SoFAIR

First report on communication and dissemination activities

Submission Date	14.01.2025
Due Date	31.12.2024
Version	1
Deliverable Title	First report on communication and dissemination activities
Deliverable No.	2.2
Lead beneficiary	IBL-PAN (BUT)
Contributing WP	2
Funding id	EP/Y036352/1

Project Full Title	Making Software FAIR: A machine-assisted workflow for the research software lifecycle
Project Acronym	SoFAIR
Project No.	EP/Y036352/1
Start Date	2024
End Date	2025
Duration	24
Project Website	https://sofair.org/
Authors	Patryk Hubar, Tomasz Umerle
Abstract	The first Communication and Dissemination Report for the SoFAIR project outlines 2024 activities in alignment with the Communication and Dissemination Plan. It details internal measures, including the adoption of standardised document templates, dedicated mailing lists, and real-time collaboration tools like Slack and Discord, ensuring efficient communication. The report highlights two consortium meetings, structured workshops on workflow integration, and tracking of conference participation. Additionally, the document describes the presentation of SoFAIR at seven events, the development of three

	key publications, and the planning of a webinar. Remaining sections cover dissemination efforts via workshops, blog posts, social media, and the project website, meeting Year 1 targets.

Version and Revision History

Version	Date	Author/Reviewer/Contributors	Comments
0.1	17.10.2024	Patryk Hubar-Kołodziejczyk	First draft
0.2	18.12.2024	Petr Knoth, Laurent Romary	Reviews
1	31.12.2024	Patryk Hubar-Kołodziejczyk, Tomasz Umerle, David Pride	Final version

List of Acronyms

C&D	Communication and Dissemination	
CDP	Communication and Dissemination Plan	
D	Deliverable	
FAIR	Findable, Accessible, Interoperable, and Reusable	
PIDs	Persistent identifiers	
ML	Machine learning	
NGOs	Non-Governmental Organisations	
KPIs	Key Performance Indicators	

TABLE OF CONTENTS

Executive Summary

The first Communication and Dissemination Report for the SoFAIR project covers activities undertaken in 2024, following the guidelines set forth in the Communication and Dissemination Plan. Internally, the project established standardised document templates, introduced dedicated mailing lists, and leveraged Slack and Discord for real-time collaboration, ensuring efficient coordination and tracking of participation in conferences and events. Two consortium meetings—in Milton Keynes (kick-off) and Brno—further strengthened team alignment, with focused workshops on workflow integration, research documentation, and dissemination strategies.

Externally, SoFAIR was presented at seven conferences and seminars—such as the CHISTERA Project Seminar, Open Repositories, and RDA events—reaching over a thousand participants. Noteworthy external engagement also included a webinar (scheduled for January 2025) that will showcase the SoFAIR workflow for enhanced software reusability and reproducibility, as well as a set of three key publications: an updated SoFAIR dataset, SoFAIR workflow documentation outlining machine-assisted software citation, and a preliminary scientific paper. Additional dissemination took place through workshops (attended by around 50 participants), nine blog posts, active involvement on X (Twitter), and a regularly updated project website (sofair.org), fulfilling or surpassing most Year 1 Key Performance Indicators.

1. Introduction

This report provides a detailed description of the communication and dissemination activities carried out during the first year of the SoFAIR project, structured according to the Key Performance Indicators planned for 2024 in the Communication and Dissemination Plan (available at: https://sofair.org/results). The report is divided into two main sections: Internal Communication and External Communication.

The Internal Communication section outlines the tools and strategies implemented to support collaboration and coordination within the project team. Key activities include the establishment of mailing lists, the use of Slack and Discord channels for communication within work packages, regular team meetings, and the development of standardised document templates to ensure consistency. This section also details internal tracking mechanisms for team attendance at conferences and meetings, and the workflows established for efficient information sharing and documentation.

