Application form: Open Science Infrastructure (2024) - full proposal

Section 1: Applicants

Name of the main applicant	Dr. Daniel Bangert
Affiliation – institution	TU Delft
Affiliation – department	4TU.ResearchData
Position	Director
End date of contract	Permanent position
Guarantee added	n.a.
E-mail address	[redacted]
ORCID ID	https://orcid.org/0000-0003-4981-2870

Name of the co-applicant	Dr. David Groep
Affiliation – institution	Nikhef
Affiliation – department	Physics Data Processing
Position	Programme Leader
End date of contract	Permanent position
Guarantee added	n.a.
E-mail address	[redacted]
ORCID ID	https://orcid.org/0000-0003-1026-6606

Name of the co-applicant	Dr. Jason Maassen
Affiliation – institution	Netherlands eScience Center
Affiliation – department	n.a.
Position	Technology Lead
End date of contract	Permanent position
Guarantee added yes/no/n.a.	n.a
E-mail address	[redacted]
ORCID ID	https://orcid.org/0000-0002-8172-4865

Name of the co-applicant	Dr. Frank Ostermann
Affiliation – institution	University of Twente
Affiliation – department	Geographic Information Processing, Faculty of Geographic Information Science and Earth Observation
Position	Associate Professor
End date of contract	Permanent position
Guarantee added	n/a
E-mail address	[redacted]
ORCID ID	https://orcid.org/0000-0002-9317-8291

Name of the cooperation partner (person)	Dr. Daniel Katz
E-mail address	[redacted]
Organisation	San Diego Supercomputer Center, UC San Diego (representing the FAIR in ML, AI Readiness, & Reproducibility Research Coordination Network)

Name of the cooperation partner (person)	Dr. Leyla Jael Garcia Castro
E-mail address	[redacted]
Organisation	ZB MED Information Centre for Life Sciences (representing NFDI4DataScience)

Name of the cooperation partner (person)	Dr. Curtis Sharma
E-mail address	[redacted]
Organisation	4TU.ResearchData, TU Delft (representing the Research Data Alliance FAIR for Machine Learning Interest Group)

Name of the cooperation partner (person)	Dr. Michiel de Boer
E-mail address	[redacted]
Organisation	University Medical Center Groningen (representing the Dutch Reproducibility Network)

Name of the cooperation partner (person)	Dr. George A.K. van Voorn
E-mail address	[redacted]
Organisation	Wageningen University & Research

Name of the cooperation partner (person)	Ivar Janmaat
E-mail address	[redacted]
Organisation	SURF

Section 2: Summary

Proposed project title	**Enabling FAIR for AI: Infrastructure for Transparency and Reproducibility of Research Software**
Project duration (in months)	36 Months

English public summary

This project aims to make research data and software easier to reuse and reproduce, especially in AI. Currently, shared software often lacks the information needed to replicate results. The project will improve transparency and long-term usability by enhancing Djehuty, the repository platform used by 4TU.ResearchData and Nikhef. This will benefit researchers, data professionals and policymakers by providing better tools and guidelines for sharing and using software responsibly, particularly software associated with AI models. The project will also align with ongoing efforts to improve research practices and meet new regulations, enabling AI research to be more open, FAIR, ethical and reproducible.

Word count EN-SUM (max 100): 100 words

Dutch public summary

Dit project heeft als doel om onderzoeksdata en software gemakkelijker herbruikbaar en reproduceerbaar te maken, vooral op het gebied van AI. Op dit moment ontbreekt het gedeelde software vaak aan de benodigde informatie om resultaten te repliceren. Het project verbetert Djehuty, het repositoryplatform van 4TU.ResearchData en Nikhef, om transparantie en langdurige bruikbaarheid te waarborgen. Dit komt onderzoekers, dataprofessionals en beleidsmakers ten goede door tools en richtlijnen te bieden voor verantwoord delen en gebruiken van software, vooral bij AI-modellen. Het sluit aan bij inspanningen om onderzoekspraktijken te verbeteren en regelgeving na te leven, zodat AI-onderzoek opener, FAIR, ethisch en reproduceerbaar wordt.

