CELSA DMP

- Initieel DMP – binnen 6 maanden na toekenning financiering

1. General Information	
1.1. Name of the project lead (PI)	**Ellen Decaestecker**
1.2. Project Number & Title	**CELSA/22/012** **Microbial diversity within an Eco-Health context: effects of anthropogenic stress and host phylogeny on the microbiome and microparasites of a keystone freshwater invertebrate.**
2. Data description	
2.1. Will you generate/collect new data and/or make use of existing data?	New data will be collected.
2.2. 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).	**Digital data:** - data on the main environmental characteristics of studied streams in a spreadsheet (.csv format) - data on chemical characteristics of water and sediment collected from studied streams in a spreadsheet (.csv format) - data on host lineage in a spreadsheet (.csv format) - photo documentation of processed gammarid specimens as a picture (.jpg format) - metagenomics data gut microbiota in FASTQ format **Physical data:** Gammarid specimens from natural systems (streams) will be collected.
3. Ethical and legal issues	
3.1. Will you use personal data? If so, shortly describe the kind of personal data you will use AND add the reference to your notification file with the privacy commission.	No

3.2. 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).	No
3.3. 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?	No
3.4. 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?	No

4.   Documentation and metadata

4.1. What documentation will be provided to enable understanding and reuse of the data collected/generated in this project?	**Digital data:** We will maintain a record of the following: - sampling protocol (.doc file) - abbreviations used (.doc file) - the structure of the data (.doc file) **-** steps involved in data analysis and relevant analysis scripts (R scripts, MOTHUR/QIIME scripts) - raw data (specific file format according to the data type) - analysed data (specific file format according to the data type) - index file/read me file (.txt file) linking the name, location (folder and subfolder on BOX/server/hard disk), and description of the above-mentioned files. In the concluding stage of the project, a master index file containing the combined information will be compiled which will be archived and stored on the personal hard drives/PC of each PI. **Physical data:** Gammarid specimens taken from natural systems (streams) will be documented and stored for up to five years after the end of the project. Depending on the kind of sample, they will be stored

	either in fixative (like ethanol) or deep-frozen and stored in the freezer.
4.2. 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.	The experiments are unique, but the data will be standardized according to data-type across experiments to make it easier to interpret the structure. Below, we list the metadata standards applicable to this project: 1) Ecological Metadata Language (EML), a metadata standard specific for the discipline of ecology will be used as a standard for community metadata for this project. We will use the application Morpho to create, edit and search metadata files following the EML specifications. Such files can be used to exchange information within and among research groups. 2) For metagenetics and metagenomics data, metadata will be organized and reported according to standard developed by the Genomic Standards Consortium (GSC) for reporting marker gene sequences— MIxS – Minimum information about any (x) sequence

5. Data storage & backup during the CELSA project	
5.1. Where will the data be stored?	On the Server space at Department of Biology and KU Leuven Enterprise BOX. There is additional backup possibility at the K-drive of the KU Leuven KULAK campus where server space is paid by Ellen Decaestecker.
5.2. How will the data be backed up?	We will make a copy on external hard discs. Server space and BOX are centrally backed-up.
5.3. 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.	We expect that the available server space will be enough. We will also buy additional external hard disks to have back-up copies safely stored in an easy to access way.
5.4. What are the expected costs for data storage and backup during the project? How will these costs be covered?	Buying additional external disks and extra space on Box can be covered by the consumables money of the C-project
5.5. Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?	The data on Box will be protected by a password. For extra safety we will for the smaller data sets also have back-ups on external hard disks that are kept at the office of the PI.

6.   Data preservation after the end of the CELSA project KU Leuven 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.

6.1. Which data will be retained for the expected 5-year period after the end of the project? If only a selection of the data can/will be preserved, clearly state why this is the case (legal or contractual restrictions, physical preservation issues, ...).	**Digital data:** We will retain all data for the expected 5-year period. For most publications, we expect that we will make the data publicly available on data repositories. Sequencing data will be submitted to public databases (EBI-ENA/NCBI-SRA), where they will be permanently archived to preserve access to the public. For community composition and phenotype data we will consider data repositories like DRYAD, GBIF and Zenodo. **Physical data:** Gammarid specimens taken from natural systems (streams) will be documented and stored for up to five years after the end of the project. Depending on the kind of sample, they will be stored either in fixative (like ethanol) or deep-frozen and stored in the freezer.
6.2. Where will these data be archived (= stored for the long term)?	Digital data will be archived on external hard disks, BOX and the K-drive of the server of the Department of Biology at KU Leuven and KULAK. For most published data sets also on data repositories listed in section 6.1.
6.3. What are the expected costs for data preservation during these 5 years? How will the costs be covered?	For external hard disks and most data repositories there are no extra costs. When using DRYAD, one-time submission cost will be covered with the CELSA project fund.
7. Data sharing and reuse	
7.1. 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 or because of IP potential)?	No
7.2. Which data will be made available after the end of the project?	Published data will be made available. Unpublished results will be made available after an embargo period (3 years; exceptionally 5 years after the project).
7.3. Where/how will the data be made available for reuse?	Data repositories and data papers
7.4. When will the data be made available?	Upon publication, or 3 years (exceptionally 5 years) after the end of the project to enable us to publish our results. In most cases whenever data are used for a publication they will become available through DRYAD and sequence databases.
7.5. Who will be able to access the data and under what conditions?	Upon publication everybody can access (EBI-ENA/NCBI-SRA, GBIF, Zenodo, DRYAD); conditions may apply e.g. information on what will be done with the data. We will attach Creative Commons Licences (CC BY) to the data deposited to enable researchers to access, mine, and reproduce our data.

7.6. What are the expected costs for data sharing? How will these costs be covered?	No expected costs for public data repositories except for DRYAD (if we decide to use it, one-time submission cost will be covered by the project fund)
8. Responsibilities	
8.1. Who will be responsible for the data documentation & metadata?	All PIs (Ellen Decaestecker, Adam Petrusek and Luc De Meester) and the day-to-day manager of the CELSA project (PhD student Nataša Katanić, technician Isabel Vanoverberghe)
8.2. Who will be responsible for data storage & back up during the project?	All PIs (Ellen Decaestecker, Adam Petrusek and Luc De Meester) and the day-to-day manager of the CELSA project (PhD student Nataša Katanić, technician Isabel Vanoverberghe)
8.3. Who will be responsible for ensuring data preservation and sharing?	All PIs (Ellen Decaestecker, Adam Petrusek and Luc De Meester) and the day-to-day manager of the CELSA project (PhD student Nataša Katanićn technician Isabel Vanoverberghe)
8.4. Who bears the end responsibility for updating & implementing this DMP?	All PIs (Ellen Decaestecker, Adam Petrusek and Luc De Meester) and the day-to-day manager of the CELSA project (PhD student Nataša Katanić, technician Isabel Vanoverberghe)