The External Communication section describes the initiatives aimed at raising awareness and promoting the SoFAIR project to external audiences. These include the publication of blog posts, active engagement on X (Twitter), representation at conferences, the launch of a project website and various events aimed at different stakeholders both the academic and wider research communities.

2. Communication and dissemination activities

2.1. Internal Communication

Standardised templates have been created for key project communications, including presentations, internal documents, external reports and blog posts. These templates ensure consistency in format and style, streamlining the development of materials across all project activities. The presentation template provides a consistent structure for sharing project updates with external audiences, while the internal and external document formats support clear and professional reporting. The blog post templates guide the creation of content for project updates, ensuring structured and accessible communication with stakeholders. In addition to document templates, the project's internal communication setup includes mailing lists established for the overall project and specific work packages. Slack and Discord channels facilitate real-time discussions within work packages and tasks. To effectively track dissemination activities, spreadsheets have been developed to track team members' participation in national and international conferences, workshops, seminars and other relevant meetings. Regular meetings are held both for the whole project team and for smaller, work package-focused groups to enable ongoing coordination, progress updates and problem-solving.

2.1.1. PROJECT MEETINGS

The project consortium held 2 project meetings in this period. The project kick-off meeting took place in Milton Keynes 9-10.01.2024 and was hosted by The Open University. There were 12 physical participants and 2 virtual participants from across the consortium. Each partner prepared presentations related to their project's responsibilities with the objective to present and align the partners around the strategy and methodology of the delivery.

Partners also agreed on regular monthly meetings for the consortium and on specific monthly meetings for WP3, WP4 + T5.1 and WP6. All of these meetings have a recorded agenda and action points.

Photo 1. Participants of the kick-off meeting on January 9.01.2024, in front of Bletchley Park (in Milton Keynes), where Alan Turing and other scientists famously cracked the WW2 german Enigma code using computational methods.

The second consortium meeting took place on 17th-18th December in Brno, Czech Republic. 10 participants joined in person and there were 4 additional online participants. The meeting was organised into three workshop workstreams dedicated to: (1) Research and experiments write-up; (2) Workflow and platform integration and (3) Communication and dissemination.

Photo 2. Participants of the SoFAIR annual meeting on 17.12.2024 at the Brno Technical University.

2.2. External communication

2.2.1. Conferences and seminars presentations

In the first project year, the SoFAIR Project was presented at 7 conferences and external seminars, reaching more than 1 000 people. These activities allowed for targeted communication aimed at different stakeholders, laid out in the Communication and Dissemination Plan.