Word count NL-SUM (max 100): 100 words

Section 3: Alignment with the scope of the call

3.1 Vision for the project and alignment with the aim of this call

4TU.ResearchData provides a trusted digital repository for research data and software in science, engineering and design. The repository supports FAIR data and software by offering curation, sharing, long-term access and preservation. The software underlying the repository system is open source, developed by 4TU.ResearchData and Nikhef. While the service supports open sharing, deposited software or models are often irreproducible due to insufficient documentation and the lack of computationally executable instructions. We aim to extend the infrastructure by implementing transparency measures and long-term reproducibility guarantees for research software with an emphasis on machine learning tools. We will provide a software reproducibility service with a detailed Software Bill of Materials, as well as a runnable version of the software in multiple forms. Further, we will implement maximum transparency and optimise findability by capturing rich metadata, including ethics, using a model card schema. We will pay special attention to components associated with AI models, e.g. modelling data and AI training algorithms, because of the vast amounts of data, computational power and complex processes and specific conditions involved, and not least the impact AI models can have on society. These enhancements will enable software and models to be transparently reused, facilitating open and reproducible science.

Word count SEC31 (max. 200): 200

3.2 User communities, challenges and needs

A significant challenge facing open science is the reuse and reproducibility of research data and software. This is even more of a concern with the rise of machine learning and artificial intelligence systems, heightening the urgency for infrastructure to enable software and model sharing in ways that support greater transparency and reproducibility. It is essential to support the implementation of open science principles in AI as far as possible, not only to enable FAIR for AI, but to reduce reliance on private entities for AI research and capabilities. While researchers working with AI recognise the need for FAIR and ethical AI, there is a dearth of clear guidelines as well as trusted, sustainable infrastructure (Huerta et al., 2023; Skills4EOSC, 2025).

There are limited options for researchers who want to share their software or models in a way that enables not just access, but also sufficiently supports FAIRness, reproducibility and preservation. While researchers are increasingly encouraged to share their research software by funders and institutions (Technopolis Group, 2024; Research Software Alliance, 2025), this is an emerging practice that requires further support, including infrastructure to guarantee reusable software environments. Such environments are currently not readily provided by researchers or within commonly used infrastructures. While the proposed infrastructure focuses on AI as a primary use case, the work is more widely applicable to software and models in general.

Communities that will benefit from this project include researchers actively producing and sharing software and models (such as groups at the Dutch technical universities), as well as data, software and model stewards at Dutch institutions who support the responsible management and sharing of such outputs. In addition, policymakers who develop or enact guidelines also benefit from infrastructure that can respond to and support national or regional regulations, such as the European AI Act (EU, 2024).

Word count SEC32 (max. 300): 300

3.3 Alignment of project with OS principles

Our project expands and extends the capabilities of the 4TU.ResearchData and Nikhef repositories, which run on Djehuty. Djehuty is an open source platform developed by 4TU.ResearchData, a longstanding infrastructure with an established and active community of researchers and data professionals. The software is available on Github and repository users are invited to contribute ideas for enhancing existing and new services, as well as to recommend

changes to the code to implement these. All data and software published and preserved in 4TU.ResearchData are reviewed, curated, and open and FAIR by default.

In line with the Principles of Open Scholarly Infrastructure (Bilder et al., 2000), 4TU.ResearchData is stakeholder-governed via representatives from partner institutions, and makes the software, assets and data of its infrastructure openly available under open licences. In addition, to implement reproducible software environments, we will contribute to GNU Guix by ensuring that software published in 4TU.ResearchData will be made available in GNU Guix, a Free and open source software project that implements the Purely Functional Software Deployment Model (Dolstra, 2006; see section 4.1).

