

IDCC27 Accepted Paper Abstracts

A Playbook for Centring Community Partners in the Curation of Non-Research Data
As partnerships between universities, community-based organizations, and local governments expand, the curation of non-research data for research purposes has grown significantly. While invaluable for understanding community needs, data produced through administrative processes is rarely analysis-ready and its curation is often treated as a technical task isolated from the needs of data producers. This contribution introduces a newly developed playbook that reframes administrative data curation as a participatory process designed to centre people and build sustainable organizational capacity for data stewardship.

Developed through partnership between a university library and a state-level social services agency in the United States, the playbook reflects insights from curating decades of historical child development data alongside six staff-led focus groups with frontline clinicians. Centring the human infrastructure of stewardship, the playbook adopts a ""braided knowledge"" framework. Instead of treating data producers as passive depositors, it maps curation activities along two interwoven axes where the tasks of the university curator and the community data provider are mutually dependent.

By adjusting the curation activities from the Data Curation Network's CURATE(D) model to fit this participatory lens, the playbook shifts the focus from curation as a means to an end of usable data toward equally important outcomes related to organizational data culture, the recognition of staff expertise, and the establishment of independent, long-term stewardship capacity. This approach offers data curators a roadmap for supporting researchers in engaging more equitably with the data stewardship challenges of their community partners.

A Tale of Two FAIRnesses: Science Diplomacy for FAIR DOs and FAIR Doers
FAIR Digital Objects (FAIR DOs) are designed to make research outputs findable, accessible, interoperable, and reusable. Yet these properties do not arise from the objects themselves but, rather, depend on FAIR doers: the people who plan data management, create metadata, maintain repositories, interpret standards, preserve digital content, and make decisions about access and reuse. While FAIR DOs are expected to meet common technical requirements, FAIR doers operate within non-uniform national, institutional, and disciplinary conditions.



This paper examines the gap between the uniformity expected of FAIR DOs and the different conditions in which FAIR doers undertake stewardship and curation. It argues that science diplomacy has a key role in bridging this gap by extending international cooperation from the circulation of shared principles to the governance of their implementation. This requires attention to who contributes to standard-setting, whose expertise informs infrastructure development, how implementation work is supported, and how communities retain agency over the DOs they steward.

The paper draws on professional experience with digital research infrastructures, insights from the FIDELIS project, and the ISC–INGSA programme on science advice to policy. These perspectives connect differences in implementation capacity within Europe with historical asymmetries between and within the Global North and Global South. AI-assisted metadata generation acts as a stress test, exacerbating some existing differences between FAIR doers.

The paper proposes a FAIR-doer framework pivoting around four areas: participation in standard-setting, institutional capacity, support for curation work, and community authority and stewardship. The framework provides a lens to analyse the gap and identify where action is needed. By bringing FAIR DOs and FAIR doers within the same analytical frame, the paper considers how science advice can bring implementation conditions into policy and how science diplomacy can support international interoperability while keeping in view the people, labour, and institutions on which it depends.

Academic Library DMP Support in Transition: A US-Based Snapshot

Academic research libraries, as central providers of research data management (RDM) services throughout the research lifecycle, have been providing support for researchers in creating data management and sharing plans (DMSPs) for more than 15 years. However, in the US, the DMSP landscape is shifting quickly. First, sweeping changes from the largest federal funding agencies expectations have resulted in new DMSP requirements, including short-form DMSPs. These are marked diversions from previous DMSP narrative formats, and are therefore likely to change how librarians provide DMSP consultations moving forward. Secondly, the emergence of Artificial Intelligence (AI) has raised questions about AI-ready data, AI-generated DMSPs, and related repository and policy uncertainties. Combined, these provided significant motivation to capture a baseline snapshot of academic library DMSP support and better understand the impacts of these changes in the RDM ecosystem.

The authors conducted a cross-sectional survey in 2026 of DMSP practices before recent US federal policy changes took full effect. The survey asked respondents about their current practices along with their libraries' anticipated changes to RDM services. In particular, this paper will demonstrate: recent changes in the volume of DMSP consultations, implications of AI on RDM services, and the shift in library RDM topical focuses as funders, publishers, and disciplinary norms continue to evolve. Findings suggest a redirection of library RDM effort rather than a reduction of it. In other words, RDM will continue to be essential for academic libraries, though the shift in emphasis may change from supporting DMSP consultations to broader conversations about the access and sharing of research data.

AI Workflow Provenance Records in Virtual Research Environments for Reproducible AI-Assisted Research

AI-assisted research is difficult to reproduce when only inputs and final outputs are preserved. In lightweight workflows using pretrained models, results are shaped not by the model alone, but by human–AI interactions and workflow context, including task framing, prompts, category sets, reference examples, generated code, deployment settings, score interpretation, and human review. This work proposes an AI Workflow Provenance Record (AIWPR), a lightweight provenance layer for documenting how AI-assisted workflows are constructed and how their outputs are produced, interpreted, reviewed, and reused. We demonstrate AIWPR through a cultural heritage image-curation pilot in a virtual research environment, where Codex Agent assists in constructing a Python/OpenCLIP/Gradio workflow and OpenCLIP suggests iconographic labels for Buddhist statue images under zero-shot and few-shot settings. Rather than treating these labels as final metadata, AIWPR links each suggestion to construction traces, model, prompt/reference settings, execution and deployment context, score semantics, and review status. The case shows that preserving human–AI interaction history and workflow context makes AI-assisted research more reproducible, reviewable, and reusable.

Auditing Structural Missingness Before AI Reuse: A Pilot Application of Data Construction Audit to Paternal AIDS Diagnosis & Mortality Documentation

This paper introduces the Data Construction Audit (DCA) as a novel methodological approach for identifying and analysing structural missingness in data before it enters AI-enabled reuse contexts. Rather than treating missing data as an issue within completed datasets, DCA shifts attention upstream to the documentation systems (forms, policies, coding rules, metadata schemas, and reporting infrastructures) that determine what can become data in the first place. The method is demonstrated through a pilot study examining paternal AIDS diagnosis and mortality documentation in Cook County, Illinois

(1985–1995), a case that highlights how kinship-related information was inconsistently captured or entirely absent.

The study investigates whether documentation practices created durable opportunities to record paternal status, parent–child relationships, informant roles, and next-of-kin structures for men who died of AIDS. Using a structured audit of documentation artifacts, the pilot applies three core instruments. Together, these tools trace how key relational fields are enabled, constrained, or lost across data construction pathways.

DCA contributes methodologically by linking observed data gaps to system design features rather than treating absence as incidental. This distinction is critical for AI governance, where computational systems often interpret available data as complete representations of reality. Structural absences, if undocumented, can be reproduced or amplified in machine learning and historical analysis applications.

The paper argues that FAIR data principles must address not only the usability of existing data but also the preconditions that shape what is captured. By making absence auditable, DCA supports digital curation, metadata practice, and responsible AI by documenting the provenance of omission. Ultimately, the approach offers a transferable framework for centring people in data stewardship by recognizing that what is missing is itself a critical part of the record.

Beyond Dataset-Centred RDM: Evidence for the Evolving Role of Data Stewards in AI-Enabled Engineering Research

Artificial intelligence (AI) is often treated as a new, discrete governance issue. In engineering research, however, it enters workflows already shaped by software, external data, privacy, intellectual property and access restrictions. This study asks how AI/ML became visible in engineering Data Management Plans (DMPs) between 2023 and 2026, what research objects and stewardship requirements accompanied it, and what these patterns imply for faculty-embedded data stewardship in engineering.

