SoFAIR

Final report on communication and dissemination activities with an exploitation plan

Submission Date	16.12.2025
Due Date	31.12.2025
Version	1
Deliverable Title	Final report on communication and dissemination activities with an exploitation plan
Deliverable No.	2.3
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 Final Report on Communication and Dissemination with an Exploitation Plan for the SoFAIR project outlines project activities in alignment with the Communication and Dissemination Plan, and supplements it with the exploitation planning. 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 consortium meetings, structured workshops on workflow integration, and tracking of conference participation. Additionally, the document describes the presentation of SoFAIR at external events, the development of key publications, and organisation of webinars. Finally, exploitation plan is defined, in alignment with periodic reporting.

Version and Revision History

Version	Date	Author/Reviewer/Contributors	Comments
0.1	17.12.2025	Patryk Hubar-Kołodziejczyk, Tomasz Umerle	First draft
0.2	23.12.2025	Petr Knoth, Laurent Romary	Reviews
1	31.12.2025	Patryk Hubar-Kołodziejczyk, Tomasz Umerle	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 final 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 SoFAIR project, structured according to the Key Performance Indicators planned in the Communication and Dissemination Plan (available at: https://sofair.org/results). The report is divided into three main sections: Internal Communication, External Communication and Exploitation Plan.

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 and management 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

As planned, the consortium held two in-person general meetings for all partners. 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. Additionally, in 2025 partners met in-person in specialised task forces to discuss project progress and work towards meeting project goals.

Two consortium meetings took place in 2024. 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.

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.

In 2025 partners met in-person in specialised task forces, to track progress of the works and ensure meeting project goals:

- Paris…
- (…)?
- 11.11.2025, Prague – project management and communication and dissemination meeting; attending: Petr Knoth, Tomasz Umerle, David Pride (remotely).

2.2. External communication

2.2.1. Conferences and seminars presentations

In the first project year, the SoFAIR Project was presented at 16 conferences and external seminars, reaching more than 1 500 people. These activities allowed for targeted communication aimed at different stakeholders, laid out in the Communication and Dissemination Plan.

2.2.2. 2024

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.

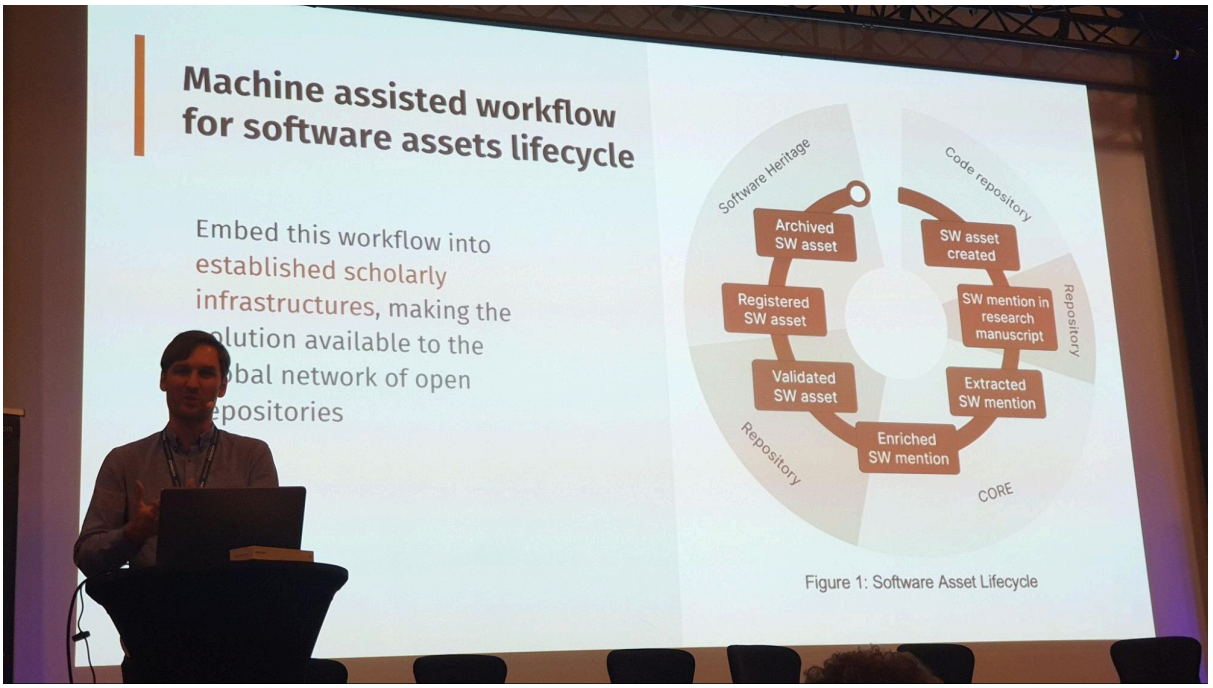

Photo 3: Presentation of SoFAIR at Open Repositories 2024

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**.