SoFAIR was presented at the CHISTERA Project Seminar which allowed to showcase SoFAIR amongst other **research institutions** (awarded initiatives) and **funders**. Special attention was paid to targeting the **libraries and open repositories community**. At the annual Open Repositories (OR) Conference, Professor Knoth represented the SoFAIR consortium, discussing our mission to streamline and semi-automate research software management processes and contribute to increased reproducibility and recognition for software assets within academia.
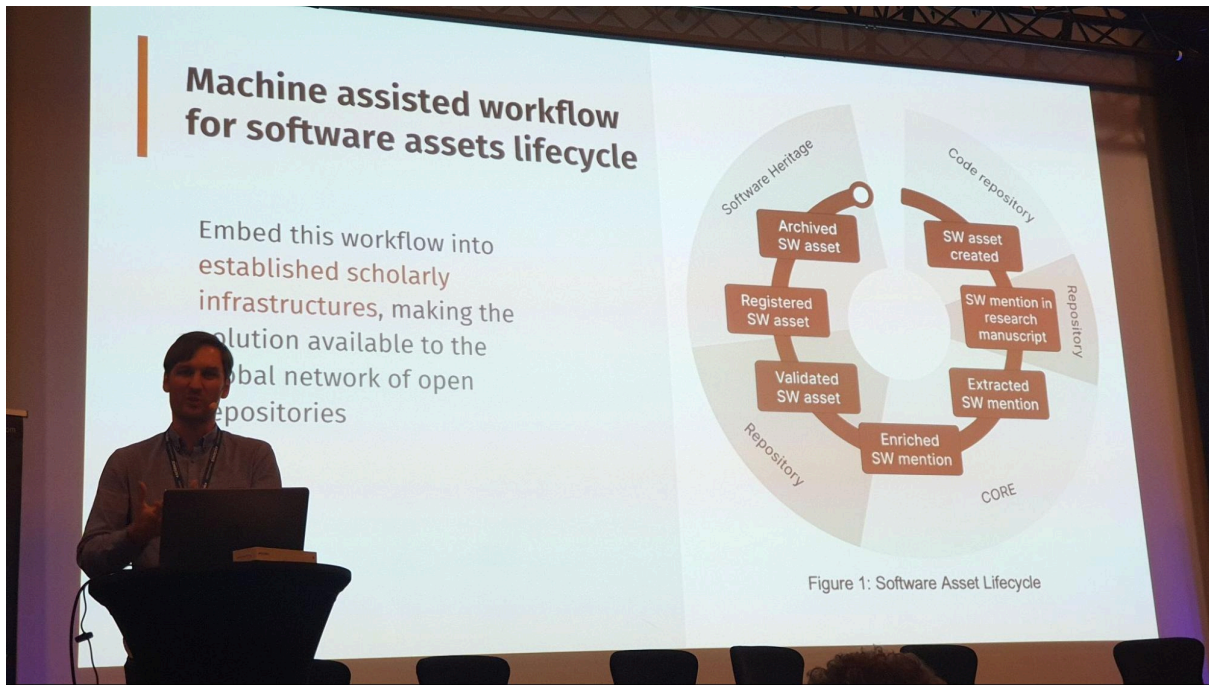
Photo 3: Presentation of SoFAIR at Open Repositories 2024

This event, known for drawing librarians, developers, and repository managers, allowed us to share SoFAIR work with a diverse audience focused on open knowledge, digital preservation, and the FAIR principles (Findable, Accessible, Interoperable, Reusable). SoFAIR was also present at the HAL Partners Assembly to present its role for development of repositories potential for identifying research software mentions. To showcase the role of SoFAIR for the **Open Science (OS) community** SoFAIR was presented during key OS events: Open Science Week 2024 and Open Science Days UGA. Finally, the Project was presented at the RDA Plenary Meeting, allowing it to reach a specific group of **research data experts**.

No	Dissemination item name	Presenters (partner)	Date	Dissemination channel	Content	Audience size	Notes
1.	SoFAIR presentation at the CHISTERA Project Seminar 2024. Program	Petr Knoth (OU)	16-18 April 2024	Project seminar	LINK	~20 detailed SoFAIR presentation for the ORD group AND ~100 the whole	1x SoFAIR presentation for the ORD Call group and 1x presentation of the ORD call group to the rest

No	Dissemination item name	Presenters (partner)	Date	Dissemination channel	Content	Audience size	Notes
					CHISTERA event		
2.	SoFAIR presented at Open Repositories 2024	Petr Knoth (OU)	4th June 2024	Conference Presentation	LINK	>120	
3.	Innovative library in digital era 2024 (ILIDE 2024)	Petr Knoth (OU)	30th September - 2nd October 2024	Networking	EVENT LINK	>80	Networking activities
4.	SoFAIR presented at Open Science Week 2024 Program	Petr Knoth (OU)	22nd October 2024	Webinar / conference presentation	LINK	>100	
5.	SoFAIR presented at HAL Partners Assembly	David Pride (OU)	20th November 2024	Conference Presentation	LINK	>400	
6.	Open Science Days UGA (France)	Roberto Di Cosmo (INRIA)	5th April 2024	Conference Presentation	LINK	>60	
7.	Research Data Alliance 22nd Plenary Meeting (RDA VP22)	Morane Gruenpeter (INRIA)	15th May 2024	Conference Presentation (online session)	LINK	>75	

Figure 1. Conferences and seminar presentations

Additionally, a SoFAIR poster presentation *Preserving Cultural Heritage in the Digital Age: The Role of Software Heritage in Safeguarding Research Software* for the largest global digital humanities conference (DH2025 in Lisbon in July of 2025) was submitted and a submission for the Open Repositories 2025 conference is planned for early 2025.

