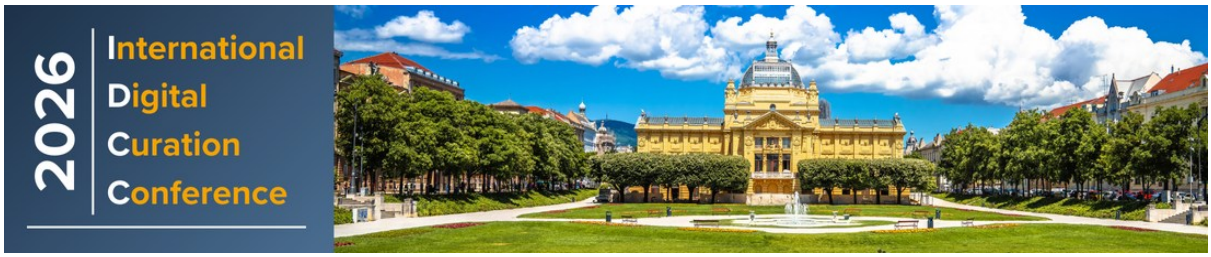


Accepted Lightning Talks

AI And Open Collections: Developing A Position And Policy For Repositories

In response to the technical, legal, and philosophical challenges posed by of the use of institutional repository collections by generative AI, the Bodleian Libraries at the University of Oxford have begun work developing a position and policy to articulate the issues we and other GLAM institutions face, and to describe positively to members of our community and external audiences how we wish our collections to be used.

As one of the first pieces of work towards this, we surveyed the policy and guidance landscape across the GLAM and publishing sector, to understand and learn from steps others have taken – and to see the questions that they might also have but don't have answers to.

In this lightning talk, we will describe our approach to this work, summarise our findings, and outline how we have shaped the questions that we plan to address in our position and policy. We'll then update colleagues about how we plan to take this work forward, to address the actual and perceived risks for our researchers, staff, and us as a library service.

AI in the Archive or an AI-Driven Archive?

In this lightning talk, we will describe two on-going development projects at the State Archives of Belgium (SAB) and the UK Data Service (UKDS) addressing the use of artificial intelligence (AI) tools to support the ingest and archiving of data. We will also discuss the potential roles of AI in the secure trusted research environments (TREs), and raise the question of whether AI has a place in the secure environment.

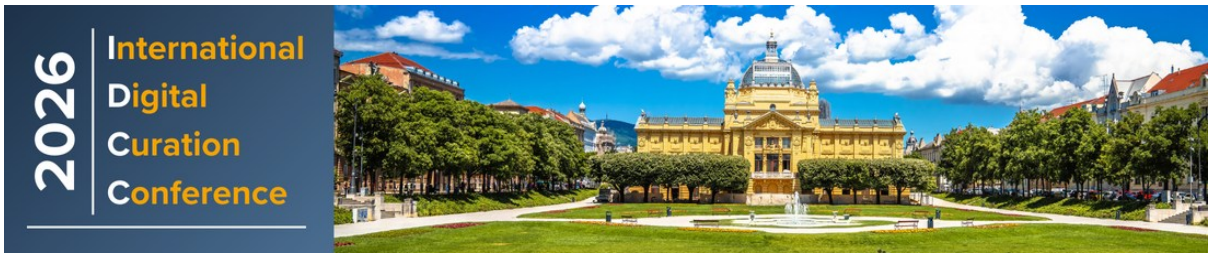
Archival Information Packages: a Data-Centric Approach to Preservation

Drawing on Artefactual's conceptual model for digital preservation systems, this talk focuses on the importance of Information Packages preserved by an Open Archival Information System (OAIS). In software design, there are core design principles for digital preservation systems that, taken together, represent the best chance to ensure data can persist over time and space. This talk will focus on one principle: the AIP must be the system's primary concern as the source of truth.