2.2.3. 2025

The SoFAIR project dedicated its final year, 2025, to ensuring that its results were presented directly to the communities responsible for their long-term adoption and integration into the global research infrastructure. SoFAIR maintained a significant presence at external events throughout the year. Key presentations included the 2025 Software Heritage Symposium and Summit (January 29, 2025), Open Repositories 2025 (April 4, 2025), and the CLARIN Annual Conference 2025 (October 2, 2025), the latter involving a large multi-partner team showcasing broad integration. Furthermore, SoFAIR engaged with the Digital Humanities community, presenting a poster at the ADHO 2025 (July 15–17, 2025) that highlighted the importance of research software preservation in safeguarding cultural and scientific memory. **Finally, specific focus was dedicated to networking and exploitation-focused activities, with the Schmidt Sciences Workshop.**

The presentation at 2025 Software Heritage Symposium and Summit was delivered by Petr Knoth and focused strategically on the preservation and archival of research software. Given that Software Heritage (INRIA) is one of the operational environments for the core software solution, this event was vital for affirming the project's role in the long-term archival of source code.

Photo 4: Presentation of SoFAIR at 2025 Software Heritage Symposium and summit

Open Repositories 2025 serves as a crucial platform for engaging librarians, developers, and repository managers, a core target audience for the SoFAIR workflow. SoFAIR-related presentations were delivered by a multi-person team, including Petr Knoth, David Pride, and Matteo Cancellieri from OU. This engagement focused on sharing SoFAIR's work regarding open knowledge, digital preservation, and upholding the FAIR principles (Findable, Accessible, Interoperable, Reusable) within repository environments.

Finally, the SoFAIR presentation at the CLARIN Annual Conference 2025 demonstrated the broad infrastructural potential of SoFAIR through significant multi-partner participation. The large team included representatives from PWr/CLARIN-PL, INRIA, OU, and IBL PAN. The collective presentation highlighted how SoFAIR integrates and supports the specific needs of infrastructure providers, such as those within the CLARIN community.

SoFAIR's engagement with the Digital Humanities community took the form of a poster session presented by a team from IBL PAN. The poster, titled "Preserving Cultural Heritage in the Digital Age: The Role of Software Heritage in Safeguarding Research Software," emphasized that research software is a critical, yet often invisible, component of digital humanities research.

Photo 4: Presentation of SoFAIR at ADHO 2025

2.2.4. Table of events

No	Dissemination item name	Presenters (partner)	Date	Dissemination channel	Content	Audience size
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 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
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.	SoFAIR presented at Open Science Days UGA (France)	Roberto Di Cosmo (INRIA)	5th April 2024	Conference Presentation	LINK	>60
7.	SoFAIR presented at Research Data Alliance 22nd Plenary Meeting (RDA VP22)	Morane Gruenpeter (INRIA)	15th May 2024	Conference Presentation (online session)	LINK	>75
8.	SoFAIR presented at 2025 Software Heritage Symposium and summit	Petr Knoth (OU)	29th January 2025	Conference Presentation	LINK	?
9.	SoFAIR presented at Sustaining and Innovating Open Research Infrastructures	Tomasz Umerle (IBL PAN), Cezary Rosiński	25th March 2025	Workshop presentation	LINK LINK	20
10.	SoFAIR presented at CHIST-ERA Projects Seminar 2025	Petr Knoth (OU)	28th April 2025	Conference Presentation	LINK	?
11.	SoFAIR presented at	Petr Knoth	4th April	Conference	LINK	?

