

---

## AGI4AGING: An AI-boosted worm-based platform to explore transcriptomics in aging and accelerate healthspan screening

*A Data Management Plan created using template from DMPonline.be*

**Creator:** Peter Antal and Walter

**Luyten Affiliation:** KU Leuven

**(KUL) Template:** KU Leuven

**BOF-IOF ID:** CELSA/23/014

**Start date:** 01-10-2023

**End date:** 30-09-2025

### Project abstract:

Healthspan is the part of our lives that we spend in good health. Although the average lifespan has increased all over the world in recent decades, the share of healthspan in this has barely increased. Citizens, governments, health professionals and payers are not so much interested in living longer as in staying healthy for longer. Healthspan is determined by genetic and environmental factors, and we would like to find treatments that act on them. In a previous EU project (<https://www.h2020awe.eu/>), we found genes that extend healthspan in both humans and a model organism (*C. elegans*), as well as natural products with similar effect in *C. elegans*.

We now want to map the metabolic pathways that determine healthspan and, using artificial intelligence and bioinformatics, build a model that allows us to analyze and predict effects on healthspan. To this end, we will measure in detail the mRNA changes elicited by manipulation of candidate genes or treatment with the above-mentioned natural products at different time points throughout the life of *C. elegans*. Based on this, we want to derive a measure that indicates what the biological age is. Unlike chronological age (how old you are in years), biological age indicates how aged you are, i.e. how well you still function compared to your chronological age. As a 50-year-old you can therefore function as a 45- or 55-year-old. Such a measure of your biological is very useful because it allows to evaluate in the short term whether a treatment will have a beneficial effect on healthspan, without having to wait for the end of life. Indeed, the lower a person's biological age compared to his chronological age, the longer his healthspan will normally be.

**Last modified:** 29-12-2023

### 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

## Research Data Summary (0/7)

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

Dataset name / ID	Description	New or reuse	Digital or Physical data	Data Type	File format	Data volume	Physical volume
UK Biobank (UKB)	UK Biobank data resource (Tier 2 level)	Old data	Digital	Numerical Textual	.txt .csv	<5TB	/
AwE RNAi (AwEs)	Ageing with elegans (Nr.633589) RNAi screening data sets using the WormCAMP utility	Old data	Digital	Numerical Textual	.xlsx .txt	<1GB	/
WormCAMP image(iWoC)	Raw image data from the WormCAMP experiments	New data	Digital	Image data	.bin	>5TB	/
WormCAMP markers (mWoC)	Automatically extracted healthspan descriptors and biological aging metrics based on the image data from the WormCAMP experiments	New data	Digital	Numerical and textual data	.csv .txt	<1GB	/
WormCAMP biobank(bbWoC)	Storage of biological samples for later manual deep phenotyping and RNAseq	New data	Physical	PaxGene(?)	/	/	10 L
Aging markers (aWoC)	Manually extracted healthspan descriptors and biological aging metrics based on the image data from the WormCAMP experiments	New data	Digital	Numerical and textual data	.csv .txt	<1GB	/
RNAseq (xWoC)	RNAseq data for the selected WormCAMP experiments	New data	Digital	Numerical Textual FASTQ		<1TB	/

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

The UK Biobank data can be accessed after approval from its standard storage. The AwEs data set is an in-house data set described in doi: <https://doi.org/10.1101/2021.10.18.464905>.

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.

- Yes, UK Biobank has approval from the North West Multi-centre Research Ethics Committee (MREC) as a Research Tissue Bank (RTB) approval. This approval means that researchers do not require separate ethical clearance and can operate under the RTB approval human subject data (<https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/about-us/ethics>).

Experiments with *C. elegans* do not have ethical concerns as this is an invertebrate.

Will you process personal data? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).

- Yes, we will analyze genetic data from the UK Biobank governed and covered by the UK Biobank ethical approval (<https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/about-us/ethics>).

