

A comparative analysis of ORganic and Conventional Agriculture's impact on aquatic biodiversity (ORCA; BR/175/A1/ORCA) Data Management Plan

De Wever, A.^{1,2}, Cours, M.¹, Milotic, T.², Adriaens, D.², Lemmens, P.³, Denys, L.², Martens, K.¹, De Meester, L.³

¹ Operational Directorate Natural Environment, Royal Belgian Institute of Natural Sciences, Brussels, Belgium – Aaike De Wever, orcid.org/0000-0001-8804-8321; Marie Cours; Koen Martens, orcid.org/0000-0001-8680-973X

² Research Institute for Nature and Forest, Brussels, Belgium – Tanja Milotic, orcid.org/0000-0002-3129-619

³ Ecology Evolution Biodiversity Conservation, KULeuven, Leuven, Belgium – Pieter Lemmens, orcid.org/0000-0002-3135-9724; Luc De Meester, orcid.org/0000-0001-5433-6843

Plan details

Grant number: BR/175/A1/ORCA and BR/175/A1/ORCA-2.

Project Name: A comparative analysis of ORganic and Conventional Agriculture's impact on aquatic biodiversity (ORCA)

Project website: bio.kuleuven.be/eeb/ldm/ORCA

Principal Investigator:

Luc De Meester, KULeuven, Ecology Evolution Biodiversity Conservation, Leuven, Belgium.

Plan data contacts:

Marie Cours <mcours@naturalsciences.be>

Pieter Lemmens <pieter.lemmens@kuleuven.be>

Tanja Milotic <tanja.milotic@inbo.be>

Description: Data Management Plan of the Belspo BRAIN project “A comparative analysis of ORganic and Conventional Agriculture's impact on aquatic biodiversity” (ORCA). This project investigates the combined effects of agriculture type (organic compared to conventional) and land use intensity (extensive versus intensive) on aquatic biodiversity in ponds and shallow lakes. The data underlying the planned analyses is obtained from stratified sampling of farmland ponds, a common garden experiment, and through access to existing datasets.

Funder: Belgian Science Policy Office (Belspo)

Status of plan: Final version, March 2020



RESEARCH INSTITUTE
NATURE AND FOREST



Data set reference and name

ORCA freshwater database

The title “ORCA freshwater database” is a working title for the database that is constructed in the course of the Belspo funded ‘ORCA’ project “A comparative analysis of ORganic and Conventional Agriculture’s impact on aquatic biodiversity”.

The project data consist of different datasets that have been collected in relation to the ponds sampled during the ORCA project. However, the land use dataset also includes information for a set of ponds that were sampled during previous projects (cfr. SAFRED database, see below).

This data management plan aims to provide an overview on how these data will be organized and made publicly available following the FAIR (Findable, Accessible, Interoperable and Reusable) principle (Wilkinson et al. 2016). The data management plan does not cover topics such as field protocols and sample labeling in the field since these aspects of data management were already fixed at an earlier stage of the project.

Planned update cycle

The present version of the document is the final version of the ORCA Data Management Plan. This file is an update of the initial version (DOI:<https://doi.org/10.6084/m9.figshare.7546694.dv1>).

Data set description

Project background

Agriculture is one of the most pervasive human activities on earth, impacting key natural resources (Rockström et al. 2009, Steffen et al. 2015), e.g. through eutrophication and the use of pesticides (Smith et al. 1999, Beketov et al. 2013). The impact of organic farming is expected to differ from that of conventional farming because only organic fertilizers are used, the strongly restricted use of a limited set of pesticides, and the protection of more natural elements such as ponds from adverse effects by larger buffer zones. In the ORCA project we use pond systems to investigate and assess the impact of agricultural on aquatic biodiversity in a comparative analysis of organic and conventional agricultural practices.

For this purpose, we combine multiple existing databases with newly collected data;

- (1) Sampling localities for collecting additional data were selected to ensure sufficient coverage of each type of agriculture (organic versus conventional) and a range in land-use intensity (extensive versus intensive) to enable comparative analysis of their combined effects on aquatic biodiversity in ponds. More specifically, we have sampled ponds along a gradient of cropland versus grassland-dominated areas embedded (to various degree) in organic (26 ponds) or conventional farmland (22 ponds; i.e. a total of 48 sampled ponds). We collected data on a broad range of environmental variables (>20 variables, cf. De Bie et al. 2012), amphibians, fish, macrophytes, macro-invertebrates, and zooplankton.
- (2) In addition, we use existing data collected in previous projects using similar sampling techniques, in particular those compiled during the Belspo BRAIN project SAFRED (Saving Freshwater Biodiversity Research Data).