We conducted a structured content analysis of 145 sampled DMPs from computer science (CS), electrical engineering (EE), mathematics (MATH) and interdisciplinary research (spanning CS, EE and MATH or application-oriented collaborations with external domains). We coded documented AI roles, primary and supporting outputs, data sources, personal data, software production, confidentiality/IP restrictions and sharing status/willingness.

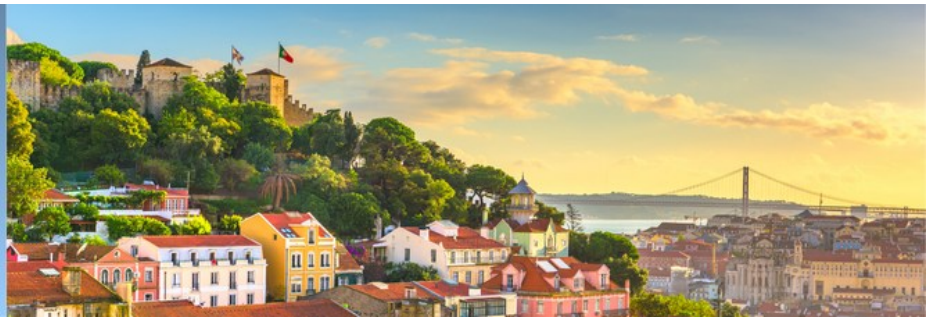
AI/ML was identified in 1 of 35 DMPs (2.9%) in 2023 and 11 of 34 (32.4%) in 2026; this upward pattern remained after disciplinary sensitivity analyses. Across the 23 AI-related DMPs, AI appeared as a research object in 13, a research method in 9 and in multiple roles in 1. AI-related projects were more often software-producing than non-AI projects (87.0% versus 58.2%) and more often combined at least two established stewardship domains—software, external data, personal data and confidentiality/IP restrictions (78.3% versus 53.3%). More broadly, 62.8% of all projects produced software, but software was a supporting rather than primary output in two-thirds of these cases.

The findings suggest that AI does not simply create a separate category of "AI data"; it intensifies an existing shift towards heterogeneous digital objects and compound governance requirements. We argue that people-centred, faculty-embedded stewardship in engineering should move beyond dataset-focused planning towards research-object mapping, early governance triage and coordination across software, privacy, legal, security, ethics and infrastructure expertise.

Beyond Frameworks: Sustaining a Data Culture for FAIR Marine Data Stewardship - Building on the Foundations of Framework-Led Data Culture Implementation
This paper examines the evolution of the FAIR (Findable, Accessible, Interoperable, Reusable) data practices within the Marine Institute (MI), Ireland, highlighting the transition from framework implementation to the cultivation of a sustained organizational data culture. Over the past decade, advances in digital curation, supported by repositories, metadata standards, and frameworks, have strengthened the technical foundations of data stewardship. The MI's adoption of internationally recognised frameworks, including the IOC-IODE accredited Data Management Quality Management Framework (DM-QMF) and CoreTrustSeal-certified Digital Preservation Framework, has enabled consistent, auditable, and interoperable management of complex marine data central to Marine Spatial Planning.

However, this study demonstrates that technical frameworks alone are insufficient to realize the full value of FAIR data. Building on earlier work, the paper explores how the MI has embedded data stewardship into organizational practices by strengthening governance, integrating data management into program delivery, and prioritizing staff engagement, training, and support. Efforts to enhance interoperability, improve metadata standards, and link curated data to analytical outputs have further reinforced the value of FAIR data in policy and operational contexts.

Our case study emphasizes the importance of human-centered approaches, highlighting the roles of collaboration, communication, and skills development in sustaining data



culture. Initiatives such as SONRAI The Irish Data Stewardship Network and evolving microcredential training opportunities, illustrate the importance of community and capacity building. This work advances the original contribution by demonstrating how FAIR principles can be sustained in practice through the ongoing alignment of people, processes, and organizational change. The findings show that sustainable FAIR ecosystems depend on aligning frameworks with organizational culture, recognizing data stewardship as a shared responsibility, and investing in the people who enable data to be trusted, reusable, and impactful.

Beyond Principles: The Institutional Realities of Academic Data Stewardship
Research data management (RDM) and Open Science have become central components of contemporary research infrastructures. While substantial attention has been devoted to policies, repositories, and technical solutions, considerably less is known about the institutional realities of implementing these principles through embedded data stewardship.

This paper reflects on four years of experience building and scaling a university-wide data steward network at the University of Vienna. Beginning with a small pilot of embedded data stewards and growing into a multidisciplinary network, we examine a deceptively simple mandate frequently given to data stewards: “Make RDM better.”

Drawing on practical experience across diverse disciplinary environments, we organize our reflections around three interconnected questions. First, what does meaningful improvement in RDM and Open Science actually look like in practice? Second, how can data stewards encourage researchers to engage with support services in environments where compliance is rarely enforced and participation depends on trust, relevance, and perceived value? Third, what conditions enable data stewards themselves to succeed in a profession that remains relatively new, lacks established educational pathways, and requires a unique combination of technical knowledge, institutional awareness, and interpersonal skills?

Rather than presenting a normative blueprint, we offer a practice-based perspective on the social, organizational, and cultural dimensions of data stewardship. We argue that sustainable progress in RDM depends less on technical infrastructure alone and more on the ability of institutions to cultivate supportive environments, foster trusted relationships, and create realistic pathways for change. Our findings contribute practical lessons for institutions seeking to establish or expand data stewardship programs and highlight the human infrastructure required to make RDM successful.



Beyond the Seal: Certification as a Catalyst for Organisational Stewardship in Federated Infrastructures

Trustworthy digital repositories depend on more than technical infrastructure. Policies, governance, preservation planning, metadata, documentation and organisational coordination all rely on people working across technical, disciplinary and institutional boundaries. Yet much of this work remains largely invisible, particularly within federated research infrastructures where responsibility is distributed across curation, technical and governance teams.

This paper examines how repository certification can act not only as an external assessment of trustworthiness, but also as a catalyst for organisational stewardship. Drawing on three successive certification exercises within SWISSUbase, Switzerland's federated national research data platform for the social sciences, linguistics and humanities – including two CoreTrustSeal (CTS) certification cycles for FORS – the Swiss Centre of Expertise in the Social Sciences, one CTS certification for the Language Repository of Switzerland (LaRS), and the subsequent Common Language Resources and Technology Infrastructure (CLARIN) B-Centre certification – we reflect on how certification shaped organisational practices beyond compliance.

Rather than focusing on certification outcomes alone, we examine four recurring effects. First, certification increased the visibility of long-term stewardship, particularly digital preservation, within technical teams and governance bodies. Second, it created a shared operational understanding of roles, responsibilities and dependencies across organisational actors through collaborative documentation and community engagement. Third, it translated findings from multiple assessment frameworks into shared organisational priorities, a preservation roadmap and concrete development backlog items, making preservation activities more actionable. Finally, repeated certification fostered continuous organisational learning by embedding reflection and improvement within routine repository development.

We argue that the value of repository certification lies not only in demonstrating trustworthiness, but in making the often invisible work of organisational stewardship visible, actionable and sustainable – particularly in federated research infrastructures, where certification provides a common framework for coordinating stewardship priorities across technical, operational and governance communities.



Building Research Data Stewardship Capacity in East Africa: Experiences from the RDM Kwanza Project in Tanzania and Uganda

Research data management (RDM) and data stewardship are fundamental to research integrity, reproducibility, and open science. Despite increasing global emphasis on RDM, it remains an emerging practice in many East African universities, constrained by limited infrastructure, weak institutional policies, and insufficient capacity. The VLIRUOS-funded RDM Kwanza project addresses these challenges by training 16 librarians at 13 universities as institutional data stewards in Tanzania and Uganda.

