



Universiteit Utrecht

Application form for Fostering Open Science Practice Fund

Closing date: 20 May 2022

The Open Science Fund is an opportunity for **Utrecht University** and **University Medical Centre Utrecht** employees to access small grants with which they can apply Open Science principles into their research. This funding amounts to € 10.000 (minimum) - € 15.000 (maximum) per application.

Contact and information

If you are considering an application and you would like to discuss this with a member of the Open Science Programme team, please send a mail to openscience@uu.nl or contact [Judith de Haan](#), programme director.

More information, such as selection criteria, who can apply and the selection process, can be found on the [fund website](#).

Names	Gerko Vink and Hanne Oberman		
Position/role	Associate Professor / Junior Researcher		
Department	Methodology and Statistics		
Faculty	Social and Behavioural Sciences		
Email address	g.vink@uu.nl		
Telephone number	0624111054		
Title of proposed project	Dessert: An automated tool for transparent and reproducible scientific reporting.		
Project start date	01/09/2022	Project end date	31/08/2023
WBS number	SA.130402.101		

THIS PROPOSAL HAS AN OPEN DEVELOPMENT REPOSITORY:
<https://github.com/gerkovink/OpenScienceFund2022>



Universiteit Utrecht

Please provide a summary of your project (max. 100 words):

(to describe the project on our website)

We can see the process of data analysis as a recipe. Every time we follow a recipe minutely, the same dish results: the process of generating the dish is thereby very standardized. What is not standardized, however, is the evaluation of the quality and flavour of the dish. The same holds for data analysis: although the computational process may be standardized, the interpretation and evaluation of the quality of the results often is not. We propose a generic tool for automatic standardized report generation and outline how to apply the tool on missing value analysis and missing data imputation.

Please outline the proposed project, including the *purpose* of Open Science Practice, the specific [*topic*](#) it addresses, the *approach* being taken and the *links* to research' (max. 500 words):



We develop **<dessert>**, an automated report generator for statistical analyses in scientific studies. **<dessert>** is a play on the widely accepted analysis pipeline in R, where the package **<recipes>** is used to standardize the data throughput procedure that leads to analysis models. Our proposed **<dessert>** comes after the analysis 'main course' and standardizes the computational evaluation and interpretation of model inference. **<dessert>** can be used to generate reports for any statistical analysis – as long as a corresponding **<dessert> course** (i.e. function) is provided. To illustrate the use and the development standard we design the **<dessert>** functions for generating reports for incomplete data analysis and imputation evaluation. We believe that there is a great need for radical openness and advocate the use of standardized report generating to add to the transparency and validity of evaluating analysis processes. **<dessert>** falls under the topic **"FAIR data and software"**.

Computational evaluation of analysis processes is paramount, especially since statistical methodology is increasingly used to guide decision making in practice. New methodologies are constantly proposed, implemented and applied in the literature. This poses challenges for authors, reviewers, publishers and readers, as a proper evaluation process is often lengthy and would take up valuable space in manuscripts. Space that is often limited, thereby leading to partial or complete omission of the necessary information to realize a sufficient evaluation of procedural and methodological validity, reproducibility and the option for replication. We develop a toolkit that standardizes such evaluations by automatically generating reports for a given analysis. The resulting report can be added as a supplement to any manuscript. With our **<dessert>** menu, anyone can interpret the utilized analysis procedure and evaluate its validity, even when access to the original data or the computer code is restricted.

We realize that full disclosure of data and code, when possible, would allow for a proper evaluation, reproduction and even replication. However, such full disclosure is not always possible. **<dessert>** will add to the transparency of research processes, whilst limiting the risk for statistical disclosure and protecting the privacy of respondents.

Purpose of open science practice with **<dessert>**

- The report protects privacy of respondents;
- The report promotes the core values of open science and dissemination;
- The resulting methodology will be **open, community-driven** and implemented in the **open-source software R**, thereby allowing developers to formulate new desserts for statistical analyses functions in R and resulting objects in other statistics processors, such as JASP, SPSS and STATA.

