Plan Overview

A Data Management Plan created using DMPonline

Title: PANORAMIX - TTR Bioassay related outputs on individual cord blood sample extracts

Creator:Maria Margalef

Principal Investigator: Marja Lamoree, Maria Margalef, Timo Hamers

Data Manager: Maria Margalef

Project Administrator: Marja Lamoree

Affiliation: Vrije Universiteit Amsterdam

Funder: European Commission

Template: Horizon 2020 Template

ORCID iD: 0000-0002-7373-7738

ORCID iD: 0000-0001-7932-9770

ORCID iD: 0000-0002-5733-3290

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: 195496

Start date: 01-11-2021

End date: 30-04-2026

Last modified: 21-01-2026

Grant number / URL: 101036631

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

PANORAMIX - TTR Bioassay related outputs on individual cord blood 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 dataset contains experimental results from TTR (transthyretin) binding competition assays, generated within the PANORAMIX project. The purpose is to quantify thyroid hormone displacement by environmental chemicals or biological samples and derive dose–response parameters (EC50, Hill slope).

Generated file TTR_PhaseII.cvs (two sheets) :

a. Assay results (T4_Fit sheet):

- Sample identifiers: date + plate well
- Fitted parameters: A, D, EC50, HillSlope
- 95% confidence intervals (profile likelihood)
- Goodness-of-fit statistics: R^2, Sy.x, Sum of Squares, Degrees of Freedom
- Constraints
- Number of datapoints per curve

b. Sample-specific results (Samples Fit sheet):

- Sample IDs CB#001–CB#750
- Fitted EC50 values:
- Hill slopes:
- TTR IC10 and IC20 values (mLblood/mLbioassay)
- Confidence intervals, replicate numbers
- Stability flags
- Quality metrics:

No external proprietary datasets were used.
The dataset supplements and extends existing PANORAMIX toxicological profiles.

Data will be useful for:

- Toxicology and endocrine disruption research
- Mixture effect modelling
- Benchmark dose derivation
- Validation of computational mixture prediction models

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 includes:

- Sample identifiers (CB numbers or plate/date codes)
- Assay conditions
- Curve fitting model
- QC notes (e.g., “Unstable fit”)
- Variable dictionary (EC50, HillSlope, CI95, R^2, Sy.x)
  Derived from dataset structure summary.

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

A DOI will be assigned upon deposition (Zenodo)

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?

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

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).

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.

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

Creative Commons Attribution (CC BY 4.0) license
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

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 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

Deliverable D5.1 of the project

6. Other

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

Not applicable

PANORAMIX - TTR Bioassay related outputs on individual cord blood sample extracts - Detailed DMP

1. Data summary

State the purpose of the data collection/generation

The dataset contains experimental results from TTR (transthyretin) binding competition assays, generated within the PANORAMIX project. The purpose is to quantify thyroid hormone displacement by environmental chemicals or biological samples and derive dose–response parameters (EC50, Hill slope).

Explain the relation to the objectives of the project

The dataset supports PANORAMIX objectives related to:

- Hazard characterization of complex mixtures
- Identification of thyroid-disrupting potencies
- Development of physiologically anchored mixture models
- Reference data for benchmark dosing and quantitative mixture toxicity modelling

Specify the types and formats of data generated/collected

Generated file TTR_PhaseII.cvs (two sheets) :
a. Assay results (T4_Fit sheet):

- Sample identifiers: date + plate well (e.g., “20-09 p1”)
- Fitted parameters: A, D, EC50, HillSlope
- 95% confidence intervals (profile likelihood)
- Goodness-of-fit statistics: R^2, Sy.x, Sum of Squares, Degrees of Freedom
- Constraints: A=0, D=100 for most curves
- Number of datapoints per curve (mostly 14–16)

b. Sample-specific results (Samples Fit sheet):

- Sample IDs CB#001–CB#750
- Fitted EC50 values: ~2.4×10^{-3} – 3.95×10^{-2}
- Hill slopes: ~0.50 – >10 (some unstable fits)
- TTR IC10 and IC20 values (mLblood/mLbioassay)
- Confidence intervals, replicate numbers
- Stability flags ("Unstable", "Very wide", “???” for missing intervals)
- Quality metrics: R^2 range ~0.42–0.99

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

No external proprietary datasets were used.
The dataset supplements and extends existing PANORAMIX toxicological profiles.

