**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](mailto:openscience@uu.nl) or contact [Judith de Haan](https://www.uu.nl/medewerkers/JJdeHaan), programme director.

More information, such as selection criteria, who can apply and the selection process, can be found on the [fund website](https://www.uu.nl/onderzoek/open-science/fostering-open-science-fund).

|  |  |  |  |
| --- | --- | --- | --- |
| **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](mailto:g.vink@uu.nl) | | |
| **Telephone number** | 0624111054 | | |
| **Title of proposed project** | automatic report generation for incomplete data analysis | | |
| **Project start date** | 01/09/2022 | **Project end date** | 30/08/2023 |
| **WBS number** | SA.130402.101 | | |

**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 the 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***](https://www.uu.nl/en/research/open-science/topics) ***it addresses*, the *approach* being taken and the *links* to research’ (max. 500 words):**

|  |
| --- |
| [Zie ook onderstaand onder SMART!]   * Open Science Topic: FAIR data and software. * Purpose of the Open Science Practice: Facilitate transparent reporting of data analyses, and with that, the replication of research. * Approach: Radical openness! We will create a FAIR open source software package, { supplementr}. In principle we’ll focus on automatic report generation for incomplete data analysis. The package may be extended to include more kinds of analyses (e.g., by contributors on GitHub, yay collaboration!). * Links to research: Missing data is ubiquitous, and there is extensive work on missing data handling strategies. We’ll focus on two popular data analysis tools: SPSS and R. The {supplementr} package will work seamlessly with the widely-used imputation package {mice}. For applied researchers, the {supplementr} package may ultimately function like a consultation with a statistician.   We develop an automated report generator for studies wherein imputation techniques are used to solve for missing values. With imputation, the process of evaluating the validity and plausibility of the resulting imputations and algorithms is vital for the interpretation of submitted manuscripts and scientific results. 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. We implement a method that standardizes such evaluations by automatically generating a report for a given analysis model. The resulting report can be added as a supplement to any manuscript wherein imputation has been used. With our method, anyone can interpret the utilized imputation procedure and evaluate its validity, even when access to the original data or the computer code is restricted.  Purpose of open science practice   * 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 package mice*, which is the de facto standard for the analysis of incomplete data sets. Voortschrijdend inzicht kan de reports beter maken. * Nog iets meer over purpose met \*dit moet je echt vertellen bla-bla praatje\*   Approach being taken   * The resulting methodology will be open, community-driven and implemented in the open-source software package mice, which is the de facto standard for the analysis of incomplete data sets. * Deliverables formuleren: convergentie/densityplots/marginals en conditionals, correlatietabellen/nieuwe plaatje/analysemodelplot/pooled en individual results * Welke software ondersteunen we vanaf begin: mice::mids object en .txt in long format en iteration\_history. Dan hebben we alle software zo’n beetje te pakken * In R; evt met shiny facility for ease of use. * TOPIC is FAIR data and software |

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

|  |
| --- |
| Progress en outcomes   * Deliverables formuleren: convergentie/densityplots/marginalen en conditionals, correlatietabbellen/nieuwe plaatje/analysemodelplot/pooled en individual results   Impact:   * Downloads on CRAN en GitHub   Dissemination   * In R – perhaps not in mice because of impact monitoring. Perhaps separate package to aid visibility with function in mice that runs the relevant package code: REPORTER package with REPORTER::imputation() as function. REPORTER allows for more standardized implementations than imputation alone.   SMART   * Specific: an R package with functions that generate reports (output in html, pdf and docx file types) based on incomplete data and imputation objects. The output contains tables, metrics and figures that can be included as supplement to manuscripts, which will standardize reporting and facilitate replication of analyses. * Measurable: By the end of this project there will be a GitHub repo with an R package and a vignette. Measure impact by CRAN downloads/GitHub stars. Benchmarks are the development of lower-level R functions for the creation of figures and tables, wrapper functions for specific input types (mice::mids objects, SPSS exports, etc.), and the bundling of these into an R package. * Achievable: Work with milestones (see above). E.g., expand to SPSS export data later. * Relevant: Openness in scientific reporting e.g. w.r.t. size of missing data problem and quality of imputations. Reproducibility of research practices. More open source software published by the UU (showcase/forefront of the open science movement). * 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):**

|  |
| --- |
| * Opportunity to develop a FAIR open source software project from scratch (e.g., first version will already have a Zenodo registration). * Opportunity to practice the guidance/supervision of a student-assistant in open science principles. * This project may become a showcase to encourage colleagues/applied researchers to use open science principles? |

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

|  |
| --- |
| Urenraming maken (bijv. 307 uur SA was €8.000 bij SIG aanvraag) |

**Please send the completed application form to** [**openscience@uu.nl**](mailto:openscience@uu.nl) **by 20 May 2022.**