Using the newly collected data —following the stratified sampling design outlined above— in combination with existing data, we assess the effects of different types of agricultural practices on

taxonomic, functional and intraspecific genetic diversity of multiple aquatic organism groups at multiple spatial scales.

Newly generated data

We performed detailed environmental characterization of the sites selected for sampling and quantify community composition and diversity of five different aquatic organism groups: zooplankton, macro-invertebrates, macrophytes, amphibians, and fish. These groups are both key targets for biodiversity conservation and important determinants of the ecological functioning and integrity of standing waters. For instance, macrophytes, zooplankton and fish are pivotal players in the context of regime shifts among alternative stable states in ponds and shallow lakes (Scheffer et al. 1993, Scheffer et al. 2001). Macro-invertebrates, zooplankton and macrophytes are sampled and quantified using well-established methods. Fish and amphibian community composition and diversity are analyzed using environmental DNA (eDNA) metabarcoding as this provides a more reliable and less intrusive way to screen for their presence and relative abundance.

In summary, the following data are generated:

- Geographic information of sampling locations
- General overview of selected ponds and contact information of pond owners
- Field observations on pond characteristics such as surrounding land use and management
- Field measurements of pH, EC, temperature and oxygen
- Photographic documentation of ponds and sampling sites
- Landscape and land use characteristics of selected ponds, as derived from existing GIS layers
- Data on presence and characteristics of bufferstrips
- Laboratory measurements of dry and suspended matter
- Laboratory measurements of major ions and nutrients
- Zooplankton community composition and diversity – counts
- Macro-invertebrate community composition and diversity – counts
- Macrophytes community composition and diversity – field survey, cover data
- Fish community composition and diversity – eDNA analyses
- Amphibia community composition and diversity – eDNA analyses
- Data related to the common garden experiment
- Population genomic and quantitative genetic data

Details on each of these datasets including the variables measured and responsible partners/persons are provided in Annex 1: Overview of different data types.

Existing data

The key existing data on which the project builds falls into two categories; (1) the data from earlier pond projects compiled in the course of the Belspo BRAIN project SAFRED (Saving Freshwater Biodiversity Research Data) will be used in the integrated data analysis, and (2) information on the land use at different perimeters (from 50 to 3000 m) around the newly sampled water bodies and the ponds studied in previous research projects that are currently included in the SAFRED database. Land use information is derived from multiple existing GIS layers. For the Flemish region, these include the Biological Valuation Map, yearly agricultural crop registration, rural development program/agri-environment measures, NARA landgebruikskaart (Poelmans & Van Daele, 2014) and other land use and land cover inventories, depending on the time of sampling. The GIS-layer Watervlakken 1.0 - which includes all standing waterbodies >1.5 m² in the Flemish region (Packet et al. 2018) – is used as a basemap for pond location. In the Brussels and Walloon region pond contours were derived from topographic maps or large-scale geographic data layers (a.o. PICC), and land use

information from La Carte d'Occupation du Sol de Wallonie and LifeWatch land cover mappings. The data from previous projects including the BELSPO projects Manscape, Pondscape and Tommelen, the EU funded project BIOMAN (EVK-CT-1999-00046), the *Toegepast Wetenschappelijk Onderzoek Leefmilieu (TWOL)* project Midden-Limburg, National Fund of Scientific Research Flanders project De Maten (grant G.0358.01), and the BiodivERSA project Tippingpond were compiled and standardised (to the Darwin Core standard; tdwg.org/standards/dwc) in the course of the SAFRED project. These datasets were made available through data.freshwaterbiodiversity.eu/ipt and the Global Biodiversity Information Facility (GBIF; gbif.org) network, thus making them publicly available for use in overarching analyses of biodiversity patterns.

Centralizing and integrating new and existing data

To ensure the efficient centralization and integration of newly generated data and existing data we discussed the data flows from the different labs and data types. Based on this we implemented a workflow to ensure that information collected in the field, during lab analyses, and those extracted from multiple existing GIS layers becomes integrated in an overall database, which is made available to all partners.