This project aligns with community-based standards by connecting with and building on several efforts that are already underway to address transparency and reproducibility, particularly related to research software and models. These include the Model Stewardship Programme at Wageningen University & Research, Netherlands; the Certification Working Group of the Open Modelling Foundation; the Model Openness Framework (Linux Foundation); the FAIR in ML, AI Readiness & Reproducibility Research Coordination Network (FARR RCN); NFDI4DataScience, Germany; the Dutch Reproducibility Network; the Research Data Alliance FAIR4ML Interest Group; the Skills4EOSC project; the EOSC EVERSE project; the Research Software Directory; and the CODECHECK project and community (including the NWO-funded CODECHECKing goes NL). Through the involvement of co-applicants and cooperation partners representing such initiatives, the project will actively engage with these networks and communities.

Word count SEC33 (max. 300): 300

3.4 Access policy

4TU.ResearchData makes all deposited data and software accessible to everyone, without cost or distinction between user groups. 4TU.ResearchData also doesn't restrict who can publish data or software, offering up to 5GB per year without cost and higher upload limits for researchers at partner institutions. Depositors are required to log in and agree with a Deposit Agreement.

The repository platform, Djehuty, supports different access types: the default is open access, but data and software that cannot be shared openly can be shared under restricted access or with a temporary embargo. Requests for access to restricted datasets are made via a button on the landing page and automatically sent to the author/depositor. If approved, the depositor sends a private link to the requester through which the relevant files are accessible. The conditions under which a restricted dataset is accessible are determined by the author/depositor and specified in an End User Licence Agreement.

Word count SEC34 (max. 150): 150

3.5 (Inter)national ecosystem

4TU.ResearchData has been part of the (inter)national landscape of data infrastructures since its founding in 2010. It works closely with (inter)national partners, including as part of EOSC projects (e.g. Skills4EOSC) and the EOSC Association. At national level, 4TU.ResearchData hosts the Thematic DCC for Natural and Engineering Sciences and this project will support the need identified in the TDCC NES roadmap (2022) to improve computational reproducibility practices.

In the EOSC context, this project aligns with objectives in the Multi-Annual Roadmap, specifically the need to further support the development, archiving, sharing and reuse of research software, focusing on reusability and the use of standards that facilitate the development and uptake of software (EOSC Association, 2025). Further, the work aligns with the EOSC EVERSE project, which has the goal to develop a community led framework for evaluating research software quality and improve reproducibility. This project will reuse the software quality indicators developed by

EVERSE to develop accreditation criteria, extending the set of indicators to support models and reproducibility. We will work with EVERSE to integrate the outcome of this project.

Finally, the infrastructure has the potential to contribute to the emerging EOSC Federation via EOSC nodes at the national or thematic level.

Word count SEC35 (max. 200): 199

Section 4: Feasibility

4.1 Project plan

The work of this project is planned across 5 work packages:

WP 1: Project management and user engagement

WP1 is led by the Project Coordinator and facilitates both internal project management between partners as well as engagement with user communities and stakeholder groups. At least four in person Community Workshops will be held to consult and collaborate with user communities, co-organised with and promoted through networks related to reproducibility, AI, software and models (see section 4.2). These will be held at regular intervals, but may shift to align/co-locate with relevant conferences (e.g. RDA, ReSA, FARR RCN, NFDI, Dutch open science/research data events).

Deliverables:

D1.1 Stakeholder engagement plan
D1.2 Report on community workshops

WP 2: Accreditation criteria

WP2 develops the criteria for a graded badge (together with the Research Software Directory) to assign a quantitative value to metadata and documentation of data, software and models, adopting and extending the software quality indicators from the EVERSE project (European Virtual Institute for Research Software Excellence, 2024), using a model card schema (Mitchell et al, 2019). These criteria would consider information on data collection, data (pre)processing, software versions, software dependencies, licensing, FAIR attributes, detailed instructions on running code, etc. We will also develop a method to quantify risk and potential impact of a model through classification with reference to the EU AI Act (EU, 2024). In summary, the tasks are to: 1) develop a comprehensive metadata schema based on a model card approach; 2) quantify and assign weights to all metadata fields; 3) create grade ranges for badges based on completeness and quality of metadata and other components (instructions, dependencies, etc).

