Enhancing Trust, Integrity, and Efficiency in Research through Next-Level Reproducibility Impact Pathways

Deliverable D4.2 – Pilot implementation and assessment plans

28 June 2024

Lead Beneficiary: GESIS

Author/s: Fakhri Momeni, Eleni Adamidi, Sven Arend Ulpts, Hajira Jabeen, Taimoor Khan, Thomas Klebel, Eva Kormann, Barbara Leitner, Allyson Lister, Elli Papadopoulou, Haris Papageorgiou, Tony Ross-Hellauer, Susanna-Assunta Sansone, Jesper W. Schneider, Joeri Tijdink, Thanasis Vergoulis

Reviewer/s: Nikol Yovcheva, Thomas Klebel

Funded by the European Union

Prepared under contract from the European Commission
Grant agreement No. 101094817

EU Horizon Europe Research and Innovation action

Project acronym:	**TIER2**
Project full title:	**Enhancing Trust, Integrity, and Efficiency in Research through Next-Level Reproducibility Impact Pathways**
Start of the project:	January 2023
Duration:	36 months
Project coordinator:	Dr. Tony Ross-Hellauer

Deliverable title:	Pilot implementation and assessment plans
Deliverable n°:	D4.2
Version n°:	1.3
Nature of the deliverable:	Report
Dissemination level:	Public

WP responsible:	WP4
Lead beneficiary:	GESIS

TIER2 Project, Grant agreement No. 101094817

Due date of deliverable:	Month 18
Actual submission date:	Month 18

Deliverable status:

Version	Status	Date	Author(s)
1.0	Draft/Final	29 05 2024	Fakhri Momeni (GESIS)
1.1	Review	04 06 2024	Nikol Yovcheva (PENSOFT)
1.2	Review	24 06 2024	Thomas Klebel (KNOW)
1.3	Final version	25 06 2024	Fakhri Momeni (GESIS)

Table of Contents

List of Abbreviations

CWL – Common Workflow Language
DMP – Data Management Plan
EU – European Union
FAIR – Findable, Accessible, Interoperable and Reusable
KIPs – Key Impact Pathways
RMP – Reproducibility Management Plan
RPOs – Research Performing Organisations
SMP – Software Management Plan
UX – User Experience
WP – Work Package

1. Executive Summary

The TIER2 project seeks to understand and address the challenges of research reproducibility across diverse contexts, focusing on social, life, and computer sciences, as well as research publishers and funders. The project aims to increase awareness, build capacity, and propose innovative solutions tailored to various research cultures. Central to TIER2's strategy are eight Pilot activities designed to develop, implement, and evaluate new reproducibility-related tools and practices. These Pilots emphasize stakeholder engagement and collaboration throughout the project's duration until December 2025.

Task 4.2 (T4.2) is a key component of TIER2, responsible for overseeing the procurement of user requirements and the co-design of interventions aimed at enhancing reproducibility across different research methodologies. T4.2 builds on the priority areas identified in Task 4.1 (T4.1) and leverages insights from previous scoping work. It collaborates with Pilot community members to specify new interventions that span all phases of the research lifecycle, from ideation to assessment.

This deliverable outlines the progress made in T4.2, highlighting our commitment to rigorous planning, stakeholder engagement, and coordination across work packages. Key activities include initial planning and coordination, tool evaluation, bilateral discussions, Pilot design and template creation, document refinement, timeline development, and the organization of two Pilot workshops. Each step has been meticulously planned to ensure the success of our initiatives.

The outcomes of T4.2 will guide the development efforts in subsequent work packages, ensuring that the interventions are well-suited to the unique needs of diverse research communities. Future steps involve implementing and assessing the refined Pilot plans based on community feedback, with continued collaboration and stakeholder involvement being essential to achieving the broader goals of the TIER2 project.

2. Introduction

The TIER2 project aims to better understand the causes, consequences, and possible solutions to the perceived poor levels of research reproducibility across various contexts. With a focus on social, life, and computer sciences, as well as research publishers and funders, the project seeks to increase awareness, build capacity, and propose innovative solutions sensitive to diverse research cultures.

Central to this approach are eight Pilot activities designed to develop, implement, and evaluate new reproducibility-related tools and practices. These activities emphasize stakeholder engagement and collaboration throughout the project's duration.

The objective of T4.2 is to oversee procurement of user-requirements and co-design for these Pilots, to increase research reproducibility across different methodologies and epistemic contexts. Building upon the priority areas identified in T4.1 and leveraging insights from previous scoping work, T4.2 collaborates with Pilot community members to specify new interventions that span all phases of the research lifecycle, from ideation to assessment. This involves creating detailed action plans, mapping technical and social requirements, and considering both new contexts for existing tools and the development of tools in existing contexts. The task's ultimate goal is to ensure that these interventions are well-suited to the unique needs and particularities of diverse research communities, thereby steering the development efforts in subsequent work packages.

This section outlines the key activities undertaken to date, highlighting our commitment to rigorous planning, stakeholder engagement, and coordination across work packages. From initial planning and coordination to the refinement of Pilot documents and the organization of Pilot workshops, each step has been carefully orchestrated to ensure the success of our endeavours. Let's delve deeper into the specifics of each activity:

Initial Planning and Coordination: We commenced T4.2 by establishing clear goals and aligning them with the priorities outlined in T4.1. Our focus was on fostering effective channels for gathering feedback from Pilot communities to guide our development process accurately.

Pilot Design and Template Creation: Recognizing the need for a structured approach, we initiated discussions on creating a Pilot template. This template serves as a blueprint for outlining Pilot aims, activities, stakeholders, assessment and development strategies.

Document Refinement and Coordination: Continuous refinement and updates of documents and templates was undertaken to ensure clarity, conciseness, and alignment with overarching objectives of different Pilots. Coordination efforts were taken to ensure coherence across work packages.

Timeline Development and Governance: We placed emphasis on delineating a clear timeline for Pilot preregistration and implementation, incorporating milestones and governance mechanisms to ensure success. Collaboration with other ongoing work, particularly Task 4.3, was crucial in this endeavour.

Pilot Workshop 1: We organized a Pilot meeting to map synergies between Pilots and discussed the timeline for Pilot preregistration and implementation. The workshop was conducted on September 21st, 2023, based on the timeline presented in **Table 1** where all Pilot leads participated, presented their Pilots and engaged in the discussion afterwards. The workshop was concluded with interesting common areas to collaborate on, open questions to discuss in coming planning meetings.

Table 1: *Workshop 1 was conducted on September 21st where all Pilot leads participated and presented their Pilot plan.*

Pilot #	Pilot Title	Presenter(s)
1	Decision Aid	Sven Arend Ulpts (Jesper Wiborg Schneider)
2	Reproducibility Management Plan (RMP)	Elli Papadopoulou
3	Reproducible Workflows	Thanasis Vergoulis, Eleni Adamidi
4	Reproducibility Checklists for Computational Social Science Research	Taimoor Khan, Hajira Jabeen
5	Reproducibility Promotion Plans for Funders	Barbara Leitner (Joeri Tijdink)
6	Reproducibility Monitoring Dashboard	Haris Papageorgiou
7	Editorial Workflows to Increase Data Sharing	Thomas Klebel, Tony Ross-Hellauer
8	An Editorial Reference Handbook for Reproducibility and FAIRness	Tony Ross-Hellauer

Pilot Workshop 2 (Table 2): We organized the second Pilot workshop, where the Pilots' progress was presented in three sessions. Pilot leaders were encouraged to include timeline expectations in their presentations. **Table 2** shows the timeline for each presentation.

Table 2: *Workshop 2 Timeline. The second workshop on Pilot progress discussion is structured to be across 4 monthly meetings. Two-three Pilots are presented each month, having 10min presentations followed up with 10mins discussion.*

Pilot #	Pilot Title	Presenter(s)	Date & Time
1	Decision Aid	Sven Arend Ulpts Jesper Wiborg Schneider	15th Nov. 23 12:00 – 12:20pm
2	Reproducibility Management Plan (RMP)	Elli Papadopoulou	15th Nov. 23 12:20 – 12:40pm
3	Reproducible Workflows	Thanasis Vergoulis Eleni Adamidi	20th Dec. 23 12:00 – 12:20pm
4	Reproducibility Checklists for Computational Social Science Research	Taimoor Khan Hajira Jabeen	20th Dec. 23 12:20 – 12:40pm
5	Reproducibility Promotion Plans for Funders	Barbara Leitner Joeri Tijdink	20th Dec. 23 10:40 – 11:00pm
6	Reproducibility Monitoring Dashboard	Haris Papageorgiou	17th Jan. 24 10:00 – 10:20am
7	Editorial Workflows to Increase Data Sharing	Thomas Klebel, Tony Ross-Hellauer	17th Jan. 24 10:20 – 10:40am

8	An Editorial Reference Handbook for Reproducibility and FAIRness	Tony Ross-Hellauer/ Susanna Sansone	17th Jan. 24 10:40 – 11:00am

Meeting Structure and Content: A regular meeting slot has been established to address organisational matters related to Pilot preparation and implementation.

Pilot Review: Pilots' plans have been reviewed by different community members to assess relevance, completeness, and feasibility. This ensures that each Pilot is aligned with its goals and can achieve its intended outcomes within the designated timeline.
Each Pilot plan has been reviewed by two reviewers from the project to assess the maturity of the Pilot from three aspects:

1. Relevance
    a. The Pilot's objectives and scope are aligned with its goal and to the objectives outlined in the TIER2 grant agreement.
2. Completeness
    a. A brief literature review that balances breadth of the domain and depth through recent approaches.
    b. The evaluation plan can assess the outcome in clearly measurable metrics.
3. Feasibility / Achievability
    a. The scale of the Pilot is appropriate to produce the intended outcomes in the given timeline.
    b. The proposed method has flexibility to adapt to change in circumstances while still focusing on the outcomes.

After reviewing, the Pilots' leaders were asked to revise their plan based on the reviewers' feedback and prepare a document to describe how they addressed each comment. The revised version of the Pilots' plan is available in the next section.

Advisory Board Meeting: In addition to the plan review, during the annual meeting with the advisory board on April 18, 2024, the pilots were presented and received substantial input and feedback.

2.1.Overview of Pilots

1. This section provides a summary of the Pilots in TIER2. **Pilot 1 - Decision Aid:** TIER2's Decision Aid will provide clarity on the meaning, relevance, and feasibility of 'reproducibility' for researchers to aid them in identifying what type of reproducibility is relevant for their research and indicate what they must consider regarding how feasible such 'reproducibility' would be for them. The tool will be piloted with two researcher groups (qualitative and machine learning researchers).
   Stakeholders: Researchers, publishers, funders

2. **Pilot 2 - Reproducibility Management Plan (RMP):** The Reproducibility Management Plan (RMP) Pilot aims to create a prototype of key thematic subjects and questions that will serve as the starting point to support reproducibility at the planning stage of research. Work involves defining what an RMP is, integrating it into the ARGOS service, and testing its effectiveness with feedback from the community. The Pilot addresses researchers, beneficiaries and funders for its adoption.
   Stakeholders: Researchers, research communities, funders, and service providers

3. **Pilot 3 - Reproducible Workflows:** The Reproducible workflows Pilot focuses on enhancing reproducibility in life sciences and computer sciences, by adapting the SCHEMA open-source platform, using technologies like software containerisation, workflow description languages, and experiment packaging specifications to fit specific epistemic needs.
   Stakeholders: Life scientists, computer scientists

4. **Pilot 4 - Reproducibility Checklists for Computational Social Science Research*:*** In this Pilot, we aim to provide a structure of well-defined checklists and templates that can help review data and code reproducibility for computational social scientists. The checklists and review templates cater for the specific needs of the three research phases, i.e., planning and data collection, process and analysis and finally sharing and archiving the research resources. It results in building trust and authority in the social science research community.
   Stakeholders: Computational Social Scientists (Research Producers and consumers)

5. **Pilot 5 - Reproducibility Promotion Plans for Funders:** The Pilot will develop a policy template with recommendations for funders to foster reproducible practices both in the research they fund (evaluation and monitoring) and their internal practices.
   Stakeholders: Funders

6. **Pilot 6 - Reproducibility Monitoring Dashboard:** The Reproducibility Monitoring Dashboard Pilot aims to develop tools that enable funding agencies to track and monitor the reusability of research artifacts across various projects, programs, topics, and disciplines. This auto-generated dashboard assesses the impacts of policies related to data and code sharing. Furthermore, we are establishing essential requirements to make the dashboard user-friendly for publishers.

Stakeholders: Research Performing Organisations (RPOs), Funders, Publishers and Researchers

7. **Pilot 7 - Editorial Workflows to Increase Data Sharing:** This Pilot is aimed at increasing data sharing in published work. Data sharing is an important building block for increased reproducibility & transparency, but current rates of sharing are low.
   Stakeholders: Publishers

8. **Pilot 8 - An Editorial Reference Handbook for Reproducibility and FAIRness:** This Pilot will co-create and test an Editorial Reference Handbook that contributes towards a common understanding of what is required to assist reproducibility and FAIRness. The Handbook, identified as a priority in a workshop with publishers, will include two components. A structured section will include educational and practical set of checks, defined by reviewing existing material, harmonising and operationalising them. Some journals have internal checks, but the type, richness and stringency vary, and there is little/no consensus among publishers. A narrative component with a general framework will help improve internal processes, defined by describing an ideal process where checks should be applied. There are a variety of internal processes, and how, when and by whom these checks are done vary, and this can also affect the results.
   The Pilot includes representatives of Cambridge University Press, Cell Press, EMBO Press, F1000 (Taylor & Francis), GigaScience Press, Lancet, Oxford University Press, PLOS, Springer Nature, Wiley.
   Stakeholders: Publishers

3. Pilots Plans

In this section we present the final version of the Pilot plans.

3.1.Pilot 1 - Decision Aid

	Pilot	ADDITIONAL DETAILS
Pilot Number & Title (Institution / Focal Person)	Pilot 1 – Decision Aid CFA/Aarhus / Jesper W. Schneider & Sven Ulpts	
Short description	Based on the framework developed in T3.1, we have developed a prototype tool which is intended to inform, support and aid stakeholders in deciding whether specific types of 'reproducibility' are relevant for different kinds of research, and if so, to what extent 'reproducibility' 'would be practically feasible considering the unique research situation. The prototype is constructed as a survey including a decision tree, relational survey and a simple scale gradings to specific questions relating to feasibility. The aid will provide final simple gradings that will aid stakeholders make decisions. The prototype is developed and the aim with the so-called Pilot is to examine if and to what extent such an aid is pertinent, useful and feasible. This is a highly exploratory process where we will rely on feedback from stakeholders. The limited resources available for further developments mean that the Pilot should provide overly positives responses. In other words, this is a high-risk piloting. If sufficiently successful, the prototype will be adjusted to users' feedback, finalized and made public available for all to use. If unsuccessful, we will write the idea up and provide an information sheet with a checklist that may still inform stakeholders about the relevance and feasibility of 'reproducibility'. The Decision Aid is planned to be an online tool.	
Objectives	1. To explore (not test) to what extent the Decision Aid is pertinent, useful and feasible. At this stage we are not interested in effectiveness. 2. The ultimate goal is to acknowledge epistemic diversity. The idea of the Decision Aid should be deemed both pertinent and useful by the targeted group of users. It should support 'reproducibility' where it is relevant and indicate where it is most likely irrelevant and/or less feasible. Risk 1. There is a high degree of risk that the Decision Aid will fail. The main challenge is the compromise needed between a simplistic tool and the demands of background knowledge from its users.	