No	Dissemination item name	Presenters (partner)	Date	Dissemination channel	Content	Audience size
	Open Repositories 2025	(OU), David Pride (OU), Matteo Cancellieri (OU)	2025	Presentations		
12.	SoFAIR presented at II Scientometrics Poznan Conference	Tomasz Umerle (IBL PAN), Cezary Rosiński (IBL PAN)	9th April 2025	Conference Presentation	LINK LINK	>100
13.	SoFAIR presented at the OpenCitations 2025	Samuel Scalbert (INRIA)	28th May 2025	Conference Poster Session	LINK	>100
14.	SoFAIR presented at the ADHO 2025	Tomasz Umerle, Cezary Rosiński, Nikodem Wołczuk, Patryk Hubar (IBL PAN)	15-17th July 2025	Conference Poster Session	LINK	>100
15.	SoFAIR presented at the Schmidt Sciences Workshop	?	?	Workshop presentation		?
16.	SoFAIR presented at the CLARIN Annual Conference 2025	Ewa Rudnicka, Marcin Oleksy, Tomasz Naskręt (PWr, CLARIN-PL), Luca Foppiano (INRIA), Petr Knoth (OU), David Pride (OU), Cezary Rosiński and Tomasz Umerle (IBL PAN)	2nd October, 2025	Conference Presentation	LINK LINK	>100

Figure 1. Conferences and seminar presentations

2.2.5. Webinars

During the project's lifetime, SoFAIR organised three webinars aimed at disseminating the project's goals, workflow, and results to a wide range of stakeholders, including research infrastructures, librarians, repository managers, software preservation initiatives, and the broader open science community.

Title	Content	Date	Registered	Attending
Towards Reusability and Reproducibility of Research: A Scalable Workflow for Research Software Archival with CORE and Software Heritage.	LINK	15.01.2025	85	45
SoFAIR and DARIAH ERIC Knowledge Exchange	LINK	19.11.2025	n/a	8
Integrating research software into open scholarly infrastructures: results of the SoFAIR project.	LINK	11.12.2025	81	52

Figure 2. Summary of SoFAIR webinar activities.

- **Webinar 1: "Towards Reusability and Reproducibility of Research: A Scalable Workflow for Research Software Archival with CORE and Software Heritage"**

Date: 15 January 2025
Registered: 85
Attending: 45
Duration: 1 hour
Format: Online

Event agenda:

- Overall project mission (Petr Knoth, SoFAIR PI, The Open University)
- Outline of the SoFAIR workflow (Morane Gruenpeter, Software Heritage)
  AI-based software extraction in CORE (David Pride, The 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

The webinar targeted key stakeholder groups, including research infrastructures, libraries, open repositories, and research software organisations. Dissemination activities were carried out through social media, mailing lists, and dedicated Slack channels. The session introduced the full SoFAIR workflow and offered participants insight into machine-assisted software extraction, validation processes, and long-term archival mechanisms. The webinar was recorded and made available to participants for later reference.

- **Webinar 2: "SoFAIR and DARIAH ERIC Knowledge Exchange"**

Date: 19 November 2025

Registered: invitation-based (DARIAH management and DARIAH services management)
Attending: 8, including Laure Barbot from DARIAH ERIC and Tomasz Parkoła, vice-chair of DARIAH ERIC's Joint Research Committee
Duration: 1,5 hour
Format: Online

The webinar targeted a key stakeholder for the SoFAIR digital humanities case study – DARIAH ERIC as a large European research infrastructure. It was aimed to align the results of SoFAIR research and development with DARIAH's infrastructure development plans. The meeting included presentations from the SoFAIR project team and discussion on potential implementation of solutions in the DARIAH's SSH Open Marketplace or/and software catalogue.

Webinar 3: "Integrating research software into open scholarly infrastructures: results of the SoFAIR project"

Date: 11 December 2025
Registered: 81
Attending: 52

Time: 16:00 CET
Duration: 1 hour 15 minutes
Format: Online

The third webinar acted as a capstone dissemination event, summarising the project's final results and showcasing the mature SoFAIR services and use cases developed during 2024–2025. The session focused on the integration of research software into existing

scholarly infrastructures and demonstrated the SoFAIR AI-assisted pipeline for automated software detection, validation, and archival.

The event was designed for researchers, repository managers, research software engineers, and open infrastructure providers, emphasising real-world applications of the SoFAIR workflow across CORE, HAL, and Software Heritage.

