performed with Belnet Filesender using the encryption option.
How will the data be backed up?	See our answer to the previous question.
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 ☐ No If no, please specify:
What are the expected costs for data storage and backup during the project? How will these costs be covered? *Although FWO has no earmarked budget at its disposal to support correct research data management, FWO allows for part of* ***the allocated project budget*** *to be used to cover the cost incurred.*	Allocated project budget will be used for costs related to data storage.
Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?	Digital data will be stored in a restricted network share on OneDrive, which can only be accessed by the involved researchers. Multi-factor authentication is activated for the KU Leuven, respectively UCLouvain, login of all researchers having access to the data.

6. Data preservation after the end of the FWO project

FWO expects that data generated during the project are retained for a period of minimally 5 years after the end of the project, in as far as legal and contractual agreements allow.

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, ...).	To be determined.
Where will these data be archived (= stored for the long term)?	After the research we will store the data in three locations: (1) on a password-protected and encrypted external hard drive, (2) in OneDrive for Business, and (3) the Server Back-End Storage of KU Leuven, a data archiving infrastructure.
What are the expected costs for data preservation during these 5 years? How will the costs be covered? *Although FWO has no earmarked budget at its disposal to support correct research data management, FWO allows for part of* ***the allocated project budget*** *to be used to cover the cost incurred.*	To be determined; allocated project budget will be used for costs related to data storage.

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)?	☒ Yes ☐ No If yes, please specify: data sets subject to NDA with insurance company or the Belgian Cancer Registry can not be shared.
Which data will be made available after the end of the project?	To be determined.

Where/how will the data be made available for reuse?	☐ In an Open Access repository ☐ In a restricted access repository ☐ Upon request by mail ☒ Other (specify): dedicated repositories on GitHub
When will the data be made available?	The data will be made available upon publication of the research results.
Who will be able to access the data and under what conditions?	The research data (if possible) and the documentation will be made available. When research data can not be shared (due to NDA with insurance companies) a simulated toy data set will be made available as alternative. Our research essentially puts focus on the design of new models, estimators and algorithms. Hence, illustrations with toy data sets are also highly valuable.
What are the expected costs for data sharing? How will these costs be covered? *Although FWO has no earmarked budget at its disposal to support correct research data management, FWO allows for part of* ***the allocated project budget*** *to be used to cover the cost incurred.*	To be determined; allocated project budget will be used for costs related to data sharing.

8. Responsibilities	
Who will be responsible for the data documentation & metadata?	Joint responsibility, coordinated by PI Katrien Antonio
Who will be responsible for data storage & back up during the project?	Idem
Who will be responsible for ensuring data preservation and sharing?	PI Katrien Antonio

Who bears the end responsibility for updating & implementing this DMP? *Default response: The PI bears the overall responsibility for updating & implementing this DMP*	PI Katrien Antonio