<table>
<tr><td>Current literature/state of play</td><td>We do not know of any such conceptual tool available at the moment for funders, publishers, and researchers. However, there are some mostly quantitative and statistics-based suggestions for replication target selection with small qualitative components that were formulated in the context of psychological science or related fields [1], [3], [4], [6], [8]. Importantly, to our knowledge no tool or systematic guidance for the assessment of the appropriateness of ‘reproducibility’ in the realm of epistemic diversity currently exists. There is only theoretical and conceptual literature on the topic, see for instance, [2], [5], [7].</td></tr>
<tr><td>Existing tools related to the Pilot</td><td>For replication target selection see e.g.:
https://www.sciencedirect.com/science/article/pii/S0010945223002691
https://link.springer.com/article/10.1007/s40279-022-01749-1
https://online.ucpress.edu/collabra/article/5/1/46/113017/When-and-Why-to-Replicate-As-Easy-as-1-2-3
https://royalsocietypublishing.org/doi/10.1098/rsos.210586
https://osf.io/preprints/metaarxiv/knjea/
https://psycnet.apa.org/doiLanding?doi=10.1037%2Fmet0000438</td></tr>
<tr><td>Overview of existing tools/resources (software/platforms/instruments)</td><td>There are none.</td></tr>
<tr><td>Methods used for piloting (methods used for tools or practices)</td><td>We currently have a prototype tool which we need to advance on.

The prototype will be scrutinized in some cognitive testing with relevant stakeholders (e.g. researchers using qualitative methods and machine learning)

Next, we will explore and discuss the tool with focus groups. Here we may terminate the process if feedback is overly critical.

Finally, if sufficiently successful, we will adjust the tool and explore its usefulness further in an online format with stakeholders.</td></tr>
<tr><td>Stakeholder groups affected and/or included</td><td><table>
<tr><th>Stakeholder Group (role: affected/included/beneficiaries)</th><th>details (expected outcome for the stakeholder)</th></tr>
<tr><td>Researchers (included/ beneficiary)</td><td>Researchers should be enabled to assess what type of reproducibility is relevant for them and how feasible it is. Furthermore, a short one-page information sheet can be used during peer review to assess relevance and feasibility of different types of reproducibility in different situations.</td></tr>
</table></td></tr>
</table>

	Publishers & funders (included/ affected)	A policy brief will support publishers & funders in installing reproducibility policies and guidelines that are in accordance with the appropriateness of different types of reproducibility for diverse kinds of research and research situations.
Stakeholder engagement plan	This is still an open question. The idea of the tool grew out of the conceptual work in T3.1 and was not initially planned to be part of the project. It is therefore developed ad hoc, and on the fly while we were at the same time were developing the framework. We are therefore still early in the planning phase of how we will engage with our stakeholders and whom they might be.	
Detailed timeline until Pilot end	The exploration of the Decision Aid will be an ongoing process that most likely will stretch over the remaining period of the project unless we decide to terminate it beforehand. Three important milestones are currently planned. 1) Finalizing the prototype (done by ultimo June); 2) cognitive explorations of the Decision Aid (done by medio October); 3) Focus group discussions (done by ultimo December); and 4) final explorations of the tool with stakeholders during the Spring of 2025. Pre-registration is clearly not relevant. The assessment of pertinence and usefulness will be done in a qualitative manner based on the feedback we get along the way. There will be assessments after each round, and we will consider whether it is worthwhile going from step 3 to 4.	
Evaluation / Implications of the Pilot (i.e., what effects do we want to examine, what are the confounding factors, etc.)		
Domain Coverage	Pilot with Qualitative research and Machine Learning research	
Evaluation Plan	Qualitative assessments will be done after each of the three phases where users are involved. Users' feedback and experience will be evaluated. For the cognitive phase such feedback will be used to adjust the tool. Feedback from the focus groups will be used to assess whether the tool is worthwhile, and if so, potentially final adjustments. Feedback from the final online exploration will be the basis for the final assessment of the tool.	
Evaluation Methods	In the cognitive exploration the users will be given an initial briefing and then provided some cases to go through. We will apply think-aloud and debriefing methods. For the focus groups will provide a briefing, demonstrate the intention and thereafter let the group go through a number of cases which we will discuss afterwards. For the online exploration, users will be presented with an initial briefing, then the tool such be able to instruct them what to do and they will then go through a number of cases after which they will get a small debriefing with some survey questions.	

	Assessments will be based on whether users find this 1) pertinent and worthwhile; 2) to what extent they think the Decision Aid is useful (cognitive and practical workload, difficulty).
Evaluation Metrics	None
Key Results /Outcomes	• aid researchers in purposefully linking the practices related to 'reproducibility' to their specific function(s) • prevent funders and publishers from demanding types of 'reproducibility' from research for which such types are irrelevant and/ or unfeasible • enable funders and publishers in facilitating types of 'reproducibility' where those are relevant and feasible
Key Performance Indicators (KPIs)	There will be no KPIs – we are perfectly capable of figuring out to what extent this will work or not.

Pilot activities timeline	**Timeline**
Pilot implementation and assessment plan template shared with timeline till M18	26th Oct 2023
Pilot plan using the Pilot implementation and assessment plan template	14th Nov 2023
Literature review / collecting evidence on existing tools and practices	Dec 2023
Piloting progress and preassessment of user studies and/or KPIs	April 2024
Pilot Pre-registration	n/a
First Pilot test with stakeholders	Autumn 2024
Reporting preliminary results	May 2025
Pilot presentation and documentation	June 2025
Second Pilot test (in 3 years, needed for funders)	

References (Guidelines (IEEE)**):**

[1] S. M. Field, R. Hoekstra, L. Bringmann, and D. Van Ravenzwaaij, "When and Why to Replicate: As Easy as 1, 2, 3?," *Collabra: Psychology*, vol. 5, no. 1, p. 46, Jan. 2019, doi: 10.1525/collabra.218.

[2] S. Guttinger, "The limits of replicability," *Euro Jnl Phil Sci*, vol. 10, no. 2, p. 10, May 2020, doi: 10.1007/s13194-019-0269-1.

[3] P. M. Isager, D. Lakens, T. Van Leeuwen, and A. E. Van 'T Veer, "Exploring a formal approach to selecting studies for replication: A feasibility study in social neuroscience," *Cortex*, vol. 171, pp. 330–346, Feb. 2024, doi: 10.1016/j.cortex.2023.10.012.

[4] P. M. Isager, A. E. Van 'T Veer, and D. Lakens, "Replication value as a function of citation impact and sample size." Aug. 24, 2021. doi: 10.31222/osf.io/knjea.

[5] S. Leonelli, "Rethinking Reproducibility as a Criterion for Research Quality," in *Research in the History of Economic Thought and Methodology*, vol. 36, L. Fiorito, S. Scheall, and C. E. Suprinyak, Eds., Emerald Publishing Limited, 2018, pp. 129–146. doi: 10.1108/S0743-41542018000036B009.

[6] J. Murphy, C. Mesquida, A. R. Caldwell, B. D. Earp, and J. P. Warne, “Proposal of a Selection Protocol for Replication of Studies in Sports and Exercise Science,” *Sports Med*, vol. 53, no. 1, pp. 281–291, Jan. 2023, doi: 10.1007/s40279-022-01749-1.
[7] Penders, Holbrook, and de Rijcke, “Rinse and Repeat: Understanding the Value of Replication across Different Ways of Knowing,” *Publications*, vol. 7, no. 3, p. 52, Jul. 2019, doi: 10.3390/publications7030052.
[8] M.-M. Pittelkow *et al.*, “The process of replication target selection in psychology: what to consider?,” *R. Soc. open sci.*, vol. 10, no. 2, p. 210586, Feb. 2023, doi: 10.1098/rsos.210586.

3.2.Pilot 2 - Reproducibility Management Plan (RMP)

	Pilot	ADDITIONAL DETAILS
Pilot Number & Title (Institution / Focal Person)	2. Reproducibility Management Plans (RMPs) OpenAIRE / Elli Papadopoulou	
Short description	The aim of the Pilot is to produce a 'Reproducibility Management Plan' (RMP) prototype that could serve as supporting material to public research funders' policies addressed to researchers / beneficiaries. The Pilot covers both theoretical and technical aspects that are concerned for its realisation. Specifically, it will define the new concept of RMPs and its content to then onboard it to ARGOS (argos.openaire.eu) open-source outputs management plans service. The Pilot employs co-creation activities to co-design the questions that need to be asked in the RMP prototype as well as to validate it in real life scenarios with the community.	
Objectives	The main objectives of the RMP Pilot are to: O1. Highlight the reproducibility activities involved in a research outputs management lifecycle. O2. Streamline reproducibility practices in public research funded projects. O3. Provide configurable tools and guidance to support the adoption of reproducibility best practices. O4. Deliver case studies to build a shared understanding of reproducibility in different domains. O5. Extend the DMP Common Standard. Below, we provide a preliminary overview of the tasks associated with fulfilling the objectives of the Pilot. A. <u>Theoretical to develop the RMPs definition</u> 1. Extend the DMPs concept, content and guidelines to cover reproducibility elements a. Create a prototype template (addressing funders) i. Co-define the questions of the RMP with the community 2. Adapt the RMP template to the diverse needs of scientific domains a. Create domain protocols i. Tailor questions and guidance per case study domain (addressing social sciences, life sciences, computer science) B. <u>Technical to support RMPs as open and FAIR outputs</u> 1. Onboarding of RMP templates to ARGOS (argos.openaire.eu): a. Configure custom APIs, enrich with semantics for "qualified references", link with MONITOR 2. Make RMPs interoperable (following the DMP Common Standard	

	a. Expose RMP outputs as machine actionable exports
Current literature/state of play	The Pilot builds on the concept of Data Management Plans (DMPs) for they are a great tool to ensuring transparency, reproducibility, and responsible conduct of research. Public research funders are among the primary stakeholders that have largely adopted the DMP concept by requiring it as a deliverable in the projects they fund. To support their uptake by the research community, they have developed DMP templates, i.e. pre-defined structured documents that contain instructions and questions to guide the input of beneficiaries in a coherent way. Science Europe, the association of major European research funding and research-performing organizations, has attempted to harmonise those templates across countries and domains with the publication of the Practical Guide to the International Alignment of Research Data Management [1] and Guidance Document Presenting a Framework for Discipline-specific Research Data Management [2]. Traditionally, DMPs refer to the handling of data, and occasionally might contain references to software that used this data as input in data management processes. Lately, the scope of DMPs is broadened to encompass the oversight of software, machine learning algorithms and other outputs generated through research [3] [4]. In parallel, services emerge to support researchers in writing and actively managing their plans, but the publishing of planning outputs in scholarly communication channels falls short affecting discovery of valuable information contained in them [5], although efforts are made to enable interoperability between those services [5]. Lastly, studies acknowledge DMPs as one of the tools for reproducibility [6], yet they target good practices to be followed outside DMPs or focus on the limitations of information covered by DMPs, e.g. in terms of data sharing [7], and not on elevating traditional DMPs to collectively plan reproducibility activities and follow them throughout the project lifecycle.
Existing tools related to the Pilot	There are different tools that can be characterized as “reproducibility tools” based on the scope they serve in a research management lifecycle. Such examples include, but are not limited to, platforms supporting DMPs and SMPs, pre-registrations, electronic notebooks, research objects management, sharing of negative results, etc. Our Pilot focuses on the planning and management of reproducibility activities for which there are no services to enable them.
Overview of existing tools/resources (software/platforms/instruments)	Our Pilot will employ ARGOS service for open, FAIR and machine actionable outputs management as it is the closest type of service which can be customized to fit the RMP work with workflows already target the planning and management of research outputs. In addition, ARGOS will be enhanced with external resources, such as FAIRsharing, for the automation of information retrieval from its registries supporting proper referencing by researchers filling in the RMP.

<table>
<tr><td>Methods used for piloting (methods used for tools or practices)</td><td>Our Pilot bases its work on existing prototypes-templates that are available for the planning of diverse types of outputs (e.g. DMPs, Datasheets for Datasets, Software Management Plans, etc) and management activities in scientific domains (e.g. Domain Data Protocols) to identify the reproducibility elements prevailing today. It also takes into consideration existing efforts to organize reproducibility practices in research projects available in literature. The Pilot will combine those findings with the literature review conducted in the context of the TIER2 project (T3.1 Conceptual framework for reproducibility across contexts) to specify reproducibility as it is expressed across different epistemic contexts. Outreach to diverse stakeholder groups will support understanding the perceptions and potential impact of the RMP Pilot outputs at different levels (policy – technical - practice). Co-creation activities and consultations with reproducibility experts and researchers will lead to the formulation of the questions that should be asked in an RMP which will be further analysed to specify their significance upon general or (multi)disciplinary use. The outcomes of this work will be combined to realise a multi-layered RMP prototype that contains general questions with the possibility to specialise the content by answering questions tailored to a discipline / domain, where applicable. The questions in the RMP will be aligned with the questions in the Decision Aid Pilot as the latter will be used in the RMP prototype to offer guidance to researchers. The final prototype, consisting of community-designed questions and instructions, will be onboarded to ARGOS [8]. Through the ARGOS service, RMPs will be enhanced with metadata and qualified references in preparation of their publication as open and FAIR outputs [9]. Further connections will be sought with 5.3.1 Reproducibility Promotion Plans to leverage outreach activities to funders and 5.3.3 Reproducibility monitoring dashboard to include RMPs in their pool of data.</td></tr>
<tr><td>Stakeholder groups affected and/or included</td><td>
<table>
<tr><th>Stakeholder Group (role: affected/included/beneficiaries)</th><th>Details (expected outcome for the stakeholder)</th></tr>
<tr><td>Researchers (affected; included; beneficiaries)</td><td>RMP prototype to enhance reproducibility practices in the context of a project.
Co-design of RMP questions and validation of the RMP prototype.
ARGOS to support the writing and publishing of RMPs.</td></tr>
<tr><td>Research communities (affected; included; beneficiaries)</td><td>RMP prototype to enhance reproducibility practices in the context of a project with domain specific elements / interests.
Co-design of RMP domain specific questions and validation of the RMP prototype.</td></tr>
</table>
</td></tr>
</table>

<table>
<tr><td></td><td>
<table>
<tr><td></td><td>ARGOS to support the writing and publishing of RMPs.</td></tr>
<tr><td>Funders (beneficiaries)</td><td>RMP prototype in support of Responsible Research and Innovation, including Open Science, policies.</td></tr>
<tr><td>Service providers (affected)</td><td>Metadata to extend the RDA DMP Commons Standard to include reproducibility elements.
Prototype / Template to adapt to.</td></tr>
</table>
</td></tr>
<tr><td>Stakeholder engagement plan</td><td>
Stakeholders' engagement happens both internally for project alignment and validation purposes, and externally for co-creation, validation, and uptake of outputs.