Augmented Intelligence and Ethical Curation: Lessons from the FamilySearch-AUC Arabic Registers Project

This presentation explores the ethical integration of AI into digital curation through a collaboration between FamilySearch and the Digitization Center, AUC, with a focus on digitizing historic Arabic death registers. These fragile manuscripts contain vital family history but present challenges for automated text extraction. These challenges include the complexity of the Arabic script, cultural naming conventions, and poor physical condition.

The project used a combination of customized AI models for Arabic script and human validation from Arabic linguists. Importantly, culturally relevant metadata fields, including lineage and death status, were integrated to maintain the contextual integrity of the records. This hybrid approach reduced errors while proactively addressing ethical concerns like data ownership and contextual accuracy.



The project shows that ensuring ethical AI in cultural heritage involves ongoing human oversight, inclusive practices, and deep cultural sensitivity, all of which can be achieved even with limited budgets. By incorporating transparency at every stage of the curation process, the project effectively maintained a balance between efficiency and ethical rigor. This model offers organizations a pathway to preserve not just historical records but also the dignity of the individuals they represent, proving that ethical AI prioritizes people alongside technical innovation.

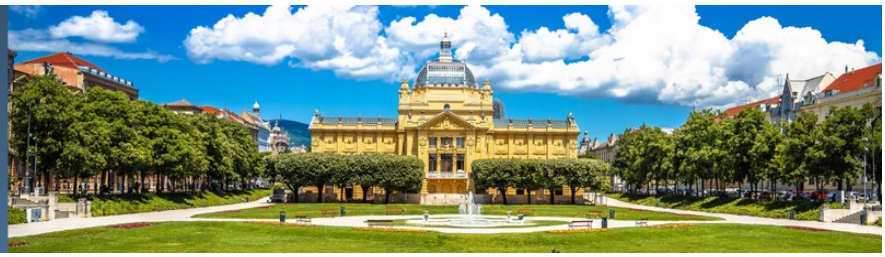
Automating Curation To Support FAIR Data Publication: The KU Leuven RDR Review Dashboard
KU Leuven's institutional data repository (RDR), built on Dataverse and CoreTrustSeal certified, plays a vital role in supporting researchers to make their data FAIR. Since its launch in 2022, the number of datasets submitted for review has increased consistently, thereby requiring a more efficient review process to avoid duplicated efforts and manage administrative workload. The RDR team developed an open-source review dashboard that streamlines the curation workflow: reviewers can track assignments, add notes, and use a checklist to generate consistent feedback. This provided additional time for more personalized support, training, and consultation.

The second version introduced automated checks for common issues such as incomplete metadata and missing or empty README files. Many of these checks proved easy to implement and effective in catching errors early. Crucially, the dashboard was made to complement human judgment - automated checks are visualized, and reviewers can override them when needed. The RDR team is currently working on making these automated checks available as an external tool. Researchers will be able to validate their draft datasets before submission, improving quality and reducing review time. This presentation shares our development journey and invites discussion on how automation can empower both curators and researchers.

Big Graphs, AI Black-boxes, And Human-out-of-the-Loop: Regaining Human FAIRness Through Human-Centric Curation And Preservation Of Intelligent Artefacts
The acknowledgement of the role played by machines in production and discovery of knowledge has underpinned the FAIR principles as a key-tenet since their inception. At the core of the principles lie the concepts of machine-readability and machine-actionability, and the pledge to satisfy, in a digital research environment, the specific information needs and information behaviours of machines, considered as e-Science ‘stakeholders’ and primary data users.

Given the extensive upstream and downstream deployment of artificial intelligence systems in the current e-research landscape, however, a FAIR machine-centric approach risks to contribute to the ongoing decentring of human-agents in knowledge discovery processes.

The need to mitigate such decentring has kindled the recent plea for a FAIREr approach to actionability of data and information in automated knowledge environments (Vogt 2023; Vogt et al. 2024), to deliver systems pivoting on the idea of “cognitive interoperability” (Naudet et al. 2023) between machine- and human-users.



This lightning talk will discuss the properties of cognitively interoperable artificial intelligence artefacts (openness, transparency, explainability) and the challenging factors that impair them. An overview of computational stewardship approaches and tools that promise to support cognitive interoperability will be presented and assessed against standardised data curation and preservation practices.

