FWO DMP Template - Flemish Standard Data Management Plan

Version KU Leuven

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.

The DMP template used by the Research Foundation Flanders (FWO) corresponds with the Flemish Standard Data Management Plan. This Flemish Standard DMP was developed by the Flemish Research Data Network (FRDN) Task Force DMP which comprises representatives of all Flemish funders and research institutions. This is a standardized DMP template based on the previous FWO template that contains the core requirements for data management planning. To increase understanding and facilitate completion of the DMP, a standardized **glossary** of definitions and abbreviations is available via the following link.

1. General Project Information	
Name Grant Holder & ORCID	**Jan Verbakel 0000-0002-7166-7211**
Contributor name(s) (+ ORCID) & roles	
Project number [1] & title	**Temporal trends and drivers of antibiotic failure in the general population**
Funder(s) GrantID [2]	G076723N
Affiliation(s)	☒ KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☐ Other: ROR identifier KU Leuven: 05f950310
Please provide a short project description	Antimicrobial resistance is a serious and growing concern for health systems worldwide. Although the link between antimicrobial use and resistance is well-established, this association is complex and much remains unknown about the exact drivers of antimicrobial resistance. Aim: to investigate temporal trends of first-line antibiotic failures in the general population as well as possible risk factors associated with increased incidence of first-line antibiotic failures, by type of infection and antibiotic class. Methods: Analyses will be performed on one of the largest and richest datasets available for research purposes in the world - anonymised electronic health records with >100 million patient-years from 1995 to 2020 from the Clinical Practice Research Datalink (CPRD). Expected Outcomes: A better understanding of the factors associated with antibiotic failures in routine clinical practice may inform the development of more targeted and effective prevention measures to limit the emergence of antimicrobial resistance. Identified risk factors as well as differences between infection types and antibiotic classes may further point towards biological functions underlying these complications and inform the design of future research.

[1] "Project number" refers to the institutional project number. This question is optional. Applicants can only provide one project number.

[2] Funder(s) GrantID refers to the number of the DMP at the funder(s), here one can specify multiple GrantIDs if multiple funding sources were used.

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 [3].

				ONLY FOR DIGITAL DATA	*ONLY FOR DIGITAL DATA*	*ONLY FOR DIGITAL DATA*	*ONLY FOR PHYSICAL DATA*
Dataset Name	Description	New or Reused	Digital or Physical	Digital Data Type	Digital Data Format	Digital Data Volume (MB, GB, TB)	Physical Volume
CPRD	anonymised electronic health records with >100 million patient-years from 1997 to 2022 from the Clinical Practice Research Datalink	☐ Generate new data ☒ Reuse existing data	☒ Digital ☒ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☒ Textual ☐ Model ☐ Software ☐ Other:	.txt	☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	

[3] Add rows for each dataset you want to describe.

GUIDANCE: *The data description forms the basis of your entire DMP, so make sure it is detailed and complete. It includes digital and physical data and encompasses the whole spectrum ranging from raw data to processed and analysed data including analysis scripts and code. Physical data are all materials that need proper management because they are valuable, difficult to replace and/or ethical issues are associated. Materials that are not considered data in an RDM context include your own manuscripts, theses and presentations; documentation is an integral part of your datasets and should described under documentation/metadata.* *RDM Guidance on data*	
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.	CPRD GOLD June 2023: https://doi.org/10.48329/4wmr-x234 CPRD AURUM March 2023: https://doi.org/10.48329/91gg-3834
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.	☐ Yes, human subject data; provide SMEC or EC approval number: ☐ Yes, animal data; provide ECD reference number: ☐ Yes, dual use; provide approval number: ☒ No Additional information: The CPRD Group has ethical approval from the National Research Ethics Service Committee (NRES) for all purely observational research, namely, studies that do not include patient involvement, which is the case of the present proposal, and hence no separate ethical approval is required - see https://www.cprd.com/isac/otherinfo.asp. Approval from CPRD's scientific advisory committee, who review protocols for scientific quality, will be required and a request will be filed in the coming months.