Deliverables:

D2.1 Report containing the criteria and scoring scale(s) for transparency/reproducibility badge
D2.2 Guidelines for implementation

WP 3: Software reproducibility service

WP3 implements the Purely Functional Software Deployment Model (Dolstra, 2006) so that deposited software and its environment can be archived in a bundle, and a starting point for continuing its development is preserved. The tasks are to: 1) set up a build service which implements the Purely Functional Software Deployment Model (PFSDM); 2) build the software environments for published software in 4TU.ResearchData using the build service; 3) adapt the software submission form to integrate with the build service; 4) and add distributable artifacts from the build service in the CodeMeta API, for use in the RSD (NL eScience Center). Included in this last task will be a machine-readable description of software dependencies that forms the basis for a Software Bill of Materials document.

Deliverables:

D3.1 Build descriptions for software deposited in 4TU.ResearchData
D3.2 Code contributions to implement the CodeMeta API extension in the 4TU.ResearchData repository
D3.3 Revised software submission form in the 4TU.ResearchData repository

WP 4: Integration of accreditation criteria

WP4 implements the User Interface workflow to capture the required metadata (WP2) to determine the quantitative value in the software underlying the 4TU.ResearchData and Nikhef data repositories. The tasks required are a technical implementation in the software as well as an adaptation of the metadata curation workflow by 4TU.ResearchData.

Deliverables:

D4.1 Badge implementation in the data repositories (4TU.ResearchData & Nikhef) and in the RSD
D4.2 Revised dataset and software submission forms to include complete metadata and documentation

WP 5: Curation and reproducibility workflows

A crucial requirement for widespread uptake of this infrastructure is that the functionality and outcomes match the needs and workflows of the diverse user (research) communities. WP5 will work with CODECHECK, an initiative that has collected a substantial body of knowledge on community needs and accumulated valuable experience with community-driven curation workflows. This WP will firstly further inform the development of feasible and useful

accreditation criteria in WP2. The main activity of WP5 will then build on the continuing output of WP3 and WP4: we envision a sample validation of the output from the software reproducibility service to check whether it is working as intended and delivering reproducible workflows and whether badges have been correctly assigned. This will inform an optimised review process for the repositories, balancing automated processes with human checks. WP5 will also explore and implement integration of existing lightweight CODECHECK infrastructure with this infrastructure.

Deliverables:

D5.1 Report on sample validations and reproducible workflows
D5.2 Revised CODECHECK infrastructure

Timeline:

	Year 1						Year 2						Year 3					
	M1	3	5	7	9	11	13	15	17	19	21	23	25	27	29	31	33	35
Work Package 1		D1.1	CW1				CW2				CW3				CW4			D1.2
Work Package 2					D2.1		D2.2											
Work Package 3					D3.1			D3.2				D3.3						
Work Package 4														D4.1			D4.2	
Work Package 5														D5.1			D5.2	

*CW = Community Workshop

Word count SEC 41 (max. 750): 750

4.2 Team composition

Name	Affiliation	Expertise	Role/contributions
Dr. Daniel Bangert	4TU.Research Data/TU Delft	Director 4TU.ResearchData	4TU.ResearchData coordination; main applicant
Roel Janssen	4TU.Research Data/TU Delft	Senior Software Engineer	WP3 lead & WP4 co-lead; lead technical development of Djehuty and PFSDM implementation
	4TU.Research Data/TU Delft	Project Coordinator	WP1 & WP5 lead; lead overall project management, workshop organisation, communications and dissemination; administration of project finances and reporting
	4TU.Research Data/TU Delft	Application Standards Developer	WP2 lead & WP4 co-lead; lead development of accreditation criteria
	4TU.Research Data/TU Delft	Data Steward	Advise on storing ML/AI models and input data (WP3-5)
Dr. Jason Maasen	Netherlands eScience Center	RSD Product Lead & Community Manager	Contribute to accreditation criteria (WP2); link to EVERSE project and RSD community
Ewan Cahen	Netherlands eScience Center	RSD Developer	Integrate reproducibility and accreditation metadata into RSD (WP3 & WP4)
Dr. David Groep	Nikhef	Physics Data Processing Programme Leader	Nikhef coordination
	Nikhef	Software Infrastructure Engineer	Integrate reproducibility and accreditation metadata into Nikhef repository (WP3 & WP4)
Dr. Frank Ostermann	University of Twente	Associate Professor Geo-Information Science	CODECHECK coordination (WP5); link to CODECHECK initiative and community
Angelina Momin	University of Twente	CODECHECK Software Developer	Update CODECHECK infrastructure (WP5)
Dr. Daniel Katz	San Diego Sup	Chief Scientist, National	Workshop participant/co-organiser; link to the FAIR in