Combining Machine Learning tools and Terminology Services to support FAIR Data Management in ESS

Integrating diverse data in Earth System Sciences (ESS) is hindered by semantic heterogeneity, as different disciplines use inconsistent terminologies. The BITS project addresses this challenge by providing a dedicated Terminology Service (TS) with over 40 vocabularies, maintained by the TIB for long-term use. A practical use case at the Senckenberg Institute combines AI tools, edge computing, and the TS to automatically annotate and process digitized natural science collections efficiently. This workflow improves data findability and interoperability, particularly for multilingual and handwritten records, while ensuring data security through near-data processing. The Lightning Talk highlights potential risks of irresponsible AI use, despite its clear benefits when integrated with other services.

Data Management Planning in a Cross-Disciplinary Research Environment: More Than a Tick Box Exercise?

In cross-disciplinary and data-intensive research environments, effective data management is essential. We have reimagined data management planning as a collaborative, evolving process that underpins the generation of FAIR data and the responsible, sustainable use of AI, transforming it from a compliance task into a valuable tool in the research lifecycle.

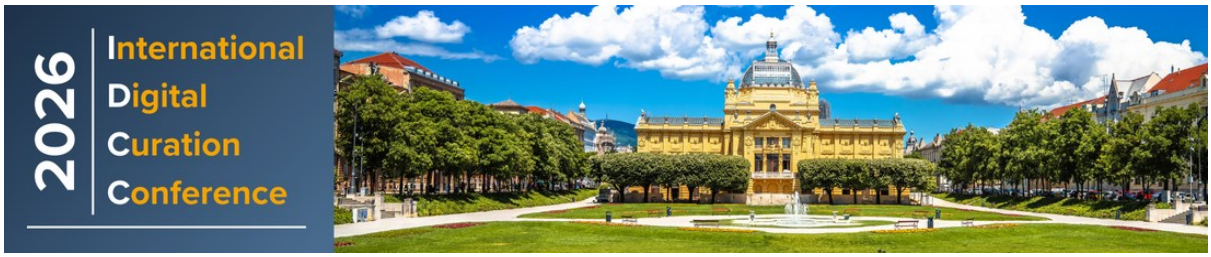
Developing Trusted Preservation Services for Sensitive Research Data: A Case Study

This abstract outlines the development of services for the long-term preservation of sensitive research data. Initially, the institution's infrastructure was not equipped to handle high-risk datasets, prompting a comprehensive effort to enhance both technical and procedural capabilities. The work focused on five key areas: reviewing legal and ethical requirements, strengthening a multidisciplinary expert network, enhancing internal competencies, refining user workflows, and upgrading technical platforms. Collaboration with legal, IT, and data protection professionals was essential, as was hands-on engagement with researchers to assess real datasets.

Notably, the team introduced structured protocols such as Data Access Protocols and Data Protection Impact Assessments to ensure compliance and accountability. Despite challenges—particularly in defining responsibilities and balancing security with usability—the strong demand from researchers proved a powerful driver. This initiative aligns with themes of sensitive data curation, tool development, and community building, offering a practical model for institutions seeking to support secure, sustainable research data preservation.

Enhancing the Benefits of Machine-Actionable DMPs with Generative AI

This talk will explore how advancements in large language models can help review, write, and improve connections of machine-actionable data management plans. It will review ongoing



initiatives at the California Digital Library to improve the benefits and easy burden of writing DMPs, while still keeping the human author as the authority. It will cover learnings and progress and tips for others exploring similar areas.

Establishment of a Certification Program for Research Data Management Support Personnel at Kyushu University, Japan

This talk will introduce an overview of a certification program for research data managment support personell at Kyushu Unversity, Japan, show student feedback from questionnaire surveys, and discuss future directions and potential international collaboration.