Event agenda:

- Overall project mission and overview of results (Petr Knoth, SoFAIR PI, The Open University)
- SoFAIR Prototype Demo (David Pride, The Open University)
- Overview of other SoFAIR use cases
  Panel discussion on research software in scholarly infrastructures (with invited guests)
- Q&A

2.2.6. Publications

1. **SoFAIR Project Dataset: An annotated dataset of software mentions in full text scientific articles.** Full Citation: Pride, David; Cancellieri, Matteo; Knoth, Petr; Rosinski, Cezary; Umerle, Tomasz; Rudnicka, Ewa; et al. (2025). SoFAIR Project Dataset - An annotated dataset of software mentions in full text scientific articles. The Open University. Dataset. Available at https://doi.org/10.21954/ou.rd.30374830.v2. Abstract: This is a multidisciplinary dataset of research papers, annotated with software mentions, built by the SoFAIR project in collaboration with the CLARIN-PL team. The SoFAIR project aims to support the reproducibility of research and the reuse of data and to recognise the contribution of software developers. The dataset was created through a workflow consisting of text selection, automatic pre-annotation of documents using the Softcite software-mentions tool and manual gold standard annotation. The dataset contains more than 9,000 software mentions, divided into 10 categories, as well as more than 2,000 relationships between mentions, which come from almost 500 texts belonging to 18 scientific disciplines. The dataset presented here contains detailed data and statistics that can be used to evaluate language models or to further

develop software extraction tools. The dataset is available in TEI XML format and has been made available under a CC-BY license.

2. **SoFAIR Dataset: A Multidisciplinary Dataset of Research Papers Annotated with Software Mentions**. Full citation: Rosiński, Cezary; Rudnicka, Ewa; Pride, David; Monteil, Alain; Oleksy, Marcin; Foppiano, Luca; Docekal, Martin; Cancellieri, Matteo; Scalbert, Samuel; Romary, Laurent; Umerle, Tomasz and Knoth, Petr (2025). SoFAIR Dataset: A Multidisciplinary Dataset of Research Papers Annotated with Software Mentions. Submitted to Nature Scientific Data, November 2025. Abstract: This publication describes the "SoFAIR Dataset: A Multidisciplinary Dataset of Research Papers Annotated with Software Mentions". Developed by the SoFAIR project in collaboration with the CLARIN-PL team, this multidisciplinary dataset contains research papers annotated specifically with software mentions. The project's goal in creating this resource is to support the reproducibility of research, facilitate data reuse, and ensure that the contributions of software developers are recognized. The dataset workflow involved selecting text and utilizing the Softcite tool for the automatic pre-annotation of documents. Authored by a large team including Rosiński, Rudnicka, and Knoth, the data paper was submitted to the journal "Nature Scientific Data" in November 2025.
3. **Identifying and Classifying Software Mentions in Full-Text Scholarly Documents**. Full Citation: Pride, David; Guenci, Matteo; Docekal, Martin; Peroni, Silvio and Knoth, Petr (2025). Identifying and Classifying Software Mentions in Full-Text Scholarly Documents. In: 25th ACM/IEEE Joint Conference on Digital Libraries (JCDL 2025), 15-19 December 2025. Available at: Abstract: The paper "Identifying and Classifying Software Mentions in Full-Text Scholarly Documents" evaluates several large language model (LLM)-based approaches using three gold-standard corpora. The study compares different prompting strategies and configurations against established baselines. The authors, including Pride and Knoth, provide the first systematic evaluation of LLMs for software mention extraction, analyze their strengths and weaknesses relative to prior techniques, and discuss the implications for reproducibility and open science. The results show that LLMs significantly enhance extraction accuracy and adaptability, thus advancing efforts to integrate software

into the scholarly record. This work was presented at the 25th ACM/IEEE Joint Conference on Digital Libraries (JCDL 2025).

4. **Interoperable verification and dissemination of software assets in repositories using COAR Notify**. Full Citation: Cancellieri, Matteo; Docekal, Martin; Pride, David; Gruenpeter, Morane; Douard, David and Knoth, Petr (2025). Interoperable verification and dissemination of software assets in repositories using COAR Notify. In: The 20th International Conference on Open Repositories (Pride, David; Cancellieri, Matteo and Knoth, Petr eds.), 18-21 Jun 2025, Chicago, Illinois, USA. Available at: https://oro.open.ac.uk/105529/. Abstract: The publication "Interoperable verification and dissemination of software assets in repositories using COAR Notify" addresses how the obscurity of open research software within academic manuscripts hinders its discoverability, attribution, and reusability. The SoFAIR project introduces a comprehensive machine learning-leveraging workflow to extract software mentions from papers, integrating repository systems, authors, and services such as HAL and Software Heritage, ensuring the software adheres to FAIR principles. To achieve interoperable communication across these diverse systems, the authors, including Cancellieri and Knoth, present the integration of the COAR Notify Protocol. This paper was presented at The 20th International Conference on Open Repositories in Chicago, Illinois, in June 2025.
5. **Making Software FAIR: A machine-assisted workflow for the research software lifecycle**. Full Citation: Knoth, Petr; Laurent, Romary; Lopez, Patrice; Di Cosmo, Roberto; Smrz, Pavel; Umerle, Tomasz; Harrison, Melissa; Monteil, Alain; Cancellieri, Matteo and Pride, David (2025). Making Software FAIR: A machine-assisted workflow for the research software lifecycle. In: 19th International Conference on Open Repositories (OR2024), 3-6 Jun 2024, Göteborg, Sweden. Available at: https://oro.open.ac.uk/102429/. Abstract: The paper "Making Software FAIR: A machine-assisted workflow for the research software lifecycle" highlights the central issue that the existence of open research software is often concealed within the manuscript of research papers, which impedes its discoverability, attribution, and reusability. For these software assets to become "first-class bibliographic records," they must first be identified and subsequently registered with persistent identifiers