We write the first of the SMART objectives as follows, where the remaining objectives are considered in the following box.

- Specific: An R package with functions that generate reports (output in .html, pdf and .docx file types) based on specific data formats and analysis objects. The resulting output of the functions will be relevant information in text, tables, metrics and



figures that can be included as supplement to manuscripts, which will standardize evaluation and reporting and facilitate reproduction of analyses results and, ultimately, replication of the complete inference process.

**How will you evaluate the progress, outcomes and impact of your project?
How will these results be shared? (max. 300 words)**

We continue with the remaining SMART objectives:

- Measurable: By the end of this project there will be an open-source software initiative with a corresponding public development repository, R package, instructional videos, development instructions and a package vignette. The project will operate under a GNU GPL-3 license, which prevents closed source distribution. We can measure the impact of this project by CRAN downloads, GitHub forks and stars, development contributions by other scientists and scientific referencing (long-term)
- Achievable: We work with manageable deliverables (D) that build up to milestones (M):
 - o D1.1: Open repository and website aimed at open development
 - o D1.2: Dessert for the standardized analysis of incomplete data in R
 - o D1.3: Extending the dessert to other data formats (SPSS, JASP, STATA, SAS, MPLUS)
 - o D1.4: Writing a package vignette
 - o D1.5: Building a package and submitting to CRAN [M1]
 - o D2.1: Dessert for the standardized evaluation of imputed data with mice in R
 - o D2.2: Extending the dessert to other statistical software (SPSS, STATA, MPLUS)
 - o D2.3: Extending the package vignette
 - o D2.4: Recording instructional videos
 - o D2.5: Extend CRAN package with function 2 [M2]
 - o D3: Add more deserts for common analyses
- Relevant: Openness in scientific reporting is optimized to the published analyses and evaluations. We go beyond that by advocating a standard that considers the whole analytical process and which can prevent that the subjectivity of decisions in the evaluation of scientific analyses may go unnoticed. We aim to instigate discussion about the transparency and documentation of analysis processes. We believe that this would fit well into the UU-wide ambitions on remaining a progressive leader in open science.
- Timebound: Start and end dates are clear. Open Science Festival or Open Science Community Utrecht presentation proposal as midway point (March/April 2023).



Please describe the potential for learning and/or development for researchers (max. 150 words):

This project has potential for enormous impact and may revolutionize the standard of transparent and open scientific reporting. The <dessert> software will facilitate applied researchers in transitioning towards open science, by lowering the threshold for sharing data analysis pipelines. We provide a safe and easy 'stepping stone' for those who are not (yet) able to take on a radically open approach in sharing data and code (e.g., due to privacy issues). With that, we provide a learning opportunity for colleagues and collaborators. Moreover, the output generated by <dessert> may prevent that the subjectivity of decisions in the evaluation of analytical processes goes unnoticed. Finally, this open-source project offers learning opportunities for the research team (in collaboration with [RDM support](#)) by centring FAIR software principles from day one. The project may ultimately become a showcase for open software development at Utrecht University.

Please detail the amount of funding applied for and justify the costs requested:

The total budget requested is €15.000, which would facilitate

75h - Gerko Vink

75h - Hanne Oberman

300h – Student Assistant capacity for 8h/wk for 40 weeks.

GV will act as project leader, package and website maintainer and will actively reach out to other package developers and maintainers to disseminate and advocate the use of <dessert> in analysis pipelines. HO will act as programming lead and core developer. GV and HO will create and streamline documentation and invite other developers to contribute their <dessert> functions. GV, HO and SA's will work together on deliverables 1 and 2. SA's will work on deliverable 3 under supervision of HO and GV. Weekly meetings will be held with GV, HO and SA's

Funding would enable us to reach the projects potential: i.e. to cover more <dessert> functions, reach a larger audience and thereby generate more impact and advance the transparency of scientific reporting.

Please send the completed application form to openscience@uu.nl by 20 May 2022.