Workflow to successfully obtain the ORCA freshwater database:

- Responsibilities and timeline of different datasets described in Annex 1
- Background information on sampling localities and protocols is recorded along with other metadata in a separate excel worksheet or in ReadMe-files
- As central repository for data storage and exchange (Microsoft OneDrive) was established at RBINS
- We use the following naming convention based on the recommendations found in the Data Management guidelines of SAFRED (tinyurl.com/y7olom3e):
 - YYYY_IN_ORCA_Name.csv
Which includes the capture date/year (YYYY), initials of the investigator (IN), project acronym (ORCA), and the dataset name (Name)
- Integrating the compiled datasets in an overall “ORCA freshwater database” and producing export queries for publishing biodiversity data in Darwin Core format (RBINS)

In order to integrate the “ORCA freshwater database” with data from previous projects (cfr. SAFRED database), datasets and variables are streamlined. This allows us to perform integrated analyses and produce products such as the “priority map on areas where biodiversity gains can be achieved by land use change” (Work package 4).

Standards and metadata

Information on the datasets (metadata)

As outlined in the project proposal, background information on sampling localities and protocols are recorded along with other metadata and kept together with the actual data. In practice, the information on the (sub)datasets will initially be captured in universally readable ReadMe-files (.txt, .md or .rft) covering the topics outlined below and will be compiled and provided along with the raw data for integration in the overall ORCA freshwater database. Metadata of the overall ORCA freshwater database or selected components considered for (separate) publication will be entered or transferred to the Freshwater Metadatabase at data.freshwaterbiodiversity.eu/metadb (which is part of the SAFRED infrastructure and supports metadata export in EML standard) and will be published in the Freshwater Metadata Journal (<http://data.freshwaterbiodiversity.eu/fmj/>). For

existing datasets (i.e. those in the scope for the SAFRED project), metadata for component datasets will be imported/integrated from existing sources where available. If this is not the case, metadata will be encoded by project partners.

Minimal required information on dataset (metadata) includes:

- Title/Name of the dataset
- Corresponding filename of the dataset
- Short description
- Keywords
- Contact person(s)
- Description of fields/column names including measurement units
- Sampling methods/protocol or reference of paper containing method description
- Time frame: sampling period
- Version

Use of standard controlled vocabularies

We use the same controlled vocabulary for pond characterization as used in the projects which are part of the SAFRED database. In this way the overarching analysis is facilitated. The unique identifiers of Watervlakken 1.0 provide a standard reference for identifying the water bodies and linking data from different analyses. We use these unique identifiers to refer to the individual ponds sampled.

Standards for organizing and sharing data

The database structure of the ORCA freshwater database will be reviewed in the course of the project, considering the need for creating export queries matching the Darwin Core standard (tdwg.org/standards/dwc) at a later stage. In parallel, we will identify the Darwin Core terms/fields (rs.tdwg.org/dwc/terms) which are most relevant in the context of this database.

Biodiversity data that will be released in the framework of ORCA will be exported in Darwin Core format and mapped on the GBIF Freshwater Network through the FIP(BioFresh) Integrated Publishing Toolkit (IPT) software (see details under data sharing).

Data sharing

As explained under “Existing data”, the data from former projects as compiled during the SAFRED project are already publicly available. Newly generated data have been made available to project partners in the framework of joint analyses as outlined in the section “Centralizing and integrating new and existing data”. The construction of this Data Management Plan is considered as a tool to discuss and streamline the integration of the newly generated data and facilitate its integration in the SAFRED infrastructure at RBINS and public release of the data (after an embargo period allowing for processing and scientific publication based on the data). Note however, that the data publication will be prepared before the end of the project to assure that all data can be made available in parallel to the associated scientific papers or at latest after the embargo period (minimum 2 year). Preparatory work for publication of data will be done by the end of the project. Data publication will focus on biodiversity related data using the tools available through the Global Biodiversity Information Facility (gbif.org) network, thereby referring to the workflow worked out during SAFRED and adopting the supported options for data citation and licensing. We consider additional data archiving through a general-purpose repository such as Dryad (datadryad.org) for the integrated

dataset and the common garden experiment data in parallel to the publication of the peer-review scientific paper.