(PIDs) to be made FAIR (Findable, Accessible, Interoperable and Reusable). The publication describes the machine-assisted workflow developed by the SoFAIR project specifically designed to address this challenge, as much open research software currently fails to meet FAIR principles and is not explicitly linked from its introductory manuscripts. This paper, authored by a team including Knoth, Romary, and Umerle, was presented at the 19th International Conference on Open Repositories (OR2024) in Göteborg, Sweden, in June 2024.

6. **Automated detection of software mentions as an instrument for research and infrastructure in the humanities**. Full Citation:; Umerle, Tomasz; Rosiński, Cezary; Hubar, Patryk; Jarzyńska, Katarzyna; Rokicka, Renata (2025). Submitted to "Poznanskie Studia Slawistyczne", December 2025. Abstract: The study examines digital transformation (DT) in the humanities by employing the Softcite tool for the automated detection of software mentions in scholarly publications. Analyzing journals in Traditional Linguistics and Literary studies (TLL) and Digital Humanities (DH) reveals a clear upward trend in software use across both domains. DH journals show substantially higher saturation, frequently reaching 30% to almost 60%, while TLL journals generally remain below 10%. This research validates the potential of automated software detection to support Social Sciences and Humanities (SSH) research infrastructures, though human validation is necessary due to notable false-positive rates and the finding that abstracts rarely contain software references. This effort is central to the SoFAIR project, which is developing a machine-assisted workflow to make research software FAIR (Findable, Accessible, Interoperable, Reusable). The SoFAIR workflow integrates tools like Softcite with platforms such as CORE, HAL, and Software Heritage to streamline the extraction of software mentions, facilitate validation, and ensure long-term archival with persistent identifiers (PIDs). The consortium has actively disseminated its results and documentation through nine blog posts and presentations at seven conferences in the first year, engaging diverse stakeholders in the open science community.

2.2.7. 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.

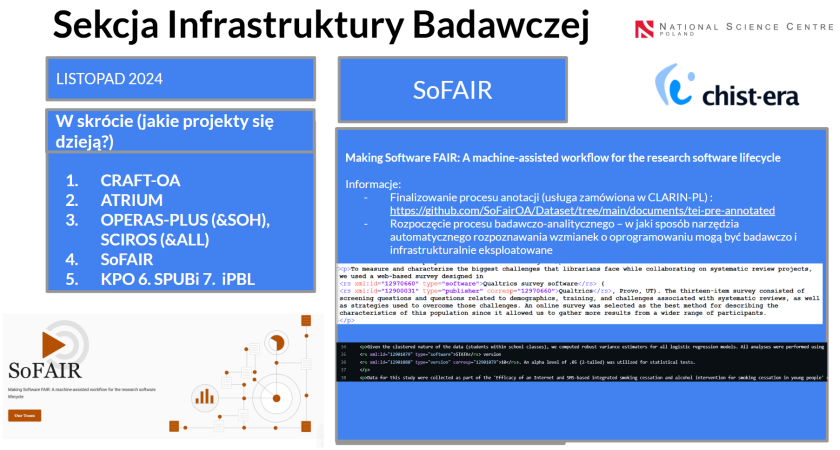

Figure 3. Screenshot of presentation for Digital Humanities Centre at IBL PAN.

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.

2.2.8. Blog Posts

During this reporting period, 17 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 project results, while posts on partners channels, especially CORE and OU, showcased participation participation at important conferences and seminars, allowing to reach wide audiences, already built up the project leader organisation.