A participatory evaluation approach was used to assess institutional RDM practices and the impact of capacity-building activities. Data were collected through three complementary methods: a baseline online survey of postgraduate students, academic staff, and research fellows at 13 universities (n = 893); an online survey of participating librarians evaluating knowledge gains and readiness to support RDM (n=16); and evaluations of two week-long in-person training workshops held in South Africa and Uganda (n=16).

Baseline survey findings showed that RDM is not yet institutionalised, with fragmented practices, limited infrastructure, inadequate policies, and insufficient researcher support. Training substantially improved librarians' understanding of RDM, data stewardship, FAIR principles, data management planning, and research support. Participants reported applying their knowledge to institutional policy development, repository planning, researcher training, advocacy, and awareness-raising activities. The project also stimulated institutional discussions on RDM, strengthened recognition of librarians' roles, and fostered regional collaboration. Key implementation challenges that remain are inadequate funding, limited technical infrastructure, lack of practical implementation experience, and incomplete institutional policy frameworks.

The RDM Kwanza project demonstrates that librarian-centred capacity building can catalyse institutional change and strengthen research data stewardship in East Africa. Sustainable implementation, however, requires continued investment in practical training, infrastructure, leadership commitment, and institutional policy development. The project provides a scalable model for advancing responsible RDM and open science across resource-constrained higher education institutions while supporting long-term regional collaboration and research excellence.



Centring Community Research Participants in Ethical Data Curation

Community-researcher partnerships are an increasingly popular research approach across disciplines: from humanities to health sciences. Yet while these partnerships can produce important research data, community participants are often left without control or ownership of their information and stories. Within recent years, efforts to move relationships between researchers and community groups from extractive to reciprocal have appeared in scholarly journals as well as social justice-oriented think tanks and consortiums (Boilevin et al., 2019; Data Equity Framework, 2021; Kidd 2018; Paris et al. 2022). However, we still lack practical approaches to more equitably centre research participants in data curation and data management within a given research relationship. As a set of co-determined governance policies that could be co-authored and shared across research networks, these best practices would go a long way in safeguarding community research participants and their data. In March of 2024, we organized a colloquium that led to the creation of a toolkit designed to support both community participants in research projects and the data created through these initiatives. We invited experts in the fields of data curation and management along with community organizations so that, together, we could isolate different stakeholders' needs and work to outline strategies to address them. It is crucial to include community groups in the development of data curation and management best practices given that, while they may not have dedicated staff to manage data curation themselves (often due to funding constraints), their voices are the ones that we most need to listen to. Our co-developed toolkit outlines a flexible, relational, and ethically nuanced approach to data management that centres data curation as an act of community care and control (MacQueen et al., 2001) through prioritizing reciprocity, accountability, and agency of the parties involved in research partnerships.

Centring People in FAIR Research Data Management: Building Human Capacity for Digital Stewardship Through Data Librarianship and Open Science

Even though the FAIR data principles have left their mark on the expectations for handling data in various fields, most attention to date has been paid to the technical infrastructure, metadata standards, and repository architecture that enable the implementation of these principles. What is called for in the paper is a new paradigm of people as the only driving force of the sustainable and equitable research data management (RDM). Based on the current practice in data librarianship and capacity development programmes for Open Science, we reflect on how the human element, relational trust and institutional culture determine if the principles of FAIR can be taken up into real research impact vs. policy statements.

We propose a conceptual model that connects three interrelated capacity-building components: (1) professional development for data librarians as boundary objects between technical data curation and disciplinary data use; (2) training for researchers that enables data stewardship confidence and data literacy in a way that does not overburden the researchers; and (3) institutional models, communities of practice, mentoring networks, and embedded data-seeking librarians that help sustain data use skills after an initial training. Human capacity investment in academic library RDM services is more effective in boosting the quality of data documentation, data reuse potential, and researcher uptake than will tooling alone, as illustrated by a few examples.

Equity aspects are also taken into account: If workforce development is not seen as part of basic infrastructure, but rather additional components, it is possible that a two-tier FAIR system prevails in the long term. We make practical suggestions for the conference for conference stakeholders, such as competency statements for data librarians, scalable train-on-the-job models, which can be adapted to a variety of disciplines and resources, and metrics for assessing human-centred RDM support beyond repository deposit counts.

Community Curation Across Languages: Strategies and Tools

Representing the experiences and languages of non-English speakers is an ongoing challenge for scholars and practitioners working in digital curation. For community archivists–the subfield in which we work–many of the highest regarded projects focusing on marginalized and underrepresented groups are presented in English. Some prominent examples include the South Asian American Digital Archive (SAADA) and the Plateau People's Web Portal. The invisibility of community work in the archive is only exacerbated by groups working in languages outside of English. This paper explores the process of co-creating an open-source platform and co-curating collections for Puerto Rican community groups, asking "How can we build a community archive that represents materials in multiple languages?", "How can we build a tool that allows other groups to build and maintain their own multilingual community archives?", and "How can we build an easy-to-use tool that allows our groups to have autonomy over their sites while maintaining connections to a central archive?" Using an approach we call participatory postcustodialism–which means that community contributors continue to participate in the curation activities after the initial hand-off of materials–we developed the Omeka S theme, "Multilingual." "Multilingual" offers new opportunities for community curation work across languages, opening the door to richer and more vibrant archiving praxis. While this paper focuses on the development of a multilingual template for community curation work utilizing Omeka, it also brings attention to the importance of multilingual curation projects

and demonstrates one way scholars can be better partners to non-English speaking communities.

Curating Personal Email as FAIR Digital Objects: A Metadata-Based Governance Approach

Born-digital personal archives are among the most significant yet fragile components of twenty-first-century documentary heritage. Among them, personal email collections are especially vulnerable, not only because they rely on technological infrastructures – messages often remain stored on service providers' servers and depend on authentication mechanisms that may become inaccessible after the creator's death – but also because their preservation and accessibility depend on human actors, including creators, heirs, and archivists.

This paper proposes EMaG (Email Metadata-based Governance) – a metadata-driven methodology that preserves the authenticity and integrity of personal email archives while enabling appraisal and computational analysis exclusively at the metadata level – which separates original informational objects (EML files and attachments), derived static representations (PDF), and structured management metadata (a machine-readable JSON manifest).

The methodology was developed and validated through the preservation of the born-digital archive of the Italian politician Guido Bodrato (1933–2023). The case study highlights the practical challenges of digital legacy, and underline the decisive role of timely human intervention and governance in ensuring long-term accessibility of personal digital records.

EMaG is implemented through the Epistula platform, which generates structured JSON manifests linking original EML files, access surrogates, attachments, descriptive metadata, and integrity information. The platform supports relational, temporal, and content-based analyses while maintaining a strict human-in-the-loop approach in which AI and natural language processing assist, rather than replace, archival appraisal. All preservation, appraisal, and analytical activities are recorded as metadata, leaving the original digital objects unchanged.

This workflow establishes the conditions for creating FAIR Digital Objects while remaining independent of any specific long-term preservation format. Its architecture is fully compatible with emerging standards such as EA-PDF, enabling future migration without redesigning the preservation model. The proposed approach offers a practical and scalable

framework for ensuring the long-term preservation, accessibility, and reuse of born-digital personal email archives.

Data Sharing Agreements and How to Get Them Signed Off in Less Than Eight Years: Findings and Future Directions from the ACCoRD Project

The process of drafting and agreeing data sharing contracts, enabling the use of third-party data in research projects, is notoriously fraught with difficulties and delays. We estimate that the UK Governments wastes millions of pounds of research funding every year as researchers struggle to get access to the data they need. Data contracting requires input from a variety of stakeholders who often work in silos and lack understanding of each other's priorities and concerns, which results in back-and-forth correspondence, avoidable delays, and contracts that are not always entirely clear to all parties.