Ethics and legal compliances

Most of the ponds sampled within the ORCA project are situated on private properties. To assure privacy of the landowners we will not share the exact location of the ponds. Therefore, coordinates (decimal degrees, EPSG 4326) will be truncated to 2 decimal places.

Biodiversity related data will be published under an Open Data License (Creative Commons Attribution License (CC-BY) or Creative Commons Attribution-Noncommercial License (CC-BY-NC)).

Being a BELSPO-BRAIN project, the data management within ORCA needs to comply with its regulations for projects and the [BELSPO Open Research Data Policy](#). As one of the project partners the Research Institute for Nature and Forest (INBO) has its own [Open Data Policy](#). The data management in ORCA is also in line with this policy.

Archiving and preservation (including storage and backup)

Backup of (sub)datasets

It is the project partners' responsibility to ensure that the data are properly backed up during the compilation of the raw data. As explained in the Data Management guidelines worked out during SAFRED (tinyurl.com/y7olom3e – see Data security section), we recommend applying the 3-2-1 rule: 3 copies, 2 media, 1 offsite copy. As data files for most types of data are typically <10Mb (with the possible exception of the eDNA analysis results), file size and storage capacity are unlikely to pose any problems.

Archiving of overall ORCA freshwater database

Organizing the back-up and archival of the centralized files and overall ORCA freshwater database is the responsibility of RBINS. Long-term archival at an institutional archive will be done by each project partner using existing data storage infrastructure (for KU Leuven: Box Cloud Storage, for RBINS: Microsoft OneDrive). The integrated dataset will be additionally stored at a general purpose repository (such as Dryad—datadryad.org for data associated with scientific papers) or Zenodo (zenodo.org) after acceptance of the associated peer-review scientific paper.

The raw eDNA data and zooplankton population genomic data will be stored in one of the [INSDC \(International Nucleotide Sequence Database Collaboration\) databases](#) (such as the EMBL-EBI or NCBI-GenBank). Metadata associated with the raw eDNA data will be stored using MIxS (Minimum Information on any (x) Sequence) standard for genomic metadata.

Archiving of ORCA freshwater samples

Physical macroinvertebrate and zooplankton samples collected during the ORCA project will be incorporated in the collections of the Royal Belgian Institute of Natural Sciences (RBINS). To this end, the collection number (Coll. RBINS IG 33.632) has been assigned to this collection and a label with this number will be attached to the ORCA samples during processing. Once analysis of the samples in

the framework of ORCA has been finalised, they can be transferred to the collection for long-term archival.

References

Beketov, M. A., B. J. Kefford, R. B. Schäfer, and M. Liess. 2013. Pesticides reduce regional biodiversity of stream invertebrates. *Proceedings of the National Academy of Sciences USA* 110:11039-11043.

De Bie, T., L. De Meester, L. Brendonck, K. Martens, B. Goddeeris, D. Ercken, H. Hampel, L. Denys, L. Vanhecke, K. Van der Gucht, J. Van Wichelen, W. Vyverman, and S. A. J. Declerck. 2012. Body size and dispersal mode as key traits determining metacommunity structure of aquatic organisms. *Ecology Letters* 15:740-747.

Declerck, S., T. De Bie, D. Ercken, H. Hampel, S. Schrijvers, J. Van Wichelen, V. Gillard, R. Mandiki, B. Losson, D. Bauwens, S. Keijers, W. Vyverman, B. Goddeeris, L. De Meester, L. Brendonck, and K. Martens. 2006. Ecological characteristics of small farmland ponds: Associations with land use practices at multiple spatial scales. *Biological Conservation* 131:523-532.

Packet J., Scheers K., Smeekens V., Leyssen A., Wils C., Denys L. 2018. Watervlakken versie 1.0: polygonenkaart van stilstaand water in Vlaanderen. Een nieuw instrument voor onderzoek, water-, milieu- en natuurbeleid. Rapporten Instituut voor Natuur- en Bosonderzoek INBO.R.2018.14, 43 p.

Poelmans L. and T. Van Daele. 2014. Landgebruikskaart NARA-T 2014. Rapporten Vlaamse Instelling voor Technologisch Onderzoek. 2014/RMA/R/45, 55 p.

