FWO DMP Template - Flemish Standard Data Management Plan

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	Isabelle Cleynen (ORCID https://orcid.org/0000-0003-0857-7683)
Contributor name(s) (+ ORCID) & roles	Máté Varga (ORCID https://orcid.org/0000-0003-4289-1705) Boris Rogelj (ORCID https://orcid.org/0000-0003-3898-1943) Tomaž Bratkovič (ORCID https://orcid.org/0000-0001-8367-5465)
Project number[1] & title	CELSA/22/026; "The role of snoRNAs in the etiology of inflammatory bowel disease"
Funder(s) GrantID[2]	CELSA/22/026
Affiliation(s)	X KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel X Other: ELTE Eötvös Loránd University; University of Ljubljana Provide ROR[3] identifier when possible:

[1] "Project number" refers to the institutional project number. This question is optional since not every institution has an internal project number different from the GrantID. 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.

[3] Research Organization Registry Community. https://ror.org/

Please provide a short project description	Inflammatory bowel disease (IBD) constitutes a group of progressive and debilitating disorders characterized by chronic inflammation of the intestine, with poorly understood etiology. We recently found that expression of small nucleolar RNAs (snoRNAs) – a family of short non-coding RNAs traditionally implicated in chemically modifying other RNA types – marks recurrent Crohn's disease (a form of IBD). We also found that lack of dyskerin, a well-known protein partner of snoRNAs, leads to IBD-like symptoms in a zebrafish model, further supporting a potential link between snoRNAs and IBD. To examine the role of snoRNAs in IBD we will combine the complementary expertise from three research groups. Specifically, at KU Leuven we will analyse differential expression of snoRNAs in intestinal biopsies and peripheral blood immune cells of IBD patients vs healthy controls using short RNA sequencing. Using a similar approach, we will identify perturbed snoRNAs in chemically inducible zebrafish and intestinal cell models of IBD at ELTE and UL, respectively. These data will set ground for mechanistic studies, where individual dysregulated snoRNAs will be knocked-out or overexpressed in both disease models using advanced molecular biology methods. Furthermore, we will use our state-of-the-art methods of RNA interactome interrogation to identify molecular targets of dysregulated snoRNAs in live cells, assisting in revealing their link to IBD symptoms. The overarching goal of the proposed project is to find reliable snoRNA diagnostic biomarkers with functional relevance to IBD, which in future should aid in accurate disease diagnosis, thereby allowing early and effective treatment. Last but not least, the animal and cell models generated in this project might find use for preclinical screens to identify new or repurposed therapeutics for IBD.

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[4].

				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
Tissue samples	Gut biopsies/blood samples from IBD patients and controls		Physical				2-4 biopsies/sample 10 mL blood
RNA	RNA extracted from tissue samples		Physical				~20µL
RNA	RNA extracted from zebrafish gut samples		Physical				~25 µL
RNA	RNA extracted from cells (Caco-2, THP-1) grown in vitro		Physical				~25 µL
Sequencing data	Raw snoRNAseq data (IBD patients and	☒ Generate new data ☐ Reuse existing	☒ Digital ☐ Physical	☐ Observational ☒ Experimental ☐ Compiled/	☐ .por ☐ .xml ☐ .tab	☐ < 100 MB ☐ < 1 GB ☒ < 100 GB	

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

	controls; zebrafish; human cell lines); PARIS/COMRADES seq data from human cell lines	data		aggregated data ☐ Simulation data ☐ Software ☐ Other ☐ NA	☐ .csv ☐ .pdf ☐ .txt ☐ .rtf ☐ .dwg ☐ .tab ☐ .gml ☒ other: fastq files ☐ NA	☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	
Read count files	Mapped and QC'ed RNAseq data	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Observational ☒ Experimental ☐ Compiled/ aggregated data ☐ Simulation data ☐ Software ☐ Other ☐ NA	☐ .por ☐ .xml ☐ .tab ☒ .csv ☐ .pdf ☒ .txt ☐ .rtf ☐ .dwg ☐ .tab ☐ .gml ☐ other: ☐ NA		
Data analysis results	Result files from statistical analyses (text/figures)	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Observational ☒ Experimental ☐ Compiled/ aggregated data ☐ Simulation data ☐ Software ☐ Other	☐ .por ☐ .xml ☐ .tab ☐ .csv ☒ .pdf ☒ .txt ☐ .rtf ☐ .dwg	☒ < 100 MB ☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB	