2.2.2. WEBINARS

All relevant preparations and planning have been carried out to prepare a webinar *Towards Reusability and Reproducibility of Research: A Scalable Workflow for Research Software Archival with CORE and Software Heritage* on January 15th, 2025.

Event agenda:

➢ Overall project mission (Petr Knoth, SoFAIR PI, Open University)

➢ Outline of the SoFAIR workflow (Morane Gruenpeter, Software Heritage)

➢ AI-based software extraction in CORE (David Pride, Open University)

➢ Management of software assets in open and institutional repositories (Martin Dočekal, Brno University of Technology)

➢ Archival of software assets with Software Heritage (Morane Gruenpeter, Software Heritage)

➢ Q&A

This online event, which was open to current and potential collaborators, students, and researchers in open science and digital data processing, included a detailed demonstration of the project's Workflow. Dissemination of the webinar included outreach towards key stakeholder groups: Research Infrastructures, Libraries and open repositories, Research software organisations. SoFAIR used social media, mailing and communication tools (e.g. dedicated Slack channels) to reach the audience. During the session, participants learned about the key stages of the SoFAIR workflow, allowing them to gain insight into our practical approaches. The webinar was recorded and will be made available to participants for future reference.

Title	Date	Registered	Attending
Towards Reusability and Reproducibility of Research: A Scalable Workflow for Research Software Archival with CORE and Software Heritage	15.01.2025	85	45

Figure 2. Summary of SoFAIR webinar activities.

2.2.3. PUBLICATIONS

During 2024, the SoFAIR team produced three major outputs. First, a **SoFAIR dataset** was made available and iteratively updated. As of the end of 2024, it contains a carefully selected collection of scholarly articles sourced from CORE and offered in multiple TEI XML formats

(pre-annotated, manually annotated and fully validated). It contains only CC-BY articles (listed in the SoFAIR-corpus-first-subset-cc-by.json file) as only these were eligible for annotation, while non-CC-BY articles had to be excluded. The dataset is available at https://github.com/SoFairOA/Dataset.

Second, the **workflow documentation** *Making Software FAIR: A machine-assisted workflow for research software* describes how the SoFAIR workflow treats research software mentions as first-class bibliographic records, enabling proper attribution and long-term archiving. In this document, the authors outline how software citations are detected, validated and securely preserved, with particular emphasis on the overarching goals and procedural steps of the SoFAIR workflow. The process begins with researchers depositing their code in a designated repository and then submitting a manuscript containing explicit or implicit references to that software. CORE then harvests the manuscript and applies advanced machine learning techniques (GROBID/Softcite) to identify relevant software mentions. Once identified, these mentions are validated via the CORE repository dashboard, either by the repository manager or directly by the authors. Following validation, the repository initiates a request to permanently archive the software and secure a persistent identifier from Software Heritage. This approach will be validated against two disciplinary use cases - Life Sciences (in collaboration with Europe PMC) and Digital Humanities (in collaboration with DARIAH) - and a multidisciplinary use case (in collaboration with HAL). The documentation for this workflow is primarily intended for developers responsible for implementing SoFAIR's machine-aided processes. Full description of workflow is available at https://sofairoa.github.io/documentation. Further communication and dissemination activities are being realised around this output, with the 1st SoFAIR webinar on January 15th as the primary tool for dissemination.

