DMP title

Project Name FWO DMP Teun Everts - DMP title

Project Identifier 3E211267

Grant Title 1S01822N

Principal Investigator / Researcher Teun Everts

Project Data Contact 0495795619, Distelsstraat 44 (bus 8b) 1030 Schaarbeek, teun.everts@inbo.be, teun.everts@kuleuven.be

Description This strategic project aims at further developing and tailoring eDNA-based methods, and ultimately embeding them as a corner stone in a decision suport system that can be used by managers to prevent invasion or facilitate eradication of an aquatic alien invasive species. More specifically, using the American bullfrog invasion in Belgium as a case study, environmental DNA (eDNA)-based methods are developed and combined with other techniques to (1) detect and quantify the density of aquatic alien invasive species (AIS), (2) reconstruct the invasion history, (3) determine the ecological impact on indigenous communities, and (4) forecast future spread.This project involves a close interplay between pure scientific research and the application of the generated results in practice, which is facilitated by the joint collaboration between the University of Leuven (KUL) and the Research Institute for Nature and Forest (INBO).

Institution KU Leuven

1. General Information

Name applicant

Teun Everts

FWO Project Number & Title

Project number: 1S01822N

Project title: A stitch in time saves nine - Developing an aquatic invasive alien species management framework based on eDNA methods

Affiliation

- KU Leuven
- Other

Second host institute: Research Institute for Nature and Forest (Instituut voor Natuur en Bosonderzoek; INBO)

2. Data description

Will you generate/collect new data and/or make use of existing data?

- Generate new data
- Reuse existing data

Describe in detail the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a table (see example) or as a data flow and per WP or objective of the project. If you reuse existing data, specify the source of these data. Distinguish data types (the kind of content) from data formats (the technical format).

Type of data	Format	Estimated volume	How created/source
DNA concentrations from geo-referenced environmental and tissue samples (all WP's)	.qlp & .csv	1 - 2 GB	Droplet digital polymerase chain reaction (ddPCR) of eDNA obtained from DNA extractions at the INBO facilities
Backup purified DNA isolated from environmental and tissue samples (all WP's)	physically in 2 mL eppendorf tubes	3-5 kg, 1-2 m²	DNA obtained from extractions from environmental samples executed and stored in temperature-logged freezers at the INBO facilities
Observational,geo-referenced environmental data of sampled water bodies (all WP's)	.xls	200-300 kB	Upon sampling the water bodies, a number of predefined habtitat characteristics are measured (either visually or instrumentally) and noted.
Exhaustive descriptions of protocols and workflows (including scripts and metabarcoding pipelines) developed and applied (all WP's)	pdf, R.files	50 - 100 kB	Written during the development and application by the applicant and his team
Historical observation records (WP1)	.xls	300 - 500 kB	Existing data publically generated and stored (https://waarnemingen.be/), available on request via host institution (INBO)
Raw DNA sequences of DNA fragments for population genetics (WP1) and metabarcoding (WP2)	FASTQ	1-2 TB	Generated from environmental and tissue samples with a MinIOn and Illumina MiSeq sequencer at the INBO facilities.

3. Legal and ethical issues

Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to your file in KU Leuven's Register of Data Processing for Research and Public Service Purposes (PRET application). Be aware that registering the fact that you process personal data is a legal obligation.

- No

Privacy Registry Reference: N/A
Short description of the kind of personal data that will be used: N/A

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

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

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

What documentation will be provided to enable reuse of the data collected/generated in this project?

1. All sampled water bodies will be labeled with a unique 10-symbol code with reference to the "watervlakkenlaag", published online by Geopunt Vlaanderen (https://www.geopunt.be/catalogus/datasetfolder/61c4245b-a177-4fe8-a5cc-455475d7b40f), associated with their respective coordinates.

2. Every sampling method and laboratory protocol (including Standard Operating Procedures; SOP's), but also data processing and analysis files, will be linked with a word-document comprehensively explaining the workflow and describing the labels for all variables, variable types, units of measurements and key identifiers.

3. All generated datasets will be coupled with a ReadMe file, carefuly describing and refering to the sampling method, laboratory protocol, data processing method, and data analysis workflow.

4. Metadata will be embedded in R-scripts consisting of step-by-step expkanations of what is being done and why.

Will a metadata standard be used? If so, describe in detail which standard will be used. If no, state in detail which metadata will be created to make the data easy/easier to find and reuse.

- Yes

Dublin Core will be used as a metadata standard, which is a general standard to describe resources, based on 15 core elements, and will be specifically adapted to this project by adding additional elements. Data associated with metadata will be deposited in online data repositories, such as Dryad and NCBI Sequence Read Archive, and the published manuscripts and reports will be deposited online in the open access repositories of KUL (Lirias) and INBO (PURE) and on social networking sites such as ResearchGate.