				☐ NA	☐ .tab ☐ .gml ☒ other: .jpg or .tiff or … ☐ NA	☐ NA	
Microarray data	Previously obtained microarray data that will be re-analysed	☐ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☐ Observational ☒ Experimental ☐ Compiled/ aggregated data ☐ Simulation data ☐ Software ☐ Other ☐ NA	☐ .por ☐ .xml ☐ .tab ☐ .csv ☐ .pdf ☒ .txt ☐ .rtf ☐ .dwg ☐ .tab ☐ .gml ☐ other: .jpg or .tiff or … ☐ NA	☐ < 100 MB ☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	
phenodata	Basic phenotypic information of included individuals	☒ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☒ Observational ☐ Experimental ☐ Compiled/ aggregated data ☐ Simulation data ☐ Software ☐ Other ☐ NA	☐ .por ☐ .xml ☐ .tab ☒ .csv ☐ .pdf ☒ .txt ☐ .rtf ☐ .dwg ☐ .tab ☐ .gml ☐ other: .jpg or .tiff or … ☐ NA	☒ < 100 MB ☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	

Guidance: *Data can be digital or physical (for example biobank, biological samples, …). Data type: Data are often grouped by type (observational, experimental etc.), format and/or collection/generation method.* *Examples of data types: observational (e.g. survey results, sensor readings, sensory observations); experimental (e.g. microscopy, spectroscopy, chromatograms, gene sequences); compiled/aggregated data[5] (e.g. text & data mining, derived variables, 3D modelling); simulation data (e.g. climate models); software, etc.* *Examples of data formats: tabular data (.por,. spss, structured text or mark-up file XML, .tab, .csv), textual data (.rtf, .xml, .txt), geospatial data (.dwg,. GML, ..), image data, audio data, video data, documentation & computational script.* *digital data volume: Please estimate the upper limit of the volume of the data per dataset or data type.* *physical volume: Please estimate the physical volume of the research materials (for example the number of relevant biological samples that need to be stored and preserved during the project and/or after).*	
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.	Microarray data: The mRNA microarray data were deposited to the Gene Expression Omnibus database according to Minimum Information About a Microarray Experiment [MIAME] guidelines [series accession GSE102133, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE102133]. This data is also available on the internal servers of KU Leuven (Archive K drive) as it is data previously generated by the involved group at KU Leuven).

[5] These data are generated by combining multiple existing datasets.

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, please describe these issues further and refer to specific datasets or data types when appropriate.	☒ Yes, human subject data ☒ Yes, animal data ☐ Yes, dual use ☐ No If yes, please describe: We use samples collected from human individuals (patients). Approval for this is obtained from the Ethics Committee Research UZ / KU Leuven (EC Research). If necessary/applicable, further approval regarding patient sample use will be sought there as well. All animal procedures will be approved by the Government Office of Pest County and/or the local ethics committee of ELTE Eötvös Loránd University.
Will you process personal data[6]? If so, briefly describe the kind of personal data you will use. Please refer to specific datasets or data types when appropriate. If available, add the reference to your file in your host institution's privacy register.	☒ Yes ☐ No If yes: - Short description of the kind of personal data that will be used: clinical information (diagnosis (IBD versus control)); sex; age (dataset: phenodata) - Privacy Registry Reference: TBD
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: identified snoRNAs could have potential as biomarkers in which case IP could be applicable (combination of markers subject of patent)

[6] See Glossary Flemish Standard Data Management Plan

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:
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: An MTA/DTA and contract agreement will need to be set up between the different partners of the project.