Will you process personal data[4]? 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).	☒ Yes (provide PRET G-number or EC S-number below) ☐ No Additional information: The dataset used in this study undergoes rigorous anonymization procedures before it can be accessed by researchers. Recital 26 of the General Data Protection Regulation (GDPR) leaves open the interpretation that record-level personal data can be considered anonymized if subject to sufficient de-identification, and hence whether it is to be considered 'personal data'. To take the most cautious approach, the data will be handled as 'personal data' under the GDPR and will be processed following the principles outlined in Article 5 of the GPRD. The safeguards that will be taken in that regard are outlined below. This study will re-use anonymized data from the Clinical Practice Research Datalink (CPRD), which contains patient related features (e.g. occupation or ethnicity), clinical signs, diagnoses, laboratory tests, medical procedures, or drug prescriptions. The dataset undergoes rigorous anonymization and third parties using this data for research purposes have no access to identifiers and are not able to link the information back to individuals. More information can be found here: www.cprd.com
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.	☐ Yes ☒ No If yes, please comment:
Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements, research collaboration agreements)? If so, please explain to what data they relate and what restrictions are in place.	☐ Yes ☒ No If yes, please explain:

[4] See Glossary Flemish Standard Data Management Plan

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 to what data they relate and which restrictions will be asserted.	☐ Yes ☒ No If yes, please explain:

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). *RDM guidance on documentation and metadata*.	CPRD has standard data dictionaries available. CPRD dictionaries are provided as text files that can be imported into standard statistical software to enable code searching. The dictionaries are also available through the CPRD Code Browser. The CPRD Code Browser and a user guide were requested by contacting enquiries@cprd.com. Furthermore, we will create a codebook for the new created computed variables, using the "codebook" package in R.
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. *REPOSITORIES COULD ASK TO DELIVER METADATA IN A CERTAIN FORMAT, WITH SPECIFIED ONTOLOGIES AND VOCABULARIES, I.E. STANDARD LISTS WITH UNIQUE IDENTIFIERS.*	☒ Yes ☐ No If yes, please specify (where appropriate per dataset or data type) which metadata standard will be used: DataCite in accordance with CPRD's guidance. If no, please specify (where appropriate per dataset or data type) which metadata will be created:

4. Data Storage & Back-up during the Research Project	
Where will the data be stored? *Consult the interactive KU Leuven storage guide to find the most suitable storage solution for your data.*	☐ Shared network drive (J-drive) ☐ Personal network drive (I-drive) ☒ OneDrive (KU Leuven) ☐ Sharepoint online ☐ Sharepoint on-premis ☒ Large Volume Storage ☐ Digital Vault ☐ Other:
How will the data be backed up? *WHAT STORAGE AND BACKUP PROCEDURES WILL BE IN PLACE TO PREVENT DATA LOSS?*	☒ Standard back-up provided by KU Leuven ICTS for my storage solution ☐ Personal back-ups I make (specify) ☐ Other (specify)
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:

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons? *CLEARLY DESCRIBE THE MEASURES (IN TERMS OF PHYSICAL SECURITY, NETWORK SECURITY, AND SECURITY OF COMPUTER SYSTEMS AND FILES) THAT WILL BE TAKEN TO ENSURE THAT STORED AND TRANSFERRED DATA ARE SAFE.* *Guidance on security for research data*	Data will be stored exclusively in secure environments overseen by KU Leuven's information technology services. This will be a secure data server on the university network (Large Volume Storage in combination with Onedrive (KU Leuven)). To ensure security of the data during its transfer from the data provider to KU Leuven, the transfer will be completed on a password-encrypted device. Access to the data will be restricted to the absolute necessary. The supervisor Prof. Verbakel will oversee data access requests, and only grant access to those persons directly involved in the research. The data will be used exclusively for the purpose of the proposed study and kept only as long as required by the study and associated regulations.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	No additional costs for data storage and backup are to be expected as these are part of the KU Leuven infrastructure.

