Plan Overview

A Data Management Plan created using DMPonline

Title: PANORAMIX - Thyroid Hormone Bioassay related outputs on pooled real life sample extracts

Creator:Maria Margalef

Principal Investigator: Timo Hamers, Maria Margalef, Marja Lamoree

Data Manager: Maria Margalef, Marja Lamoree

Project Administrator: Marja Lamoree

Contributor: Veronica von Saint Paul

Affiliation: Vrije Universiteit Amsterdam

Funder: European Commission

Template: Horizon 2020 Template

ORCID iD: 0000-0002-5733-3290

ORCID iD: 0000-0001-7932-9770

ORCID iD: 0000-0002-7373-7738

Project abstract:

While exposure to multiple chemicals raises concern, mixtures are only slowly making their way into regulatory risk assessment. It remains unknown which and how many chemicals drive mixture effects on the environment and on humans. The EU-funded PANORAMIX project will develop an innovative experimental path based on whole mixture assessments for identifying and quantifying the risk of chemical mixtures extracted from real-life samples representing the environment, food and humans. The project will use mixture modelling, case studies and experimental data to deliver a web-based, ready-to-use and practical tool for chemical mixtures risk assessment, contributing to the EU's goals for a toxic- and pollution-free environment.

The toxicological impact of exposure to chemical mixtures is a matter of undisputed concern, but mixtures are only slowly making their way into regulatory risk assessment. Critical knowledge gaps are which and how many chemicals drive mixture effects in the environment and in humans. Scientific uncertainty remains on the validity of the dose addition principle for complex mixtures of large numbers of chemicals at low concentrations as they occur in our bodies. The PANORAMIX consortium addresses these challenges by showcasing a novel experimental path based on whole mixture assessments for identifying and quantifying the risk of chemical mixtures extracted from real-life samples representing environment and food as well as humans. We provide ready-to-use and practical tools for mixture risk assessment of several chemicals with a diverse range of adverse health outcomes. The applied methodologies, including a panel of in vitro assays coupled with effect-directed analyses and large-scale suspect and non-targeted chemical profiling are innovative in their combinatorial approach. Specifically, we will take advantage of a well-studied human cohort of new-borns, in whom adverse health outcomes related to developmental toxicity originating from chemical mixture exposure will be identified. PANORAMIX will use mixture modelling, case studies and experimental

data to deliver a web-based interface for calculating risks to chemical mixtures and to define effect-based trigger values for in vitro effects that can be directly measured in water, food, and blood to identify when mixture exposure is posing a health threat. By involving regulatory and scientific stakeholders throughout the project, we support the implementation of existing mixture risk assessment and management approaches to reduce the most critical exposures and assist in optimizing regulatory approaches to yield evidence-based policies, contributing to EU's zero-pollution ambition for a toxic free environment in the future.

ID: 122427

Start date: 01-11-2021

End date: 30-04-2026

Last modified: 21-01-2026

Grant number / URL: 101036631

PANORAMIX - Thyroid Hormone Bioassay related outputs on pooled real life sample extracts - Initial DMP

1. Data summary

Provide a summary of the data addressing the following issues:

- State the purpose of the data collection/generation
- Explain the relation to the objectives of the project
- Specify the types and formats of data generated/collected
- Specify if existing data is being re-used (if any)
- Specify the origin of the data
- State the expected size of the data (if known)
- Outline the data utility: to whom will it be useful

The data collected in this project consist of bioassay responses, dose–response metrics, and sample metadata derived from environmental, food, and human matrices (e.g., wastewater, surface water, fish, milk, and human serum/breast milk). These data were generated to:

- Characterize the biological activity of real-life chemical mixtures using a panel of in vitro assays, including endocrine-related, developmental toxicity, neurotoxicity, oxidative stress, and receptor-based modes of action.
  (These methods directly align with PANORAMIX's emphasis on effect-based and whole-mixture testing using bioassays.) [panoramix-h2020.eu]
- Quantify mixture effects and identify samples posing potential health risks, using metrics such as EC10/20/50 and IC10/20, as part of effect-directed analysis.
- Provide data inputs for mixture modelling and risk characterization, supporting the development of new tools for exposure and risk estimation.
- Enable cross-matrix comparison (environment → food → humans), reflecting real-world exposure pathways.
- Support the development of effect-based trigger values that can be used to screen water, food, and blood samples for potential mixture toxicity.
  (A stated methodological goal of the PANORAMIX project.)

In summary, the primary purpose is to measure, compare, and interpret the biological activity of complex chemical mixtures extracted from real samples, bridging hazard assessment and regulatory needs.

Types of data:

This dataset comprises in vitro bioassay results and QA/QC metadata from the PANORAMIX Phase I study investigating thyroid-related and broader toxicological activities in environmental and food matrices. The Excel workbook contains 7 sheets:

- T1 Samples: codes and sample types (wastewater influent/effluent, surface & drinking water, bottled water, multiple fish categories, cow milk, human breast milk, and serum panels S1–S5, plus blanks and procedural controls).
- T2 Bioassays: list of effect groups, bioassays and QA/QC parameters (e.g., Z'-factors, reference chemicals, robustness indices).
- T3 THR-TA / T4 TTR / T5 TBG: processed dose–response outcomes (e.g., EC10/20/50, IC10/20), curve fits and goodness-of-fit metrics for thyroid-related endpoints.
- T6 chemical and mixtures TTR: single chemicals & defined mixtures evaluated in TTR binding.