5. Data storage and backup during the FWO project

Where will the data be stored?

1. A master data set will be compiled per work package, with reference to comprehensive raw databases which were used to construct this data set, and with a ReadMe file containing the storage location, database name, associated processing description documents, etc. of the associated data. This will be stored in duplicate: on an external hard drive and on INBO's Google Drive (enterprise license with unlimited storage).

2. Raw DNA sequences will be uploaded to the NCBI Sequence Read Archive and individual sequences in NCBI's GenBank.

3. Only when the volume of generated data exceeds the foreseen data storage capacity specifically devoted to this project by INBO, the remaining part of the data will additionally be stored on the KU Leuven SharePoint.

How is backup of the data provided?

The data will be stored on the university's central servers with automatic daily back-up procedures, and weekly backups will be made on both INBO's Google Drive as well as an external hard drive.

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

Two 2 TB external hard drives are available to the PhD student for data storage, which easily covers the expected volume of data that will be generated. Additionally, 120 GB cloud storage capacity is provided by INBO, which can be extended with the Sharepoint possibilities provided by the KU Leuven.

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

KU Leuven and INBO cloud storage costs are covered by both host institutes independetly from the research project. One 2 TB hard drive is personal property of the applicant, the other will be bought from the bench fee for €100.

Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

The cloud storage of data will be protected by a password that is changed twice a year, and is firmly secured by both host institutes. Further, a selection will be made of datasets to which only controlled access is granted, such as 'read only', 'read and write', or 'administrator only' permission, to avoid loss of valuable data.

6. Data preservation after the FWO project

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, ...).

All generated digital and physical data will be stored for at least five years after completion of this research on the external hard drives and on the cloud storage locations. This data storage period is already foreseen for each PhD project at both host institutes.

Where will the data be archived (= stored for the longer term)?

The data will be stored on the university’s central servers with autmatic back-up procedures and on the cloud storage of INBO for at least 10 years, conform KU Leuven & INBO’s RDM policy, respectively. All research outputs will additionally remain publicaly available and thus archived on ResearchGate, and on INBO’s PURE and KUL’s Lirias platforms.

What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?

Data storage in the cloud and on the university’s central servers will be covered by the KU Leuven and INBO, respectively, as part of their general data management plan.

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. Specify:

The historical observation records retrieved from data portal waarnemingen.be, hosted by Natuurpunt Studie vzw, can only be used when explicitely stating the origin of data, followed by the following statement “these data cannot be used without permission”. INBO can only use these data in the explicit context in which they were requested.

Which data will be made available after the end of the project?

All generated datasets will be uploaded in a csv format in online data repositories such as Zenodo, Dryad... under a CC-BY license. Raw DNA sequences will be uploaded to the NCBI Sequence Read Archive and individual sequences in NCBI's GenBank. All output of the research will be made publicaly available by uploading to Research Gate, and will be stored on INBO’s PURE and KUL’s Lirias. The historical observation records used as part of this research will not be uploaded online, as we only have the right to use and not the spread the data.

Where/how will the data be made available for reuse?

- In an Open Access repository

1. The full dataset with documentation will be uploaded in csv format in Zenodo and Dryad under a CC-BY license.
2. The used metabarcoding pipelines, if modified based from earlier published pipelines, will be uploaded on GitHub.

3. Raw DNA sequences will be uploaded to the NCBI Sequence Read Archive and individual sequences in NCBI's GenBank.

When will the data be made available?

- Immediately after the end of the project

The full generated dataset will be uploaded in a csv format in Zenodo and Dryad under a CC-BY license immediately upon publication of manuscripts, and at the end of the project (after publication).

Who will be able to access the data and under what conditions?

Anyone will be able to access and use the data under the CC-BY license, provided that they give appropriate credit to the creators.

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

Besides €100 for a 2 TB external hard drive that will be taken from the bench fee, no further costs are expected.

8. Responsibilities

Who will be responsible for data documentation & metadata?

Teun Everts (applicant)

Aaike De Wever (data manager INBO)

Who will be responsible for data storage & back up during the project?

Teun Everts (applicant)

Hans Jacquemyn (promotor, KUL)

Rein Brys (promotor, INBO)

Aaike De Wever (data manager INBO)

Who will be responsible for ensuring data preservation and reuse ?

Teun Everts (applicant)

Hans Jacquemyn (promotor, KUL)

Rein Brys (promotor, INBO)

Aaike De Wever (data manager INBO)

Who bears the end responsibility for updating & implementing this DMP?

The PI bears the overall responsibility for updating & implementing this DMP