The ACCoRD Project, funded by the UK's National Federated Compute Services (NFCS) NetworkPlus programme, has been working with the various stakeholder groups involved in drafting research data sharing agreements to identify pain points, good practice, and ways in which contracting can be improved, both at the national and the institutional level. This paper will describe the work of the project, its findings, and the recommendations that have resulted from it. The development of a research data contracting community has been one of the key objectives of ACCoRD, and details of how IDCC delegates can get involved in this community will be provided.

Development of an Artificial Intelligence Tool to Assist with Self-Deposit and Curation of Research Data in Digital Repositories

The project aims to present the development of an Artificial Intelligence (AI) agent designed to assist self-deposit and curation of research data in digital repositories designer in Dataverse software. The tool addresses the complexity of the manual curation process, which often discourages researchers or results in data without proper standardization. To mitigate these obstacles, the agent integrates analytical AI and generative AI technologies (based on large language models, or LLMs) to perform structural inspections, detect inconsistencies such as gaps or outliers, and suggest appropriate metadata for different fields of knowledge. In addition, the system’s rule engine will promote semantic enrichment by linking the data to controlled vocabularies and persistent identifiers, such as ORCID. With the exponential growth in the volume of data produced in scientific environments and the introduction of open science policies, data curation has become essential for the advancement of collaborative and interdisciplinary research. Digital data curation is indispensable for maintaining the quality, visibility, and usability of informational resources, allowing data to be effectively findable, accessible, interoperable,

and reusable (FAIR principles). Furthermore, curation involves recognizing the need to adapt to different disciplinary and institutional contexts, considering the ethical, legal, and technical specificities of each area of knowledge.

Exploring Possibilities for Primary Source Digital Objects in a FAIR+CARE Stewardship Framework

This paper contributes to FAIR+CARE conversations about cultural heritage data through practice-based work with born-digital and digitized primary sources as they move into shared digital infrastructure, aggregation platforms, and AI-assisted workflows. The talk begins from a concrete stewardship problem: archival and special collections materials are increasingly available as digital objects, but their research value as primary sources depends on evidentiary and relational conditions that do not necessarily travel with the file. Provenance, context, access practice, and human judgment are not ancillary to reuse; they are part of what allows these materials to remain intelligible as evidence.

FAIR+CARE are useful because they bring two pressures into view at once. FAIR sharpens the question of how primary-source digital objects become discoverable, accessible, interoperable, and reusable within the broader research ecosystem. CARE presses on what happens when that movement through systems is separated from responsibility, authority, ethics, and the people whose knowledge gives the materials meaning. Taken together, the frameworks help shift the question from whether archival materials can circulate as research objects to whether they can circulate without losing the conditions that make them accountable as primary sources.

The paper develops this argument through connected strands of practice research: metadata assessment, accessibility planning, contextual guidance concept testing, AI-generated alt text and image description experimentation with a large post-custodial digital collection, and field-facing engagement with repository practitioners and leaders. That field-facing dimension extends the paper beyond a single workflow or collection. It shows that FAIR+CARE stewardship is also a leadership and maintenance problem for a field with strong frameworks for responsible digital stewardship but uneven ways to resource the continuing work those frameworks imply.

From Ethical Principles to Research Priorities: A Future Agenda for Operationalising FATE in AI-Enabled Data Curation Within Academic Libraries

Artificial Intelligence (AI) is increasingly transforming data curation practices across academic libraries, introducing new opportunities for scalability, automation and enhanced data stewardship. However, these developments simultaneously raise complex

questions concerning fairness, accountability, transparency and ethics (FATE), particularly as institutions seek to integrate AI systems into research data infrastructures and curatorial workflows. While existing scholarship has advanced normative discussions surrounding responsible AI, comparatively limited attention has been devoted to identifying the research priorities necessary to translate ethical principles into sustainable institutional practice.

Building upon previous systematic evidence synthesis in the field of AI-enabled data curation, this paper proposes a structured research agenda to support the future operationalisation of FATE principles within academic libraries. Rather than introducing an additional ethical framework, the study addresses a critical implementation gap by examining what forms of evidence, governance mechanisms, operational artefacts and institutional capabilities remain underdeveloped.

The paper advances three contributions. First, it identifies key research domains required to progress ethical AI from conceptual aspiration to curatorial practice, including governance and accountability, curatorial infrastructures, evaluation and validation mechanisms and human-centred implementation. Second, it proposes a research roadmap designed to guide future empirical and interdisciplinary investigation. Third, it introduces a maturity matrix that situates emerging research according to conceptual development and operational readiness.

By reframing ethical data curation as an implementation challenge rather than solely a normative concern, this work contributes a forward-looking agenda for responsible AI adoption in academic libraries and positions data curation as a strategic site for institutional trust, stewardship and long-term digital sustainability.

From FAIR Data Creation to Long-Term Preservation: Centring the People That Make the Workflows Happen

After roughly a decade since the launch of the FAIR principles (Wilkinson et al., 2016) we have seen a significant increase in the number of datasets being published. At the same time, institutional policies around data management and data publication have been formulated and infrastructures that support these practices have matured in many research performing institutions. While doing so, the focus has remained on the research data life cycle during research, as well as on data publication and possible reuse during the minimal retention period, often set at 10 years. However, while some datasets need to be kept beyond the minimal retention period for further research or as digital heritage, some

can be deaccessioned after the retention period, and some data must be deaccessioned after a given period of time.

Appraisal and reappraisal are key activities necessary for the responsible stewardship of research data as the increasing volume of data produced, published and archived creates a heavy burden on research institutes and repositories storing and maintaining access to the data, with an increasing impact on institutional budgets. Good (re)appraisal practices are an important component of the curation process as they can guide current and future preservation activities and retention decisions.

From FAIR to CARE: Demonstrating the Value of Community-Based Research Infrastructure in the Age of AI

As artificial intelligence (AI) becomes embedded in research infrastructures, conversations often focus on automation and machine actionability. However, our work with the DataCite UK community continually reminds us that AI only functions because people create the conditions that make research objects understandable in the first place. Metadata, persistent identifiers (PIDs), governance frameworks, and curated collections are not abstract technical components. Rather, they are the result of human judgement, labour, and care.

In this paper, we respond to the IDCC27 theme by showing how community based, people centred stewardship strengthens the reuse potential of FAIR digital objects. While the FAIR Principles provide essential technical guidance, the CARE Principles highlight relationships, responsibility, and ethics. Together, they reveal that research infrastructures are socio technical systems shaped, built and maintained by people.

Drawing on our engagement with nearly 125 organisations in the DataCite UK Consortium, we illustrate how shared infrastructure and collective funding models expand equitable access to PIDs, particularly for under resourced institutions. Communities of practice, such as the Heritage Samples Registry, demonstrate how local support, shared standards, and collaborative curation make research objects more discoverable, citable, and interoperable. These examples show that embedding PIDs into everyday workflows is not simply a technical act but a community effort that increases the value and reuse of digital objects.

As AI systems depend on rich, interconnected metadata to support discovery, provenance tracking, and knowledge graph construction, human centred stewardship becomes even more critical. Activities such as ethical decision making, cultural sensitivity assessment,

and community engagement cannot be automated. Instead, AI highlights the importance of the people whose expertise makes digital infrastructures trustworthy.

In this paper, we argue that building AI ready research ecosystems requires recognising, supporting, and investing in the human labour that keeps FAIR and CARE principles alive in practice.