How are the data generated?
Bioassay outputs follow standard procedures per sheet T2 and include internal references (e.g., T3 THR-TA uses T3 as reference agonist; T4/T5 use T4 in competitive binding).

Relationships to other data:
Sheets share sample codes (e.g., WW, EFF, SW, OCM, HBM, S1–S5) enabling cross-assay comparison.

Data is not being re-used, and is produced at Vrije Universiteit Amsterdam, De Boelelaan 1085, 1081 HV Amsterdam (The Netherlands) www.vu.nl

Total size of numeric bioassay data is in the range of 0.2 MB.

2. FAIR data

2.1 Making data findable, including provisions for metadata:

- Outline the discoverability of data (metadata provision)
- Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?
- Outline naming conventions used
- Outline the approach towards search keyword
- Outline the approach for clear versioning
- Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

Metadata is typically written in lab notebooks or in readme.txt files saved in the same folder as the data.

The file structure of the project is as follows:

- Raw data is stored in the project folder at the VU and can be available upon request.

- The project folder contains different sub-folders (study name)

- Raw data is accompanied by readme.txt (Experimental details) files with the specifics of each experiment.

- Analysis (graphpad) files are also included with the raw data (generally excel files) if applicable, and can be available upon request.

2.2 Making data openly accessible:

- Specify which data will be made openly available? If some data is kept closed provide rationale for doing so
- Specify how the data will be made available
- Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?
- Specify where the data and associated metadata, documentation and code are deposited
- Specify how access will be provided in case there are any restrictions

-All processed data will be made openly available once the article is published in the Zenodo Project Environment. A DOI will be created for it.

Data will be archived in the VU data repository (Yoda):

- SurfDrive for processed data and metadata for:

- Bioassay responses
- Fractionation
- Mass spectrometry identification of active fractions
- Aligment of bioassay responses and chemical analysis

- Scistor for high resolution mass spectrometry raw data files that are actively being processed or not published.

- YODA for high resolution mass spectrometry raw data files that have been processed and published and need to be archived.

2.3 Making data interoperable:

- Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.
- Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

Since all data sets will be accompanied by metadata files explaining how data is processed and what kinds of data are shown, the interoperability of the data will be possible.

2.4 Increase data re-use (through clarifying licenses):

- Specify how the data will be licenced to permit the widest reuse possible
- Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed
- Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why
- Describe data quality assurance processes
- Specify the length of time for which the data will remain re-usable

All processed data will be made openly available for reuse once the article is published.

3. Allocation of resources

Explain the allocation of resources, addressing the following issues:

- Estimate the costs for making your data FAIR. Describe how you intend to cover these costs
- Clearly identify responsibilities for data management in your project
- Describe costs and potential value of long term preservation

The PANORAMIX Phase I dataset requires relatively modest resources to achieve full FAIR compliance because:

- The data are already structured in tabular form (Excel), which reduces preparation effort.
- FAIRification steps consist mainly of:
  - cleaning headers and column names,
  - exporting open formats (CSV/MD/PDF),
  - generating metadata (README, data dictionary),
  - preparing repository-ready documentation.

Estimated personnel time:

- 6–10 hours of researcher/data steward time for:
  - data curation (format normalization, QA),
  - metadata creation (README, variable dictionary),
  - preparing the repository deposit (DOI, license selection).
- No specialized software costs: all tasks rely on standard institutional tools or open-source formats (CSV, Markdown).

Estimated financial cost:

- €0–€300, depending on institutional policies:
  - Many repositories (e.g., Zenodo, 4TU.ResearchData) provide free deposit.
  - If a publisher requires a specific data journal submission, costs may increase, but this is outside core FAIRification.

How these costs will be covered:

- FAIR-related work will be performed by the research team (data curator + PI) within existing project time.
- Repository hosting costs, if any, will be supported by institutional RDM services or existing project budget allocations.

Primary responsibility:

- Dr. Maria Margalef Jornet (PI / VU Amsterdam) — responsible for overall data governance, FAIR compliance, and final approval of repository-ready materials.

Operational data management tasks:

- Data curation and standardisation: led by the project researcher generating the dataset, including cleaning, ensuring consistent column naming, file conversion to open formats, and creation of metadata.
- Storage, backup, and security oversight: handled by the institution’s IT services via secure network storage with automatic backups.
- Documentation and metadata development: shared between the PI and the data curator, with optional review by the

institutional Data Stewardship or RDM Support team.
- Repository submission and DOI assignment: performed by the PI or designated team member, following repository guidelines.

4. Data security

Address data recovery as well as secure storage and transfer of sensitive data

While processing, data is stored in the university one drive archive system. Security copies are made regularly. Only authorised personnel has reading access to the data.
After processing data will be made available, and a DOI will be adjudicated for publication and accessibility. The location of storage is still under discussion within the University section.
For the specific case of the PANORAMIX project, processed data, metadata and its linked DOI will be registered in the PANORAMX platform as well.