	ercomputer Center, UC San Diego	Center for Supercomputing Applications; FARR RCN co-investigator	ML, AI Readiness, & Reproducibility Research Coordination Network (FARR RCN)
Dr. Leyla Jael Garcia Castro	ZB MED Information Centre for Life Sciences	Team Leader Semantic Technologies; NFDI4DataScience consortium	Workshop participant/co-organiser; link to NFDI4DataScience (national research data infrastructure for data science and AI in Germany)
Dr. Curtis Sharma	4TU.Research Data/TU Delft	Project Manager Skills4EOSC; RDA FAIR4ML co-chair	Workshop participant/co-organiser; link to the Research Data Alliance FAIR for Machine Learning Interest Group
Dr. Michiel de Boer	UMC Groningen	Associate Professor Epidemiology; Head of NLRN Steering Committee	Workshop participant/co-organiser; link to the Dutch Reproducibility Network (NLRN)
Dr. George A.K. van Voorn	Wageningen University & Research	Associate Professor Applied Modelling & Machine Learning	Workshop participant/co-organiser; link to WUR modelling initiatives (e.g. WUR Modelling Group and the Model & Data Day)
Ivar Janmaat	SURF	Manager SURF Research Cloud	Workshop participant/co-organiser; link to reproducibility initiatives at SURF

Word count SEC42 (max. 350): 349

4.3 Risk management

Risk	**Likelihood and impact**	**Risk mitigation strategy**
Lack of trust in the infrastructure developed, resulting in lower than expected uptake.	Likelihood: Medium Impact: Medium	We will work closely with user communities and (inter)national networks to build trust and validate the proposed infrastructure. Outcomes will be transparently communicated and be verifiable (e.g. transparent accreditation criteria and scoring).
Unable to converge to a single metadata standard for AI models; making implementing such metadata standard more time-consuming for the data repository.	Likelihood: Medium Impact: Low	If this occurs, the Application Standards Developer can continue their work. The permanent software developers at 4TU.ResearchData can continue implementing the standards after the project funding runs out.
Relying on or publishing proprietary software becomes more popular than publishing fully open source software; The nonfree software cannot be upstreamed into GNU Guix.	Likelihood: Low Impact: Low	4TU.ResearchData can maintain its own software “channel” to provide limited support on making reproducible software environments available for these cases.

Word count SEC43 (max. 250): 158

4.4 Budget

Position	HOT scale	Institu-tion	Total FTE	Years active	Amount
Project Coordinator	[redacted]	TU Delft (4TU.ResearchData)	3.00	3	[redacted]
Software Engineer	[redacted]	TU Delft (4TU.ResearchData)	1.50	3	[redacted]
Accreditation Standards Developer	[redacted]	TU Delft (4TU.ResearchData)	1.80	3	[redacted]
Data Steward	[redacted]	TU Delft (4TU.ResearchData)	0.60	3	[redacted]
Software Infrastructure Engineer	[redacted]	Nikhef	0.90	3	[redacted]
Software Developer	[redacted]	Netherlands eScience Center	0.90	3	[redacted]
Research Software Directory Product Lead	[redacted]	Netherlands eScience Center	0.15	3	[redacted]
CODECHECK Coordinator	[redacted]	University of Twente	0.42	3	[redacted]
CODECHECK Software Developer	[redacted]	University of Twente	0.29	3	[redacted]