Rockström, J., W. Steffen, K. Noone, Å. Persson, F. S. Chapin, E. F. Lambin, T. M. Lenton, M. Scheffer, C. Folke, and H. J. Schellnhuber. 2009. A safe operating space for humanity. *Nature* 461:472-475.

Scheffer, M., S. Carpenter, J. A. Foley, C. Folke, and B. Walker. 2001. Catastrophic shifts in ecosystems. *Nature* 413:591-596.

Scheffer, M., S. H. Hosper, M. L. Meijer, B. Moss, and E. Jeppesen. 1993. Alternative equilibria in shallow lakes. *Trends in Ecology & Evolution* 8:275-279.

Smith, V. H., G. D. Tilman, and J. C. Nekola. 1999. Eutrophication: impacts of excess nutrient inputs on freshwater, marine, and terrestrial ecosystems. *Environmental Pollution* 100:179-196.

Steffen, W., K. Richardson, J. Rockström, S. E. Cornell, I. Fetzer, E. M. Bennett, R. Biggs, S. R. Carpenter, W. de Vries, and C. A. de Wit. 2015. Planetary boundaries: Guiding human development on a changing planet. *Science* 347:1259855.

Wilkinson, M. D., M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*. 3: 160018.

Annex 1: Overview of different data types

Data types: description of available/expected data & responsible parties

Description	Type	File format	File/Folder name	Responsible partner	Responsible/Contact person
Geographic information of sampling locations	Sampling event details	Excel	2017_MC_ORCA_PondCoord.xlsx	RBINS	Marie Cours
General overview of selected ponds and contact information of the pond owners	Sampling event details	Excel	2017_MC_ORCA_ContactInformation.xlsx	RBINS	Marie Cours
Field observations on pond characteristics such as surrounding land use and management and Field measurements (pH, EC, Temp, oxygen)	Sampling event details/ measurements	Excel	2017_MC_RA_ORCA_environmental.xlsx	RBINS/ KU Leuven	Marie Cours/ Rafaela Almeida
Photographic documentation of ponds and sampling sites - images from 2017	Sampling event details		2017_MC_ORCA_Images	KBIN	Marie Cours
Photographic documentation of ponds and sampling sites - images from 2018	Sampling event details		2018_RB_ORCA_Images	INBO	Rein Brys
Landscape and land use characteristics of selected ponds, as derived from existing GIS layers	Measurements	Excel	2017_DA_ORCA_SAFRED_LandUse.xlsx	INBO	Dries Adriaens
Data on presence and characteristics of bufferstrips (field + GIS)	Measurements / Sampling event details	Excel	2017_JP_DA_ORCA_BufferStrips.xlsx	INBO	Jo Packet/ Dries Adriaens
Laboratory measurements of dry and suspended matter (KULEuven)	Measurements	Excel	2017_RA_ORCA_SuspendedMatter_Rawdata.xlsx	KU Leuven	Rafaela Almeida
Laboratory measurements of major ions and nutrients (RBINS)	Measurements	Excel	2017_MC_ORCA_WaterChemistry_Rawdata.xlsx	RBINS	Marie Cours
Zooplankton community composition and diversity – counts	Occurrences	Excel	2017_RA_ORCA_Zooplankton.xlsx	KU Leuven	Rafaela Almeida
Macro-invertebrate community composition and diversity – counts	Occurrences	Excel	2017_MC_ORCA_MacroInvertebrates.xlsx	RBINS	Marie Cours
Macrophytes community composition and diversity – field survey, cover data	Occurrences	Excel	2017_JP_LD_ORCA_Macrophytes.xlsx	INBO	Luc Denys/Jo Packet
Fish community composition and diversity – eDNA analyses	Occurrences	Excel	2018_RB_ORCA_eDNA_Fish.xlsx	INBO	Rein Brys
Amphibia community composition and diversity – eDNA analyses	Occurrences	Excel	2018_RB_ORCA_eDNA_Amphibia.xlsx	INBO	Rein Brys
Data related to the common garden experiment			YYYY_RA_ORCA_CommonGarden	KU Leuven	Rafaela Almeida
Population genomic and quantitative genetic data	Occurrences	FASTQ	YYYY_RA_ORCA_GENzooplankton	KU Leuven	Rafaela Almeida