5. Ethical aspects

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

Not applicable

6. Other

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

Not applicable

PANORAMIX - Thyroid Hormone Bioassay related outputs on pooled real life sample extracts - Detailed DMP

1. Data summary

State the purpose of the data collection/generation

The data collected in this project consist of bioassay responses, dose–response metrics, and sample metadata derived from environmental, food, and human matrices (e.g., wastewater, surface water, fish, milk, and human serum/breast milk). These data were generated to:

- Characterize the biological activity of real-life chemical mixtures using a panel of in vitro assays, including endocrine-related, developmental toxicity, neurotoxicity, oxidative stress, and receptor-based modes of action.
  (These methods directly align with PANORAMIX's emphasis on effect-based and whole-mixture testing using bioassays.) [panoramix-h2020.eu]
- Quantify mixture effects and identify samples posing potential health risks, using metrics such as EC10/20/50 and IC10/20, as part of effect-directed analysis.
- Provide data inputs for mixture modelling and risk characterization, supporting the development of new tools for exposure and risk estimation.
- Enable cross-matrix comparison (environment → food → humans), reflecting real-world exposure pathways.
- Support the development of effect-based trigger values that can be used to screen water, food, and blood samples for potential mixture toxicity.
  (A stated methodological goal of the PANORAMIX project.) [explore.openaire.eu]

In summary, the primary purpose is to measure, compare, and interpret the biological activity of complex chemical mixtures extracted from real samples, bridging hazard assessment and regulatory needs.

Explain the relation to the objectives of the project

PANORAMIX aims to "identify and quantify the risk of chemical mixture exposures across the environment–food–human continuum" and evaluates real-life scenarios using mixtures extracted from relevant samples.
The dataset you provided directly reflects these sampling origins and methodological approaches.

Specify the types and formats of data generated/collected

Types of data:
This dataset comprises in vitro bioassay results and QA/QC metadata from the PANORAMIX Phase I study investigating thyroid-related and broader toxicological activities in environmental and food matrices. The Excel workbook contains 7 sheets:

- T1 Samples: codes and sample types (wastewater influent/effluent, surface & drinking water, bottled water, multiple fish categories, cow milk, human breast milk, and serum panels S1–S5, plus blanks and procedural controls).
- T2 Bioassays: list of effect groups, bioassays and QA/QC parameters (e.g., Z'-factors, reference chemicals, robustness indices).
- T3 THR-TA / T4 TTR / T5 TBG: processed dose–response outcomes (e.g., EC10/20/50, IC10/20), curve fits and goodness-of-fit metrics for thyroid-related endpoints.
- T6 chemical and mixtures TTR: single chemicals & defined mixtures evaluated in TTR binding.

How are the data generated?
Bioassay outputs follow standard procedures per sheet T2 and include internal references (e.g., T3 THR-TA uses T3 as reference agonist; T4/T5 use T4 in competitive binding).
Relationships to other data:
Sheets share sample codes (e.g., WW, EFF, SW, OCM, HBM, S1–S5) enabling cross-assay comparison.
Data is not being re-used, and is produced at Vrije Universiteit Amsterdam, De Boelelaan 1085, 1081 HV Amsterdam (The Netherlands) www.vu.nl

Specify if existing data is being re-used (if any)

Not at the moment

Specify the origin of the data

The data originate from the PANORAMIX H2020 project, which investigates chemical mixture toxicity using real-life samples collected across the environment–food–human continuum. PANORAMIX focuses on chemical mixtures present in water, food, and the human body, aiming to develop new scientific tools for mixture risk assessment. [slu.se]

Environmental and food samples

The dataset includes extracted mixtures from:

- Wastewater influent and effluent,
- Surface water and drinking water,
- Fish samples (fatty, lean, aquaculture and wild),
- Cow milk and human breast milk,
  representing real exposure sources for European citizens. These sample categories reflect the project objective of analysing mixtures present in water and food.

Human samples

Some data originate from human serum and breast milk extracts, consistent with PANORAMIX's aim to assess mixtures in the human body and to understand health outcomes related to exposure. The project explicitly studies mixtures occurring in humans and evaluates their potential developmental or endocrine effects.

Bioassay-generated data

All biological activity data (e.g., THR-TA, TTR, TBG, AREc32, MMP, DIO assays, AhR, PPARγ) were generated using in vitro whole-mixture bioassays, which are a core methodological pillar of PANORAMIX. The project uses whole mixture testing to link observed effects to real-life exposures and inform risk assessment.

State the expected size of the data (if known)

Total size of numeric bioassay data is in the range of 0.2 MB.

Outline the data utility: to whom will it be useful

Fellow scientist (toxicologists) with a research focus on the effect of complex chemical mixtures and their thyroid hormone system disruption capacity

2.1 Making data findable, including provisions for metadata [FAIR data]

Outline the discoverability of data (metadata provision)

Metadata is typically written in lab notebooks or in readme.txt files saved in the same folder as the data.

The file structure of the project is as follows:

- Raw data is stored in the project folder at the VU and can be available upon request.
- The project folder contains different sub-folders (study name)
- Raw data is accompanied by readme.txt (Experimental details) files with the specifics of each experiment.
- Analysis (graphpad) files are also included with the raw data (generally excel files) if applicable, and can be available upon request.

Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?