Material/IT Type	Description	Cost calculation	Total (€)
national symposium/conference/workshop organised by the project itself;	Organisation of workshops in collaboration with cooperation partners	4 workshops (catering, venue hire: 10K per workshop)	€ 40,000.00
(international) travel and accommodation expenses incl. conference visit	Attendance of project personnel at relevant international conferences (e.g. RDA, ReSA, FARR RCN, NFDI); attendance of international cooperation partners and other speakers/participants at project workshops	1.25K per conference x 3yrs x 3p; 1.25K per workshop x 4 workshops x3p	€ 26,250.00
Hardware	Server hardware and housing	Hardware 40K; housing at Nikhef 25K	€ 65,000.00
knowledge dissemination costs	Communication and dissemination	Digital and physical project materials costs	€ 20,000.00

Total Personnel requested: 1.250.177,20
Total Material/IT requested: 151.250,00
Total budget requested: 1.401.427,20

4.5 Budget justification

Personnel costs: At 4TU.ResearchData, there are four roles funded by the project, including a Project Coordinator (1FTE) that will be hired to lead the overall project management as well as workshop organisation and project communications and dissemination. This position will lead WP1 and WP5, and collaborate with cooperation partners and others on the Community Workshops. An Application Standards Developer (0.6FTE) will be hired to lead WP2 and co-lead WP4. A Software Engineer (0.5FTE) will lead WP3 and co-lead WP4, and a Data Steward (0.2FTE) will be hired to contribute and provide advice for WPs 3-5. At Nikhef, a Software Infrastructure Engineer (0.3FTE) will work on WP3 and WP4. At the Netherlands eScience Center, a Software Developer for the Research Software Directory (0.3FTE) will work on WP3 and WP4, and the RSD Product Lead will contribute 0.05FTE. The University of Twente will

provide input from CODECHECK for WP5, involving a Coordinator (0.14FTE) and a Developer (0.1FTE). There will also be in kind contributions, for example from 4TU.ResearchData to review policies and curation workflows to help implement and maintain the outcomes of the project.

Material and IT costs: Non-personnel costs are requested for the organisations of Community Workshops (€40K), travel for participation in the workshops and other conferences (€26.250), server hardware and housing required for the infrastructure (€65K), and dissemination costs (€20K). As the project seeks to build on and engage widely with (inter)national initiatives, engagement with users and other stakeholders are crucial to all elements of the work described. Where possible, workshops will be co-located with (inter)national events of relevance to user communities (e.g. reproducibility, AI, software, models).

Word count SEC45 (max. 300): 267

4.6 Impact plan

This project will facilitate FAIR for AI research, and software and models more broadly, enabling researchers to share, reproduce and preserve their work using trusted infrastructure. ML/AI can have a large impact on society and this project will demonstrate ways forward and set standards for these processes and outputs that encourage open science, preserve digital sovereignty and comply with relevant regulations.

Through open source development and open standards, the work can be readily connected or adopted by others. For example, the software reproducibility service will enable ready to run environments on a national level, meaning that all software deposited in 4TU.ResearchData can be reused and run on infrastructures provided by others (e.g. SURF or an EOSC node). The proposed work on software reproducibility directly adds to the Free and open source software community by means of the GNU Guix project. 4TU.ResearchData can liaise for other institutions in the Netherlands to do similarly.

The infrastructure will support greater reusability of software, AI/ML models and related data for the benefit of all (including scientists and businesses). By leveraging an existing platform with established links to communities and institutions, the outcomes will be of broad national benefit and sustained into the future.

