Application form NWO – Open Science Fund – 2023

Section 1a – Details of the applicant	
Name	*Dr. W.C.E.P. (Wilco) Verberk*
Affiliation – institution	**Radboud University**
Affiliation – department	*Department of Animal Ecology and Physiology*
Position	*Associate Professor*
End date of contract	*permanent position*
E-mail address	*w.verberk@science.ru.nl*
ORCID ID	*https://orcid.org/0000-0002-0691-583X*

Section 1b – Details of the team member(s)	
Name team member 1	*prof. dr. J Ellers*
Affiliation	*Vrije Universiteit Amsterdam, Department of Ecological Sciences – Animal Ecology*
E-mail address	*j.ellers@vu.nl*
ORCID ID	*https://orcid.org/0000-0003-2665-1971*

Name team member 2	*prof. dr. Matty Berg*
Affiliation	*Vrije Universiteit Amsterdam, Department of Ecological Sciences – Animal Ecology*
E-mail address	*m.p.berg@vu.nl*
ORCID ID	*https://orcid.org/0000-0001-8442-8503*

Name team member 3	*Michiel Punt*
Affiliation	*Netherlands eScience Center - Section Head Life Sciences*
E-mail address	*m.punt@esciencecenter.nl*
ORCID ID	*https://orcid.org/*

Section 2 - Public summary

English public summary

Energy metabolism, development and fecundity are key traits by which species respond to a changing environment. However, data on these key traits is scattered throughout the literature. In the NWO ShareTrait-project we have developed a platform for the collection, reformatting and standardization of trait data. We now want to upscale and transform this platform into a valuable resource the wider scientific community. To do so, we envision a top-down approach by partnering with key journals to mandate deposition of data on these traits, as well as a bottom-up approach to entice and facilitate researchers to deposit their data.

Word count (max 100): 98

Dutch public summary

Eigenschappen van soorten die inzicht geven in de relatie met hun leefomgeving zijn energiehuishouding, ontwikkelingssnelheid en aantal nakomelingen. Echter, data over deze eigenschappen is versnipperd. In het NWO ShareTrait-project hebben we een platform ontwikkeld voor het verzamelen, standaardiseren en omrekenen van degelijke data. We willen nu dit platform opschalen en transformeren tot een waardevolle resource voor andere onderzoekers. Via een top-down aanpak, in samenwerking met biologische tijdschriften, willen we het archiveren van deze data onderdeel te maken van het publicatieproces. Via een bottom-up aanpak willen we onderzoekers stimuleren en faciliteren om hun data beschikbaar te stellen.

Word count (max 100): 96

Section 3 – Project proposal	
3.1 The details of proposal	
Proposed project title and acronym	ShareTrait: an open data resource to achieve a synthesis of the evolution and ecology of traits
Project duration (in months)	12 months
The project will primarily address	*Open platforms or tools*
The project will secondarily address	*Culture change towards Open Science*
Relevance for a specific discipline	22.60.00　　Dierkunde 22.40.00　　Ecologie 21.70.00　　Fysiologie

3.2 The vision for your project (Criterion: Alignment with the aim of the Call for proposals)

A longstanding goal in ecology is to understand species-environment relationships. Species traits are attributes that explain how species interact with, and have adapted to their environment[1,2]. Knowledge about trait diversity is therefore crucial for addressing current ecological challenges, such as understanding the vulnerability and resilience of ecosystems, managing the biodiversity crisis, and biocontrol of crop pests[3,4].

Although traits are increasingly measured on a wide variety of animals, their use is limited by poor data infrastructure and lack of FAIR practices. Our vision is to create a common platform for exchanging trait information and set it up in such a way that it will be suitable for collating, searching, accessing, and extracting data and their metadata for the challenging questions in the field.

In a previous NWO Open Science project we developed ShareTrait, a platform to collate, reformat and standardize data on three traits: metabolic rate, development time and fecundity. These traits are universal to animal life and drive population trends. A unique feature of ShareTrait is that a wealth of metadata is incorporated, which is key to resolve barriers to data interoperability and reusability[5,6]. The ShareTrait database holds 33,123 records from 56 submissions, demonstrating its potential[7]. This initial round of submissions has debugged and fool-proofed the platform's submission and data curation process. The next step is to upscale.