All processed data will be made openly available once the article is published in the Zenodo Project Environment. A DOI will be created for it.

Data will be archived in the VU data repository (Yoda):
- SurfDrive for processed data and metadata for:
Bioassay responses
Fractionation
Mass spectrometry identification of active fractions
Aligment of bioassay responses and chemical analysis
- Scistor for high resolution mass spectrometry raw data files that are actively being processed or not published.
- YODA for high resolution mass spectrometry raw data files that have been processed and published and need to be archived.

Outline naming conventions used

1. File naming conventions

All files use descriptive and machine-readable names:

- Lowercase or consistent TitleCase
- Words separated by underscores (_)
- No spaces, accents, or special characters
- Version numbers included when relevant

- T1_Samples_v1.0.csv
- T2_Bioassays_v1.0
- T3_THR_TA__v1.0.csv
- T4_TTR_v1.0.csv
- T5_TBG_v1.0.csv
- T6_ chemical_and_mixtures_TTR_v1.0.csv
- README.md

This ensures compatibility with software pipelines and repository systems.

2. Sheet / dataset naming conventions

The dataset preserves the structured table identifiers used within the PANORAMIX project:

- T1_Samples
- T2_Bioassays
- T3_THR_TA
- T4_TTR
- T5_TBG
- T6_Chemical_and_Mixtures_TTR

Each name starts with a table number (T1–T6) and ends with a short, descriptive label of the content or assay, and the version.

3. Variable naming conventions

Variable names:

- Use short, clear descriptors
- Avoid spaces and special characters
- Use consistent case formatting
- Do not embed units in column names

Examples:

- sample_code
- sample_type
- EC10, EC20, EC50

- IC10, IC20
- hillslope
- goodness_of_fit

Such conventions ensure seamless import into R, Python, MATLAB, and regulatory modelling tools.

4. Sample code conventions

Sample codes follow structured short alphanumeric identifiers established in PANORAMIX:

- Water: WW, EFF, SW, DW, BW
- Fish: FFW, FFA, LFW
- Milk: OCM, CCM, BM1–BM4, HBM
- Serum: S1–S5
- Blanks / controls: BWSPE, BWF, BF1, BF2, BHBM, BSSPE, etc.

These codes ensure clear linkage between sample metadata (T1) and assay outputs (T3–T6).

5. Versioning conventions

The project uses semantic versioning for all dataset releases:

- v1.0 — First public release
- v1.1, v1.2 — Minor updates or documentation refinements
- v2.0 — Major structural updates or additional data

Versions are included in filenames, changelogs, and repository metadata.

Outline the approach towards search keyword

Question not answered.

Outline the approach for clear versioning

The project uses semantic versioning for all dataset releases:

- v1.0 — First public release
- v1.1, v1.2 — Minor updates or documentation refinements
- v2.0 — Major structural updates or additional data

Versions are included in filenames, changelogs, and repository metadata.

Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

Question not answered.

2.2 Making data openly accessible [FAIR data]

Specify which data will be made openly available? If some data is kept closed provide rationale for doing so

The data that will be made available will be:

- Raw data
- Final results after data processing
- Statistical data obtained from Graphpad Prism (dose response curves, non linear regression results).

Specify how the data will be made available

The data will be made available upon publication via a DOI directly related to the ZENODO open repository.

Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?

Excel, or R

Specify where the data and associated metadata, documentation and code are deposited

Question not answered.

Specify how access will be provided in case there are any restrictions

Question not answered.

2.3 Making data interoperable [FAIR data]

Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.

The dataset is designed to be interoperable across research domains by using open, non-proprietary formats and following community-recognised metadata standards. Interoperability is further supported by the scientific framework of the PANORAMIX project, which uses harmonised methodologies for assessing real-life chemical mixtures across the environment–food–human continuum.
To ensure that the data can be easily exchanged, combined, and processed across systems:

- All datasets are provided in CSV, an open, widely supported, machine-readable format.
- Documentation is provided in Markdown (MD) and PDF, both human-readable and platform-independent.
- The original Excel file is preserved for traceability but is not required for interoperability.

Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

Controlled vocabularies and semantics
Where applicable, standardised vocabularies will be used:

- Assay terminology based on common toxicology and pharmacology usage (e.g., THR-TA, TTR, TBG).
- Sample type descriptors aligned with environmental and food science use (surface water, wastewater influent/effluent, fish tissue, serum, etc.).
  These categories match the PANORAMIX project's structured definitions of mixtures across environment, food, and humans.

2.4 Increase data re-use (through clarifying licenses) [FAIR data]

Specify how the data will be licenced to permit the widest reuse possible

Creative Commons Attribution (CC BY 4.0) licence

Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed

Data will be made available once the publication of the related manuscript is accepted in a peer reviewed journal

Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why

Under CC BY 4.0:

- Users may copy, redistribute, transform, and build upon the dataset for any purpose, including commercial use.
- Reusers must provide proper attribution to the creators, link to the licence, and indicate if changes were made.
- No additional restrictions (e.g., non-commercial clauses) are imposed, ensuring maximal interoperability and reuse potential, consistent with H2020 expectations for open access.
  (H2020 emphasises rights to “read, download, copy, distribute, search, link, crawl and mine” open data.)