Evolving with the Community: Updates to the CoreTrustSeal Requirements for 2026-2028

This presentation outlines the latest update to the CoreTrustSeal certification requirements (v04.00), which will guide repository certification from 2026 to 2028. Reflecting extensive community consultation, the update introduces minimal but meaningful revisions to maintain stability while emphasizing active preservation as a core feature of trustworthy repositories. Attendees will learn about the update process, key changes, and how to stay involved as CoreTrustSeal evolves to meet the future of trustworthy data stewardship.

Exploring the Use of LLMs to Support Research Data Documentation in the Social

This presentation examines how large language models (LLMs) like ChatGPT can assist researchers in generating documentation such as codebooks and README files, helping meet Open Science and FAIR data requirements. It explores how LLMs can streamline workflows, align outputs with repository standards, and reduce curation burden. Alongside these opportunities, the talk also addresses the ethical, legal, and environmental challenges posed by AI integration, especially in contexts involving sensitive social science data.

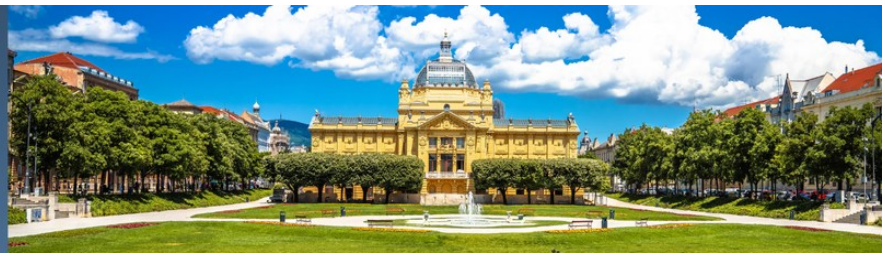
FAIRVault, An Inter-University Federated Dataverse Pilot In Flanders

In 2023, four Flemish universities (Ghent University, Hasselt University, University of Antwerp, and Vrije Universiteit Brussel) initiated the FAIRVault project to provide researchers with a secure solution for preserving and providing (controlled) access to research data. It especially targets cases where external repositories are less suitable, such as for sensitive or large datasets, ensuring proper data retention and security.

GenAI For Research Support: A Case To Dmps Pre-Filling

Researchers find preparing data management plans (DMPs) time consuming and see data stewards as an appealing solution for filling them in on their behalf. At Eindhoven University of Technology (TU/e), data stewards support in filling DMPs is limited by the knowledge on the research topic. Advances in technological solutions are perhaps the most feasible way to support researchers in filling their DMPs before the data stewards final approval.

At TU/e different research support groups, such as the Research Data Infrastructure (RDI) and the Data Stewards (DS), have initiated a collaboration with an external party to develop a GenAI tool that would directly and automatically extract the relevant information from research proposals and to pre-fill the DMPs. The tool is able to extract structured metadata from research proposals, based



on a user-defined schema specifying the desired metadata attributes, e.g., research goals, presence of human participants, data origin and size, etc. (including a description per metadata attribute).

Helpdesk on Research Data Management: AI-based Chatbot for FAIR Research Data in Portugal
This lightning talk concentrates on presenting the project's helpdesk system, a service that blends artificial intelligence with distributed human expertise. At its core is an AI-driven chatbot powered by state-of-the-art natural-language processing. The bot understands free-text questions, retrieves curated, context-aware guidance and automatically escalates complex or discipline-specific issues to domain specialists, thereby reducing first-response times and releasing staff to focus on higher-value curation and stewardship tasks.