We envision a three-pronged approach for ShareTrait to stimulate a wider adoption of data sharing:
1) Build data infrastructure to interconnect other trait databases to facilitate data exchange;
2) Create bottom-up incentives to entice and facilitate researchers to deposit data;
3) Develop top-down standards for trait data submission by partnering with key journals to mandate the deposition of trait data upon acceptance of articles.

Each of the three actions combine to achieve **our aim of a wider uptake of our ShareTrait platform, making it the primary go-to database for researchers working on the ecology, physiology and evolution of animal life.**
　By working on the interconnection between our ShareTrait platform and other trait databases, we will **adapt the ShareTrait platform for a wider use by the research community**. Instead of individual resarchers starting to pull together data in separate databases we can channel those efforts and energy under the ShareTrait umbrella.
　By working together with biological journals to integrate data deposition in ShareTrait as part of the publication process, we can **transform the way researchers publish**. The value of individual papers will be beyond the data reported on a specific species, but will increase our knowledge base of the tree of life
　By showcasing the success and usefulness of ShareTrait at conferences and synthesis papers we will further stimulate a **wider adoption of Open Science practice among researchers in our field and beyond**.
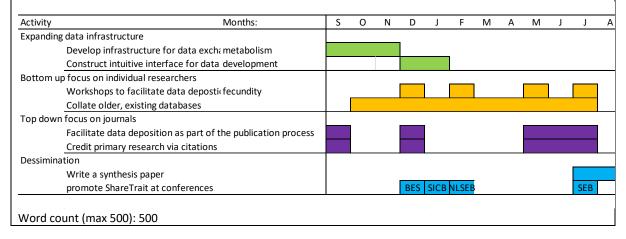
Word count (max 450): 450

3.3 Project plan (Criterion: Feasibility of the project plan)

To upscale our ShareTrait platform we will focus on three aspects:

1. We will **expand the current data infrastructure of ShareTrait** to facilitate data exchange with other, existing databases. These databases are often dedicated to specific taxonomic groups or traits, and set up without considering data interoperability. Differences in the data structure and the way information is coded are major obstacles for unlocking and connecting these databases. To resolve these obstacles, we will partner with the eScience Center, which provides the technical support for developing communication protocols to enhance the capacity of the databases to *interoperate* with recognizable metadata. To enable researchers to easily extract data from ShareTrait, we will also expand the current infrastructure of ShareTrait with an intuitive interface. For example, through filtering with key words on trait, taxonomic group, and metadata properties.

2. We will use a **bottom-up approach to encourage individual researchers to commit** to the ShareTrait initiative. We will run online and in-person workshops at annual meetings of key societies to facilitate researchers to deposit their data. One of the services provided by ShareTrait will be to provide quality checks of uploaded data (e.g. completeness of metadata, identify outlying data points), which enhances correct re-use of data. To make older, published data on traits also available via ShareTrait we will hire and train student assistants to help extract and collate this data. We also want to attract ongoing PhD students from the trait research community to be involved in ShareTrait as part of their thesis research. This will provide them with training in database management and allow them to compile a dataset to write a synthesis paper on a topic of their interest.

3. We will use a **top-down approach by partnering with key journals** in ecology, evolution and physiology. We will work together to make trait data deposition a standard practice for published research for example by mandating deposition of trait data via ShareTrait upon acceptance of articles. The user-friendly questionnaire we have developed in ShareTrait could be employed to facilitate researchers to deposit their trait data and the associated meta-data in a standardized manner. We already contacted several editors and only received positive responses to working with ShareTrait (e.g. Functional Ecology, Current Research on Insect Science, Oikos, Physiological and Biochemical Zoology). In addition, we will work with the journals to better reward and recognize good Open Science practice by enabling the citation of primary research papers in meta-analyses and synthesis papers. This practice is already adopted by Global Ecology and Biogeography and increasingly recognised as the way forward to ensure giving appropriate credit to primary research efforts[8].

To further publicize the ShareTrait platform and its uses to the general scientific community, towards the end of the project we will write a data synthesis paper using the data compiled so far, focussing on major patterns in energy usage and expenditure across the animal kingdom. All these activities will allow **the database to grow and the platform to become the primary go-to for researchers**.