There are no legal or ethical constraints on reusing the dataset, as it does not contain personally identifiable information. Extracts derived from human matrices (e.g., serum, breast milk) are anonymised and aggregated. Any components that cannot be shared publicly (if applicable) will be described in the DMP but excluded from the open deposit, as permitted under the flexible H2020 Open Research Data Pilot

Describe data quality assurance processes

The dataset already includes QA/QC metrics in T2 (e.g., Z'-factors, reference compounds, robustness indices) and fit diagnostics (R^2, Sum of Squares, Sy.x) within assay sheets (T3–T5).We will document:

- plate acceptance criteria and any excluded runs,
- calculation methods for EC/IC values and confidence/SE,
- cross-checks across replicates and controls (blanks, REF, etc.).

Any subsequent corrections will trigger a minor version update and a CHANGELOG entry.

Specify the length of time for which the data will remain re-usable

Question not answered.

3. Allocation of resources

Estimate the costs for making your data FAIR. Describe how you intend to cover these costs

The PANORAMIX Phase I dataset requires relatively modest resources to achieve full FAIR compliance because:

- The data are already structured in tabular form (Excel), which reduces preparation effort.
- FAIRification steps consist mainly of:
  - cleaning headers and column names,
  - exporting open formats (CSV/MD/PDF),

    - generating metadata (README, data dictionary),
    - preparing repository-ready documentation.

Estimated personnel time:

- 6–10 hours of researcher/data steward time for:
    - data curation (format normalization, QA),
    - metadata creation (README, variable dictionary),
    - preparing the repository deposit (DOI, license selection).
- No specialized software costs: all tasks rely on standard institutional tools or open-source formats (CSV, Markdown).

Estimated financial cost:

- €0–€300, depending on institutional policies:
    - Many repositories (e.g., Zenodo, 4TU.ResearchData) provide free deposit.
    - If a publisher requires a specific data journal submission, costs may increase, but this is outside core FAIRification.

How these costs will be covered:

- FAIR-related work will be performed by the research team (data curator + PI) within existing project time.
- Repository hosting costs, if any, will be supported by institutional RDM services or existing project budget allocations.

Clearly identify responsibilities for data management in your project

Primary responsibility:

- Dr. Maria Margalef Jornet (PI / VU Amsterdam) — responsible for overall data governance, FAIR compliance, and final approval of repository-ready materials.

Operational data management tasks:

- Data curation and standardisation: led by the project researcher generating the dataset, including cleaning, ensuring consistent column naming, file conversion to open formats, and creation of metadata.
- Storage, backup, and security oversight: handled by the institution’s IT services via secure network storage with automatic backups.
- Documentation and metadata development: shared between the PI and the data curator, with optional review by the institutional Data Stewardship or RDM Support team.
- Repository submission and DOI assignment: performed by the PI or designated team member, following repository guidelines.

Describe costs and potential value of long term preservation

- Repository hosting costs, if any, will be supported by institutional RDM services or existing project budget allocations.
- Raw data and processed data will be archived for a minimum of 5 years. This will allow reusability of data in potential retrospective analysis

4. Data security

Address data recovery as well as secure storage and transfer of sensitive data

While processing, data is stored in the university one drive archive system. Security copies are made regularly. Only authorised personnel has reading access to the data.
After processing and publication data will be made available, and a DOI will be adjudicated for publication and accessibility.

Published data will be archived in the University Server

For the specific case of the PANORAMIX project, processed data, metadata and its linked DOI will be registered in the PANORAMX platform as well.

5. Ethical aspects

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

Not applicable

6. Other

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

Not applicable

PANORAMIX - Thyroid Hormone Bioassay related outputs on pooled real life sample extracts - Final review DMP

1. Data summary

State the purpose of the data collection/generation

The data collected in this project consist of bioassay responses, dose–response metrics, and sample metadata derived from environmental, food, and human matrices (e.g., wastewater, surface water, fish, milk, and human serum/breast milk). These data were generated to:

- Characterize the biological activity of real-life chemical mixtures using a panel of in vitro assays, including endocrine-related, developmental toxicity, neurotoxicity, oxidative stress, and receptor-based modes of action.
  (These methods directly align with PANORAMIX's emphasis on effect-based and whole-mixture testing using bioassays.) [panoramix-h2020.eu]
- Quantify mixture effects and identify samples posing potential health risks, using metrics such as EC10/20/50 and IC10/20, as part of effect-directed analysis.
- Provide data inputs for mixture modelling and risk characterization, supporting the development of new tools for exposure and risk estimation.
- Enable cross-matrix comparison (environment → food → humans), reflecting real-world exposure pathways.
- Support the development of effect-based trigger values that can be used to screen water, food, and blood samples for potential mixture toxicity.
  (A stated methodological goal of the PANORAMIX project.)

In summary, the primary purpose is to measure, compare, and interpret the biological activity of complex chemical mixtures extracted from real samples, bridging hazard assessment and regulatory needs.