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, ...)? If so, please comment per dataset or data type where appropriate.

- Yes, adaptive study design schemes using biological clocks can accelerate healthspan screening.

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material or Data transfer agreements, Research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- No, regarding *C. elegans* data (UK Biobank MTA excludes further dissemination of the data).

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

## Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g. in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, codebook.tsv etc. where this information is recorded).

We will use international standards to analyze the UK Biobank data sets (ICD-10, ATC). The *C. elegans* data will be harmonized according to the WormBase resource (<https://doi.org/10.1093/nar/gkz920>, <https://doi.org/10.1093/genetics/iyac003>), especially its phenotype ontology (<https://github.com/obophenotype/c-elegans-phenotype-ontology>).

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify which metadata standard will be used.

If not, please specify which metadata will be created to make the data easier to find and reuse.

- Yes, we will use the phenotype ontology from WormBase as a baseline.

## Data Storage & Back-up during the Research Project

Where will the data be stored?

- The UK Biobank data will be downloaded and analyzed at the BME (Budapest University of Technology and Economics). The temporary image data (iWoC) generated during the WormCAMP experiments are stored locally and will be archived at zenodo (<https://zenodo.org/>). The derived data sets (mWoC, aWoC) will be stored using standard back-up provided by KU Leuven ICTS and reposit at WormBase. The raw RNAseq data will be processed at the BME and archived at zenodo (<https://zenodo.org/>). The derived data sets (xWoC) will be stored using standard back-up provided by KU Leuven ICTS and reposit at EMBL-EBI Expression Atlas/GEO linked to WormBase.

How will the data be backed up?

- Standard back-up provided by BME and KU Leuven ICTS. Reported data sets will be archived at zenodo (<https://zenodo.org/>), WormBase (<https://wormbase.org/>), and EMBL-EBI Expression Atlas/GEO.

Is there currently sufficient storage & backup capacity during the project?

If no or insufficient storage or backup capacities are available, explain how this will be taken care of.

- Yes, there is sufficient storage & backup capacity during the project.

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

Access will be limited to authorised people of the research team using the standard IT protocols.

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

No costs.

## Data Preservation after the end of the Research Project

Which data will be retained for 10 years (or longer, in agreement with other retention policies that are applicable) after the end of the project?

In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

UK Biobank data can be accessed in a restricted period (three years by default). The zenodo and EMBL-EBI Express Atlas are official long-term repositories maintained by the EU/CERN.

-

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

Finalized data sets will be archived at zenodo (<https://zenodo.org/>), WormBase (<https://wormbase.org/>), and EMBL-EBI Expression Atlas/GEO.

- 

-

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

None

## Data Sharing and Reuse

Will the data (or part of the data) be made available for reuse after/during the project?

Please explain per dataset or data type which data will be made available.

- Reported and finalized data sets will be archived at zenodo (<https://zenodo.org/>), WormBase (<https://wormbase.org/>), and EMBL-EBI Expression Atlas/GEO.

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

Experimental data will be publicly accessible after publication.

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

Please explain per dataset or data type where appropriate.

- UK Biobank data cannot be shared.

Where will the data be made available?

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

- Finalized data sets will be archived at zenodo (<https://zenodo.org/>), WormBase (<https://wormbase.org/>), and EMBL-EBI Expression Atlas/GEO.

When will the data be made available?

- Depending on publications.

Which data usage licenses are you going to provide?

If none, please explain why.

- CC BY-NC-ND (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Do you intend to add a persistent identifier (PID) to your dataset(s), e.g. a DOI or accession number? If already available, please provide it here.

- Yes

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

- None

## Responsibilities

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

András Gézsi, Péter Antal

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

András Gézsi, Péter Antal

**Who will manage data preservation and sharing?**

András Gézsi, Péter Antal

**Who will update and implement this DMP?**

András Gézsi, Péter Antal, Liesbet Temmerman, Walter Luyten