Timeline:

Activity	Months:	S	O	N	D	J	F	M	A	M	J	J	A
Expanding data infrastructure													
Develop infrastructure for data excha metabolism		■	■	■									
Construct intuitive interface for data development					■	■							
Bottom up focus on individual researchers													
Workshops to facilitate data depositi fecundity					■		■			■	■		
Collate older, existing databases					■	■	■	■	■	■	■	■	
Top down focus on journals													
Facilitate data deposition as part of the publication process		■			■				■	■	■		
Credit primary research via citations		■			■				■	■	■		
Dessimination													
Write a synthesis paper												■	
promote ShareTrait at conferences					BES	SICB NLSEB						SEB	

Word count (max 500): 500

3

3.4 Project roles and expertise (Criterion: Feasibility of the project plan)

Wilco Verberk (RU Nijmegen) – Main applicant with a track record in coordinating NWO grants. Expertise focussed on species traits and physiological ecology in diverse groups of ectotherms (fish, insects, crustaceans, molluscs) and as such well-informed about the difficulties of integrating trait information from different disciplines. Experience with metabolic rate measurements, and compiling, curating, and analysing large datasets for synthesis (e.g. metabolic rate in fish, dive duration in animals ranging from water beetles to whales). Lead author on a perspectives paper highlighting the challenges in transforming trait-based ecology into a predictive science (currently cited 200+ times) and embraced open science (currently published 19 datasets).

Jacintha Ellers (VU Amsterdam) – Broad expertise in trait plasticity and other causes of intraspecific variation in species traits. Long-standing experience in measuring behavioural, life history and physiological traits to understand the impact of global change on the outcome of species interaction. Senior author on the handbook of protocols promoting standardized measurements of invertebrate functional traits across taxa and trophic levels.

Matty P. Berg (VU Amsterdam) – Demonstrable experience with life-history trait measurements of soil fauna. Application of trait-based approaches in understanding animal responses to environmental stress and their effects on ecosystem processes. Experience with database construction as curator of the national database on soil fauna distribution and investigator of the global Collembola distribution database.

Michiel Punt (Netherlands eScience Center) – Software developer, heading a team of 18 programmers that specifically work closely with scientists in the field of Life Sciences to generate software solutions for scientific challenges.

Word count (max 250):248

3.5 Budget table

Type of costs	Short description	Costs in euros
Personnel	*postdoc, 0.2 FTE, 12 months*	*€ 16,751*
Implementation costs	*Student assistants (200 hours a EUR 56,25)*	*€ 11,250*
Project-related goods/services	*Software development eScience Centre*	*€ 20,000*
Dissemination	*conference expenses (4x 500)*	*€ 2,000*
Own contribution	*Mostly in kind contribution to project*	*- € 1*
Total request from NWO		**€ 50,000**

3.6 Budget justification (Criterion: Feasibility of the project plan)

We ask for budget for a coordinator (0.2 fte) to organise the different activities, such as liaise with the journal editors and the technical support from eScience Center, organise periodic workshops, train students, approach researchers to submit data and provide feedback to these that do. An enthusiastic postdoc or PhD student in one of our labs will take this on as an additional opportunity to enhance his/her research experience.
We also ask for budget for technical support to expanding the database infrastructure.
In addition, we will spend about a quarter of the budget on data extraction and collating by student assistants (200 hours to help in extracting data from published papers and help integrate existing databases* into ShareTrait) Finally, we ask for support to promote ShareTrait at conferences and meetings.

* DEB Energetics data base (https://www.bio.vu.nl/thb/deb/deblab/add_my_pet/), Reptile Development (https://repdevo.com/team/), AmphiBIO (https://doi.org/10.1038/sdata.2017.123), Copepod Trait database (http://dx.doi.org/10.5194/essd-9-99-2017)

Word count (max 200):145

Section 4 – Open Science track record of the applicant

I embrace open science which I believe opens new avenues for collaborative research. For example, I see great potential to improve upon how meta-analyses are currently done. This led me to launch the NWO - ShareTrait initative together with profs Ellers and Berg, which enabled us to take a first step towards having researchers collaborate towards a common goal of synthesizing trait data. I also encourage my collaborators to make their

data available in data repositories. As a result, datasets have been deposited in open repositories for 19 recent studies. In each of these studies, I was involved in either as a principal investigator or as a co-authored for the manuscript. In addition, in a recent paper[2], we highlighted the value of open data. As a case study we quantified trait relationships in aquatic macroinvertebrates at the scale level of Europe and showed that variation in environmental conditions across large distances could be most consistently related to variation in species traits. Finally, as a reviewer I always check the availability of the data and suggest to the authors to deposit the data underlying their manuscript in an open respository (e.g. when it is added as supplementary material).