Types of data:

This dataset comprises in vitro bioassay results and QA/QC metadata from the PANORAMIX Phase I study investigating thyroid-related and broader toxicological activities in environmental and food matrices. The Excel workbook contains 7 sheets:

- T1 Samples: codes and sample types (wastewater influent/effluent, surface & drinking water, bottled water, multiple fish categories, cow milk, human breast milk, and serum panels S1–S5, plus blanks and procedural controls).
- T2 Bioassays: list of effect groups, bioassays and QA/QC parameters (e.g., Z'-factors, reference chemicals, robustness indices).
- T3 THR-TA / T4 TTR / T5 TBG: processed dose–response outcomes (e.g., EC10/20/50, IC10/20), curve fits and goodness-of-fit metrics for thyroid-related endpoints.
- T6 chemical and mixtures TTR: single chemicals & defined mixtures evaluated in TTR binding.

How are the data generated?
Bioassay outputs follow standard procedures per sheet T2 and include internal references (e.g., T3 THR-TA uses T3 as reference agonist; T4/T5 use T4 in competitive binding).

Relationships to other data:
Sheets share sample codes (e.g., WW, EFF, SW, OCM, HBM, S1–S5) enabling cross-assay comparison.

Data is not being re-used, and is produced at Vrije Universiteit Amsterdam, De Boelelaan 1085, 1081 HV Amsterdam (The Netherlands) www.vu.nl

Total size of numeric bioassay data is in the range of 0.2 MB.

Explain the relation to the objectives of the project

PANORAMIX aims to "identify and quantify the risk of chemical mixture exposures across the environment–food–human continuum" and evaluates real-life scenarios using mixtures extracted from relevant samples.
The dataset you provided directly reflects these sampling origins and methodological approaches.

Specify the types and formats of data generated/collected

Types of data:

This dataset comprises in vitro bioassay results and QA/QC metadata from the PANORAMIX Phase I study investigating thyroid-related and broader toxicological activities in environmental and food matrices. The Excel workbook contains 7 sheets:

- T1 Samples: codes and sample types (wastewater influent/effluent, surface & drinking water, bottled water, multiple fish categories, cow milk, human breast milk, and serum panels S1–S5, plus blanks and procedural controls).
- T2 Bioassays: list of effect groups, bioassays and QA/QC parameters (e.g., Z'-factors, reference chemicals, robustness indices).
- T3 THR-TA / T4 TTR / T5 TBG: processed dose–response outcomes (e.g., EC10/20/50, IC10/20), curve fits and goodness-of-fit metrics for thyroid-related endpoints.
- T6 chemical and mixtures TTR: single chemicals & defined mixtures evaluated in TTR binding.

How are the data generated?
Bioassay outputs follow standard procedures per sheet T2 and include internal references (e.g., T3 THR-TA uses T3 as reference agonist; T4/T5 use T4 in competitive binding).

Relationships to other data:
Sheets share sample codes (e.g., WW, EFF, SW, OCM, HBM, S1–S5) enabling cross-assay comparison.

Data is not being re-used, and is produced at Vrije Universiteit Amsterdam, De Boelelaan 1085, 1081 HV Amsterdam (The Netherlands) www.vu.nl

Specify if existing data is being re-used (if any)

Data is not being re-used

Specify the origin of the data

The data originate from the PANORAMIX H2020 project, which investigates chemical mixture toxicity using real-life samples collected across the environment–food–human continuum. PANORAMIX focuses on chemical mixtures present in water, food, and the human body, aiming to develop new scientific tools for mixture risk assessment. [slu.se]

Environmental and food samples

The dataset includes extracted mixtures from:

- Wastewater influent and effluent,
- Surface water and drinking water,
- Fish samples (fatty, lean, aquaculture and wild),
- Cow milk and human breast milk,
  representing real exposure sources for European citizens. These sample categories reflect the project objective of analysing mixtures present in water and food.

Human samples

Some data originate from human serum and breast milk extracts, consistent with PANORAMIX's aim to assess mixtures in the human body and to understand health outcomes related to exposure. The project explicitly studies mixtures occurring in humans and evaluates their potential developmental or endocrine effects.

Bioassay-generated data

All biological activity data (e.g., THR-TA, TTR, TBG, AREc32, MMP, DIO assays, AhR, PPARγ) were generated using in vitro whole-mixture bioassays, which are a core methodological pillar of PANORAMIX. The project uses whole mixture testing to link observed effects to real-life exposures and inform risk assessment.

State the expected size of the data (if known)

0.2 MB

Outline the data utility: to whom will it be useful

1. Scientific utility

- The dataset contains bioassay-based whole-mixture activity profiles, which are essential for understanding mixture effects that cannot be explained by evaluating chemicals one-by-one.
  This aligns with PANORAMIX's whole-mixture approach to risk assessment. 2
- The wide range of biological endpoints (e.g., endocrine, developmental, metabolic, oxidative stress) supports mechanistic insights into mixture toxicity.
- Dose–response parameters (EC10/20/50, IC10/20) and QA/QC metrics enable robust modelling and comparative analysis across matrices.

2. Regulatory and risk-assessment utility

- PANORAMIX aims to deliver ready-to-use tools for mixture risk assessment, including the development of effect-based trigger values that can be applied to water, food, and blood.
- This dataset provides the underlying quantitative biological activity data needed to:
  - define mixture-specific thresholds for concern,
  - evaluate real-life exposure scenarios,
  - support future regulatory proposals and EU zero-pollution goals.
- It can be integrated into regulatory workflows (e.g., EFSA/ECHA mixture frameworks) to enhance evidence-based policy actions.