Third, a presentation was given at the 19th Open Repositories conference held in Gothenburg, Sweden. This presentation provided a detailed introduction to the SoFAIR project and the infrastructures being leveraged to deliver the project's goals. The presentation was given by Professor Petr Knoth to an audience of researchers, research managers and repository managers. This submission is now made available on arXiv: https://arxiv.org/abs/2501.10415

2.2.4. OTHER DISSEMINATION ACTIVITIES, INCLUDING NETWORKING

In addition to our external representation on conferences and webinars, the SoFAIR project organised educational events for students and academic staff and networking efforts. On September 30th, Petr Knoth, the project's coordinator, attended the ILIDE 2024 where many librarian stakeholders were reached through informal networking.

For academic staff, a presentation was held on December 2 at the Institute of Literary Research (IBL PAN) within the Digital Humanities Centre for the Digital Humanities Centre staff. This presentation was aimed at engaged lecturers and researchers interested in exploring applications of data annotation in their work.

Figure 3. Screenshot of presentation for Digital Humanities Centre at IBL PAN.

On December 10, a workshop was held for students at the University of Warsaw, aimed at teaching them how to annotate data in accordance with the established SoFAIR workflow. The workshop combined a theoretical overview of the annotation workflow with a hands-on session focused on annotating textual data to identify mentions of software, aligning with the project's objectives. Additionally, participants were trained to enhance their skills in data annotation, which resulted in increased awareness of the problematics associated with research software management. All these activities allowed SoFAIR to reach approximately 50 additional people.

2.2.5. BLOG POSTS

During this reporting period, 9 blog posts have been published to inform stakeholders and the general public about the progress of the project and the roles of the various project

partners. Initial posts were dedicated to inform the public about the key information on the project's goals and were published through partners' dissemination channels. Posts published on the SoFAIR website highlighted the contributions of specific partners and how their expertise is essential to achieving the overall goals of the SoFAIR project and were focused around the project's main activities (such as participation at important conferences and seminars and planned SoFAIR webinar) and outputs. An in-depth blog post detailing the SoFAIR workflow was published providing a comprehensive overview of the workflow designed and adopted in the project.

No	Title	Topics & connected WPs	URL
1.	IBL PAN w projekcie SoFAIR	Key information on project (partners, goals, planned results)	https://ibl.waw.pl/pl/strona-glowna/aktualnosci/ibl-pan-w-projekcie-sofair
2.	SoFAIR: The Open University to coordinate new international project to facilitate the reproducibility of research studies	Key information on project (partners, goals, planned results)	https://blog.core.ac.uk/2023/11/07/sofair-the-open-university-to-coordinate-new-international-project-to-facilitate-the-reproducibility-of-research-studies/
3.	Connecting publications and software with SoFAIR	Key information on project (partners, goals, planned results)	https://www.softwareheritage.org/2023/12/08/connecting-publications-software-sofair/
4.	Výzkumníci z FIT VUT pracují v rámci projektu SoFAIR na nových metodách propojování publikací a softwaru	Key information on project (partners, goals, planned results)	https://www.fit.vut.cz/fit/press/3752/.cs
5.	SoFAIR is live! Kickoff, progress and website launch	Key information on project, projects's kickoff	https://sofair.org/kickoff-meeting/

No	Title	Topics & connected WPs	URL
6.	Meet the SoFAIR project partners	Detailed project partners' introduction	https://sofair.org/meet-the-sofair-project-partners/
7.	SoFAIR Project at Open Repositories and CHIST-ERA Projects Seminar	External communication; conferences	https://sofair.org/sofair-project-at-open-repositories-and-chist-era-projects-seminar/
8.	Towards Reusability and Reproducibility of Research: A Scalable Workflow for Research Software Archival with CORE and Software Heritage webinar announcement	Webinar informations	https://sofair.org/towards-reusability-and-reproducibility-of-research-a-scalable-workflow-for-research-software-archival-with-core-and-software-heritage-webinar-announcement/
9.	The SoFAIR documentation: Managing the lifecycle of software assets - A workflow guide for developers	Project workflow presentation	https://sofair.org/the-sofair-documentation-managing-the-lifecycle-of-software-assets-a-workflow-guide-for-developers/