Internally

Consultation with consortium partners who are experts on reproducibility

Provide input that will inform the draft prototype as part of the co-creation activity.
Align with other project activities that the RMP is dependent on or has links with.

Externally:

Content specific

Focus groups to get input and feedback from scientists, research investigators, reproducibility initiatives and professionals on:

The reproducibility elements that they are able to identify in a traditional DMP template.
The reproducibility practices that they perform as part of the research process.
The reproducibility questions that they believe should be asked at the planning stage of the research output management lifecycle.

Focus groups with policymakers, research administration and funders to collect perceptions and seek adoption of RMPs:

As a concept (literature)
As a policy supporting tool

Workshops engaging research communities and researchers to increase the RMP adoption by HE projects.

Tool specific

Argos monthly community calls to get input and feedback on the usability and completeness of the RMP template.

</td></tr>
</table>

Detailed timeline until Pilot end	Preliminary timeline of the RMP Pilot activities (project deliverables are in bold): M6 Stakeholders identified in DMP. M7 Stakeholder data collected at the 1st workshop during the CERN x NASA Open Science Summit. M11 FAIRsharing API configured in ARGOS. M12-22 Other APIs configuration; Analysis of literature review findings from TIER2 project deliverables. M15-18 Stakeholder data collected (focus groups and workshops). Completion of data collection. Prototype structure drafted and questions analysed. M19-22 Prototype validation and alignment with Decision Aid Pilot; Usability testing of the RMP template; Prototype refinements and publication; TIER2 DMP update. **M22 M5.1 & M5.3: First releases of toolkits.** M23-31 Usability testing of the RMP template; User assessment survey; Finalisation of iterations and refinement of prototype; Publication of final version; Publication of case studies. M32-33 Update TIER2 DMP. **M34 D5.1 & D5.3: Delivery of toolkits.**
Evaluation / Implications of the Pilot (i.e., what effects do we want to examine, what are the confounding factors, etc.)	
Domain Coverage	The main part of the RMP prototype targets all researchers irrespective of their domain. From there, some parts of the RMP will be tailored according to the identified epistemic contexts and domain specific best practices. TIER2 has immediate access to social sciences, life sciences, and computer sciences, due to its partners competences and/or the thematic coverage of their respective Pilots. Hence, these domains will be explored further in the context of this Pilot.
Evaluation Plan	Evaluation of the plan will use both qualitative and quantitative methods. • Qualitative: understanding community perceptions and practices as well as identifying themes/patterns by analysing the open-ended questions and focus group discussions. • Quantitative: analysis of closed-ended responses from focus groups and workshops' activities using descriptive statistics.

Evaluation Methods	• User requirements assessment
	  ○ Develop feedback form(s) to be shared with workshop participants and beyond targeting the effectiveness and completeness of the draft template in terms of reproducibility.
	• Usability testing of the published RMP template.
	  ○ Develop a usability form with questions that capture the user experience (UX) to improve how the template is structured and how users navigate its content.

	▪ Example of questions: Why would you choose this template (drop down: makes writing easier; provides clear guidance on reproducibility; it applies best practices etc)? What are your thoughts from using the template (drop down: easy to complete; easy to navigate; etc)? What are the strongest points (drop down: the questions; the documentation; the tool features etc)? Would you recommend it to peers (likert scale)? Did you experience any challenges while using the template (open ended)? What can be improved (open ended)? • Peer Review of technical specifications ○ Present the new entities and properties to the RDA DMP group and seek feedback and adoption.
Evaluation Metrics	• Innovation – This is the first prototype for reproducibility that addresses the research output management lifecycle at the planning stage. • Inclusivity – The prototype is co-created with researchers and professionals beyond the TIER2 consortium. • Reproducibility – Completeness of information needed to replicate or reproduce findings. • Validity – addressing the requirements and practices of users in real life • Adoption Rate – The number of stakeholders (namely funders and projects) that will adopt the prototype.
Key Results /Outcomes (expected outcome)	O1: At least 25 reproducibility questions introduced in the prototype. At least 6 qualified references supported by the template (data-publications-software-methods-workflows-researchers). O2: Engagement of at least 3 funders for the uptake of the RMP prototype. O3: RMP prototype in TRL5 (ARGOS). O4: 15 projects including all identified domains. O5: At least 5 new entities and properties proposed to extend the DMP Common Standard.
Key Performance Indicators (KPIs)	Please see above.

Pilot activities timeline	**Timeline**
Pilot implementation and assessment plan template shared with timeline till M18	26th Oct 2023
Pilot plan using the Pilot implementation and assessment plan template	14th Nov 2023

Literature review / collecting evidence on existing tools and practices	Dec 2023
Piloting progress and preassessment of user studies and/or KPIs	Jan 2024
Pilot Pre-registration	Jan 2024
First Pilot test with stakeholders	Feb 2024
Reporting preliminary results	Mar 2024
Pilot presentation and documentation	May 2024
Second Pilot test (in 3 years, needed for funders)	

References

[1] S. Europe, “Practical Guide to the International Alignment of Research Data Management,” Science Europe. https://scienceeurope.org/our-resources/practical-guide-to-the-international-alignment-of-research-data-management/ (accessed May 23, 2024).

[2] S. Europe, “Guidance Document Presenting a Framework for Discipline-specific Research Data Management,” Zenodo, Jan. 2018, doi: https://doi.org/10.5281/zenodo.4925907.

[3] T. S. S. Institute, “Checklist for a Software Management Plan,” Zenodo, Dec. 10, 2018. doi: https://zenodo.org/record/2159713.

[4] T. Gebru et al., “Datasheets for Datasets,” arXiv (Cornell University), Mar. 2018, doi: https://doi.org/10.48550/arxiv.1803.09010.

[5] T. Miksa, P. Walk, and P. Neish, “RDA DMP Common Standard for Machine-actionable Data Management Plans,” GitHub, Apr. 29, 2024. https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard.

[6] S. Soundararajan and S. Mishra, “Data Management: The First Step in Reproducible Research,” Indian Journal of Occupational and Environmental Medicine, vol. 27, no. 4, p. 359, Dec. 2023, doi: https://doi.org/10.4103/ijoem.ijoem_342_22.

[7] M. Williams, J. Bagwell, and M. Nahm Zozus, “Data management plans: the missing perspective,” Journal of Biomedical Informatics, vol. 71, pp. 130–142, Jul. 2017, doi: https://doi.org/10.1016/j.jbi.2017.05.004.

[8] “Argos,” argos.openaire.eu. https://argos.openaire.eu/ (accessed May 23, 2024).

[9] E. Papadopoulou, A. Bardi, G. Kakaletris, Diamadis Tziotzios, Paolo Manghi, and N. Manola, “Data management plans as linked open data: exploiting ARGOS FAIR and machine actionable outputs in the OpenAIRE research graph,” Journal of biomedical semantics, vol. 14, no. 1, Nov. 2023, doi: https://doi.org/10.1186/s13326-023-00297-5.

3.3.Pilot 3 - Reproducible Workflows

Glossary for Pilot 3	
Software containerization	**Definition**: Containerization is the packaging of software code with just the operating system (OS) libraries and dependencies that are required to run the code to create a single lightweight executable—called a container—that runs consistently on any infrastructure [1]. **Significance:** Containerization helps encapsulate software in a way that is reproducible and consistent, regardless of the deployment environment. This is critical in research workflows where replicating computational environments is necessary to ensure reproducibility. **Example:** Docker, a popular containerization technology, has been extensively used in projects to package applications and their dependencies in a portable container that can run across different computing environments.
Workflow description languages	**Definition:** Workflow description languages are formal languages used to describe data analysis workflows including the steps and conditions involved in processing data. **Significance**: These languages are essential for ensuring that scientific workflows are reproducible. **Example:** Common Workflow Language (CWL) is an open standard for describing how to run command line tools and connect them to create workflows [2]. CWL is used as a workflow language in bioinformatics, helping standardize the description and execution of complex genomic analyses which can be reproduced across different systems.
Experiment Packaging Specifications	**Definition:** Experiment packaging specifications provide a structured, machine-readable format for packaging research data, software, and metadata that describe the digital objects and their relationships. This helps encapsulate all elements of an experiment. **Significance:** Packaging specifications ensure that all components of research (data, software, environment settings) are bundled together in a way that others can understand and reproduce the results. This is particularly important for the integrity and reproducibility of scientific experiments. **Example:** RO-Crate [3] is used in various research fields to package and share research outputs seamlessly, ensuring that all necessary metadata are included to replicate the study.

	Pilot	ADDITIONAL DETAILS
Pilot Number & Title (Institution / Focal Person)	3. Reproducible workflows (ARC / Thanasis Vergoulis, Eleni Adamidi)	
Short description	Pilot 3 aims to customize and evaluate tools/practices for reproducible workflows in life sciences and computer sciences. More specifically, we will adapt the open-source platform, SCHEMA, to explore its applicability in the aforementioned epistemic contexts. By	

	leveraging software containerization, workflow description languages (e.g., CWL, Snakemake), and experiment packaging specifications (e.g., RO-crate), we will extend SCHEMA to support reproducibility in life sciences and computer sciences (specifically for Machine Learning).
Objectives	**Main Goal:** Customize and evaluate tools/practices for reproducible workflows in the fields of life and computer sciences. **Underlying Objectives:** 1) Extend the open-source platform, SCHEMA, to support reproducibility in life science by leveraging software containerization, workflow description languages (e.g., CWL, Snakemake), and experiment packaging specifications (e.g., RO-crate). 2) Extend the open-source platform, SCHEMA to support reproducibility in computer science, particularly in the domain of Machine Learning.
Current literature/state of play	Faced with the complexity of analysis pipelines, the large number of computational tools, and the enormous amount of data to manage, there is compelling evidence that the reproducibility of computational workflows is of paramount importance [4], [5], [6]. Without advanced workflow systems, scripts that work on a single computer are often not scalable to larger or cloud-based systems without significant modification. Workflow systems like Galaxy and CWL provide scalable solutions that maintain the integrity and reproducibility of workflows across different computational environments [7]. Specifically, regarding the reproducibility in the computational research, the absence of systematic methods for managing data manipulation and version control, can lead to non-reproducible outcomes. Automated and documented workflows help avoid these pitfalls by ensuring that all data manipulations are traceable and reproducible [8]. Moreover, Detailed documentation and version control are critical for reproducibility, especially in computational research where outputs are highly dependent on specific software versions and configurations. Systems that track changes and manage versions of scripts and software settings help in maintaining the integrity of research outcomes [8].
Existing tools related to the Pilot	• ReproZip [9] is a tool designed for creating reproducible experiments by capturing and packaging the computational environment. It helps in sharing and reproducing research findings. • MLflow [10] is a platform for managing the end-to-end machine learning lifecycle, including experiment tracking, packaging code into reproducible runs, and sharing and deploying models.
Overview of existing tools/resources (software/platforms/instruments)	1. SCHEMA is an open-source platform that facilitates the execution of computational analysis on heterogeneous clusters, leveraging containerization methods to support the computational needs of research communities. Built upon SCHEMA's open-source framework, our Pilot extends its

<table>
<tr><td></td><td>functionality to support reproducibility in two epistemic contexts.
2. Common Workflow Language (CWL) and Snakemake provide a standardized language for describing computational workflows, contributing to the reproducibility framework. Exploration of additional workflow languages (e.g. Nextflow) will be performed.
3. RO-crates incorporated as an experiment packaging specification will enhance metadata standards, ensuring comprehensive reproducibility practices within the computational workflows tested in the Pilot.</td></tr>
<tr><td>Methods used for piloting (methods used for tools or practices)</td><td>The primary goal is to extend SCHeMa's functionalities based on stakeholder feedback, ensuring alignment with the diverse needs of the target communities in life sciences and computer sciences. The following methodology is set for this Pilot:
1. Stakeholder Engagement methods will be implemented including two rounds of questionnaires to explore special requirements and needs of stakeholder communities and to capture both quantitative and qualitative insights. The stakeholder engagement plan is further described in the related field.
2. Demonstration and feedback collection will be achieved through dedicated webinars. The first webinar will be conducted to demonstrate the core functionalities of the 1st TIER2 SCHEMA deployment release. The second webinar will be organized to present updates and enhancements based on the initial feedback and receive final feedback from the stakeholders.
3. Throughout the Pilot we will establish a ticketing mechanism within the GitHub repository for SCHEMA to encourage stakeholders to actively engage by reporting feature suggestions and bugs as well as to regularly monitor and categorize GitHub issues in the ongoing development process.
4. The 1st TIER2 SCHEMA Deployment Release (May 2024) will include core functionalities such as running tasks/workflows, saving outputs, and creating RO-crates.
5. Following the first release we will analyse both quantitative and qualitative data to identify, prioritize and implement improvements.</td></tr>
<tr><td>Stakeholder groups affected and/or included</td><td><table>
<tr><th>Stakeholder Group (role: affected/included/beneficiaries)</th><th>details (expected outcome for the stakeholder)</th></tr>
<tr><td>Life Scientists</td><td>Enhanced reproducibility practices tailored to the specific demands of life sciences research, leading to more</td></tr>
</table></td></tr>
</table>

<table>
<tr><td></td><td><table>
<tr><td></td><td>reliable and transparent computational workflows.</td></tr>
<tr><td>Computer Scientists</td><td>Improved reproducibility practices in computer science research, addressing challenges specific to this domain and fostering transparent and reusable ML workflows.</td></tr>
</table></td></tr>
<tr><td>Stakeholder engagement plan</td><td>Our stakeholder engagement plan involves the following steps:

1. Initial Engagement
Round 1 Questionnaire (by March 2024):
• Stakeholders will be engaged through a targeted questionnaire to identify their specific requirements and needs.
• Detailed instructions will accompany the questionnaire, providing clarity on the tool's objectives and functionalities.
• Support from WP2 is desired for the organization and facilitation of the initial stakeholder outreach, ensuring broad participation (especially for computer sciences).

2. Intermediate Engagement
1st Webinar (June 2024):
• A webinar will be conducted to demonstrate the core functionalities of the 1st TIER2 SCHEMA deployment release.
• Stakeholders will actively participate, offering immediate feedback on usability and features.
• Trainings will be integrated into the webinar to ensure stakeholders are well-versed in utilizing SCHEMA effectively.
GitHub Repo Interaction:
• A ticketing mechanism within the GitHub repository will provide stakeholders with a platform to report feature suggestions and bugs.
• Clear documentation and guides on how to engage with the GitHub repo will be provided.

3. Final Engagement
2nd Webinar (May 2025?):
• A second webinar will be organized to present updates and enhancements based on the initial feedback.
• Stakeholders will have the opportunity to offer final feedback on the improved SCHEMA functionalities.
• Trainings and support materials will be provided during the webinar to assist stakeholders in navigating updated features.
• WP2 support is desired for organizing and facilitating the final engagement webinar.