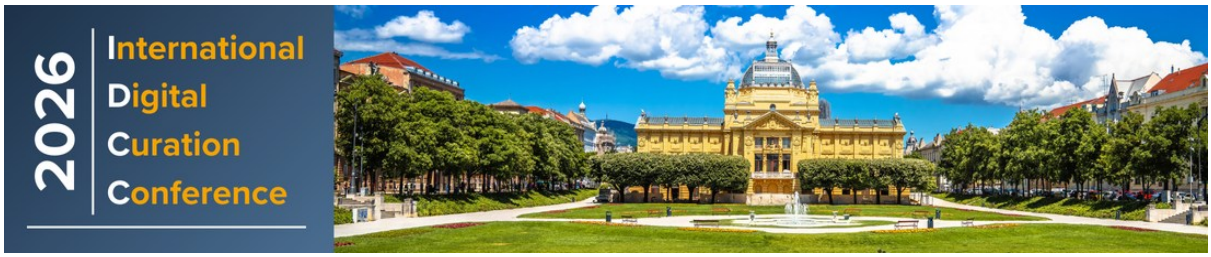
A key innovation is the knowledge base that powers the chatbot. Created in collaboration with a growing network of institutional Competence Centres, this multilingual repository contains entries such as FAQs, policy templates and decision trees covering data management planning, secure storage, legal and ethical compliance, FAIRification techniques, and long-term preservation strategies. Regular content sprints and community editorial workflows ensure the guidance remains authoritative, evidence-based and traceable.

How & When To Use A Data Steward? The Use-case Of The Czech RDM Ecosystem
The National Repository Platform of Czech Republic contains services, functionalities and trainings, such as the Data Stewards Community CZ. The role of the data steward becomes more and more important with the burst of AI implementation into research and through this lightning talk we will highlight how the Data Stewards Community in Czechia gets involved and promotes good data management practices.

How Arquivo.pt is Preserving Scientific Research Project Websites and Promoting Data Reuse
This presentation shows how Arquivo.pt is preserving websites related to scientific research projects and how web information from the past is becoming a source of data for research projects and the development of Artificial Intelligence applications. It is argued that the preserved contents of the web are also data and, as such, should be included in issues related to curation.

How Do You Calculate the Carbon Footprint of Your Digital Preservation Activities?
This talk will discuss recent work by the Digital Preservation Coalition to collaboratively create guidance and advice for the community on how to calculate the carbon footprint of digital preservation activities. Through the Carbon Footprint Task Force, which has been meeting every month during 2025, a group DPC Members has come together to share their own experiences and develop a toolkit for others within the community to use to help them work in this area. This short talk will describe this work and showcase the online resource which is due for publication in February 2026.

How Do You Make Your University FAIR-er?
Many of us are familiar with limited resources, complex organizational structures, and growing demands for openness, accountability, and long-term data stewardship. Working within those constraints, how can we make our university FAIR-er?



At the Swedish University of Agricultural Sciences, we chose to shift our focus from development of concrete infrastructure to development of operational capacity and culture. A four-year institutional data management programme was initiated which focused on soft infrastructure. Rather than introducing new systems, the programme looked to strengthen existing services, support local capacity, and foster cross-university collaboration. Among the programme activities can be mentioned the creation of processes and workflows, strategic recruitment of key expertise, development of support partnerships, and building capacity for targeted communication and outreach.

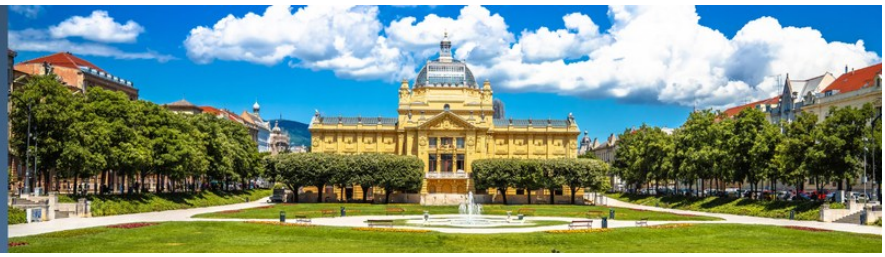
In this lightning talk, we will present the design, implementation and tangible outcomes of the programme, highlighting some key challenges and reflecting on what we have learned. We hope to spark discussion, gather feedback, and connect with others working to build FAIR-er institutions by operational means as much as technical development.