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).	**KU Leuven:** - Each researcher keeps a lab notebook (written and/or electronic) which keeps track of the experiments done, and where to find the results if not in the lab notebook itself. - The bioinformatics researchers dealing with data analysis (RNAseq data analysis) use well commented scripts (usually via Jupyter notebooks) that indicate each step of the analysis such that it can be easily reproduced. - We use a systematic file/folder system where each folder has the same naming structure and subfolders (input/output/scripts/obsolete). All raw data files are backed up to Large storage and Archive drives managed by KU Leuven, as well as to the vsc staging (also see below). A README.txt file is included in the main folder explaining the data structure. **ELTE Eötvös Loránd University:** - experimental data will be recorded in electronic notebooks on the Benchling platform - novel snoRNA-seq analysis pipelines will be documented and made available through GitHub repositories. **University of Ljubljana:** - Each researcher keeps a lab journal in physical and digital format (local and University of Ljubljana OneDrive cloud and/or Dropbox with synchronization), where all experiment details and observations/comments and experimental data location are stored.

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: If no, please specify (where appropriate per dataset or data type) which metadata will be created: Cfr above

4. Data Storage & Back-up during the Research Project

Where will the data be stored?	**KU Leuven:** - During the research, we make use of secured network folders (in blocks of 1TB, with automatic offsite back-up) as well as 2TB personal OneDrive cloud storage provided by KU Leuven IT. Data will also be loaded onto the VSC associated computer hardware (https://www.vscentrum.be/) for data analysis. Larger files will be stored there on staging space on the vsc. **ELTE Eötvös Loránd University:** - Sequencing data will be stored on personal hard drives and also on cloud backups using the Galaxy platform (where some of the analysis will be also performed). - Imaging data will be stored on physical hard drives, cloud storage of the University and flash-memory backup drives. **University of Ljubljana:** - Data will be stored on University of Ljubljana OneDrive cloud and backed up locally (hard drives). - Big data storage TBD **Data shared among partners during project:** - Where applicable, data will be shared through a Teams Team that has been set up specifically for this grant, which is a multi-center grant. Only individuals involved in the project are members of the Team. - If data is too large, files will be transferred using a cloud storage service which is then shared with partners; or using SFTP.

How will the data be backed up? *What storage and backup procedures will be in place to prevent data loss? Describe the locations, storage media and procedures that will be used for storing and backing up digital and non-digital data during research.*[7] *Refer to institution-specific policies regarding backup procedures when appropriate.*	**KU Leuven** - Cfr above: secured network folders are automatically backed-up off-site - Physical lab notebooks are kept locally **ELTE Eötvös Loránd University:** - Sequencing data will be backed up to the Sequence Read Archive (SRA) repository. - Electronic notebooks will be archived and backed up in secure ELTE network folders. **University of Ljubljana:** - Lab diaries will be kept locally and automatically synchronized to University OneDrive. - Sequencing data will be deposited to the Sequence Read Archive (SRA) repository.
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 yes, please specify concisely: If no, please specify:

[7] Source: Ghent University Generic DMP Evaluation Rubric: https://osf.io/2z5g3/

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.* [7]	**KU Leuven:** - Cfr above: KU Leuven offers secured network folders that are only accessible by registered users with granted access. **ELTE Eötvös Loránd University:** - Besides the secured network folders of ELTE, secure public repositories such as SRA will be used. Physical hard drives are stored in safe (locked) cabinets in rooms where only authorized personnel has access. - Only people participating in the research will have access to the relevant notebooks. **University of Ljubljana:** - Access to locally stored and network files and folders is password-protected and only accessible to research group members/registered cloud users. **Data shared among partners during project:** - Where applicable, data will be shared through a Teams Team that has been set up specifically for this grant. Only individuals involved in the project are members of the Team.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?	**KU Leuven** - Local network storage: 51.90 EUR/100GB/year (only stored data is charged) - Large volume storage: 569.20 EUR/5TB/year - High performance computing (Vsc): staging storage (20 EUR/TB/year; standard data storage space on vsc is free). - → Total expected cost ca 500 EUR - Cost coverage: general lab funding + CELSA project funding. **ELTE Eötvös Loránd University:** - TBD **University of Ljubljana:** - There will be no data storage costs.