Word count SEC46 (max. 200): 199

Section 5: Sustainability and software management

Section 5.1: Sustainability plan

This project will result in several outcomes that need to be sustained after the OSNL project funding period, including the accreditation criteria and mechanism (graded badge), software reproducibility service, and curation and reproducibility workflows. The organisations involved in developing these outcomes are committed to long-term sustainability in the context of their organisational plans, as outlined below:

4TU.ResearchData develops its repository service in close collaboration with the research community and aims to meet domain-specific needs for science, engineering and design. Through this project, it will test and implement support for ML/AI and software reproducibility, which it intends to sustain by integrating as part of its core services offered through the repository. The 4TU.ResearchData repository has been in operation for 15 years, with a mission and mandate as a trusted digital repository with CoreTrustSeal certification offering curation, access and long-term preservation (guaranteed for at least 15 years). 4TU.ResearchData is funded and governed by a consortium of the four technical universities in the Netherlands (4TU); it is therefore not-for-profit and supported as part of a long-term strategic vision of the 4TUs. In addition, Djehuty as the underlying repository platform is currently supported by staff based at two institutions (TU Delft and Nikhef) and its open source model encourages community contribution and shared ownership. 4TU.ResearchData expects the platform to remain key to its services and is committed to further development and maintenance to support its mission as an organisation and its role as part of the (inter)national landscape of open science infrastructures. As software and models are relevant for the domains supported by 4TU.ResearchData, it is well placed to continue the required user engagement, aided by its ongoing community and training activities, and sustain the infrastructure well beyond the OSNL project funding period.

The Research Software Directory is currently under active development and funded by the Netherlands eScience Center and the Helmholtz Association. Both organisations are committed for the coming years to support the platform as it is extensively used at both organisations, as well as some 35 other organisations and 70 research groups worldwide. Both teams administer their respective online RSD services, provide user support and (a moderate amount of) content curation, and are responsible for community management. Additional funding is frequently obtained through project proposals with various partners, and an OSNL proposal will be submitted to this call aiming to further improve the governance structure of the platform. The RSD is developed as open source on GitHub. Snapshots of the source code are frequently archived by Software Heritage, while all official releases are archived on Zenodo.

Word count SEC51 (450): 424

Section 5.2: Software management plan

1. Please provide a brief description of your software, stating its purpose and intended audience.

Djehuty is a data and software repository used by universities and research institutions to publish and archive data and software, accompanied by metadata. The purpose of the software is to provide a place where data and software is stored under the FAIR principles. The intended audience consists of researchers and others who work on data and software used for scientific publications.

2. How will you manage versioning of your software?

The software is released on a monthly schedule through Github and the 4TU.ResearchData repository.

3. How will you make your software publicly available? What licence will your software have?

The software is released under the Affero GPLv3+ licence, publicly developed on Github (with a mirror on Codeberg), and published in 4TU.ResearchData on a monthly basis.

Beyond merely making the software publicly available, the software only uses dependencies that are commonly available in GNU/Linux distributions, making it more likely that Free and open source software developers are familiar with the intricacies of its software dependencies.

4. How will users of your software be able to cite your software?

Djehuty publishes its monthly releases in the 4TU.ResearchData repository (10.4121/342efadc-66f8-4e9b-9d27-da7b28b849d2) from which citations in various formats can be copied or exported.

5. How will your software be documented? Please describe your plans for: user documentation, documentation for future developers and for installation requirements. Please provide links to documentation if already available.

The documentation can be found in the "doc" folder of the source code, and a compiled instance of that documentation is available on https://djehuty.4tu.nl where there's also a PDF version of said documentation.

6. How will contribution guidelines and governance structure of your software be documented?

Guidelines on how to contribute to the software are found at https://djehuty.4tu.nl/. The project largely follows GNU contribution guidelines, including recommended tools to (re)distribute software.

7. How will your software be tested?

Other than the integration and regression testing done by 4TU.ResearchData using a staging environment (https://next.data.4tu.nl), the formal releases are analysed using Coverity Scan for which the results can be found at https://scan.coverity.com/projects/djehuty/.