Micro-Creds: Innovative Training for Sustainable Communities
The Sonraí Irish Data Stewardship Network was initially funded under the National Open Research Forum (NORF) 2022 Open Research Fund call. Now, as we move to a self-sustaining network, we are exploring options for financial sustainability as well as delivering value to our community of data stewards—recognising the valuable role data stewards play in curation of research data. We are rolling out a fee-based micro-credential in Data Stewardship in University College Cork (UCC) in Ireland. A micro-credential Task & Finish group was established in October 2024. Using the Skills4EOSC Curriculum Outline for Data Stewards, they completed significant work in adapting the curriculum to the Irish context.

The launch of the Introduction to Data Stewardship micro-credential in 2026 is a targeted investment in skills, will increase institutional capacity to align practices with funder, policy and legislative requirements, provide a route to an accredited data stewardship career pathway, and offer Sonraí a means for longer-term self-sustainability.

Special thanks go to the Sonraí Task & Finish group, who have donated their time to this project. This lightning talk will give an overview of the platform choice, curriculum work to date, and plans for expansion through a model for developing discipline-specific curricula.

Navigating Sensitivity, Scarcity, and Scale: AI and the Future of Sensitive Data Curation in Small Social Science Archives
The curation of sensitive social science data is becoming increasingly complex in the face of rapid technological change, shrinking resources and stricter ethical and legal requirements. Small domain-specific data archives such as the Slovenian Social Science Data Archives (ADP) are under pressure as they have to balance long-term data stewardship with the growing demands for rapid and accessible data publication. While quantitative data is often anonymised by researchers, qualitative material such as interviews requires expert review to identify identifiers — an effort that requires a lot of time and resources.



AI offers potential relief by automating tasks such as entity recognition and redaction. However, its use raises significant concerns about the legal and ethical implications of machine processing, particularly in the sensitive pre-publication phase when archivists prepare records for secure transfer. For archives with limited technical capacity, the independent implementation of such tools is a challenge.

This lightning talk highlights ADP's role in CESSDA's collective efforts to explore AI in digital curation and develop collaborative solutions for small archives with ethical oversight. It advocates inclusive, federated AI approaches that ensure data integrity, privacy and trust."

Organizing a Community to Survive Research Ecosystem Instability

The instability of the current United States research landscape has required the rapid response of data curators and librarians to emerging changes. As the shifts in the research landscape continue, shouldering the weight of the changes alone diminishes our ability to provide support to researchers, keep up to date on the situation, and respond to changes in an informed manner.

Instead of placing the burden on a sole individual in an organization, a more holistic and sustainable model is developing a community focused on mutually supporting one another. This lightening talk will discuss how a data librarian built a community dedicated addressing the changes in federally funded research policies through knowledge sharing, dividing labor, and developing effective training. The goal of this presentation is to provide concrete takeaways from the successes and challenges of this use case, in order that attendees can develop their own resilient community in the face of continual fluctuations.

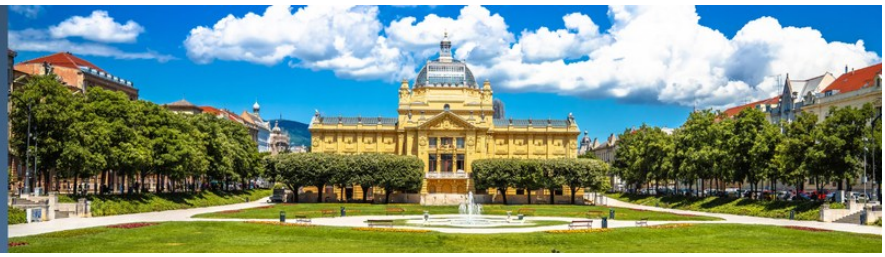
The Base4NFDI Development Framework: Growing Interoperable RDM Basic Services Across Domains

Germany's research data infrastructure is both mature and fragmented - rooted in deep institutional expertise but challenged by complexity when it comes to collaboration and shared service development. The National Research Data Infrastructure (NFDI) initiative was launched to address this by fostering shared services across disciplines. Within NFDI, the Base4NFDI project provides a technical and decision-making framework to support the structured development of "basic" Research Data Management services that are interoperable, reusable, and sustainable.