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...).	**KU Leuven:** - Raw sequencing data (fastq files) and count files + accompanying README files - Scripts used - Final analysis results/figures cfr publication **ELTE Eötvös Loránd University** - Raw sequencing data, imaging data and archived electronic notebooks. - Scripts used. - Final analysis results and figures. **University of Ljubljana:** - Raw sequencing data (fastq files) and count files. - Scripts used. - Final analysis results and figures.

Where will these data be archived (stored and curated for the long-term)?	**KU Leuven:** - Local data preservation is the responsibility of the local IT team, in coordination with the lab PI (prof Cleynen). - Large volume archive storage is available via KU Leuven IT (≥5TB storage with automatic back-up) for indefinite archiving. - Hard copy lab notebooks are being kept by the researchers involved in the project, and will remain accessible via the KU Leuven archives. - Sequencing data will be saved also to a data repository (KU Leuven RDR or domain-specific repository, likely **ELTE Eötvös Loránd University** - Sequencing data will be backed up to the Sequence Read Archive (SRA) repository. - Electronic notebooks will be archived and backed up in secure ELTE network folders. **University of Ljubljana:** - Sequencing data will be deposited to the Sequence Read Archive (SRA) repository. - Hard copy lab diaries of all the researcher involved in the project will be stored in the PI's archives. Electronic diaries will be kept on University OneDrive cloud.

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	**KU Leuven** - Large volume archive storage: 11.384 EUR/100GB/year - Local network storage: 51.90 EUR/100GB/year (only stored data is charged) - Large volume storage: 569.20 EUR/5TB/year - → Total expected cost ca 750 EUR - Cost coverage: general lab funding. 50% of large volume archive storage is paid by Group Biomedical Sciences. **ELTE Eötvös Loránd University:** - TBD **University of Ljubljana:** - There will be no data storage costs.

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, in an Open Access repository ☒ Yes, in a restricted access repository (after approval, institutional access only, …) ☐ No (closed access) ☐ Other, please specify: - Raw sequencing data Zebrafish samples (open access) - RNA sequencing data human samples (restricted) - RNA sequencing data human cell lines (treated vs. control) (open access)
If access is restricted, please specify who will be able to access the data and under what conditions.	
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: - Human patient samples: any sharing will be bound by privacy aspects and ethical agreement (if covered in informed consent). - IP rights related to potential patentability of biomarker panel.

Where will the data be made available? If already known, please provide a repository per dataset or data type.	- Zebrafish raw sequencing data: Sequence Read Archive (SRA) repository - Human RNAseq data: domain specific repository, likely https://www.ebi.ac.uk/biostudies/arrayexpress or GEO (https://www.ncbi.nlm.nih.gov/geo/) - Human cell lines RNA sequencing data: Sequence Read Archive (SRA) repository
When will the data be made available? *This could be a specific date (dd/mm/yyyy) or an indication such as 'upon publication of research results'.*	Upon publication of research results
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.* *Example Answer: E.g. "Data from the project that can be shared will be made available under a Creative Commons Attribution license (CC-BY 4.0), so that users have to give credit to the original data creators."* [8]	Data from the project that can be shared will be made available under a Creative Commons Attribution License (likely CC-BY 4.0), so that users have to give credit to the original data creators.

[8] Source: Ghent University Generic DMP Evaluation Rubric: https://osf.io/2z5g3/

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 ☐ No If yes: There will be a unique identifier for deposited data cfr guidelines SRA/Array Express/GEO.
What are the expected costs for data sharing? How will these costs be covered?	No costs

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	All partners; coordinated by Isabelle Cleynen
Who will manage data storage and backup during the research project?	All partners; coordinated by Isabelle Cleynen
Who will manage data preservation and sharing?	All partners; coordinated by Isabelle Cleynen
Who will update and implement this DMP?	Isabelle Cleynen