8. How will you check that it respects the licences of libraries and dependencies it uses?

The software is packaged for GNU Guix, a GNU distribution that has a strict no-nonfree-software policy. The dependencies used by/in the software are commonly distributed dependencies in GNU/Linux distributions. We build the software for Enterprise Linux (with the EPEL repository), for which the output can be found here: https://copr.fedorainfracloud.org/coprs/4turesearchdata/djehuty

9. How will your software be packaged and distributed?

The software is made available on PyPi, and each release includes RPM packages for EPEL-compatible GNU/Linux distributions. Each release also includes a Docker image made available through DockerHub. In the mid-to-long term, the software will be attempted to be included in GNU/Linux distributions. The first steps have been taken to adhere to the packaging guidelines of Fedora Linux, which is the first step to include it in EPEL.

10. What level of support will be provided for users of the software and how will this support be organised?

There are multiple support channels: On 4TU.ResearchData, the feedback form on the main page is used by researchers to requests features and report bugs. Nikhef has an internal Mattermost channel to capture feedback, and we've seen both contributions through pull-requests on Github and patches sent by e-mail. (See the Github history to see our track record of including contributions).

11. How do you plan to ensure long term maintenance of your software?

The codebase is rather small (~15000 lines of Python at the time of writing) and it doesn't have many dependencies. Moreso, TU Delft employs two software engineers to maintain and develop it regardless of project-based funding. At Nikhef there is time dedicated to the maintenance of the data repository powered by Djehuty by a software infrastructure engineer.

Section 6: References

4TU.ResearchData. (2025). *4TU.ResearchData Strategy 2025-2029*. https://doi.org/10.5281/zenodo.14176780

Bilder, G., Lin, J., & Neylon, C. (2020), *The Principles of Open Scholarly Infrastructure*. https://doi.org/10.24343/C34W2H

CODECHECK (2025). https://codecheck.org.uk

Dolstra, E. (2006). *The Purely Functional Software Deployment Model*. https://edolstra.github.io/pubs/phd-thesis.pdf

EOSC Association (2025). *EOSC Multi-Annual Roadmap*. https://eosc.eu/sria-mar/

European Virtual Institute for Research Software Excellence (2024). http://everse.software

EU (2024). *Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)* https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689

FARR: FAIR in ML, AI Readiness & Reproducibility Research Coordination Network (2025). https://www.farr-rcn.org

Huerta, E.A., Blaiszik, B., Brinson, L.C. et al. (2023), FAIR for AI: An interdisciplinary and international community building perspective. *Scientific Data* 10, 487. https://doi.org/10.1038/s41597-023-02298-6

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., ... & Gebru, T. (2019, January). Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency* (pp. 220-229). https://doi.org/10.1145/3287560.3287596

NFDI4DataScience (2025). https://www.nfdi4datascience.de

Ostermann, F., Nüst, D., Granell, C., Hofer, B., & Konkol, M. (2021). Reproducible Research and GIScience: An Evaluation Using GIScience Conference Papers. In 1*1th International Conference on Geographic Information Science (GIScience 2021) - Part II* (pp. 2:1–2:16). Schloss Dagstuhl – Leibniz-Zentrum fur Informatik. https://doi.org/10.4230/LIPIcs.GIScience.2021.II.2

Research Data Alliance FAIR4ML IG (2025). https://www.rd-alliance.org/groups/fair-machine-learning-fair4ml-ig/activity

Research Software Alliance (2025). *Research institution policies to support research software*. https://www.researchsoft.org/software-policies/

Skills4EOSC (2025). https://www.skills4eosc.eu

TDCC NES (2022). *Roadmap TDCC Natural and Engineering Sciences*. https://www.nwo.nl/en/researchprogrammes/implementation-plan-investments-digital-research-infrastructure/roadmaps-three

Technopolis Group. (2024). *Dataset and software citations and mentions in NWO and ZonMw funded research*. Zenodo. https://doi.org/10.5281/zenodo.14389449

Section 7: Declaration

By submitting this form, I declare that:

- I and all the individuals involved in this proposal 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).
- The research organisation has been informed of this grant application and the research organisation accepts the grant conditions of this programme.
- I have completed this application form truthfully.
- I have submitted a pre-proposal for this Call for proposals in ISAAC.