Word count (max 200): 197

Section 5 – Data management (Criterion: Feasibility of the project plan)

5.1 Will this project involve re-using existing research data?
Yes: Are there any constraints on its re-use?

Constraints relate to incomplete metadata on how the data were collected, hampering comparisons across studies. Authors that supply trait data will be asked to provide missing information on such metadata to enhance the interoperability of trait data. To facilitate this process, we have developed user-friendly questionnaire.

5.2 Will data be collected or generated that are suitable for reuse?
Yes: Please answer question 5.3 and 5.4

5.3 After the project has been completed, how will the data be stored for the long-term and made available for the use by third parties? Are there possible restrictions to data sharing or embargo reasons? Please state these here.

Description:
For our project we will aggregate, reformat and harmonize existing data. All the primary data sources will be made available online in standard repositories. We will ask researchers to deposit their data in such repositories prior to adding the data to our database via the pipelines (e.g. via Zenodo). Consequently, each data item in our database will have a DOI.
In addition, our derived data will be published online in a digital repository where we can submit updates with version tracking. As such the data will be available long-term by third parties. In all cases, documentation and meta-data will be provided to facilitate their use by the scientific community beyond our project. This includes 'readme' text files, code for the pipelines as well as code to extract data and search the dataset.
While research is ongoing, we will use data storage services provided by our organisations (local servers that are regularly backed up, SurfDrive). These data storage servers allow for managed access and sharing of data between partners and guests during the project.

5.4 Will any costs (financial and time) related to data management and sharing/preservation be incurred?
No: All the necessary resources (financial and time) to store and prepare data for sharing/preservation are or will be available at no extra cost.

Section 6 – Software sustainability (Criterion: Feasibility of the project plan)
6.1 Will software be generated during the project?
No: Skip to section 7
We will develop communication protocols to enhance the capacity of the databases to interoperate with recognizable metadata. These protocols will be developed mostly in existing software packages (R, Phyton) to allow for uptake by the wider research community. All these protocols will be made publicly available on Github (https://github.com/) under a CC-BY-4.0 licence.

Section 7 – Literature references
[1] **Verberk, W. C. E. P**., Van Noordwijk, C. G. E. & Hildrew, A. G. (2013) Delivering on a promise: integrating species traits to transform descriptive community ecology into a predictive science.
[2] Wilkes, M. A.,... **Verberk, W. C. E. P.**, et al. (2020) Trait-based ecology at large scales: Assessing functional trait correlations, phylogenetic constraints and spatial variability using open data. Glob. Change Biol. 26, 7255–7267.

[3] Deutsch, C., Penn, J.L., Seibel, B. (2020) Metabolic trait diversity shapes marine biogeography. Nature 585, 557-562.

[4] Segoli, M., Abram, P.K., **Ellers, J**., Hardy, I.C.W., Greenbaum, G., Heimpel, G.E., Keasar, T., Ode, P.J., Sadeh, A., Wajnberg, E. (in press). Trait-based approaches to predicting biological control success: challenges and prospects. Trends Ecol. Evol.

[5] Gallagher, R. V. et al. Open Science principles for accelerating trait-based science across the Tree of Life. Nat. Ecol. Evol. 4, 294–303 (2020).

[6] Culina, A. et al. (2021) Connecting the data landscape of long-term ecological studies: The SPI-Birds data hub. J. Anim. Ecol. 90, 2147-2160.

[7] https://github.com/ShareTraitProject/ShareTrait

[8] https://datacite.org/

By submitting this form, I declare that:

I and all the individuals involved in this proposals satisfy the nationally and internationally accepted standards for scientific conduct as stated in the Netherlands Code of Conduct for Research Integrity (The Universities of the Netherlands): Yes

The research organisation has been informed of this grant application and the research organisation accepts the grant conditions of this programme: **Yes**

The team members named in this form have read and agreed with the submission of this proposal and have agreed with their role and intended contribution to the project, should this be awarded: **Yes**

I have completed this application form truthfully: **Yes**