4. Dissemination Plan
Communication Channels:</td></tr>
</table>

	• Regular updates and announcements will be disseminated through project newsletters, project website, and relevant community forums. • Targeted emails will be sent to stakeholders to keep them informed about the progress and opportunities for engagement. Training Materials and Guides: • Develop comprehensive training materials and guides to accompany each phase of stakeholder engagement. • These materials will be disseminated through webinars, project documentation, and collaborative platforms. Regular coordination with WP2 is essential for leveraging their expertise in organizational and facilitation support.
Detailed timeline until Pilot end	**1. Preparation Phase** September to November 2023: • Develop and finalize a comprehensive Pilot and assessment plan. • Identify stakeholders for the initial phase. December 2023: • Draft the first round of questionnaires for stakeholder engagement. • Begin preparations for the 1st TIER2 SCHEMA deployment release. **2. Round 1 Stakeholder Engagement** March 2024: • Distribute the 1st round of questionnaires to stakeholders. • Analyse responses and identify key requirements. **3. Tool Development and Deployment** April 2024: • Develop and customize SCHEMA based on initial feedback. • Prepare training materials for the 1st TIER2 SCHEMA Deployment release. • Begin preparations for the Round 1 Webinar. May 2024 (Pilot Milestone 1): • Finalize the deployment of the 1st TIER2 SCHEMA Deployment release with core functionalities. **4. Round 2 Stakeholder Engagement** June 2024: • Conduct Round 1 Webinar to demonstrate to the stakeholders the core functionalities of the 1st TIER2 SCHEMA deployment release. Sep 2024:

	• Analyse data from the initial deployment and feedback. • Distribute the Round 2 Questionnaire for updated feedback. **5. Tool Development and Customization** March 2025: • Further develop and customize SCHEMA based on the Round 2 Questionnaire. May 2025 (Pilot Milestone 2): • Finalize TIER2 SCHEMA development based on the Round 2 Questionnaire data collection. July 2025 • Conduct Round 2 Webinar to present updates and enhancements based on the initial feedback (round 1). • Conduct additional training sessions for users as needed. • Encourage stakeholders to utilize the GitHub repo for interactions. **6. Analysis and Iterative Development** Sep 2025: • Analyse data from both rounds of questionnaires, webinars, and GitHub interactions. • Iterate on the tool based on collected feedback. Oct 2025: • Finalize iterative developments and enhancements to SCHEMA. • Prepare documentation and guides for the improved tool. • Continue engaging stakeholders in the GitHub repo for ongoing support and feedback. **7. Assessment and Reporting** Oct 2025: • Begin the assessment phase, evaluating the overall impact of SCHEMA on reproducibility. • Conduct internal reviews and assessments, identifying areas for further improvement. Nov 2025: • Present preliminary findings at relevant project meetings. • Plan for knowledge-sharing and dissemination activities. 8. **Dissemination and Pilot conclusion** Nov 2025: • Engage in knowledge-sharing activities, presenting the Pilot's outcomes at relevant conferences or workshops. • Disseminate results through project newsletters, the project website, and targeted communication channels.

Evaluation / Implications of the Pilot (i.e., what effects do we want to examine, what are the confounding factors, etc.)	
Domain Coverage	1) Life Sciences Life sciences encompass diverse fields such as biology, bioinformatics, and genetics, where computational workflows play a crucial role. The SCHeMa Pilot is specifically tailored to address the challenges of reproducibility in life sciences research. Computational methods are integral to tasks like data analysis, simulations, and modelling in life sciences, making it essential to ensure the reproducibility of these workflows. 2) Computer Sciences The extension of SCHeMa to computer sciences is motivated by the increasing reliance on computational methods, particularly in machine learning, artificial intelligence, and data science. Reproducibility is a critical concern in computer sciences to ensure the transparency and reliability of such algorithms and models.
Evaluation Plan	1) Round 1 Questionnaire (Feb 2024): Method: Distributed questionnaire targeting stakeholders (life scientists and computer scientists) to identify specific requirements and challenges related to reproducibility in computational workflows. Analysis: Qualitative analysis of responses to extract key themes and insights. 2) Round 1 Webinar (June 2024): Method: Conduct Round 1 webinar to showcase SCHEMA's core functionalities and gather initial feedback from participants. Analysis: Qualitative analysis of feedback during the webinar to understand user experiences and perceptions. 3) Round 2 Questionnaire (Sep 2024): Method: Distribute a follow-up Round 2 questionnaire to gather updated feedback and assess the evolution of user perspectives post the 1st TIER2 SCHeMa Deployment. Analysis: Qualitative analysis of responses, comparing them with Round 1, to identify changes. 4) Round 2 Webinar (July 2025): Method: Conduct a second Round 2 webinar for final feedback, focusing on the improvements made based on Round 1 feedback. Analysis: Qualitative analysis of final feedback, assessing overall user satisfaction and capturing suggestions for further enhancements. 5) Quantitative Data Analysis (Throughout): Method: Continuously collect quantitative data on tool usage, performance, and user interactions through the GitHub repository and SCHEMA platform. Analysis: Ongoing quantitative analysis of metrics such as the number of workflow executions, user engagement on GitHub.

	6) Final Assessment (Dec 2025): Method: Host a closing webinar to present the final version of SCHeMa, share outcomes, and gather last-round feedback. Analysis: Qualitative analysis of final feedback and a comprehensive review of quantitative data to measure the overall success and impact of the SCHeMa Pilot.
Evaluation Methods	1) Quantitative Metrics: Measure success through quantitative metrics such as the number of workflow executions, frequency of tool usage, and engagement on the GitHub repository. Track these metrics throughout the Pilot to assess adoption rates and user activity. 2) Webinar Feedback: Collect qualitative feedback during webinars to assess user reactions, understanding of SCHEMA functionalities, and identify areas of improvement. Use this feedback to inform iterative development. 3) GitHub Interactions: Track user interactions on the GitHub repository, including issues raised, feature requests, and contributions. Evaluate the level of community engagement and the responsiveness of the development team to user input.
Evaluation Metrics	1. Reproducibility Rate: Metric: The percentage of computational workflows that can be successfully reproduced using SCHEMA. This metric assesses the tool's impact on achieving reproducibility in comparison to existing approaches. 2. Adoption Rate: Metric: The rate of adoption of SCHEMA among the target user community, measured by the number of active users, workflow executions, and engagement on the GitHub repository. Improvement Criteria: A steady increase in the adoption rate over time, indicating the tool's acceptance and integration into users' workflows. 3. User Satisfaction Scores: Metric: Scores obtained from user assessment surveys, reflecting user satisfaction with SCHEMA's features, usability, and overall effectiveness in comparison to existing tools. Improvement Criteria: Positive trends in user satisfaction scores, indicating improvements in user experience and satisfaction over the course of the Pilot. 4. GitHub Interaction Metrics: Metric: Analysis of GitHub interactions, including the number of issues raised, feature requests, and contributions. This assesses the level of community engagement and the responsiveness of the development team.

	Improvement Criteria: Increased GitHub activity, with a diverse range of contributions and active participation from the user community.
Key Results /Outcomes (expected outcome)	1. Improved Reproducibility: Expected Outcome: SCHEMA facilitates a significant increase in the reproducibility rate of computational workflows in both life sciences and computer sciences. This outcome directly aligns with the Pilot's objective to customize and evaluate tools/practices for reproducible workflows. The significance lies in establishing SCHEMA as an effective solution for ensuring the transparency and reliability of computational research across diverse epistemic contexts. 2. Enhanced Adoption and Community Engagement: Expected Outcome: A notable increase in the adoption rate of SCHEMA, as evidenced by the growing number of active users, workflow executions, and contributions on the GitHub repository. 3. Positive User Satisfaction and Feedback: Expected Outcome: Continuous improvement in user satisfaction scores and positive feedback throughout the Pilot. This outcome is directly tied to the objective of adapting SCHEMA to different epistemic contexts. Positive user experiences indicate that SCHEMA effectively meets the needs of both life scientists and computer scientists, fostering a user-friendly environment for reproducible research practices. 4. Adherence to Reproducibility Best Practices: Expected Outcome: Successful integration of reproducibility best practices, such as the adoption of standardized workflow descriptions (e.g., CWL) and experiment packaging (e.g., RO-crate). This outcome is essential for ensuring the compatibility of SCHEMA with existing standards and guidelines, contributing to the tool's credibility and reliability in the research community.
Key Performance Indicators (KPIs)	1. KPI 1_Reproducibility Rate: Percentage increase in the successful reproduction of computational workflows using SCHEMA compared to baseline measurements or existing tools. 2. KPI2_Adoption Rate Across Domains: Rate of adoption measured by the number of active users, workflow executions, and GitHub engagement, categorized by different domains (life sciences and computer sciences). 3. KPI3_User Satisfaction Scores: Average satisfaction scores obtained from user assessment surveys at different stages of the Pilot. 4. KPI4_GitHub Interaction Metrics: GitHub activity metrics, including the number of issues raised, feature requests, and contributions from the user community.

	5. KPI5_Integration of Reproducibility Best Practices: Percentage of workflows adopting standardized descriptions (e.g., CWL) and experiment packaging (e.g., RO-crate).

Pilot activities timeline	Timeline
Pilot implementation and assessment plan template shared with timeline till M18	26th Oct 2023
Pilot plan using the Pilot implementation and assessment plan template	14th Nov 2023
Literature review / collecting evidence on existing tools and practices	Dec 2023
Piloting progress and preassessment of user studies and/or KPIs	Jan 2024
Pilot Pre-registration	Jan 2024
First Pilot test with stakeholders	Feb 2024
Reporting preliminary results	Mar 2024
Pilot presentation and documentation	May 2024
Second Pilot test (in 3 years, needed for funders)	

References

[1] "What Is Containerization? | IBM." Accessed: Feb. 12, 2024. [Online]. Available: https://www.ibm.com/topics/containerization

[2] "Common Workflow Language (CWL)." Accessed: Feb. 12, 2024. [Online]. Available: https://www.commonwl.org/

[3] P. Sefton *et al.*, "RO-Crate Metadata Specification 1.1," Oct. 2020, doi: 10.5281/ZENODO.4031327.

[4] P. DI Tommaso, M. Chatzou, E. W. Floden, P. P. Barja, E. Palumbo, and C. Notredame, "Nextflow enables reproducible computational workflows," *Nature Biotechnology 2017 35:4*, vol. 35, no. 4, pp. 316–319, Apr. 2017, doi: 10.1038/nbt.3820.

[5] T. Vergoulis, K. Zagganas, L. Kavouras, M. Reczko, S. Sartzetakis, and T. Dalamagas, "SCHeMa: Scheduling Scientific Containers on a Cluster of Heterogeneous Machines," Mar. 2021, Accessed: Nov. 17, 2023. [Online]. Available: https://arxiv.org/abs/2103.13138v2

[6] S. Cohen-Boulakia *et al.*, "Scientific workflows for computational reproducibility in the life sciences: Status, challenges and opportunities," *Future Generation Computer Systems*, vol. 75, pp. 284–298, Oct. 2017, doi: 10.1016/J.FUTURE.2017.01.012.

[7] J. M. Perkel, "Workflow systems turn raw data into scientific knowledge," Nature, vol. 573, no. 7772, pp. 149–150, Sep. 2019, doi: 10.1038/D41586-019-02619-Z.

[8] G. K. Sandve, A. Nekrutenko, J. Taylor, and E. Hovig, "Ten Simple Rules for Reproducible Computational Research," PLoS Comput Biol, vol. 9, no. 10, p. e1003285, 2013, doi: 10.1371/JOURNAL.PCBI.1003285.

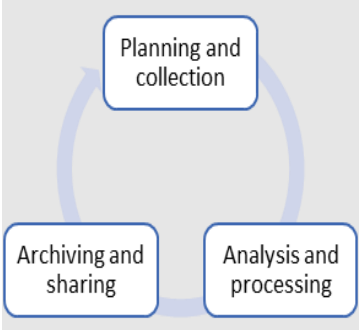
[9] “ReproZip! The Reproducibility Packer.” Accessed: Nov. 15, 2023. [Online]. Available: https://www.reprozip.org/

[10] “MLflow - A platform for the machine learning lifecycle | MLflow.” Accessed: Nov. 15, 2023. [Online]. Available: https://mlflow.org/

3.4.Pilot 4 - Reproducibility Checklists for Computational Social Science Research

	Pilot	ADDITIONAL DETAILS
Pilot Number & Title (Institution / Focal Person)	Pilot 4: Reproducibility checklists for Computational Social Science Research / Fakhri Momeni	
Short description	In this Pilot, we aim to provide a structure of well-defined checklists and templates that can help review data and code reproducibility for computational social scientists. The checklists and review templates cater for the specific needs of the three research phases, i.e., planning and data collection, process and analysis and finally sharing and archiving the research resources. It results in building trust and authority in the social science research community.	The outcomes of this Pilot are structured checklists and implementation templates for developing and reviewing data and code during the development phases of research. It can assist towards reproducibility workflows.
Objectives	The objectives of this Pilot are: • Encourage best practices (data, code) • High availability and reproducibility through review checklists • Uniformity in the research resources standards for social scientists	
Current literature/state of play	Reproducibility in scientific publications has been a concern across disciplines. In social science, computational models face the same challenges where either the information provided is not enough to reproduce results or the results are found to be different. Out of 19 publications claiming to be reproducible, only 13 were found to be mostly or fully reproducible [1]. They were evaluated by reproducing the same figures, numerical results, and conclusions. In a similar study on biomedical publications, the issues found are not enough information to reproduce the experimental environment, code with errors, and code having different results [2]. There is more of a consensus in the files and information required for reproducibility, where only sharing the data and code is clearly not sufficient [3]. The information to reproduce the research environment requires details e.g., requirements.txt file, yaml file and containerization files in some cases [1, 4, 5, 6, 7]. These files are also expected to be in the open science file format [1]. However, there are still challenges such as changing dependencies, software versions and computational	

	environments [8]. There is also scarcity of indicators crucial for replication and verification [4]. Non-computational resources are also highlighted as transparency and reproducibility related indicators in social science literature. They include sharing raw material, access protocols and statements of interest, funding, and access [4, 5]. They are referred to as meta research data indicating how reproducibility-friendly a research policy is e.g., having open access publication with availability of statements, material, and scripts [6]. There are barriers hindering reproducibility that are about the ecosystem involving institutional, data related, ethical, organizational, and infrastructural challenges [8]. Preregistrations are widely adopted in research; however, it also shows signs of misuse and potentially adverse effects in the long run [16]. Pre-publication replicability assessments may be considered by journals for submitted manuscript, where post publication reproducibility can be more strictly enforced [17]. To analyse reproducibility at a deeper level, it must be observed beyond binary as reproducible or not reproducible. A tier system is suggested that shows how to aim above reproducibility i.e., targeting replicability, robustness and generalizability, counteracting the practices of open washing [2, 15]. It is also a lesser highlighted concern that to whom is the research reproducible. Three degrees of computational reproducibility are defined where *1°CR* is reproducible by the scholar, *2°CR* is by a trusted third party e.g., journal and *3°CR* is for public [15]. Reproducibility is rather extended to 8 different levels with minimum being not reproducible while higher levels emphasizing on containerization with efforts and online connectivity required to reproduce the experiments [7]. Early researchers should be trained in reproducibility and open science practices in their daily work. However, efforts from researchers at various levels and roles are essential to promote reproducibility in research and open science practices in institutions [17]. The collaboration on data, access protocols and interdisciplinary policies and practices at an institute level can lower the barriers further [8]. The reproducibility level targeted is to be identified and efforts on reproducibility must focus on it at the user and software levels. There is a need for roadmaps, guidelines and policies that would facilitate reproducibility at different levels, however, it should not overburden the scholars [2]. The existing literature, although highlighted reproducibility but does not provide clear guidelines to achieve reproducibility across the different stages of research. Similarly, they do not indicate the effort required in attaining reproducibility at different stages of research which leads to more willingness to follow reproducibility in theory but not in practice, even more so for early researchers [19]. [22] proposed an agenda for adopting open science practices in Communication including seven suggestions. Also, TOP guidelines [20] provide a valuable framework for enhancing transparency and reproducibility in research across various disciplines. However, they do not fully address the broader challenges related to code sharing, reproducibility, and sustainability in social science research, particularly in the context of micro-level administrative data analysis [21]. These challenges include the lack of dedicated tools, data

<table>
<tr><td></td><td>retention policies, version control and code sharing practices, sustainability issues with web-based resources, and the absence of structured frameworks for research objects [21]. Introducing a web-based platform for sharing code and data, supplemented with a checklist, can provide a structured approach for computational social science researchers to share methods. This platform promotes reproducibility, facilitates code sharing, centralizes resources, and encourages community engagement, thereby addressing highlighted barriers to transparent and reproducible research.</td></tr>
<tr><td>Existing tools related to the Pilot</td><td>The existing notable work on reproducibility checklist are:
Transparency and openness promotion (TOP) guidelines (by journals, funders and societies) for researchers
Open Science Framework (OSF) offers guidelines and framework for researchers to organize and assist in planning, collecting, analysing and sharing their work
Center of Open Science (COS) offering resources and guidance on implementing reproducibility in research
DIME standards by world bank on reproducibility in research</td></tr>
<tr><td>Overview of existing tools/resources (software/platforms/instruments)</td><td></td><td></td></tr>
<tr><td>Methods used for piloting (methods used for tools or practices)</td><td>The method for developing the reproducibility checklist involves different dimensions, such as:
Research area: Social Science
Data: DBD data
Resource type: Data, Code
Research phase: e.g., Planning and Collection, Process, Preserve and Sharing 

The Methods-Hub portal assists computational social scientists in offering reproducible computational methods that would primarily work with DBD data.