From Ledger to Living Knowledge: Community-Centred Redocumentation and FAIR Digital Objects in West African Museums

As museums increasingly convert physical collections into digital objects, critical questions arise about whose knowledge is represented, whose interpretations are prioritised, and how communities connected to cultural heritage can actively shape the records that describe them. Although digitisation has improved access to museum collections, many existing digital records still reflect legacy documentation practices that marginalise community perspectives and limit the representation of indigenous knowledge systems and lived cultural meanings.

This paper examines how community-centred redocumentation can contribute to the development and stewardship of FAIR Digital Objects within museum environments. Drawing on ongoing museum documentation and digitisation practice in West Africa, it explores how museum professionals, researchers, and community knowledge holders collaboratively enrich collection records through oral histories, local terminologies, contextual narratives, and culturally grounded interpretations.

Using a qualitative case-study approach, the paper analyses the processes, opportunities, and challenges of integrating community-generated knowledge into digital curation workflows. It focuses on how participatory documentation practices influence the quality, accessibility, discoverability, and long-term usability of digital heritage resources. The study also reflects on the role of emerging digital tools, including artificial intelligence, in supporting documentation while ensuring that human judgement, cultural sensitivity, and community authority remain central to stewardship decisions.

Findings suggest that community-centred redocumentation strengthens both the technical and social dimensions of FAIR Digital Objects by improving contextual richness, enhancing metadata quality, and fostering shared ownership of cultural knowledge. The paper argues that digital stewardship is fundamentally a relational practice that depends on sustained engagement between institutions and the communities connected to collections. It

concludes with practical lessons for heritage institutions seeking to align FAIR principles with inclusive, ethical, and people-centred approaches to digital curation.

From Mention to Citation: Data Citation Practices in the Social Sciences Across Two European Data Archives

Research data are increasingly recognized as a key scholarly output, but formal data citation remains uncommon in scholarly publications. Researchers often acknowledge the datasets they use, yet these datasets are frequently mentioned informally or cited through publications describing the datasets rather than through formal dataset citations.

This paper investigates data citation practices among users of two European social science data archives: the Swiss Centre of Expertise in the Social Sciences (FORS) and the Czech Social Science Data Archive (CSDA). It is based on a systematic review of publications from 2025 using data from the Swiss Household Panel (SHP) and publications from 2021–2025 using data from the Czech Household Panel Survey (CHPS).

Using a structured coding scheme, we assessed how datasets were referenced in publications, including in-text citations, in-text mentions, reference lists, data availability statements, acknowledgements, and footnotes. Preliminary results for SHP publications show that although datasets were mentioned in 95% of articles, only 26% included a formal dataset citation with a DOI. CHPS publications showed a similar pattern: only 18% included a formal dataset citation with a DOI, while 33% cited a publication describing the dataset and 63% mentioned the dataset without formally citing it.

The findings reveal a consistent gap between acknowledging research data and formally citing them. They suggest that improving formal data citation requires more than technical repository services alone. Persistent identifiers, recommended citations, and metadata provide an important foundation, but wider adoption also depends on clear guidance, training, collaboration with publishers, journals, and research organisations, and sustained engagement with research communities. This study provides empirical evidence on current data citation practices and highlights how research infrastructures can support the wider adoption of formal data citation as part of Open Science.

From Policy to Practice: An Integrated Support Ecosystem for Sensitive Data Publication

Publishing research data in line with the principle “as open as possible, as closed as necessary” remains central to research data management policies. However, putting this principle into practice is particularly challenging when datasets contain sensitive or

personal data. While institutional policies provide clear expectations regarding responsible data sharing, legal compliance, and ethical considerations; researchers often struggle to translate these into concrete publication decisions. This can lead to uncertainty and hinder both compliance and effective FAIR data sharing.

At KU Leuven, these challenges are addressed through an integrated support ecosystem centred around the institutional research data repository, RDR. It provides a trusted infrastructure for sharing research data openly or under appropriate restrictions and operates within a policy framework that formalises expectations for responsible data publication.

This paper presents KU Leuven's approach to operationalising responsible data sharing for sensitive research data through three interconnected components. First, targeted documentation and guidance materials complemented by an interactive assessment tool, support researchers in navigating complex, legal, and ethical publication decisions. Second, active support and a structured curation workflow provide an additional layer of quality assurance, combining expert review with actionable feedback to ensure compliance with FAIR principles while improving datasets. Third, a streamlined Data Transfer Agreement workflow facilitates controlled access while reducing administrative complexity and ensuring legal compliance.

Together, these elements form a coordinated ecosystem in which guidance, human expertise, and repository processes are closely aligned and mutually reinforcing. This approach is underpinned by close collaboration between repository staff, legal services, and the technology transfer office, ensuring that legal and ethical requirements are consistently translated into practical support throughout the data publication process.

This case demonstrates how repositories can move beyond technical infrastructure to act as mediators between policy and research practice, enabling responsible and FAIR data sharing in complex and sensitive contexts.

Grounding Data: Data Value Discourse and Indigenous Structural Critique
Data curation literature describes the profession's work as "adding value" to research data, but seldom interrogates the nature of this value. This paper analyses the value of research data in political economic terms. I contrast discourse about the value of data found in the FAIR Principles (Wilkinson et al. 2016) with similar discourse from Indigenous data sovereignty (IDSov) literature, including the CARE Principles (Russo Carroll et al, 2020). IDSov discourse demands an understanding of the politics that surround open data, raising

questions about the effects that open data can produce in the material world and the value data might have outside a framework that takes Western scientific and economic progress as a good and as a given. In describing open research data as a "return on investment," the FAIR Principles depict data as a circulating commodity, alienable from its original context. Both the technical mechanisms of machine learning and the companies that run on data subsume data curators' work while extracting value from their efforts to contextualize and describe data. In contrast, Indigenous data sovereignty discourse understands data's inalienability–its connection to the peoples, lands, and worlds it describes–as core to its value. The rootedness and relationality of data in this worldview connects it to the Dene scholar Glen Coulthard's (2014) political economic theory of "grounded normativity," which understands Indigenous people's reciprocal relationships to the land as a model for nonexploitative relationships between peoples and nations. At the same time, IDSov discourse sees data's strategic potential to enhance Indigenous peoples' prosperity as part of the larger project of decolonization. I argue that Indigenous data sovereignty (IDSov) theory does not merely supplement the FAIR Principles but serves as a key political critique, demanding that curatorial labour refocus on building non-extractive relationships in the material world.

How Did We Get Here, What Are We Doing Now, What's Next? A Unified Approach to Data Stewardship 20 Years in the Making

As technology and digital research infrastructure (DRI) used in scientific research continues to evolve, data stewards are uniquely placed to advocate for FAIR and open data best practice, helping to bridge the gap between DRI, funders, and scientists. However, it can be challenging to achieve this in isolation, with data stewardship often relegated to an unofficial role within a research group. This talk discusses how the incorporation and investment in a dedicated data stewardship team (DST) can lead to an increased understanding and adoption of FAIR data principles, and provide opportunity for hands-on development of DRI and digital tools, built with people in-mind. We'll explore how the UK Centre for Ecology & Hydrology (UKCEH) DST has evolved over the past 20 years; starting as a few data focussed individuals known as 'Informatics Liaison Officers (ILOs)', into a national data repository and unified DST with discipline-specific expertise, and share some lessons learnt along the way.