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/
6.	Meet the SoFAIR project partners	Detailed project partners' introduction	https://sofair.org/meet-the-sofair-project-partners/

No	Title	Topics & connected WPs	URL
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/
10.	SoFAIR Dataset in the Making:Challenges of Human Annotation of Software Mentions	Dataset description	https://sofair.org/sofair-dataset-in-the-makingchallenges-of-human-annotation-of-software-mentions/
11.	Boost Research Reproducibility – the SoFAIR Webinar Recording!	Webinar informations	https://sofair.org/boost-research-reproducibility-the-sofair-webinar-recording/

No	Title	Topics & connected WPs	URL
12.	Making Research Software FAIR: How SoFAIR Tackles the Reproducibility Challenge	Topical discussion	https://sofair.org/making-research-software-fair-how-sofair-tackles-the-reproducibility-challenge/
13.	CORE at the CHIST-ERA Projects Seminar 2025: Turning Research Software into Reusable Knowledge	Description of SoFAIR conference presentations	https://www.linkedin.com/pulse/core-chist-era-projects-seminar-2025-turning-research-software-3ozzf/?trackingId=dDmuW%2FW2kWIgIWoDF3RB2Q%3D%3D
14.	CORE at Open Repositories 2025: Unlocking Insights and Empowering Open Access	Description of SoFAIR conference presentations	https://www.linkedin.com/pulse/core-open-repositories-2025-unlocking-insights-empowering-access-ckcff/?trackingId=RFoKvDkeSUbp0SL8XWAvcw%3D%3D
15.	SoFAIR study paper accepted to JCDL2025	Project publication discussion	https://blog.core.ac.uk/2025/11/05/sofair-study-paper-accepted-to-jcdl2025/
16.	CORE presents SoFAIR project at UNESCO	Description of SoFAIR conference presentations	https://blog.core.ac.uk/2025/02/03/core-presents-sofair-project-at-unesco/

No	Title	Topics & connected WPs	URL
17.	SoFAIR final webinar	SoFAIR results presentation	https://blog.core.ac.uk/2025/12/17/sofair-final-webinar/
16.	USE CASE1		
17.	USE CASE2		
18.	USE CASE3		
19.	SoFAIR 2025		

Figure 4. Summary of SoFAIR blog posts

2.2.9. 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	22
Number of following accounts	122
Number of tweets	100

Figure 5. Summary of SoFAIR X (Twitter) activity

2.2.10. 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.

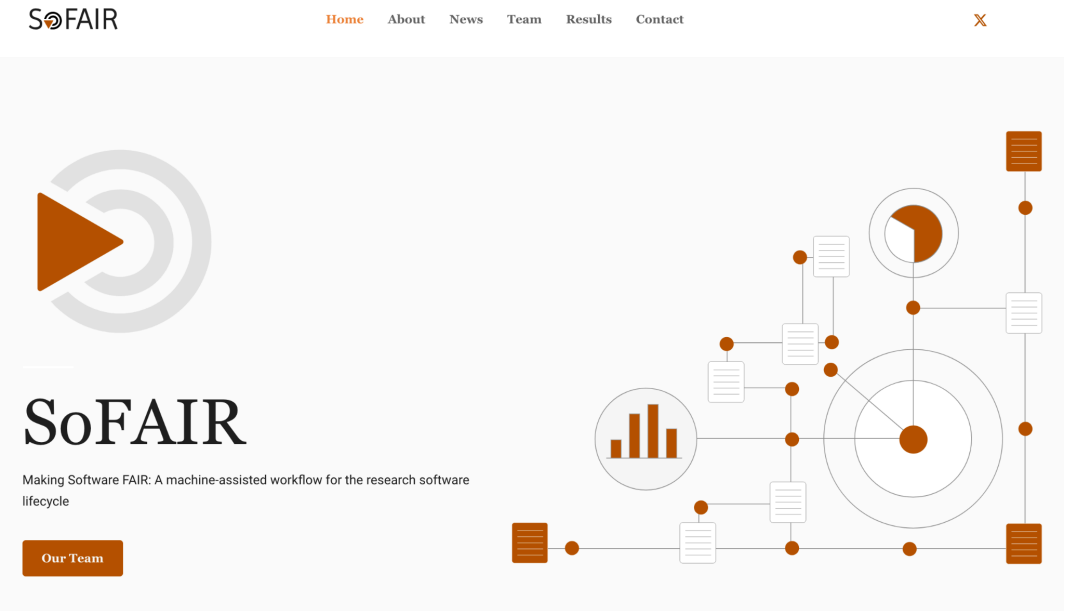

Figure 6. Project homepage

3. C&D 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	#planned at the end	# delivered
Number of blog posts	10	18
Publications submitted (preprints available to the public)	2	6 (4)