3. Modelling and computational utility

- The dataset offers structured inputs for:
  - mixture toxicity modelling,
  - benchmark dose modelling,
  - effect-based mixture characterisation,
  - integration into machine-learning or mixture-prediction tools.
- The harmonised format (CSV), clear data dictionary, and comprehensive metadata enhance interoperability and make it suitable for automated workflows and large-scale analyses.

4. Cross-disciplinary and translational utility

- Because the dataset spans environmental, food, and human matrices, it enables:
  - tracing mixture effects across exposure pathways,
  - evaluating transfer and transformation of mixture toxicity from environment → food → humans,
  - supporting epidemiological and exposome-related research.
- These applications directly address the project objective of quantifying mixture risks across the full exposure continuum.

5. Reuse potential

- The dataset has strong reuse potential for:
  - academic toxicology, environmental chemistry, and public health research,
  - regulatory science,
  - mixture modelling and risk-assessment method development,
  - meta-studies and cross-project comparisons.
- It is provided under the CC BY 4.0 licence, enabling unrestricted reuse in both academic and applied/regulatory contexts.

2.1 Making data findable, including provisions for metadata [FAIR data]

Outline the discoverability of data (metadata provision)

Metadata is typically written in lab notebooks or in readme.txt files saved in the same folder as the data.
The file structure of the project is as follows:
- Raw data is stored in the project folder at the VU and can be available upon request.
- The project folder contains different sub-folders (study name)
- Raw data is accompanied by readme.txt (Experimental details) files with the specifics of each experiment.
- Analysis (graphpad) files are also included with the raw data (generally excel files) if applicable, and can be available upon request.

Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?

All processed data will be made openly available once the article is published in the Zenodo Project Environment. A DOI will be created for it.

Data will be archived in the VU data repository (Yoda):
- SurfDrive for processed data and metadata for:
Bioassay responses
Fractionation
Mass spectrometry identification of active fractions
Aligment of bioassay responses and chemical analysis
- Scistor for high resolution mass spectrometry raw data files that are actively being processed or not published.
- YODA for high resolution mass spectrometry raw data files that have been processed and published and need to be archived.

Outline naming conventions used

1. File naming conventions

All files use descriptive and machine-readable names:

- Lowercase or consistent TitleCase
- Words separated by underscores (_)
- No spaces, accents, or special characters
- Version numbers included when relevant

- T1_Samples_v1.0.csv
- T2_Bioassays_v1.0
- T3_THR_TA__v1.0.csv
- T4_TTR_v1.0.csv
- T5_TBG_v1.0.csv
- T6_ chemical_and_mixtures_TTR_v1.0.csv
- README.md

This ensures compatibility with software pipelines and repository systems.

2. Sheet / dataset naming conventions

The dataset preserves the structured table identifiers used within the PANORAMIX project:

- T1_Samples
- T2_Bioassays
- T3_THR_TA
- T4_TTR
- T5_TBG
- T6_Chemical_and_Mixtures_TTR

Each name starts with a table number (T1–T6) and ends with a short, descriptive label of the content or assay, and the version.

3. Variable naming conventions

Variable names:

- Use short, clear descriptors
- Avoid spaces and special characters
- Use consistent case formatting
- Do not embed units in column names

Examples:

- sample_code
- sample_type

- EC10, EC20, EC50
- IC10, IC20
- hillslope
- goodness_of_fit

Such conventions ensure seamless import into R, Python, MATLAB, and regulatory modelling tools.

4. Sample code conventions

Sample codes follow structured short alphanumeric identifiers established in PANORAMIX:

- Water: WW, EFF, SW, DW, BW
- Fish: FFW, FFA, LFW
- Milk: OCM, CCM, BM1–BM4, HBM
- Serum: S1–S5
- Blanks / controls: BWSPE, BWF, BF1, BF2, BHBM, BSSPE, etc.

These codes ensure clear linkage between sample metadata (T1) and assay outputs (T3–T6).

5. Versioning conventions

The project uses semantic versioning for all dataset releases:

- v1.0 — First public release
- v1.1, v1.2 — Minor updates or documentation refinements
- v2.0 — Major structural updates or additional data

Versions are included in filenames, changelogs, and repository metadata.

Outline the approach towards search keyword

Question not answered.

Outline the approach for clear versioning

The project uses semantic versioning for all dataset releases:

- v1.0 — First public release
- v1.1, v1.2 — Minor updates or documentation refinements
- v2.0 — Major structural updates or additional data

Versions are included in filenames, changelogs, and repository metadata.

Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

Question not answered.

2.2 Making data openly accessible [FAIR data]

Specify which data will be made openly available? If some data is kept closed provide rationale for doing so

The data that will be made available will be:

- Raw data
- Final results after data processing

- Statistical data obtained from Graphpad Prism (dose response curves, non linear regression results).

Specify how the data will be made available

The data will be made available upon publication via a DOI directly related to the ZENODO open repository.

Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?

Excel, or R

Specify where the data and associated metadata, documentation and code are deposited

Question not answered.

Specify how access will be provided in case there are any restrictions

Question not answered.

2.3 Making data interoperable [FAIR data]

Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.