Figure 4. Summary of SoFAIR blog posts

2.2.6.   SOCIAL MEDIA

Our communication efforts on X (Twitter) focused on engaging the wider scientific community and sharing individual project achievements and milestones. Xs (Tweets) covered a range of topics, including updates on project progress and specific results, which helped to build interest and maintain engagement. In addition, the team actively participated in discussions surrounding open science, natural language processing and data science in research.

Type	Number
Number of followers	14
Number of following accounts	111
Number of tweets	51

Figure 5. Summary of SoFAIR X (Twitter) activity

2.2.7. WEBSITE

The SoFAIR project website was launched in March 2024 at sofair.org, providing comprehensive information about the project, its partners and recent results. The website includes a blog section where updates and news are regularly posted. The site has been developed with a consistent visual identity and colour scheme, ensuring alignment with the project's branding. It is based on WordPress, providing a flexible, maintainable structure.
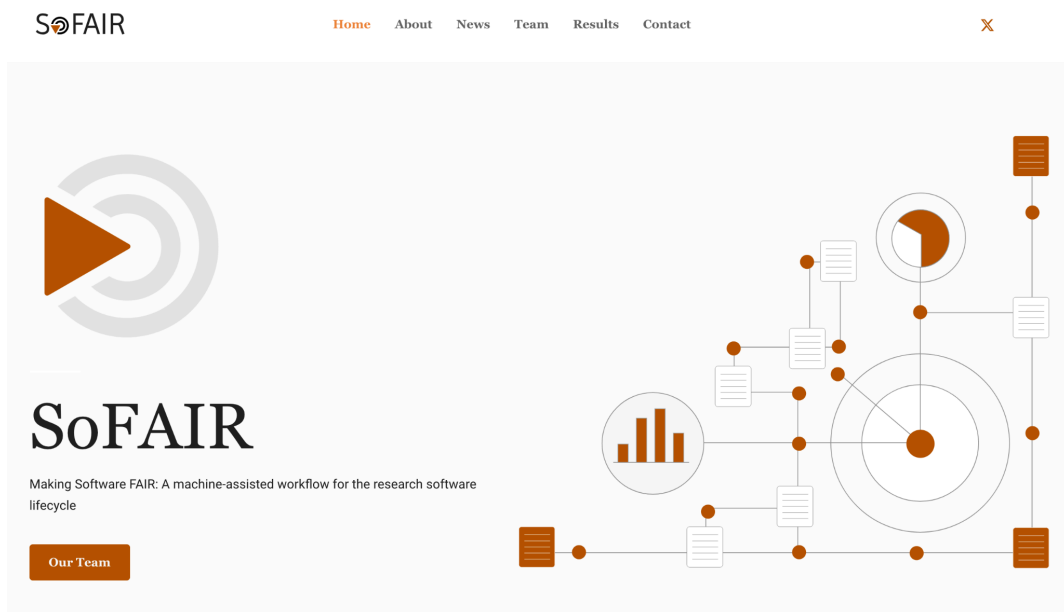

Figure 6. Project homepage

3. Key Performance Indicators

In accordance with the Communication and dissemination plan for 2024, key performance indicators were established with regard to activity on the blog, social media, and participation in conferences, publishing activity, as well as the organisation of a webinar. All targets were met, however, due to the Christmas period, it was decided to postpone the webinar until early 2025. Also, due to the early stage of the project, activity on the X (Twitter) site consisted mainly of retweets rather than the creation of original tweets.

Activity	#Year 1 plan	#Year 1 delivered
Number of blog posts	5	9
Publications submitted (preprints available to the public)	0	1
Webinars	1	1 in January 2025
Number of Xs (tweets)	50	51

Figure 7. KPI