The proposed methodology consists of developing checklists and templates that would assist researchers along different research</td></tr>
</table>

<table>
<tr><td></td><td>phases. We further will determine KPIs to evaluate our Pilot. We will conduct surveys before and after the Pilot implementation to analyse and evaluate the Pilot. The main steps are:
Determining KPIs
Designing and conducting surveys
Developing checklists and templates for data and code reproducibility
Implementing them as workflows for the three research phases
Conducting post Pilot surveys, after Pilot implementation
Evaluation, analysis and final Pilot review
The methodology steps defined would be applied across the following three research phases for research data and code.</td></tr>
<tr><td>Stakeholder groups affected and/or included</td><td><table><tr><th>Stakeholder Group (role: affected/included/beneficiaries)</th><th>details (expected outcome for the stakeholder)</th></tr><tr><td>affected</td><td>Computational Social Scientists (Research Producers and consumers)</td></tr><tr><td>included</td><td>Computational Social scientists</td></tr><tr><td>beneficiaries</td><td>Social science community</td></tr></table></td></tr>
<tr><td>Stakeholder engagement plan</td><td>We gather input from researchers across computer science, social science, and computational social science domains, incorporating feedback from diverse perspectives. Through various user studies, we assess the checklist's readability and comprehensibility among researchers with varying computational backgrounds. This iterative process ensures that the checklist evolves based on real-world usage and diverse viewpoints, improving its usability and applicability across different research contexts. Additionally, we focus on disseminating the checklist through other channels, such as conferences, workshops, and online forums, to enhance its relevance and adoption within the research community.

The Pilot development and evaluation involve engagement with stakeholders twice through surveys conducted on computational social scientists:
Baseline Survey: This survey will be conducted at the beginning of the Pilot among individuals with varying levels of familiarity with reproducibility practices. It will assess their current practices, attitudes, and awareness regarding reproducibility in their research workflows.
Follow-up Survey: After the implementation of the Pilot, we will conduct a follow-up survey specifically targeting individuals who have utilized the checklist provided on the Methods-Hub platform. This survey will assess their experiences, challenges</td></tr>
</table>

<table>
<tr><td></td><td colspan="2">encountered, and improvements observed in their adoption of reproducibility practices as a result of using the checklist.

The results will be analysed for evaluating the Pilot. Alongside that, we also use KPIs to evaluate the performance of the Pilots.</td></tr>
<tr><td>Detailed timeline until Pilot end</td><td colspan="2">• Literature review (Dec 2023)
• First draft of the general checklist for reproducibility (Mar 2024)
• First User study for the feedback on the first draft (march-April 2024)
• Baseline Survey (June 2024)
• Second draft of the checklist and focus on three parts: (June 2024)
  ○ Planning and Collection
  ○ Analysis and processing
  ○ Archiving and Sharing
• Second user study for feedback on the checklists (July-Sep. 2024)
• Revise the checklists and finalize it and embed into the Methods-Hub portal
  ○ prototype (Oct.-Nov. 2024)
  ○ First external release planned (Jan- July. 2025)
• Dissemination (workshop, conferences or online forums, hackathons) (Jan. 2025-June. 2025)
• Second round of Checklists updates based on feedback (Jan-June 2025)
• Follow-up Survey (Aug. 2025)
• Evaluation using KPIs (Oct. 2025)</td></tr>
<tr><td colspan="3">Evaluation / Implications of the Pilot
(i.e., what effects do we want to examine, what are the confounding factors, etc.)</td></tr>
<tr><td>Domain Coverage</td><td>Computational social science</td><td>The Pilot activity focuses on the needs of researchers working on various sub-domains of computational social science.</td></tr>
<tr><td>Evaluation Plan</td><td colspan="2">We plan to evaluate our Pilot implementation by putting it to practice directly to computational social scientists and learning from their experience through surveys.</td></tr>
<tr><td>Evaluation Methods</td><td colspan="2">The method of evaluation is to prepare questionnaires and align them into pre-Pilot implementation and post-Pilot implementation user studies.
Each survey would have questions to cover the three phases of research to not only measure the improvement in reproducibility of the research at the end but also during the development phases. To benefit from the resources at GESIS, we also consider expert opinion on the Pilot implementation.</td></tr>
</table>

Evaluation Metrics	The evaluation metrics we plan to use are separated across two resource types (data and code). They would determine improvement based on pre and post usage of the implementation for each KPI e.g., improvement in code sharing rate, improvement in data quality and FAIRness etc.	
Key Results /Outcomes (expected outcome)	Improved quality of research after the checklists (before, after) (KPIs: To develop) e.g., Data: data sharing rate, data quality (FAIRness), etc. Code: code sharing rate, replicability rate etc.	Better collaboration, encouragement of best practices, efficient use of resources and increased public trust in the research community
Key Performance Indicators (KPIs)	Data and Code Sharing Rate, data quality, Replicability Rate	

Pilot planning activities timeline	**Timeline**
Pilot implementation and assessment plan template shared with timeline till M18	26th Oct 2023
Pilot plan using the Pilot implementation and assessment plan template	14th Nov 2023
Literature review / collecting evidence on existing tools and practices	Dec 2023
Piloting progress and preassessment of user studies and/or KPIs	Jan 2024
Pilot Pre-registration	June 2024
First Pilot test with stakeholders	March-April 2024
Reporting preliminary results	May 2024
Pilot presentation and documentation	May 2024
Second Pilot test (in 3 years, needed for funders)	October 2025

References:

[1] Archmiller, A. A., Johnson, A. D., Nolan, J., Edwards, M., Elliott, L. H., Ferguson, J. M., ... & Fieberg, J. (2020). Computational reproducibility in The Wildlife Society's flagship journals. The Journal of Wildlife Management, 84(5), 1012-1017.

[2] Samuel, S., & Mietchen, D. (2023). Computational reproducibility of jupyter notebooks from biomedical publications. *arXiv preprint arXiv:2308.07333*.

[3] Clyburne-Sherin, A., Fei, X., & Green, S. A. (2019). Computational reproducibility via containers in psychology. Meta-psychology, 3.

[4] Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., ... & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. Annual review of psychology, 73, 719-748.

[5] Hardwicke, T. E., Thibault, R. T., Kosie, J. E., Wallach, J. D., Kidwell, M. C., & Ioannidis, J. P. (2022). Estimating the prevalence of transparency and reproducibility-related research practices in psychology (2014–2017). Perspectives on Psychological Science, 17(1), 239-251.

[6] Hardwicke, T. E., Wallach, J. D., Kidwell, M. C., Bendixen, T., Crüwell, S., & Ioannidis, J. P. (2020). An empirical assessment of transparency and reproducibility-related research practices in the social sciences (2014–2017). Royal Society Open Science, 7(2), 190806.

[7] J. Bednar. “Reproducibility: Future-proofing your python projects” Website Name. https://www.anaconda.com/blog/8-levels-of-reproducibility (accessed 11-10, 2023).

[8] Lazer, D. M., Pentland, A., Watts, D. J., Aral, S., Athey, S., Contractor, N., ... & Wagner, C. (2020). Computational social science: Obstacles and opportunities. Science, 369(6507), 1060-1062.

[9] Chan, C. H., & Schoch, D. (2023). rang: Reconstructing reproducible R computational environments. Plos one, 18(6), e0286761.

[10] Schoch, D., & Chan, C. H. (2023). Software presentation: Rtoot: Collecting and Analyzing Mastodon Data. Mobile Media & Communication, 20501579231176678.

[11] Chan, C. H., & Sältzer, M. (2020). oolong: An R package for validating automated content analysis tools. The Journal of Open Source Software: JOSS, 5(55), 2461.

[12] Chan, C. H. (2023). grafzahl: fine-tuning Transformers for text data from within R. Computational Communication Research, 5(1), 76.

[13] Ho, J. C. T., & Chan, C. H. (2023). Evaluating transferability in multilingual text analyses. Computational Communication Research, 5(2), 1.

[14] Chan, C. H., Zeng, J., Wessler, H., Jungblut, M., Welbers, K., Bajjalieh, J. W., ... & Althaus, S. L. (2020). Reproducible extraction of cross-lingual topics (rectr). Communication Methods and Measures, 14(4), 285-305.

[15] Schoch, D., Chan, C. H., Wagner, C., & Bleier, A. (2023). Computational Reproducibility in Computational Social Science. arXiv preprint arXiv:2307.01918.

[16] Pham, M. T., & Oh, T. T. (2021). Preregistration is neither sufficient nor necessary for good science. Journal of Consumer Psychology, 31(1), 163-176.

[17] Altmejd, A., Dreber, A., Forsell, E., Huber, J., Imai, T., Johannesson, M., ... & Camerer, C. (2019). Predicting the replicability of social science lab experiments. PloS one, 14(12), e0225826.

[18] Kohrs, F. E., Auer, S., Bannach-Brown, A., Fiedler, S., Haven, T. L., Heise, V., ... & Weissgerber, T. L. (2023). Eleven strategies for making reproducible research and open science training the norm at research institutions. Elife, 12, e89736.

[19] Ferguson, J., Littman, R., Christensen, G., Paluck, E. L., Swanson, N., Wang, Z., ... & Pezzuto, J. H. (2023). Survey of open science practices and attitudes in the social sciences. Nature communications, 14(1), 5401.

[20] Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S., Breckler, S., ... & DeHaven, A. C. (2016). Transparency and openness promotion (TOP) guidelines.

[21] Playford, C. J., Gayle, V., Connelly, R., & Gray, A. J. (2016). Administrative social science data: The challenge of reproducible research. *Big Data & Society*, *3*(2), 2053951716684143.

[22] Dienlin, T., Johannes, N., Bowman, N. D., Masur, P. K., Engesser, S., Kümpel, A. S., ... & De Vreese, C. (2021). An agenda for open science in communication. *Journal of Communication*, *71*(1), 1-26.

3.5.Pilot 5 - Reproducibility Promotion Plans for Funders

	Pilot	ADDITIONAL DETAILS
Pilot Number & Title (Institution / Focal Person)	5 – Reproducibility Promotion Plans for Funders (RPPF)	VUmc; Joeri Tljdink and Barbara Leitner
Short description	Pilot 5 aims to develop a reproducibility promotion plan for funders to foster change amongst researchers towards reproducibility and reproducible practices. The RPPF will be co-created by a group of stakeholders who have a strong background and motivation towards increasing reproducibility in research. During two workshops, funders co-create the themes/elements they find the most important to include in a promotion plan followed by specific recommendations for funders on how they can navigate these themes/elements. The RPP is aimed to be a skeleton promotion plan that funders can take and personalize to their own funding organization's needs. Funders will be asked to Pilot the RPP after development.	
Objectives	Objectives: 1. Co-create with a group of funders the elements and themes with specific recommendations to help foster reproducibility in researchers 2. Test these recommendations and guidance/guiding principles in the stakeholder group during several Pilots in different settings 3. Come up with a clear plan that funders can use to improve reproducibility practices	The Pilot aims to help promote reproducibility amongst funders and help them create a reproducibility promotion plan they can use and promote with researchers. The goal is for funders to create their own 'RPP4F' plans and tailor it according to their needs.
Current literature/state of play	Currently there is not a lot of evidence or literature discussing how funders can foster reproducibility. However, literature suggests incentives for researchers and research organizations to promote reproducible practices [4]. Current literature in the field has examined how to foster research	However, it is known that they fund reproducibility and mandate reproducibility practices such as Open Science practices, data management plans to name a few.

	integrity institutionally to influence researchers' behaviour [3].	
Existing tools related to the Pilot	Currently two tools are present for funders to help foster reproducibility. • One of the tools for funding agencies is the CIHR IRSC, which promotes funder's roles in improving data stewardship through program design, peer review, and policy/guidelines. The CIHR IRSC is focused on funding research (clinical and non-clinical) trials [1]. • The second tool called Research Integrity Promotion Plan (RIPP) was designed by Serge et al. (2022). The RIPP consists of 6 core topics with guidelines for each topic to help funders implement a RIPP. The topics and guidelines guide funders towards strengthening research integrity policy in their organizations and guide research organizations they fund [2].	
Overview of existing tools/resources (software/platforms/instruments)	The recommendations and documents used by the CIHR IRSC and RIPP will be used as examples to help inspire funders during the co-creation workshops.	
Methods used for piloting (methods used for tools or practices)	The Pilot has 3 methodological steps: 1. The cocreation process: with several funders, we organise 2 workshops that will determine the topics that should be included in the RPPs and will come up with guidelines (and sub guidelines) with recommendations to improve reproducibility practices in these funders 2. These recommendations will be collected, improved and refined in one promotion plan that outlines how funders can use the guideline, the topics and the recommendations 3. This will eventually result in a final version	• Creating a RPP for funders with the input from the workshops • Organise a second workshop to safeguard implementation and user friendliness of the RPP • Use 4 Pilot funding institutions to test the RPP and require feedback and suggestions for improvement • Qualitative analysis of workshops