Through unifying the organisation's data stewardship, digital research, and Environmental Information Data Centre (EIDC) repository staff, we have been able to take a whole-project-life-cycle approach to data stewardship. This includes the support and creation of data management plans (DMP) using the Data Stewardship Wizard (DSW), data publication and curation of nationally important datasets, facilitation of research data management

training, and individual project support. By working collaboratively across teams, data stewards are involved in hands-on data management and research, for example, projects to harmonise data workflow pipelines in monitoring, developing tools for assessing data quality, and incorporating research ethics into the DSW DMP questionnaire. For data stewards, being part of a diversely skilled team has also allowed for growth of our own career pathways, providing opportunities to develop in specific areas of expertise and disciplines whilst still sharing a shared team vision of a FAIR data world.

Inside the Archive: A Deposit Analysis of Social Science Data
Research data archives are central to open science infrastructure, yet little systematic evidence exists on what is actually deposited and by whom. This paper takes an empirical approach to this question, reporting findings from a quantitative analysis of metadata from 1,217 social science datasets. We present a deposit analysis of the FORS social science data archive.

We characterise deposits along several dimensions, including disciplinary composition, project type, and funding source. Sociology is the dominant discipline, with psychology, communication sciences, and political science also well represented. Deposits are typically multi-disciplinary. Institutionally funded research and projects funded by the Swiss National Science Foundation (SNSF) together account for the large majority of deposits, while doctoral theses and personal research represent a small share.

These patterns reflect the archive's development, from FORSbase (2016) to SWISSUbase (2021), the successive self-deposit platforms used by FORS for data curation and archiving. Deposit volume has grown substantially, reaching 164 new datasets in 2025. These findings are preliminary; the full paper will present a more complete deposit characterisation, covering additional variables and offering more systematic interpretation of observed patterns.

While this analysis focuses on a specific case, social science data archived in Switzerland, the findings have broader implications, both for data stewardship practice and for how archives understand their depositor communities, improve curation, and design targeted outreach. Knowing who deposits what, and what is absent, equips data stewards to better support researchers, and enables archives to tailor their services accordingly. We position collection profiling as a practical, evidence-based tool for archive management in the social sciences.



Mapping the Labor of Research Data Management and Sharing: Evidence from Twelve US Research Universities

Federal funders and publishers require managed, shared data, but academic institutions often have little information about how these mandates impact research teams and where additional support could make a difference. The work of research data management and sharing (DMS) is often undertaken by research team members, support staff, and librarians at universities. This paper examines the time spent by those groups on DMS activities, in federally funded projects, breaking them down by role and research lifecycle phase.

In 2026, a research team conducted a retrospective survey of over 500 researchers at 12 universities in the United States on federally funded projects completed between 2018 and 2024. The research team also interviewed 99 research administrators in various roles at the same 12 universities. This mixed methods approach characterizes both the time researchers spend on DMS and institutional support, including staff time and effort.

This data will help academic libraries at all funding levels and levels of research: (1) understand who is doing DMS work at universities in the US; (2) comprehend the time and effort required to make federally funded research data publicly accessible; (3) provide libraries with evidence-based strategies to empower research teams to manage and share data effectively and to begin or enhance data services needed by researchers.

The paper describes the work done to comply with federal funders' and publishers' research data management and sharing requirements, and provides academic libraries and research institutions with information to help them support the research enterprise.

Metadata Enrichment for FAIR Digital Research Objects: Evidence from a Mineral Resources Report in Indigenous Lands in the Northwest Amazon

The growing production of scientific data has increased the need for digital curation strategies capable of ensuring that research objects can be found, understood, and reused across different scientific communities. Although the FAIR principles (Findable, Accessible, Interoperable, Reusable) provide the main international framework for improving data management, they do not specify which contextual information should be incorporated across domains. This gap becomes particularly evident in the geosciences, where complex Digital Research Objects aggregate multiple heterogeneous resources—such as datasets, maps, photographs, and analytical tables, produced throughout the research lifecycle. This study addresses the research question of how metadata enrichment can improve compliance with FAIR principles and expand the reuse potential of such objects. Adopting a qualitative, applied, and exploratory approach based on

documentary analysis and a single case study, the research examines a Mineral Resources Report on Indigenous Lands in the Northwest Amazon, produced by the Geological Survey of Brazil. The analysis was conducted in four stages identification of digital resources, classification of research data types, comparison with the Dublin Core and ISO 19115 standards, and identification of enrichment requirements—guided by an analytical matrix based on the metadata quality dimensions of completeness, provenance, accuracy, consistency, conformance, timeliness, and accessibility. The results show that the report's current metadata predominantly describe bibliographic aspects of the final document, while provenance, methodological context, and relationships among digital resources remain underrepresented. The findings indicate that the main challenge for FAIR compliance lies not in data availability but in the capacity of metadata to represent the scientific processes behind the data. The study concludes that metadata enrichment constitutes an ongoing Digital Curation strategy, rather than a one-time cataloguing task, that can strengthen the findability, interoperability, and reuse of complex Digital Research Objects, offering a transferable analytical approach for other scientific domains.

Mind the Gap(s): How Well Does the Data Lifecycle Fare as a Model for Disciplinary Data Practices?

FAIR Digital Objects (FDOs) are often understood solely in technical terms. By contrast, scholars have increasingly described digital research outputs in terms of their contextuality, situatedness, and relationality, stressing that FDOs depend intimately upon judgment, documentation, translation, negotiation, quality assurance, ethical reflection, and the long-term maintenance of trust. By way of comparing data (curation) practices across epistemic cultures, the contribution documents the disciplinary affordances of making digital research objects FAIR, including a critical engagement with the FAIR Guiding Principles, the DCC Curation Lifecycle Model, and recent work on the visible and invisible craft of data curation, to ask how data professionals shape the stewardship of digital research objects.

The contribution is designed as a comparison between the life sciences (biosciences and the medical sciences) and the social sciences (sociology, communication science, and economics), to study how data (practices) are implicated in the broader “machineries of knowledge production”. Experts from different institutions and research domains identify key challenges and tasks in research data management and research infrastructure operations, from planning and data creation to metadata processing, storage, access, reuse, and preservation. For every lifecycle phase, the contribution maps roles, tasks, and challenges that arise when disciplinary experts, data stewards, infrastructure providers, and researchers collaborate, to highlight common and discipline-specific tensions,

illustrating how organizational disciplinary background and job descriptions create a highly uneven landscape for those expected to perform similar tasks. Through studying the lived experience of data intermediaries, the contribution analyses social and socio-technical dimensions of data curation, to understand what data-centric data curation models tend to obscure how FAIRness depends on situated practices of (data) curation.

New Museum Informatics and Ethical Collection Practices

Museums have been managing physical and analytic description, provenance, attribution, permissions, sensitivity and contextual information relating to physical objects and documentary information for centuries. Collection documentation and cataloguing (part of informatics practice) over that timespan has shifted from sparse entries in large registers, to card catalogues, to treating catalogues as datasets, and recently collections “as data”.

In the last two decades trusted digital repositories have been introduced to support collection and corporate digital media preservation. Repositories retain digital collection items, documentary information and content. Educational, exhibition, publications or marketing content (rich audiovisual media and software) requires digital preservation (for medium to long term use and to meet retention policies).

Increasingly data management is an enable of digital object management and of museum informatics best practice. Museum informatics design needs to encode and apply FAIR (Findable, Accessible, Interoperable, Reusable), FATE (Fairness, Accountability, Transparency, Ethics) and CARE (Collective Benefit, Ability to Control, Responsibility, Ethics) principles into metadata for the management of collection items, documentary information and content. Incorporating cultural, rights and technical requirements into museum informatics assists with meeting curatorial, legislative, community and audience needs for context, reusability, auditability, transparency and explainability.

This paper covers shifts in digital media stewardship practice at the National Museum of Australia as part of its digital transformation.