5. Data Preservation after the end of the Research Project	
Which data will be retained for at least five 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...). *Guidance on data preservation*	☒ All data will be preserved for 10 years according to KU Leuven RDM policy ☐ All data will be preserved for 25 years according to CTC recommendations for clinical trials with medicinal products for human use and for clinical experiments on humans ☐ Certain data cannot be kept for 10 years (explain)

Where will these data be archived (stored and curated for the long-term)? *Dedicated data repositories are often the best place to preserve your data. Data not suitable for preservation in a repository can be stored using a KU Leuven storage solution, consult the interactive KU Leuven storage guide.*	☒ KU Leuven RDR ☐ Large Volume Storage (longterm for large volumes) ☐ Shared network drive (J-drive) ☐ Other (specifiy):
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	No additional costs for data storage and backup are to be expected as these are part of the KU Leuven infrastructure.

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. *NOTE THAT 'AVAILABLE' DOES NOT NECESSARILY MEAN THAT THE DATA SET BECOMES OPENLY AVAILABLE, CONDITIONS FOR ACCESS AND USE MAY APPLY. AVAILABILITY IN THIS QUESTION THUS ENTAILS BOTH OPEN & RESTRICTED ACCESS. FOR MORE INFORMATION: HTTPS://WIKI.SURFNET.NL/DISPLAY/STANDARDS/INFO-EU-REPO/#INFOEUREPO-ACCESSRIGHTS*	☐ Yes, as open data ☐ Yes, as embargoed data (temporary restriction) ☐ Yes, as restricted data (upon approval, or institutional access only) ☒ No (closed access) ☐ Other, please specify:

If access is restricted, please specify who will be able to access the data and under what conditions.	Access to the data will be restricted to the absolute necessary. The supervisor Prof. Verbakel will oversee data access requests, and only grant access to those persons directly involved in the research. The data will be used exclusively for the purpose of the proposed study and kept only as long as required by the study and associated regulations.
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.	☐ Yes, privacy aspects ☐ Yes, intellectual property rights ☐ Yes, ethical aspects ☐ Yes, aspects of dual use ☒ Yes, other ☐ No If yes, please specify: Access to CPRD data, including UK Primary Care Data, and linked data such as Hospital Episode Statistics, is subject to protocol approval via CPRD's Research Data Governance (RDG) Process. The data will be used exclusively for the purpose of the proposed study and kept only as long as required by the study and associated regulations.
Where will the data be made available? If already known, please provide a repository per dataset or data type.	☐ KU Leuven RDR ☐ Other data repository (specify) ☒ Other (specify) The data will not be made available as per CPRD's regulations.
When will the data be made available?	☐ Upon publication of research results ☐ Specific date (specify) ☒ Other (specify): Never The data will not be made available as per CPRD's regulations.

Which data usage licenses are you going to provide? If none, please explain why. *A DATA USAGE LICENSE INDICATES WHETHER THE DATA CAN BE REUSED OR NOT AND UNDER WHAT CONDITIONS. IF NO LICENCE IS GRANTED, THE DATA ARE IN A GREY ZONE AND CANNOT BE LEGALLY REUSED. DO NOTE THAT YOU MAY ONLY RELEASE DATA UNDER A LICENCE CHOSEN BY YOURSELF IF IT DOES NOT ALREADY FALL UNDER ANOTHER LICENCE THAT MIGHT PROHIBIT THAT.* *Check the RDR guidance on licences for data and software sources code or consult the License selector tool to help you choose.*	☐ CC-BY 4.0 (data) ☐ Data Transfer Agreement (restricted data) ☐ MIT licence (code) ☐ GNU GPL-3.0 (code) ☒ Other (specify): None Not applicable as no data usage will be allowed in accordance with CPRD's regulations.
Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, please provide it here. *INDICATE WHETHER YOU INTEND TO ADD A PERSISTENT AND UNIQUE IDENTIFIER IN ORDER TO IDENTIFY AND RETRIEVE THE DATA.*	☐ Yes, a PID will be added upon deposit in a data repository ☐ My dataset already has a PID ☒ Not applicable
What are the expected costs for data sharing? How will these costs be covered?	Not applicable.

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	Jan Verbakel
Who will manage data storage and backup during the research project?	Jan Verbakel
Who will manage data preservation and sharing?	Jan Verbakel
Who will update and implement this DMP?	Jan Verbakel