<table>
<tr><td></td><td>of the guideline with recommendations</td><td></td></tr>
<tr><td>Stakeholder groups affected and/or included</td><td colspan="2"><table>
<tr><th>Stakeholder Group (role: affected/included/beneficiaries)</th><th>details (expected outcome for the stakeholder)</th></tr>
<tr><td>Funders</td><td rowspan="3">Useful reproducibility promotion plan that can be used in different setting/contexts and countries
They will have standards that will help them make their work reproducible.
Develop guidelines where they can aid funders to promote reproducibility</td></tr>
<tr><td>Researchers</td></tr>
<tr><td>Policymakers</td></tr>
</table></td></tr>
<tr><td>Stakeholder engagement plan</td><td colspan="2">1. We already have organised with WP2 a stakeholder workshop. In this workshop we outlined the plan of the project, what role funders can play in reproducibility and what they can improved. There we already improved their engagement and consent to be invited for our workshops.
2. Two co-creation workshops will be held, one in March and one in April with a total of eight funders. The funders will be the ones creating the content of the RPPF during these workshops. Participants after the workshop will be sent the first draft of the RPPF for validation before piloting.
3. The funders who participated in the workshops will be contacted to participate in the piloting of the RPPF.</td></tr>
<tr><td>Detailed timeline until Pilot end</td><td colspan="2">• January –March 2024: Design of plan, development of co-creation workshops, evaluation development
• March- June 2024: Preregistration of workshops, cocreation workshops to develop the RPP
• October-November 2024: Finalization of the RPP, including template for piloting
• January-October 2025: Organization of Pilots to further refine the guidelines/recommendations
• Assessment of Pilot testing
• Reporting and finalization of the RPP</td></tr>
<tr><td colspan="3">Evaluation / Implications of the Pilot
(i.e., what effects do we want to examine, what are the confounding factors, etc.)</td></tr>
<tr><td>Domain Coverage</td><td colspan="2">Not domain specific, suitable for funders for all domains</td></tr>
<tr><td>Evaluation Plan</td><td colspan="2">There will be multiple opportunities for evaluation during the Pilot process. The first occasion for evaluation occurs at the end of the two workshops. Participants are asked two evaluation questions to assess the workshop (what they thought about the workshop and what they learnt and took onboard from the discussion during the workshop). The</td></tr>
</table>

<table>
<tr><td></td><td colspan="2">second opportunity for evaluation occurs after the two workshops, all the participants are sent the first draft of the reproducibility promotion plan with recommendations for feedback. Whilst funders Pilot the RPP within their own funding organizations we will have multiple check-ins on the progress and for evaluation. Piloting funders will be interviewed after 6 months of using the RPP, then after one year, again after one and a half years from the start. And a final closing interview to evaluate the RPP for funding will be assessed after two years of use.</td></tr>
<tr><td>Evaluation Methods</td><td>1. Evaluation of the content of the workshop at the end of the workshop (happens at the end of each workshop, qualitative asked on thoughts of the workshop and what they take home from the discussion)
2. Evaluation of the cocreated content during the second workshop
3. RPP sent to all participants for final review and proofreading through MIRO
4. Interview and survey evaluation of the funders piloting the RPP within their own funding organizations</td><td>1. Document evaluation
2. Process evaluation
3. Evaluation of implementation Pilot institutions</td></tr>
<tr><td>Evaluation Metrics</td><td colspan="2">The metrics employed for evaluation of the Pilot are qualitative in nature. During the four follow-up interviews multiple qualitative metrics will be used:
• Funder satisfaction with the RPP: Funders indicate their satisfaction with the use of the RPP on scale of 1 to 5 (where 1 is unhappy and 5 is very happy)
• Funder satisfaction with change in the researcher's behaviour: Funders indicate their satisfaction with how the RPP influenced researcher's actions in regard to reproducibility and reproducible practices on scale of 1 to 5 (where 1 is unhappy and 5 is very happy)
• Compliance Rate of Researchers: 1-5 Likert scale (where 1 means no compliance with the RPP and 5 means full compliance). Funders are asked how they would rate the compliance.</td></tr>
</table>

	• Adoption Rate: 1-5 Likert scale (where 1 means not adopted at all and 5 adopted multiple times/multiple projects)
Key Results /Outcomes (expected outcome)	Finalization of the RPP ready to use for funders 1. Quantitative assessment: is there a change in policy, was the RPP implemented within the funding organization, have more funders employed it from the same organization. 2. Content evaluation with funder 3. Process evaluation with funder
Key Performance Indicators (KPIs)	Interview with Pilot institutions on the RPP and interviews with the applicants, funders indicate they can see themselves using this tool in the future through direct time points (2 years' time and again in 4 years' time). Qualitative measure of adoption rate (insights into real use).

Pilot activities timeline	**Timeline**
Pilot implementation and assessment plan template shared with timeline till M18	26th Oct 2023
Pilot plan using the Pilot implementation and assessment plan template	14th Nov 2023
Literature review / collecting evidence on existing tools and practices	Dec 2023
Piloting progress and preassessment of user studies and/or KPIs	Jan 2024
Pilot Pre-registration	Jan 2024
First Pilot test with stakeholders	Feb 2024
Reporting preliminary results and validation	April-May 2024
Pilot presentation and documentation	May 2024
Start of funders piloting RPPF	June 2024
Check in with funders (first round of interviews)	November 2024
Check in with funders (second round of interviews)	June 2025
Check in with funders (third round of interviews)	November 2025
Second Pilot test (in 3 years, needed for funders)	2027

References

[1] Goosney, D. (2014-10-16). Funding agencies and reproducibility: CIHR's role. EQUATOR Network. Funding Agencies and Reproducibility: CIHR's Role (equator-network.org)

[2] Horbach, S.P.J.M, Bouter, L.M. Gaskell, G., Hiney, M., Kouvras, P .... (2022). "Designing and implementing a research integrity promotion plan: Recommendations for research funders." *PLoS Biology,20* (8), e3001773, doi: 10.1371/journal.pbio.3001773

[3] Labib, K., Bouter, L., Roje, R., Widdershoven, G., Evans, N., Marusic, A., Mokking, L., & Tijdink, J. (2021). "Important topics for fostering research integrity by research performing and research funding organizations: A delphi consensus study." *Science and Engineering Ethics, 27* (47), doi: 10.1007/s11948-021-00322-9

[4] Nosek, B.A., Alter G., Banks, G.C., Borsboom, D., Browman, S.D., & Breckler, S.J. (2015). “Promoting open research culture”. *Science, 348* (6242), 1422-1425, doi: 10.1126/science.aab2374

3.6.Pilot 6 - Reproducibility Monitoring Dashboard

	Pilot	ADDITIONAL DETAILS
Pilot Number & Title (Institution / Focal Person)	6. Reproducibility Monitoring Dashboard (ARC / Haris Papageorgiou, OpenAIRE/Natalia Manola)	
Short description	Pilot 6 aims to develop tools that enable funding agencies in tracking & monitoring reusability of research artefacts (datasets, software, tools/systems, etc) produced within projects of interest, across different programmes, topics & disciplines. An automatically generated report (dashboard) will be developed facilitating assessment & quantification of the impact of policies for data-sharing, code-sharing, etc.	
Objectives	**Main Goal:** Produce tools that enable funding agencies in tracking & monitoring reusability of research artefacts. **Underlying Objectives:** 1) Develop/extend and test an artillery of tools for tracking major research artefacts in Computer Sciences (e.g., datasets, software, conclusions) with a focus on Artificial Intelligence, 2) Quantify and estimate Reusability indicators based on different types of artefacts, 3) Design & implement a dashboard enabling funding agencies in tracking & monitoring reusability of research artefacts (datasets, software, tools/systems, etc) created in funded projects.	
Current literature/state of play	Literature contains many datasets and systems that use different annotation schemes, metrics and definitions on the Research Artefacts (RAs) • Different ontologies and degrees of abstraction used (e.g., software vs. application, plugin, programming language), • Overlapping types of artefacts (e.g., dataset vs. supplementary material), • No general framework exists for comparisons [1] There is a distinction between **named** (explicit) and **unnamed** (implicit) research artefacts, with the latter being rather hard to successfully detect, deduplicate and link to well-defined Research Artefact entities (e.g., artifacts in a knowledge base), making reusability estimators a real challenge. The task of Research Artefact Analysis (RAA) is deeply dependent on the **discipline** of the document, publication or text that is being analysed. Most approaches treat RAA as a “Sequence Classification” task (e.g., SoMeSci [2], SciRex [4], SciErc [5], DMDD [7], BioNERds [8], etc.) using RNN and BERT-based technologies (i.e., SciBert [9]) • They do not take advantage of the technological capabilities of Large Language Models (LLMs)	

	The extraction of useful metadata alongside the research artefacts that indicate **use of FAIR principles** and contribute to measuring the reusability and reproducibility of the work is **limited** (e.g., SoMeSci [2], Softcite [3]) • Name, License, Version, URL, Developer, Citation, etc • Ownership and usage by the authors

Existing tools related to the Pilot	• DataSeer (https://dataseer.ai/)
	  ○ Identification of sections and sentences introducing datasets in a scientific article, and classification of the type of these datasets
	• Datastet (https://github.com/kermitt2/datastet)
	  ○ Identification of named and implicit research datasets and associated attributes in scientific articles
	• Softcite [3] software mention recognition service (https://github.com/softcite/software-mentions)
	  ○ Recognition of software mentions and associated attributes in scientific literature within the disciplines of life sciences and economics
	• SoMeNLP (https://github.com/dave-s477/SoMeNLP)
	  ○ Information extraction for software mentions in scientific articles within the discipline of life sciences
	  ○ Named Entity Recognition, Relation Extraction and Entity Disambiguation
	  ○ Trained on SoMeSci [2] dataset (https://dl.acm.org/doi/10.1145/3459637.3482017)
	• SoftwareKG (https://github.com/dave-s477/softwareKG/)
	  ○ Information extraction for software mentions in scientific articles within the discipline of social science
	  ○ Trained on SoSciSoCi [6] dataset (https://link.springer.com/chapter/10.1007/978-3-030-49461-2_16)

Overview of existing tools/resources (software/platforms/instruments)	• GROBID (https://github.com/kermitt2/grobid) ○ Machine learning library for extracting, parsing and re-structuring raw documents such as PDF into structured XML/TEI encoded documents with a particular focus on technical and scientific publications.

Methods used for piloting (methods used for tools or practices)	• **Plan (iterative**):
	• Collect open publications from OpenAIRE based on specific criteria (topic, FP Projects, disciplines) & build collection(s),
	• Develop user stories for targeted stakeholders,
	• Analyse the collection(s),
	• Indicative Indicators to measure & report on the dashboard:
	  ○ Documentation of new artefacts (Software, Datasets)
	  ○ Reusability of artefacts (Software, Datasets)
	  ○ Impact of Reusability
	• Run the eval processes and a webinar with stakeholders in rounds

Stakeholder groups affected and/or included	**Stakeholder Group** | **details (expected outcome for the stakeholder)**

<table>
<tr><td></td><td>
<table>
<tr><td>(role: affected/included/beneficiaries)</td><td></td></tr>
<tr><td>Research Performing Organizations (RPOs)</td><td>RPOs will benefit from enhanced visibility and traceability of their research artefacts through the new tools and dashboard. This will facilitate better showcasing of their work's impact and reusability, potentially leading to increased funding opportunities and collaboration.</td></tr>
<tr><td>Funders</td><td>Funders will gain a comprehensive tool for monitoring and assessing the impact and reusability of research artefacts. This will enable more informed decision-making, ensuring effective allocation of resources and fostering a culture of transparency and accountability in research funding.</td></tr>
<tr><td>Publishers</td><td>Publishers will benefit from spotting undocumented or partially documented research artefacts in scientific manuscripts at an early stage.</td></tr>
<tr><td>Researchers</td><td>Researchers can search for well-documented and well-cited research artefacts in the specific scientific field or area they are most interested in.</td></tr>
</table>
</td></tr>
<tr><td>Stakeholder engagement plan</td><td>
Our stakeholder engagement plan involves the following steps:

1. Stakeholder Engagement (First Phase)
Webinar:
• A webinar will be conducted to demonstrate to the stakeholders the core functionalities of
  ○ the 1st Deployment Dashboard release,
• Stakeholders will actively participate, offering immediate feedback on usability and features.
• Trainings will be integrated into the webinar to ensure stakeholders are familiar with the OpenAIRE Platform.

2. Stakeholder Engagement (Second Phase)
Workshop:
• A workshop will be organized to present updates and enhancements based on the initial feedback.
• Stakeholders will have the opportunity to offer final feedback on the improved OpenAIRE functionalities.
• Trainings and support materials will be provided during the workshop to assist stakeholders in navigating updated features.
• WP2 support is desired for organizing and facilitating the final engagement workshop.

3. Dissemination Plan
Communication Channels:
</td></tr>
</table>

	• Regular updates and announcements will be disseminated through project newsletters, project website, and relevant community forums. • Targeted communication will be sent to stakeholders to keep them informed about the progress and opportunities for engagement. Materials and Guides: • Develop comprehensive materials and guides to accompany each phase of stakeholder engagement. • These materials will be disseminated through webinar/workshop, project documentation, and collaborative platforms. Regular coordination with WP2 is essential for leveraging their expertise in organizational and facilitation support.
Detailed timeline until Pilot end	1. **Preparation Phase** **Q1 2024 [Design Phase]:** • Develop a Pilot and assessment plan, • Draft a preliminary set of requirements, • Specify and Setup the preparatory work, • Collect and Prepare Data Collections: Systematic gathering and organization of data, setting the groundwork for subsequent analysis. 2. **Development and Testing [Analysis Phase]:** **Q2-Q3 2024 (Pilot Milestone 1)**: • Specify the functional and technical requirements of the Monitoring Dashboard, • Run the first round of Data Analysis: Start of in-depth examination of sampled data, including research artefact analysis and evaluation, • Draft and estimate a first set of Reusability Indicators, • Deploy the first release of the Dashboard with core functionalities on OpenAIRE. 3. **Stakeholder Engagement [1st Phase]:** **Q3-Q4 2024 (Pilot Milestone 2)**: • Prepare webinar materials for the 1st Deployment Dashboard release, • Begin preparations for the Webinar, • Conduct **Webinar** to demonstrate to the stakeholders the core functionalities, • Analyse data from the initial deployment and qualitative feedback. 4. **Development and Testing [Update Phase]** **Q1-Q2 2025 (Pilot Milestone 3)**: • Update the functional and technical requirements of the Monitoring Dashboard,