Specify the origin of the data

State the expected size of the data (if known)

- Excel file: ~1–5 MB
- Exported CSV: ~2–6 MB
- Metadata / protocol PDF: <1 MB

Outline the data utility: to whom will it be useful

Useful for:

- Toxicology and endocrine disruption research
- Mixture effect modelling
- Benchmark dose derivation
- Validation of computational mixture prediction models

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

Outline the discoverability of data (metadata provision)

Metadata includes:

- Sample identifiers (CB numbers or plate/date codes)
- Assay conditions
- Curve fitting model
- QC notes (e.g., “Unstable fit”)
- Variable dictionary (EC50, HillSlope, CI95, R^2, Sy.x)
  Derived from dataset structure summary.

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?

A DOI will be assigned upon deposition (Zenodo)

Outline naming conventions used

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

Outline the approach towards search keyword

Metadata tags (e.g. “TTR binding”, “thyroid disruption”, “PANORAMIX”, “dose-response”) will enhance discoverability.

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

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).

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 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 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?

Question not answered.

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 (reference compoundss) and fit diagnostics (R^2, Sum of Squares, Sy.x).We will document:
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

Deliverable D5.1 of the project

6. Other

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

Not applicable

PANORAMIX - TTR Bioassay related outputs on individual cord blood sample extracts - Final review DMP

1. Data summary

State the purpose of the data collection/generation

The dataset contains experimental results from TTR (transthyretin) binding competition assays, generated within the PANORAMIX project. The purpose is to quantify thyroid hormone displacement by environmental chemicals or biological samples and derive dose–response parameters (EC50, Hill slope).

Explain the relation to the objectives of the project

The dataset supports PANORAMIX objectives related to:

- Hazard characterization of complex mixtures
- Identification of thyroid-disrupting potencies
- Development of physiologically anchored mixture models
- Reference data for benchmark dosing and quantitative mixture toxicity modelling

Specify the types and formats of data generated/collected

Generated file TTR_PhaseII.cvs (two sheets) :
a. Assay results (T4_Fit sheet):

- Sample identifiers: date + plate well (e.g., “20-09 p1”)
- Fitted parameters: A, D, EC50, HillSlope
- 95% confidence intervals (profile likelihood)
- Goodness-of-fit statistics: R^2, Sy.x, Sum of Squares, Degrees of Freedom
- Constraints: A=0, D=100 for most curves
- Number of datapoints per curve (mostly 14–16)

b. Sample-specific results (Samples Fit sheet):

- Sample IDs CB#001–CB#750
- Fitted EC50 values: ~2.4×10^{-3} – 3.95×10^{-2}
- Hill slopes: ~0.50 – >10 (some unstable fits)
- TTR IC10 and IC20 values (mLblood/mLbioassay)
- Confidence intervals, replicate numbers
- Stability flags ("Unstable", "Very wide", “???” for missing intervals)
- Quality metrics: R^2 range ~0.42–0.99

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

No external proprietary datasets were used.
The dataset supplements and extends existing PANORAMIX toxicological profiles.

Specify the origin of the data

Question not answered.

State the expected size of the data (if known)

- Excel file: ~1–5 MB
- Exported CSV: ~2–6 MB
- Metadata / protocol PDF: <1 MB

Outline the data utility: to whom will it be useful

Useful for:

- Toxicology and endocrine disruption research
- Mixture effect modelling
- Benchmark dose derivation
- Validation of computational mixture prediction models

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

Outline the discoverability of data (metadata provision)

Metadata includes:

- Sample identifiers (CB numbers or plate/date codes)
- Assay conditions
- Curve fitting model
- QC notes (e.g., “Unstable fit”)
- Variable dictionary (EC50, HillSlope, CI95, R^2, Sy.x)
  Derived from dataset structure summary.

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?

A DOI will be assigned upon deposition (Zenodo)

Outline naming conventions used

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

Outline the approach towards search keyword

Metadata tags (e.g. “TTR binding”, “thyroid disruption”, “PANORAMIX”, “dose-response”) will enhance discoverability.

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

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).

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 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 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?

Question not answered.

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 (reference compoundss) and fit diagnostics (R^2, Sum of Squares, Sy.x).We will document:
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

Deliverable D5.1 of the project

6. Other

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

Not applicable