Webinars	3	3
Number of Xs (tweets)	100	100

Figure 7. KPI

4. Exploitation plan

SoFAIR Key Exploitable Results (KERs)

SoFAIR has identified 4 KERs:

KER1: SoFAIR Repository Connector & SoFAIR Notification Module

KER2: SoFAIR dataset

KER3: SoFAIR workflow documentation

KER4: SoFAIR ML Models

Table below presents the overview of these KERs, defining the type of results, exploiting entity, target end user and any existing IP protections. We define **entities** as organizations, institutions, or businesses that actively adopt, commercialize, or integrate the project's results into their own operations, offerings, or strategies. They are the intermediaries or stakeholders who take the results to scale, implement, or transform them into products, services, or solutions. **End users** are the individuals, groups, or organizations who directly benefit from the project's results as consumers, beneficiaries, or recipients of the developed products, services, or innovations. They are the ultimate targets or intended audience for the exploitation of the results.

Key Exploitable Result (KER)	**Type**	**Exploitation entity**	**Target end user**	**IP protections**
KER1: SoFAIR Repository Connector & SoFAIR Notification Module	software solution (Open Source)	Project partners: CORE (OU) and HAL (INRIA) + any aggregators and repositories intending take up the solution in CORE environment or elsewhere	Researchers; Research software engineers	No IP protections have been applied

				for these KERs.
KER2: SoFAIR workflow documentation	Free resource	Project partners + any services exploiting KER1	Research software engineers; RI managers	
KER3: SoFAIR ML Models	Free resource (Open Source)	Research software engineers; Individual researchers	Researchers; ML/software engineers	
KER4: SoFAIR dataset	Free resource	Research software engineers; Individual researchers	Researchers; ML/software engineers	

Roadmap to Exploitation (goals and timeframe for managing KERs)

KER1: The roadmap to exploitation for KER1 **during** the project is through deployment of SoFAIR Repository Connector & SoFAIR Notification Module in the CORE environment, with HAL open repository and Software Heritage acting as operational environments. Other CORE repositories will be able to exploit this solution, leading to widespread impact of SoFAIR; CORE is the largest aggregator of open scientific publications in the world, supporting 12,000 repositories. This stage will be achieved at the end of Year 2 of project's implementation. **After** the project completion the Project Consortium will aim to scale up implementation of this KER, aiming for two key **goals**:

1) Goal no 1: wider adoption in the CORE community – communication and dissemination by Project Consortium, mainly the OU, towards the CORE repository community. (Project end + Year 1-2)

2) Goal no 2: exploring additional deployments of the software for wider exploitations, such as a plugin/feature for DSpace software, which is the largest repository software provider in the world. (Project end + Year 1-3).

KER2 accompanies KER1 as a technical overview and documentation needed to deploy KER1 on any level: by aggregators (like CORE) or repositories (HAL), and is dedicated towards expert staff: repository or aggregator managers, technical staff , research software creators and engineer in general. As such, KER2's roadmap is dependent on KER1's roadmap. **During** the project the workflow will be communicated to the community for any feedback and tested against initial deployments in the HAL repository, which will be achieved by the end of Year 2 of project's implementation. **After** the project completion, as KER1 is more widely exploited, KER2 is expected to become a more widely used piece of documentation, hence to be subject to more scrutiny, but also more

1) Goal no 1: KER2 becomes a community-driven documentation shaped by organisations implementing KER1. (Project end + Year 1-3)

KERs 3 and 4 are different in nature, as they can be exploited both as components of software solutions, but also as vehicles to drive knowledge production by any and all researchers aiming to solve the challenge of making software mentions more FAIR. KER3 could also become crucial for potential spin-off effects of the project, as they can feed into new solutions, applicable in other areas, discussed in section 2.3.

KER3: although an indispensable element of the SoFAIR software solution (KER1), will be made available as a stand-alone result as an Open Source ML model for software mentions identification, allowing any entities to leverage its further implementation, including reuse for business, commercial or other endeavours. **During the project** this KER will be published in the SoFAIR Github repository. After project's completion SoFAIR consortium will aim to:

1) Goal no 1: analyse commercialisation potential for this KER3 with relevant innovation hubs/lab and/or technology transfer centers, especially around the topic of AI Open Source software reuse in software identification and tracking (see point 2.3); (Project end + Year 2-3)

KER4 has already been made available for future reuse **during** Year 1 and will be a subject of intensive communication and dissemination efforts during Year 2, as the SoFAIR Dataset is the largest and most reliable resource in the area of software mentions identification. After the project's completion, the Consortium will continue to leverage its potential:

1) Goal no 1: communicate and disseminate the Dataset as a valuable research asset to enhance research in the area (Project end + Year 1-2)
2) Goal no 2: communicate and disseminate annotation workflow, impact best practices in annotation guidelines for software mentions (Project end + Year 1-2)
3) Goal no 3: seek further funding to grow the Dataset and encourage new use cases (Project end + Year 2-3)

Exploitation overview

Period	Planned goals	Actual exploited results
Year 1	KER1: prototype created KER2: created and disseminated KER3: a variety of zero-shot methods tested KER4: dataset created	KER1: CORE and HAL exploit the results in controlled environments KER2: partners leveraging documentation for prototype development; relevant stakeholders informed KER3: first evaluation results obtained KER4: dataset published
Year 2	KER1: deployment in operational environment of CORE, HAL and Software Heritage KER2: workflow used and disseminated	KER1: prototypes of envisaged SoFAIR modules demonstrated in relevant production environments of CORE, HAL and Software heritage (TRL6)

	KER3: submission of a research paper documenting the experiments KER4: dataset disseminated and made publicly available	KER2: workflows used in use cases/demonstrators with workflow documentation used as an implementation specification. Workflow widely disseminated at events and conferences. KER3: research paper published (JCDL) with research data in repository KER4: dataset published in data repository, data paper submitted
Project end + 1 year	KER1: Goal no 1; Goal no 2 KER2: Goal no 1 KER3: n/a KER4: Goal no 1; Goal no 2	n/a
Project end + 2 year	KER1: Goal no 1 & Goal no 2 continued; Goal no 3 KER2: Goal no 1 continued KER3: Goal no 1 KER4: Goal no 1 and Goal no 2 continued; Goal no 3	n/a
Project end + 3 year	KER1: Goal no 2, 3 continued KER2: Goal no 1 continued KER3: Goal no 1 continued KER4: Goal no 3 continued	n/a

SoFAIR Project: Potential for Spin-Off Effects of Key Exploitable Results (KERs)

The SoFAIR project has the potential to generate significant spin-off effects that extend beyond the initial research objectives. These spin-off effects can be analysed, especially in the context of the project's **KERs 1 and 3**.

KER1

- This **open-source software** solution has potential applications beyond the CORE environment. In addition to **integration with repositories and aggregators**, it could be adapted as a **plugin for popular repository platforms, such as DSpace or EPrints**, significantly increasing its reach.
- Furthermore, there is potential for **building commercial services** around this solution, such as offering **technical and integration support** for repositories. This is particularly important given the growing demand for tools that facilitate **the management and sharing of research software** in accordance with FAIR principles.
- Potential users of this solution include not only **research institutions and repositories** but also **commercial companies,** such as Digital Science, Elsevier or Clarivate, that provide **services and software for the scientific community** and other open scholarly infrastures, including e.g. OpenAlex and DataSeer.

KER3

5. The **machine learning (ML) models** for identifying software mentions, released as **open-source**, can be applied in various contexts, including both academic research and commercial sectors.
6. In addition to **supporting the identification of software mentions**, these models can be **customised and developed** for other applications, such as **sentiment analysis, information retrieval, and technology trend monitoring**.
7. **Commercialisation opportunities** for these models include **licensing, integration with SaaS platforms**, and offering **consulting services and model customisation** for specific needs.
8. **Potential users** include **researchers, ML engineers, and tech companies** seeking solutions for **natural language processing (NLP) and text analysis**.

Potential Future Directions and Actions

- **Commercialisation of Solutions**: Assessing the commercialisation potential of ML models (KER3) in collaboration with **innovation hubs and technology transfer centres**.
- **Development of New Products and Services**: Leveraging ML models and data (KER3 and KER4) to **create new products and services** in areas such as **market analysis, competitive intelligence, document analysis, and knowledge management**.
- **Building Partnerships**: Establishing collaborations with other projects, companies, and institutions to **jointly develop and commercialise** project results.
- **Platform Development**: Creating a platform based on project results that will be available to the community, enabling **collaborative development and implementation** of the solutions.
- **Securing Further Funding**: Applying for **additional funding** to expand and refine the data set (KER4) and pursue further commercialization efforts.

Conclusion

The SoFAIR project has **broad potential for generating spin-off effects** that could **significantly impact how research software is identified, managed, and shared**. The project's key exploitable results provide a strong foundation for **further development, commercialization, and community-building** around the concept of **open and FAIR research software**.