The dataset is designed to be interoperable across research domains by using open, non-proprietary formats and following community-recognised metadata standards. Interoperability is further supported by the scientific framework of the PANORAMIX project, which uses harmonised methodologies for assessing real-life chemical mixtures across the environment–food–human continuum.
To ensure that the data can be easily exchanged, combined, and processed across systems:

- All datasets are provided in CSV, an open, widely supported, machine-readable format.
- Documentation is provided in Markdown (MD) and PDF, both human-readable and platform-independent.
- The original Excel file is preserved for traceability but is not required for interoperability.

Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

Controlled vocabularies and semantics
Where applicable, standardised vocabularies will be used:

- Assay terminology based on common toxicology and pharmacology usage (e.g., THR-TA, TTR, TBG).
- Sample type descriptors aligned with environmental and food science use (surface water, wastewater influent/effluent, fish tissue, serum, etc.).
  These categories match the PANORAMIX project's structured definitions of mixtures across environment, food, and humans.

2.4 Increase data re-use (through clarifying licenses) [FAIR data]

Specify how the data will be licenced to permit the widest reuse possible

Creative Commons Attribution (CC BY 4.0) license

Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed

Data will be made available once the publication of the related manuscript is accepted in a peer reviewed journal

Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why

Under CC BY 4.0:

- Users may copy, redistribute, transform, and build upon the dataset for any purpose, including commercial use.
- Reusers must provide proper attribution to the creators, link to the licence, and indicate if changes were made.
- No additional restrictions (e.g., non-commercial clauses) are imposed, ensuring maximal interoperability and reuse potential, consistent with H2020 expectations for open access.
  (H2020 emphasises rights to “read, download, copy, distribute, search, link, crawl and mine” open data.)

There are no legal or ethical constraints on reusing the dataset, as it does not contain personally identifiable information. Extracts derived from human matrices (e.g., serum, breast milk) are anonymised and aggregated. Any components that cannot be shared publicly (if applicable) will be described in the DMP but excluded from the open deposit, as permitted under the flexible H2020 Open Research Data Pilot

Describe data quality assurance processes

The dataset already includes QA/QC metrics in T2 (e.g., Z'-factors, reference compounds, robustness indices) and fit diagnostics (R^2, Sum of Squares, Sy.x) within assay sheets (T3–T5).We will document:

- plate acceptance criteria and any excluded runs,
- calculation methods for EC/IC values and confidence/SE,
- cross-checks across replicates and controls (blanks, REF, etc.).

Any subsequent corrections will trigger a minor version update and a CHANGELOG entry.

Specify the length of time for which the data will remain re-usable

Question not answered.

3. Allocation of resources

Estimate the costs for making your data FAIR. Describe how you intend to cover these costs

The PANORAMIX Phase I dataset requires relatively modest resources to achieve full FAIR compliance because:

- The data are already structured in tabular form (Excel), which reduces preparation effort.
- FAIRification steps consist mainly of:
  - cleaning headers and column names,
  - exporting open formats (CSV/MD/PDF),
  - generating metadata (README, data dictionary),
  - preparing repository-ready documentation.

Estimated personnel time:

- 6–10 hours of researcher/data steward time for:
  - data curation (format normalization, QA),
  - metadata creation (README, variable dictionary),
  - preparing the repository deposit (DOI, license selection).
- No specialized software costs: all tasks rely on standard institutional tools or open-source formats (CSV, Markdown).

Estimated financial cost:

- €0–€300, depending on institutional policies:
  - Many repositories (e.g., Zenodo, 4TU.ResearchData) provide free deposit.
  - If a publisher requires a specific data journal submission, costs may increase, but this is outside core FAIRification.

How these costs will be covered:

- FAIR-related work will be performed by the research team (data curator + PI) within existing project time.
- Repository hosting costs, if any, will be supported by institutional RDM services or existing project budget allocations.

Clearly identify responsibilities for data management in your project

Primary responsibility:

- Dr. Maria Margalef Jornet (PI / VU Amsterdam) — responsible for overall data governance, FAIR compliance, and final approval of repository-ready materials.

Operational data management tasks:

- Data curation and standardisation: led by the project researcher generating the dataset, including cleaning, ensuring consistent column naming, file conversion to open formats, and creation of metadata.
- Storage, backup, and security oversight: handled by the institution's IT services via secure network storage with automatic backups.
- Documentation and metadata development: shared between the PI and the data curator, with optional review by the institutional Data Stewardship or RDM Support team.
- Repository submission and DOI assignment: performed by the PI or designated team member, following repository guidelines.

Describe costs and potential value of long term preservation

- Repository hosting costs, if any, will be supported by institutional RDM services or existing project budget allocations.
- Raw data and processed data will be archived for a minimum of 5 years. This will allow reusability of data in potential retrospective analysis

4. Data security

Address data recovery as well as secure storage and transfer of sensitive data

While processing, data is stored in the university one drive archive system. Security copies are made regularly. Only authorised personnel has reading access to the data.

After processing and publication data will be made available, and a DOI will be adjudicated for publication and accessibility. Published data will be archived in the University Server

For the specific case of the PANORAMIX project, processed data, metadata and its linked DOI will be registered in the PANORAMX platform as well.

5. Ethical aspects

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

Not applicable

6. Other

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

Not aplicable