	• Run Data Analysis: Start of in-depth examination of all data collections, including research artefact analysis and evaluation, • Compile the KPI methodology and estimate the final set of Reusability Indicators, • Finalize iterative developments and enhancements, • Further Refine the Dashboard based on initial implementation feedback, focusing on enhancement and optimization, • Deploy the final release of the Dashboard with all functionalities on OpenAIRE. 5. **Stakeholder Engagement [2nd Phase]** **Q2-Q3 2025 (Pilot Milestone 4)**: • Prepare Workshop materials for the 2nd Deployment Dashboard release, • Begin preparations for the Workshop, • Conduct **Workshop** to demonstrate to the stakeholders the Dashboard functionalities, • Continue engaging stakeholders for ongoing support and qualitative feedback. **6. Assessment and Reporting** **Q2-Q3 2025**: • Begin the assessment phase, evaluating the overall user experience of the Dashboard and its effectiveness in tracking and monitoring reusability of research artefacts, • Conduct internal reviews and assessments, identifying areas for further improvement, • Present preliminary findings at relevant project meetings. • Plan for knowledge-sharing and dissemination activities. 7. **Dissemination and Pilot conclusion** **Q2-Q4 2025**: • Engage in knowledge-sharing activities, presenting the Pilot's outcomes at relevant conferences or workshops. • Disseminate findings and Pilot results through project newsletters, the project website, and targeted communication channels, • Pilot Conclusion: Official closing of the Pilot, with potential identification of next steps based on the Pilot's outcomes.
Evaluation / Implications of the Pilot (i.e., what effects do we want to examine, what are the confounding factors, etc.)	
1) Domain Coverage	1) Computer Sciences/Artificial Intelligence (AI) The Pilot focuses on AI, a domain where the reusability of data, models, and tools is paramount due to rapid advancements and ethical considerations. We aim to enhance the tracking of AI research artifacts' impact, fostering progress and responsible innovation in the field.
Evaluation Plan	1) Webinar (Q3-Q4 2024): • **Method**: Conduct a webinar to demonstrate the core functionalities of the OpenAIRE Dashboard. Include a hands-on session for participants to test these

	functionalities. Afterwards, distribute a survey to collect initial feedback. • **Analysis**: Qualitative analysis of feedback received after the webinar to understand user experiences and perceptions. 2) Workshop (Q2-Q3 2025): • **Method**: Conduct a workshop where participants are guided through the OpenAIRE Dashboard to test the improved functionalities. Afterwards, distribute a survey to collect final feedback. • **Analysis**: Qualitative analysis of final feedback, capturing suggestions for further enhancements. 3) Quantitative Data Analysis (2024-2025): • **Method**: Continuously collect quantitative data on tool performance, • **Analysis**: Ongoing quantitative analysis of Reusability analysis and metrics. 4) Final webinar + Satisfaction Survey (Q3-Q4 2025): a. **Method:** Conduct a webinar to present the final version of the tool. Include presentations from at least one RPO and one funder to illustrate their dashboards. Afterward, distribute a satisfaction survey to measure KPIs. b. **Analysis**: Qualitative analysis of the User Satisfaction survey results.
Evaluation Methods	Success within the Pilot will be gauged through a blend of formal and informal methods: • Quantitative Analysis: Utilizing objective metrics such as the reuse indicators of research artefacts to evaluate the practical impact of the tools developed. • User Surveys and/or Interviews: Conducting structured surveys and interviews with the funder communities to gather feedback on the usability, effectiveness, and overall satisfaction with the dashboard and associated tools. • Performance Metrics: Monitoring specific performance metrics derived from the dashboard's KPIs to assess the success of the Pilot in enhancing the tracking of AI research artefacts.
Evaluation Metrics	1. Reusability Indicators: • **Metric**: The percentage of Research artefacts that can be successfully tagged using the tools. This metric assesses the tools impact on quantifying reusability in comparison to existing approaches. 2. User Satisfaction Scores: • **Metric**: Scores obtained from user assessment, reflecting user satisfaction with OpenAIRE Dashboard's features, usability, and overall effectiveness.

	• **Improvement Criteria**: Positive trends in user satisfaction scores, indicating improvements in user experience and satisfaction over the course of the Pilot.
Key Results /Outcomes (expected outcome)	1. Improved Reusability: Expected Outcome: The OpenAIRE dashboard streamlines and facilitates a significant increase in the reusability rate of research artefacts in computer sciences and artificial intelligence. This outcome directly aligns with the Pilot's objective to customize and evaluate tools that enable funders to monitor their funded projects along their Open Science policies. 2. Positive User Satisfaction and Feedback: Expected Outcome: Continuous improvement in user satisfaction scores and positive feedback throughout the Pilot. Positive user experiences indicate that the Dashboard with its indicators effectively meets the needs of stakeholders, fostering a user-friendly environment for monitoring reproducibility of funded projects. 3. Adherence to Reproducibility Best Practices: Expected Outcome: Successful integration of reproducibility best practices, such as the adoption of FAIR principles and documentation guidelines for research artefacts, contributing to the tool's credibility and reliability in the research community.
Key Performance Indicators (KPIs)	1. KPI1_Reusability_Analysis Rate: Percentage increase in the successful automated annotation of research artefacts using SCINOBO compared to baseline measurements or existing tools. 2. KPI2_Adoption Rate: Rate of adoption measured by the number of stakeholders' representatives. 3. KPI3_User Satisfaction Scores: Average satisfaction scores obtained from user assessment at different stages of the Pilot.

Pilot activities timeline	**Timeline**
Pilot implementation and assessment plan template shared with timeline till M18	26th Oct 2023
Pilot plan using the Pilot implementation and assessment plan template	14th Nov 2023
Literature review / collecting evidence on existing tools and practices	Dec 2023
Piloting progress and preassessment of user studies and/or KPIs	Jan 2024
Pilot Pre-registration	Jan 2024
First Pilot test with stakeholders	Feb 2024
Reporting preliminary results	Mar 2024
Pilot presentation and documentation	May 2024
Second Pilot test (in 3 years, needed for funders)	

References

[1] J. Heddes, P. Meerdink, M. Pieters, and M. Marx, “The Automatic Detection of Dataset Names in Scientific Articles,” *Data*, vol. 6, no. 8, p. 84, Aug. 2021, doi: 10.3390/data6080084.

[2] D. Schindler, F. Bensmann, S. Dietze, and F. Krüger, “SoMeSci- A 5 Star Open Data Gold Standard Knowledge Graph of Software Mentions in Scientific Articles,” in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, Virtual Event Queensland Australia: ACM, Oct. 2021, pp. 4574–4583. doi: 10.1145/3459637.3482017.

[3] C. Du, J. Cohoon, P. Lopez, and J. Howison, “Softcite dataset: A dataset of software mentions in biomedical and economic research publications,” *Asso for Info Science & Tech*, vol. 72, no. 7, pp. 870–884, Jul. 2021, doi: 10.1002/asi.24454.

[4] S. Jain, M. Van Zuylen, H. Hajishirzi, and I. Beltagy, “SciREX: A Challenge Dataset for Document-Level Information Extraction,” in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online: Association for Computational Linguistics, 2020, pp. 7506–7516. doi: 10.18653/v1/2020.acl-main.670.

[5] Y. Luan, L. He, M. Ostendorf, and H. Hajishirzi, “Multi-Task Identification of Entities, Relations, and Coreference for Scientific Knowledge Graph Construction,” in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium: Association for Computational Linguistics, 2018, pp. 3219–3232. doi: 10.18653/v1/D18-1360.

[6] D. Schindler, B. Zapilko, and F. Krüger, “Investigating Software Usage in the Social Sciences: A Knowledge Graph Approach,” in *The Semantic Web*, vol. 12123, A. Harth, S. Kirrane, A.-C. Ngonga Ngomo, H. Paulheim, A. Rula, A. L. Gentile, P. Haase, and M. Cochez, Eds., in Lecture Notes in Computer Science, vol. 12123. , Cham: Springer International Publishing, 2020, pp. 271–286. doi: 10.1007/978-3-030-49461-2_16.

[7] H. Pan, Q. Zhang, E. Dragut, C. Caragea, and L. J. Latecki, “DMDD: A Large-Scale Dataset for Dataset Mentions Detection,” *Transactions of the Association for Computational Linguistics*, vol. 11, pp. 1132–1146, Sep. 2023, doi: 10.1162/tacl_a_00592.

[8] G. Duck, G. Nenadic, A. Brass, D. L. Robertson, and R. Stevens, “bioNerDS: exploring bioinformatics’ database and software use through literature mining,” *BMC Bioinformatics*, vol. 14, no. 1, p. 194, Dec. 2013, doi: 10.1186/1471-2105-14-194.

[9] I. Beltagy, K. Lo, and A. Cohan, “SciBERT: A Pretrained Language Model for Scientific Text,” in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, China: Association for Computational Linguistics, 2019, pp. 3613–3618. doi: 10.18653/v1/D19-1371.

3.7.Pilot 7 - Editorial Workflows to Increase Data Sharing

	Pilot	ADDITIONAL DETAILS
Pilot Number & Title (Institution / Focal Person)	Pilot 7 – Editorial Workflows to Increase Data Sharing KNOW / Thomas Klebel & Tony Ross-Hellauer	
Short description	Research data sharing is an important building block for reproducibility and transparency, but current rates of data sharing remain low. To improve on the status quo, we will conduct two activities within this Pilot: (1) **Intervention study:** We will conduct a randomised controlled trial of an intervention targeting data availability statements (DAS). Currently, DAS often contain generic language stating that data is “available upon request", rather than providing direct ways to obtain the data. We will Pilot an intervention aiming to increase the rate of shared data. (2) **Delphi study:** In addition, initial exchanges with our stakeholders have unveiled a need to explore current issues around data sharing and potential solutions in greater detail. To facilitate this process, we will conduct a Delphi study to gather consensus on the most pressing issues and best paths to improve the status quo.	
Objectives	Two objectives listed below are structured by the two sub-studies. Objectives for Intervention study: 1. Assess the feasibility of implementing interventions in parallel to peer-review aiming to change author behaviour 2. Generate evidence on the efficacy of inexpensive interventions in the editorial process to increase levels of data sharing Objectives for Delphi study: 1. Develop, via a Delphi study with relevant stakeholders, a consensus on (1) the main issues impeding greater sharing of data (as well as code and other materials) associated with publications and (2) priority actions to improve this situation 2. Disseminate these findings to facilitate coherent and collaborative understanding and action amongst publishers	
Current literature/state of play	As of 2023, journals increasingly require authors to submit a data availability statement (DAS). PLOS ONE was one of the first journals to do so, and multiple studies have analysed the	

	effectiveness of the policy. Federer et al. [1] show that most authors comply with providing a data availability statement. Yet, the rate of DASs which include a direct link to the data is low. Similar findings have been made for preprints posted to medRxiv [2], and for articles published by Wiley [3]. The study by Graf et al. [3] particularly highlights the ubiquity of statements such as "data available upon (reasonable) request". Based on the available literature, policies mandating data availability can increase actual data availability [4]. Yet there is still room for improvement, especially given that overall rates of data sharing remain rather low [5]. A workshop held by the Data policy standardisation and implementation IG in February 2023 developed ideas for interventions. The targeted intervention (feedback mechanism to push back on DAS that could be improved) is one of the ideas that were discussed in the IG's workshop.
Existing tools related to the Pilot	F1000 has a checklist they apply when assessing submissions. The checklist includes the following items: • Is the repository name provided? • Is the title of the project where the data is included provided? • Has the dataset been given a DOI, is it provided with the full URL? • Are the data files contained in the project listed and is it clear what they contain? • Is the license included? • Is there a citation included? Other publishers might have similar checklists/practices, but we don't know this at this time. This will be worked out further by directly discussing with the publishers involved in the Pilot.
Overview of existing tools/resources (software/platforms/instruments)	We build on work from various groups at the Research Data Alliance, as well as our workshop with publishers earlier this year. Main inputs come from the Data policy standardisation and implementation IG, with minutes from two meetings in early 2023 (one, two)
Methods used for piloting (methods used for tools or practices)	In this section, we detail methods for three aspects: 1. The development of the intervention 2. The assessment of the intervention via a randomised controlled trial 3. The Delphi Study Development of the intervention The intervention will consist of a single page document, motivating researchers to share data and providing guidance on how to share data. To develop the intervention, we will heavily rely on our co-creation community (the publisher representatives). We will prepare a first draft of the intervention, which will be subsequently discussed

<table>
<tr><td></td><td>and further developed through feedback from the publisher representatives.

<u>Assessment of the intervention via RCT</u>
The intervention's efficacy will be evaluated with a randomised controlled trial. The full methodology for this trial is being developed in this document. The finalised protocol will be preregistered in June 2024.

<u>Delphi Study</u>
• Initial interactive workshop to gather suggestions for main challenges and solutions related to DAS
• 2 rounds of anonymised survey, distributed to group members and editorial staff/journal editors of journals (ensuring disciplinary coverage)
○ Round 1: Participants assess suggestions for challenges/solutions, give suggestions for revisions or extensions
○ Round 2: Participants reassess revised suggestions (based on feedback from last round) towards consensus
○ Round 3: Optional final round to resolve any further conflicts to achieve optimal consensus</td></tr>
<tr><td>Stakeholder groups affected and/or included</td><td><table><tr><th>Stakeholder Group (role: affected/included/beneficiaries)</th><th>details (expected outcome for the stakeholder)</th></tr><tr><td>Publishers (included)</td><td>Publishers will benefit from the Pilot in two ways: The intervention study will generate evidence as to whether the tested intervention is effective and feasible. The Delphi-Study will clarify issues and priorities, which will enable publishers to take effective action at improving rates of data sharing.</td></tr><tr><td>Researchers (affected)</td><td>Researchers will be subject to the intervention. We expect some additional effort required by researchers in case they change their data sharing practices due to the intervention.</td></tr></table></td></tr>
<tr><td>Stakeholder engagement plan</td><td>We build on personal networks to publishers, and the contacts to the RDA interest and working groups. As we will take a targeted approach, we don't require support from WP2 to disseminate more widely. We will work with specific publishers and the editors active at their journals.

Further details on the exact interaction with publishers during the RCT are described in the study protocol, sections 4.2, 4.5, 4.7.</td></tr>
</table>

Detailed timeline until Pilot end	- Buy-in from 1+ publishers to work on a joint intervention (Dec 2023) - First draft of workflow (end Feb 2024) - Feedback session with publisher(s) (March 2024) - 1-2 Workshops with publishers (April 2024) - Finalisation of workflow (May 2024) - Sample selection for implementation (June 2024) - Preregistration of intervention methodology (30. June 2024) - Intervention start (1. July 2024) - Intervention end (31. March 2025) - Post-intervention survey among editors (April 2025) - Data collection complete (Sept 2025) - Data analysis complete (Nov 2025) - Publication ready (Dec 2025)
Evaluation / Implications of the Pilot (i.e., what effects do we want to examine, what are the confounding factors, etc.)	
Domain Coverage	Domain coverage for the RCT will be determined by the participating journals. This will be finalised by the end of June 2024. The Delphi-Study will involve representatives from a wide range of publishers and journals, which will result in broad coverage across domains.
Evaluation Plan	The evaluation of the intervention will be conducted with a randomised controlled trial. Submissions at journals will be either allocated to the intervention or control arm. Authors of manuscripts in the intervention arm will receive the guidance document alongside the peer review reports. Authors of manuscripts in the control group will receive normal peer review. The length of the intervention will be determined by two factors: either by reaching the target sample size, or by reaching the specified date (31. March 2025). The cutoff is necessary to ensure that we can analyse the data within the scope of the project. Further details on the evaluation are provided in the study protocol.
Evaluation Methods	See above and the study protocol.
Evaluation Metrics	The RCT will measure four outcomes: 1. Primary outcome: The primary outcome will be difference of the % of Data Availability Statements that contain a working link to a trusted repository between the two intervention arms. 2. Secondary (final) outcome: % of DAS that state “data available on request” or similar. 3. Secondary intermediate outcomes: a. Time from initial submission to publication, measured in days. b. Time from first decision (requiring minor revisions, major revisions, or revise and resubmit) until resubmission. Details on the exact procedures to evaluate the metrics are detailed in the study protocol.