Preserving Query-Based Data Citation Across Multi-Site Research Data Repository Deployments

Research data repositories often manage data that continues to change after publication. This is especially common for relational data, where researchers may cite a subset produced by a query rather than a complete dataset release. Query-based data citation addresses this by storing the query, the execution time, versioned data, and fixity information so that the cited result can later be reconstructed.

This model is relatively clear in a single repository instance, where timestamps, tuple histories, and query records are managed in one place. It becomes more difficult when the same repository system is deployed at several sites. A query may be created at one site, while the data and citation metadata needed for re-execution must later be available at another. Standard replication mechanisms can copy data, but they do not normally preserve persisted subset queries, result hashes, row counts, landing-page metadata, or the site-specific temporal values needed for reproducible citation.

This submission describes an application-level replication approach for an open-source relational research data repository. The approach replicates repository objects such as databases, tables, views, tuple histories, timestamp metadata, and persisted subset queries. Its main purpose is not to provide a new distributed transaction system, but to preserve the information needed to re-execute cited queries across independently operated repository sites.

Initial tests in a multi-instance setup show that structural metadata, tuple changes, and persisted query records can be replicated and used for matching query re-execution at another site. The work highlights that replication for dynamic data citation must preserve the temporal and citation context of data, not only current table contents.

Relationship Continuity Across Disciplinary and PID Metadata: A Task-Oriented Comparison of DDI Lifecycle and DataCite

Research objects are commonly described through complementary disciplinary and persistent-identifier metadata layers. Although these layers support FAIR identification, interpretation, citation, and reuse, differences in their entity models and levels of granularity can weaken meaningful relationships during metadata transfer. Element-level crosswalks alone cannot determine whether the function, entities, semantics, direction, and task-level usability of a relationship remain available.

This paper introduces relationship continuity, defined as the extent to which a meaningful, machine-actionable relationship represented in disciplinary metadata remains explicit, qualified, and usable in a broader PID or citation metadata layer. A task-oriented model is developed and applied to DDI Lifecycle 3.3 and DataCite Metadata Schema 4.7. Relationship-bearing constructs are normalized as source–predicate–target triples, assigned functional codes, organized into five categories, and operationalized through competency questions and associated human and machine tasks. DDI and DataCite are

then compared across relationship function, source and target entities, predicate semantics, direction, granularity, and usability.

The schema-level analysis covers 28 purposively selected DDI relationships, all 39 DataCite relationType, and 11 dedicated DataCite properties. Preliminary findings show several continuity patterns. Some functions, such as creator and funding attribution, show close correspondence when the described entities are aligned. Others undergo granularity compression, as when variable-level derivation in DDI is represented only between registered resources in DataCite. Variable-to-concept relationships may experience both semantic weakening and granularity compression, while variable-to-question, universe, and value-representation relationships depend on linked disciplinary metadata.

The schema-level findings lead to preliminary curatorial recommendations: curators should document which relationships are represented directly, generalized, or retained in disciplinary metadata; record entity scope and expected information loss; and maintain resolvable links between PID and disciplinary metadata records.

Responsible and Sovereign Stewardship: A University Data Steward as Institutional Mediator

Discussions of digital curation increasingly recognise stewardship as a value-laden practice rather than a neutral technical function. Equitable stewardship, in particular, has drawn attention to whose interests curation decisions serve. This paper proposes a complementary lens, grounded in institutional practice rather than theory-building: responsible and sovereign stewardship. We use this framework to examine a recurring institutional tension observed in the Social Sciences DataLab, a research data infrastructure operated by NOVA SBE, Portugal.

The university Data Steward brings Open Science and FAIR principles into an institutional environment that is not neutral but shaped by a commercial data ecosystem of vendors, licences, platforms, and infrastructures that the university does not own or fully control. Stewardship in this context requires balancing two enduring responsibilities rather than resolving a single trade-off. Responsible stewardship refers to the duty of care exercised when advising where and how third-party data—and, increasingly, the AI tools that process it—should be accessed, hosted, and secured, recognising that key elements of the underlying infrastructure remain outside the DataLab's direct control. Sovereign stewardship, by contrast, is the parallel commitment to expanding institutional and public control over data by promoting the integration of public-sector data into the university on open and FAIR terms, while negotiating commercial contracts that reduce vendor lock-in

instead of simply accepting prevailing market conditions. From this perspective, AI governance is best understood not as a distinct governance challenge but as the most visible contemporary manifestation of this long-standing institutional tension.

The object of analysis is the DataLab as an institutional arrangement—its governance structures, roles, decision-making processes, contractual practices, and technical infrastructure—rather than the personal experiences of either author.

Revealing the Human Behind the Data: A High Speed 2 Case Study
This paper discusses the importance of embedding people within the dissemination process for archaeological archives, using High Speed 2 (HS2) as a case study. To address the many ethical issues surrounding the reuse of digital bioarchaeological data, the Archaeology Data Service (ADS), a CoreTrustSeal accredited repository for UK-based archaeological and heritage data, adopted a multi-phased archiving approach which centred the opinions of its user community and the specialist knowledge of its team members.

The construction of the HS2 railway required the largest historic environment investigation programme in the UK's history, resulting in an unprecedented amount of data (34TB) being deposited with the ADS. Six sites along the HS2 route produced large burial assemblages (in total, >23,000 excavated individuals and >3,400 photogrammetric burial models).

Results of a survey targeting the archaeology community demonstrated that a majority of participants wanted content advisories for archives containing images of human remains and to view 3D models of human remains alongside contextual information. It was therefore decided that the six HS2 archives would receive Special Collection interfaces including an interactive map, advanced query, and linkage with other data types, including osteological information and excavation photographs that could be obscured with content advisories. The creation of these Special Collections was a complex process that required archive-specific approaches, and the archaeological, osteological, and technical expertise of the ADS team.

This paper describes how embedding people in the data curation process has brought six HS2 archives further in line with the FAIR principles – information about past people is now more findable, accessible, and reusable. It is hoped that this approach will encourage others who face a similar challenge of disseminating a large, sometimes faceless body of bioarchaeological data to capitalise on specialist knowledge and ensure that they too strive to centre the human behind the data.



Risk and Expertise: How Professional Roles Shape Views of Repository Certification Requirements

Trustworthy Digital Repository certification processes function as risk mitigation strategies, enabling repositories to signal quality and commitment to best practices to external stakeholders. The effectiveness of these processes depends in part on a shared understanding of certification requirements among those who participate in them. However, prior research suggests that stakeholders in TDR certification do not always share the same views about what certification requirements mean or what they demonstrate. This study examines how professional role and review experience shape repository staff attitudes about the CoreTrustSeal requirements.

Drawing on a survey of 88 staff members from CoreTrustSeal certified repositories (53.98% response rate), this study asked respondents to rank the three sections of the CoreTrustSeal requirements — Organizational Infrastructure, Digital Object Management, and Technology — by importance for long-term preservation, and to reflect on a sample requirement from the section they ranked highest.

Findings demonstrate two patterns. First, respondents overall ranked Organizational Infrastructure as most important, followed by Digital Object Management and then Technology, consistent with research emphasizing the primacy of social and financial sustainability for repositories. Second, professional role and review experience were associated with differences in those rankings: respondents in IT roles were more likely to rank Technology as most important, while those with experience as CoreTrustSeal reviewers showed greater consensus in their rankings than those without. These findings suggest that expertise shapes attitudes about certification requirements, and that participation in the review process reinforces a shared understanding of those requirements — a finding with implications for how trustworthy digital repository certification communities are structured and sustained.

