Assessing the distribution and abundance of native and invasive weather loach to develop more effective management and conservation strategies: An interdisciplinary approach using eDNA, genetic profiling, telemetry and species distribution modelling

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.

Dataset name / ID	Description	New or reuse	Digital or Physical data	Data Type	File format	Data volume	Physical volume
		Indicate: ***N****(ew data) or* ***E****(xisting data)*	Indicate: **D**(igital) or **P**(hysical)	Indicate: **A**udiovisual **I**mages **S**ound **N**umerical **T**extual **M**odel **SO**ftware Other (specify)		Indicate: <1GB <100GB <1TB <5TB >5TB NA	
Field data	Hardcopy notebooks	N	P	Textual			One small box (30 x 40 x 25 cm)
Fish samples	Fins	N	P	Observational			Stored in ethanol in small containers at room temperature in dedicated closets
Water samples	Water samples for eDNA analyses	N	P	Observational			Stored in small containers
DNA	DNA extracts	N	P				Stored in Eppendorf tubes at -80°C in dedicated freezers
Sequences	Sequencing data	N	D	Textual	.FASTQ	<100GB	
R code	R code describing statistical analyses	N	D	Textual	.R	<1GB	
Environment	Summary tables and data sheets	N	D	Numerical	.xlsx	<1GB	
Papers	Word and pdf files of papers resulting from this project	N	D	Textual	.pdf .docx	<1GB	

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:

Not applicable

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.

- No

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.

- No

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.

- No

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).

For each WP, the (e-)lab books will contain information on experimental design, sampling methodology, fieldwork, sampling location (GPS coordinates), variable-level detail, and all information necessary for a secondary analyst to use the data accurately and effectively. A clear coding for all data files related to the project will be used. These will have the form: WPX_TaskY_yyyymmdd_NameExperiment. In addition, templates for this documentation will be provided to all researchers associated to the project to allow consistent documentation.

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.

- Yes

Metadata with experimental procedures (including sampling location, date and handling person, etc.), sampled species, preparation information, storage will be saved to read and interpret the data for other users in the future. Persons that perform the experiments and generate the data will document this information. In the data management system (openBIS ELN-LIMS), metadata are provided as attributes of the respective datasets. Based on the defined metadata scheme, openBIS ELN-LIMS will be configured so that the required metadata is automatically assigned to datasets and/or manually provided by the researcher.

Data Storage & Back-up during the Research Project

Where will the data be stored?

- Shared network drive (J-drive)
- Personal network drive (I-drive)
- Other (specify below)

The time-stamped master copy of the data will be kept on KU Leuven's central storage facility. In addition, copies will be made and kept on personal devices.

How will the data be backed up?

- Standard back-up provided by KU Leuven ICTS for my storage solution
- Personal back-ups I make (specify below)

The data is stored on the central servers of the university and these have automatic daily back-up procedures. In addition, personal back-ups will be made on a monthly basis on personal storage devices.

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

KU Leuven storage facilities allow to store up to 10TB, which is much more than will be collected in this project.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

The access to the folder containing all project related data is restricted to the project researchers and access can only be granted by the project coordinator.

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

None

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...).

- All data will be preserved for 10 years according to KU Leuven RDM policy

Sequencing data
Fin samples
DNA extracts
Notebooks
Data sheets

R code

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

- Other (specify below)
- Shared network drive (J-drive)

Fin samples will be kept in dedicated closets.
DNA extracts will be maintained at -80°C in the lab.
Digital data (sequencing data, pictures, data sheets, R code, pdfs) will be stored on KU Leuven's storage facilities.
Notebooks will be stored in dedicated closets of the promoter.

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

None

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, as open data

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

Question not answered.

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.

- No

Where will the data be made available?

If already known, please provide a repository per dataset or data type.

- Other data repository (specify below)
- KU Leuven RDR (Research Data Repository)

Sequencing data will be submitted to public databases, where they will be permanently archived to preserve access to the public.
Other data accompanying published papers will be archived in repositories such as DRYAD, GBIF and Zenodo.

When will the data be made available?

- Upon publication of research results

Which data usage licenses are you going to provide?

If none, please explain why.

None. Collected data can be freely used by third parties.

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.

- Yes, a PID will be added upon deposit in a data repository

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

None

Responsibilities

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

The PI (Hans Jacquemyn) and the day-to-day manager of the project (Jonathan Tibo)

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

The PI (Hans Jacquemyn) and the day-to-day manager of the project (Jonathan Tibo)

Who will manage data preservation and sharing?

The PI (Hans Jacquemyn) and the day-to-day manager of the project (Jonathan Tibo)

Who will update and implement this DMP?

The PI (Hans Jacquemyn)