Key Results /Outcomes (expected outcome)	The randomised controlled trial will improve our understanding of which interventions show promise to improve rates of data sharing. We assume that the intervention will lead to higher rates of data sharing and higher rates of data deposited in trusted repositories. Data availability is a clear precondition to computational reproducibility. Rolling out the intervention across multiple journals might have a substantial effect on data availability and thus reproducibility. In addition, the Delphi study will build consensus and gather momentum for alignment among publishers on where to focus efforts to improve the current state of data sharing.
Key Performance Indicators (KPIs)	Intervention study: Participating journals: at least 1. Randomised and assessed manuscripts: More than 200. Delphi study: At least 50 Delphi participants throughout the process One workshop and min. 2 rounds of anonymised survey Dissemination: min 5 publishers publicise the outcomes via their own channels

Pilot activities timeline	**Timeline**
Pilot implementation and assessment plan template shared with timeline till M18	26th Oct 2023
Pilot plan using the Pilot implementation and assessment plan template	14th Nov 2023
Literature review / collecting evidence on existing tools and practices	Dec 2023
Piloting progress and preassessment of user studies and/or KPIs	Jan 2024
Pilot Pre-registration	Jan 2024
First Pilot test with stakeholders	March-April 2024
Reporting preliminary results	May 2024
Pilot presentation and documentation	May 2024
Second Pilot test (in 3 years, needed for funders)	

References

[1] L. M. Federer *et al.*, 'Data sharing in PLOS ONE: An analysis of Data Availability Statements', *PLOS ONE*, vol. 13, no. 5, p. e0194768, May 2018, doi: 10.1371/journal.pone.0194768.

[2] L. A. McGuinness and A. L. Sheppard, 'A descriptive analysis of the data availability statements accompanying medRxiv preprints and a comparison with their published counterparts', *PLOS ONE*, vol. 16, no. 5, p. e0250887, May 2021, doi: 10.1371/journal.pone.0250887.

[3] C. Graf, D. Flanagan, L. Wylie, and D. Silver, ‘The Open Data Challenge: An Analysis of 124,000 Data Availability Statements and an Ironic Lesson about Data Management Plans’, *Data Intell.*, vol. 2, no. 4, pp. 554–568, Oct. 2020, doi: 10.1162/dint_a_00061.
[4] T. E. Hardwicke *et al.*, ‘Data availability, reusability, and analytic reproducibility: evaluating the impact of a mandatory open data policy at the journal Cognition’, *R. Soc. Open Sci.*, vol. 5, no. 8, p. 180448, 2018, doi: 10.1098/rsos.180448.
[5] S. Serghiou, D. G. Contopoulos-Ioannidis, K. W. Boyack, N. Riedel, J. D. Wallach, and J. P. A. Ioannidis, ‘Assessment of transparency indicators across the biomedical literature: How open is open?’, *PLOS Biol.*, vol. 19, no. 3, p. e3001107, Mar. 2021, doi: 10.1371/journal.pbio.3001107.

3.8.Pilot 8 - An Editorial Reference Handbook for Reproducibility and FAIRness

	Pilot	ADDITIONAL DETAILS
Pilot Number & Title (Institution / Focal Person)	**Pilot 8 – An Editorial Reference Handbook for Reproducibility and FAIRness** Allyson Lister (UOXF) ***Co-Leads:*** Allyson Lister, Susanna-Assunta Sansone (UOXF) Rebecca Taylor-Grant and Matt Cannon (Taylor & Francis) ***TIER2 Coordination Group:*** Tony Ross-Hellauer and Thomas Klebel (KNOW) Liz Allen (F1000) Christopher Osborne (UOXF)	

Short description	• Co-creation and testing of an **Editorial Reference Handbook** that contributes towards a common understanding and what is required to assist **reproducibility** and **FAIRness**
	  ○ The handbook should **put the requirements of the journal data policy in action**
	  ▪ Journals that already have their own internal guidance will be able to use the handbook to validate and refine their existing methodology
	  ▪ Journals that do not yet have their own internal guidance should use it as an opportunity to define their own process
	• The planned intervention will target ***in-house editorial staff*** managing the manuscripts, but also benefit *reviewers*, *authors* and *service providers* by making the requirements transparent and understandable to them
	• Release the work in a FAIR manner, and publish it as a collaborative OA article	

Objectives	**Objective 1**: An **educational** and **practical set of checks** in support of reproducibility and FAIRness
	• Some journals have internal checks, but the type, richness and stringency vary, and there is little/no consensus among publishers
	  => Harmonise and operationalise core checks: **structured component** in the Handbook
	
	Objective 2: A **general framework** to help improve **internal processes**, where needed
	• There is a variety of internal processes, and how, when and by whom these checks are done vary, and this can also affect the results
	  => Define and describe an ideal process: **narrative component** in the Handbook	

	• The understanding of existing internal processes is also essential for the implementation of the guidance, and the success of the intervention
Current literature/state of play	There a number of related resources and efforts (e.g. [1], [2], [3]), and as part of the workshops we will collate them and reviewed them for relevance.
Existing tools related to the Pilot	n/a
Overview of existing tools/resources (software/platforms/instruments)	• We will leverage elements of existing work (as listed above) • We will also solicit sharing of similar internal documents or information from other publishers that will be involved in the Pilot. • We will use FAIRsharing: to point to standards, data resources and policies, as relevant
Methods used for piloting (methods used for tools or practices) **Describe the method that will be used for creation/improvements of tool/practices**	Details of the plan for each workstream, and the methods, are here below. **WORKSTREAM 1: Educational and practical set of checks** **Output** • List of checks with **definitions**, **values**, and **implementations** **Method** • Starting point "F1000 Guidelines on Data and Software for Editors" and checks focused on data sharing; now structuring a table adding other relevant material by the community, e.g. RDA-related work and resources such as PRO-MaP • Where relevant, we use FAIRsharing to signpost standards and data resources, and indicate which checks can be automatised and how **Benefit/ficiary** • Guidance to ***in-house editors*** managing the manuscripts - primary target audience • Advice to ***reviewers***, on what compliance to the journal data policy may require • Information for ***authors*** on what is expected from them by a number of journals • Requirements source for ***developers*** to drive their service provisions to publishers **WORKSTREAM 2: General framework for internal process** **Output**

<table>
<tr><td></td><td>

A framework for situating and discussing the checks that publishers conduct on manuscripts as part of the submission, peer review and publication process.

Method

Review and discuss a generalised workflow
Consider what is checked, when in the process, how it is communicated with the author(s), editors and reviewers.

Benefit/ficiary

Increased awareness of how other publishers are corresponding with authors and what/how they are checking; being mindful that we are <u>not</u> requiring any competitive information to be shared
A more consistent process for authors and increases understanding and potential for reproducibility

METHOD: for both workstreams

Run a series of online working sessions of 1 hr with Pilot’s members

Alternating the focus on the two workstreams

Sustain an iterative process to collect feedback and complete the workstreams
With offline work in between calls, via google docs/sheet an emails
Allow ample time for publishers’ internal review and approval prior intervention
Organize a dedicated online session for the intervention

Identify in-house editors in journals willing to participate, as they may differ from the member of this Pilots

</td></tr>
<tr><td>Stakeholder groups affected and/or included</td><td>
<table>
<tr><th>Stakeholder Group (role: affected/included/beneficiaries)</th><th>details (expected outcome for the stakeholder)</th></tr>
<tr><td>(primary) Editors (and potentially reviewers) (included)</td><td>Journals that already have their own internal process will be able to use the handbook to validate and refine their existing reproducibility methodology.

Journals that do not yet have their own internal process can implement the guidance in the handbook to create reproducibility guidance.</td></tr>
<tr><td>(secondary) Authors (beneficiaries)</td><td>Will use the handbook to understand what is expected from them, and that those expectations are required by a number of journals.</td></tr>
</table>
</td></tr>
</table>

	(secondary) Tools developers and service providers (beneficiaries)	Will leverage the handbook to identify requirements to drive their service provisions, particularly to publishers.
Stakeholder engagement plan	We will build on the first workshop we ran in May 2023 as well as the FAIRsharing and RDA networks. The next steps involve the editor stakeholder group and are as outlined in the 'Methods used for piloting' section above. Once the initial draft of the handbook has been created, additional iterative feedback can be provided by the author and developer/provider stakeholder groups to inform further drafts. We would like support from WP2 on the creation of our surveys.	

Detailed timeline until Pilot end	• Preparation and planning (Dec 2023 - Feb 2024)
	• Development and iterative feedback (Mar - July 2024)
	  ○ 1st online workshop with Pilot members, to discuss workstreams and timelines
	  ○ 2nd online workshop on workstream 1: checklist and related flowchart (Mar 2024)
	  ○ 3rd online workshop on workstream 2: general framework (Apr 2024)
	  ○ 4th online workshop on workstream 2: general framework (22 May 2024)
	  ○ 5th online workshop on workstream 1: checklist and related flowchart (Jun 2024)
	  ○ 6th online workshop on intervention, and presentation of the methodology (Jul 2024)
	• Draft Handbook delivered (July 2024)
	• Internal review and editors' identification (Jul - Sep 2024)
	  ○ Publishers collect internal feedback on handbook, and identifying willing editors for the intervention
	• Final version of the Handbook delivered (Oct 2024)
	  ○ Off-line revision of the checks and general framework (or online workshop, if needed)
	• Intervention and evaluation (Nov 2024)
	  ○ 7th online workshop on intervention and post-intervention method (Nov 2024)
	  ○ Intervention starts (Nov 2024) and end (Jun 2025)
	• Post-intervention survey among editors (Jul and Aug 2025)
	  ○ Data analysis complete, and presentation to final online workshop (Sep 2025)
	• Publication ready (Oct 2025) - TIER2 ends Dec 2025	

Evaluation / Implications of the Pilot (i.e., what effects do we want to examine, what are the confounding factors, etc.)		

Domain Coverage		
Evaluation Plan	Upon completion of the Editorial Handbook, an intervention will be staged and designed with a group of interested parties. This will measure the value of the handbook in terms of required effort/input as well as outcome quality/reproducibility.	
Evaluation Methods	We will measure success through a post-intervention survey of the involved editors. We will draw from the Behaviour Change Wheel framework to develop the handbook/intervention (Michie et al., 2014); evaluate (e.g., user experience and impact on reproducibility/FAIRness).	
Evaluation Metrics	Based on the qualitative results of the surveys, an **improved internal workflow** among editors, reviewers and authors about what is expected of them with regards to data reproducibility. This will measure: • Stakeholders' perceived improvements to the publishing workflow • Stakeholders' evaluation of the improved FAIRness of the produced • Stakeholders' satisfaction with the new workflow	
Key Results /Outcomes (expected outcome)	This work will create an Editorial Handbook that contributes towards a common understanding and what is required to assist **reproducibility** and **FAIRness.** The handbook should **put the requirements of the journal data policy in action.**	
Key Performance Indicators (KPIs)		

Pilot activities timeline	**Timeline**
Preparation and planning	Dec 2023 - Feb 2024
Development and iterative feedback 1st online workshop with Pilot members, to discuss workstreams and timelines 2nd online workshop on workstream 1: checklist and related flowchart 3rd online workshop on workstream 2: general framework 4th online workshop on workstream 2: general framework *5th online workshop on all outputs: checklist, flowchart, framework, handbook* *6th online workshop on intervention, and presentation of the methodology*	Mar - July 2024 Mar 2024 Apr 2024 22 May 2024 July 3, 2024 Jul or Sept 2024
Draft Handbook delivered	July 2024
Internal review and editors' identification Publishers collect internal feedback on handbook, and identifying willing editors for the intervention	Jul - Sep 2024 Oct 2024
Final version of the Handbook delivered	Oct 2024

Off-line revision of the checks and general framework (or online workshop, if needed)	
Intervention and evaluation 7th online workshop on intervention and post-intervention method	Nov - Jun 2024
Post-intervention survey among editors Data analysis complete, and presentation to final online workshop	Jul - Aug 2025 Sep 2025
Publication ready	Oct 2025

[1] M. Macleod *et al.*, 'The MDAR (Materials Design Analysis Reporting) Framework for transparent reporting in the life sciences', PNAS, April 23, 2021, 118 (17) e2103238118 https://doi.org/10.1073/pnas.2103238118

[2] S. B. Leite *et al.*, 'Promoting Reusable and Open Methods and Protocols (PRO-MaP): Draft recommendations to improve methodological clarity in life sciences publications', OSF, https://doi.org/10.31219/osf.io/x85gh

[3] A. Kenall, *et al.*, 'Better reporting for better research: a checklist for reproducibility, *GigaScience*, Volume 4, Issue 1, December 2015, s13742–015–0071–8, https://doi.org/10.1186/s13742-015-0071-8

4. Conclusion

Deliverable 4.2 marks a significant milestone in Task 4.2, focusing on enhancing research reproducibility across various methodologies and epistemic contexts. Throughout this deliverable, we have summarised several concrete steps that have been crucial to our progress.

We began by developing for each Pilot project, establishing a structured approach to our interventions. This was followed by drafting detailed protocols that outlined the technical and social requirements needed for the development and implementation of reproducibility tools.

We also implemented a rigorous review process, with each Pilot being reviewed by two team members. This provided valuable feedback and ensured the robustness and relevance of our interventions.

In addition, we organized Pilot workshops and bilateral discussions to align all partners and tailor the tools and practices to the specific needs of our diverse research communities. These activities have been vital in refining our approaches and ensuring that our Pilots are well-prepared for implementation.

In summary, the steps we have taken—preregistration, protocol development, ethical review, a comprehensive review process, and collaborative workshops—have laid a strong foundation for the successful advancement of Task 4.2. These efforts underscore our commitment to enhancing research reproducibility and set the stage for the continued success of our project.

Looking ahead, we anticipate continued collaboration and progress in Task 4.3, which involves the preparation activities for Pilots. These activities involve meticulous planning to anticipate and address potential issues, including stakeholder familiarity with technologies and unforeseen project challenges. These steps are crucial for ensuring the successful planning, execution, and assessment of the Pilot projects.

To maintain and enhance engagement with our stakeholder communities, we will continue to leverage a co-design approach. This involves regular communication and collaboration with stakeholders to ensure their needs and feedback are integral to the development process. By organizing workshops, bilateral meetings, and continuous updates, we ensure stakeholders remain actively involved and their inputs shape the outcomes. We will compile a list of performance indicators with stakeholder collaboration to unify the assessment of tools per research type and group.

Furthermore, Task 4.4 will see the execution of Pilot actions and evaluation activities. This task involves showcasing and applying reproducibility-related tools and practices across various stakeholder communities, including researchers, publishers, and funders. Through diligent assessment methodologies, we seek insights into the effectiveness of these tools and practices across diverse epistemic contexts, ultimately aiming to successfully implement them within the research community.

In summary, Deliverable 4.2 sets the stage for the continued advancement of our project objectives, emphasizing collaboration, adaptability, and responsiveness to stakeholder needs. Through our collective efforts, we remain committed to enhancing research reproducibility and fostering positive impacts within the broader research community.