Three Dimensions of Human-Centred Digital Curation: An Interdisciplinary Approach to Preserving Public Audiovisual Memory

This paper presents results from a collaborative project between the University of Brasília (UnB) and the Brazilian Superior Labour Court (TST) to preserve and modernise a collection of more than 11,000 analogue audiovisual tapes, comprising nearly 20,000 hours of recordings threatened by technological obsolescence. Beyond technological migration, the project adopts a people-centred approach to digital curation, carried out by a multidisciplinary team spanning Communication, Information Science and Computing.

Digital objects do not become searchable, interpretable or reusable through digitisation alone; they acquire informational value through human intellectual work, including archival description, metadata creation, controlled vocabulary development, indexing and validation. The study organises the empirical work as a digital curation workflow encompassing cleaning, digitisation, archival description and organisation of the collection within the AtoM platform, using the open-source application Archivematica to ensure long-term preservation under international archival description standards. The paper argues that human-centred digital curation can be understood through three complementary and interconnected dimensions. Intellectual curation refers to the human work required to identify, describe, classify and contextualise audiovisual records. Collaborative curation recognises that this knowledge is produced collectively by a multidisciplinary team through shared protocols, metadata standards and validation procedures. Public-oriented curation highlights that preservation fulfils its democratic and communicational purpose only when organised information becomes accessible and reusable by institutions, researchers, journalists and citizens. These three dimensions are, in turn, structured by two cross-cutting principles: the FAIR principles (Findable, Accessible, Interoperable, Reusable), which anchor the intellectual and public-oriented work of making records useful for institutional memory, and the notion of data culture, which explains the shared values and practices underlying the project's collaborative dimension. Together, these dimensions and principles describe complementary processes through which human expertise transforms an endangered audiovisual collection into enduring public memory, offering a methodological approach replicable by other public audiovisual archives facing similar risks.

We Can Work It Out: Collaborating to Address the Challenges of Transversal Skills Training for Data Professionals

Research Data Netherlands (RDNL) has in 2026 presented a new competency framework for research data professionals in the Netherlands. This framework specifies seven competencies that data professionals can develop and has helped to define the scope of a Dutch national curriculum for data stewardship training. After mapping existing training for data professionals to these seven competency areas, the project team have identified one area with a significant training gap: transversal, or ‘soft’ / ‘non-cognitive’ skills.

These can be defined as ‘competencies that can be used in a wide variety of settings, relating most explicitly to the duties of a data professional, such as networking, effectively communicating, and managing RDM projects’. Many established data professionals will recognise these abilities, such as being able to ‘navigate the complex structures, rules, power dynamics, and interpersonal relationships within research organisations’. However,

many are difficult to teach. These skills are often developed ‘on the job’, through years of experience.

This paper outlines some of the existing training in transversal skills for data professionals in a research context. This includes the Skills4EOSC Data Steward Training Curriculum; the Dutch ‘DCC Spring Training Days’; RDNL’s flagship ‘Essentials 4 Data Support’ course; and the pioneering work of several individuals who are interested in these transversal skills. It also asks the audience to share knowledge about other existing training in this area. It moves on to explore future opportunities for the RDNL project, in collaboration, to rise to the challenge of filling this gap in training.

What Digital Maps Forget: Recovering the Knowledge of Disappeared Places
Disappearance is not a new concern for digital curation, but it has typically been examined through the lens of politically motivated erasure or archival silence: cases where records are deliberately destroyed, suppressed, or never created in the first place, often in relation to marginalized histories or contested political events. This paper examines a different and understudied form of disappearance: the routine loss of information as commercial digital platforms continuously update and replace existing content. The goal of this study is to understand how knowledge about disappeared places can be recovered after it is no longer readily accessible through digital maps. Through analysing discussions in the r/FoodNYC subreddit, comprising 841 posts and 18,193 comments about disappeared places in New York City, we identify disappeared places using named entity recognition, qualitatively code the types of place knowledge preserved in these discussions, and evaluate their current representation on Google Maps. Preliminary findings suggest that community discussions preserve rich contextual knowledge that is often absent from digital maps, including provenance (e.g., ownership history, founding stories), spatial memory (e.g., former location, what currently occupies the space), personal experience (e.g., menu items, specific memories), nostalgia (e.g., emotional attachment, sense of loss), and regulatory information (e.g., health violations, building code violations). These findings demonstrate the value of community-based discussions as complementary resources for recovering disappeared places and highlight the role of communities in the people-centered curation of place knowledge. The full paper will present the completed qualitative coding analysis together with a comprehensive reconstruction of disappeared places and an analysis of their recoverability on digital maps.



Why Data Stewardship Must Remain Close to Research: Evidence from the Social Sciences

European research infrastructures are increasingly expected to make research objects FAIR, support Open Science, and connect local research communities to transnational data spaces. These ambitions are often framed as technical or governance requirements, yet their implementation depends on substantial human work: planning, documenting, checking, anonymising, curating, depositing, explaining, and enabling reuse. This paper asks how social scientists perceive and encounter data curation and stewardship, what expertise they expect from these roles, and why stewardship should remain closely connected to research and teaching.

The analysis draws on [PROJECT], a two-stage study of the Austrian social sciences combining a structural mapping of 5,241 researchers with a push-to-web survey of 731 respondents. It examines secondary-data use, repository engagement, disciplinary cultures, perceived benefits and barriers, and data-management experience.

The findings reveal a persistent gap between the recognised value of secondary-data reuse and its institutional support. Although 44% of respondents describe secondary analysis as common in their field, only around 30% perceive it as actively supported or rewarded. Around 39% know or use open-access retrieval services, whereas approximately 5% report using formal archiving routes. Reported barriers combine legal uncertainty, preparation effort, concerns about control and interpretation, and doubts about data quality. Multivariate models explain only modest shares of variance, indicating that individual experience and motivation alone cannot account for formal sharing.

The paper proposes an empirically grounded model of embedded stewardship. Curators and stewards require specialist legal, ethical, technical, and preservation expertise, but should not be separated from the research environments they support. Their participation in project design, methods teaching, supervision, and disciplinary exchange enables them to translate FAIR requirements into workable, context-sensitive practices. Sustainable digital curation therefore depends on stewardship that is professionally specialised but institutionally integrated into research.

Yoda: A Collaborative Data Management Infrastructure to Safeguard Academic Values

Both funders and institutions acknowledge the importance of integrating high-quality data management practices into the research lifecycle. To achieve this, individual researchers must take active steps towards the management and longevity of their research outputs throughout their workflow. The approach taken by VU Amsterdam is to create a supportive

ecosystem that aims to smoothen the complexities often associated with research data management, as well as the archiving and publishing of research output.

Two crucial elements of a professional support organisation are the technical infrastructure, and the support services based on this infrastructure. These should be up-to-date, easy to use and adhere to data practices and connect to the requirements of researchers. The core data management and storage solution VU Amsterdam provides facilitates the research process from data discovery and sharing to archiving and publishing of research results, contributing to data findability, accessibility, interoperability and reusability.

A third important element is the sustainability of the infrastructure itself. There is a growing dependence on a limited number of dominant market players often with their own closed systems/platforms with their own terms and standards. This creates vendor lock-in, resulting in unpredictable price increases, a lack of transparency, and difficulties in switching to other suppliers. It also involves a loss of control over digital infrastructures and data, loss of intellectual property, and fear of undesirable influence from commercial parties and certain governments.

All in all, this is why VU Amsterdam has chosen for the implementation of a collaborative, open-source solution to be the heart of the data management ecosystem. This solution is called Yoda which stand for Your Data. Yoda provides researchers with a tool that ranges from safe and easy storage to sharing of data and ultimately to archiving and publishing of data.