The talk highlights the development framework which combines technical (e.g. common AAI), organisational (co-development incentives), and sustainability components (business models, training resources), and emphasizes FAIR service development and open-source licensing.

Development proceeds in three phases: Initialisation (requirement analysis, persona work, and prototyping), Integration (co-development with NFDI consortia through incubator projects), and Ramp-Up (governance, business planning, EOSC integration).

As of mid-2025, eight services are supported, with four in integration and one entering ramp-up. We will share lessons learned from early service development, including insights from service requirement analyses and incubator projects. By fostering collaboration and transparency from the



start, Base4NFDI has the potential to inform infrastructure development efforts in other contexts. We look forward to exchanging ideas and experiences with the IDCC community.

The File Check Assistant: A User-Friendly Tool for Improving Data Reusability
Supporting scientists in meeting basic data standards is essential for enabling long-term re-use. Environmental research often involves wrangling datasets that suffer from inconsistent formatting, unclear structure, or minimal metadata. These issues raise barriers to human understanding and machine readability, limiting interoperability and re-usability.

We have developed a user-friendly tool to help improve quality of archived data files for future re-use. The File Check Assistant tests CSV data files for basic principles of re-usability, such as file structure and encoding, and helps guide users towards good practice. Written in R Shiny, the tool has a point-and-click user interface to reduce technical barriers to applying it to research data.

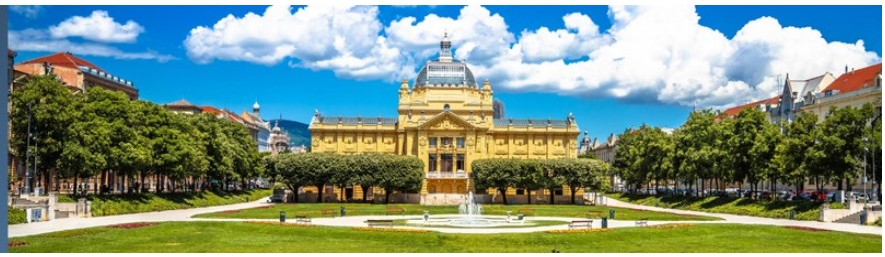
This low-barrier tool that promotes FAIR data practices will be published as an app for the wider community.. Tools such as this can play a key role in supporting the curation of sustainable and re-usable research data that is machine readable and AI-ready.

Transforming Historical Records into Digital Assets for Improved Access and Efficiency; The National Social Security Fund Records digitization Journey.
The National Social Security Fund (NSSF) Uganda, established under the NSSF Act, provides social security services to all employees in Uganda. In 2018, the Fund initiated a records digitization project to address challenges such as manual processes, long document flows, and rising operational costs. This led to the implementation of an Electronic Document and Records Management System (EDRMS), known as ADA (Advanced Digital Archival), aimed at digitizing historical records for easier access and retrieval.

Key activities included preparation, scanning, and indexing of records, resulting in a digital central repository that allows real-time access for employees, enhancing customer service and workflow tracking. The Fund achieved a 95% paperless operation, promoting a clean desk policy. However, challenges arose, including high software maintenance costs and the need for ongoing change management.

Looking ahead, the Fund aims to integrate the EDRMS with other systems, build internal capacity, and leverage data-driven decision-making. This presentation outlines the digitization journey, highlighting improvements in accessibility, operational efficiency, and customer service, while recommending strategies for other institutions to enhance their records management processes.

Trust, Types & Transparency: TIC-TAC-TOE
Transparent Trustworthy Repository Attributes Matrix (TTRAM) supports the identification of a common understanding of repositories capabilities and needs for the FIDELIS and EDEN projects and the European Open Science Cloud (EOSC). This talk examines the next steps for the Matrix in seeking to reach consensus on the characteristics of Trustworthy Digital Repositories (TDR) in terms of trust



in context, types and categories, and their implications for transparency across digital objects and organisational entities (transparent objects/entities): tic-tac-toe.

What Should Be Saved? The Impact of Austerity on Data Rescue

At a time where research institutions globally are being faced with diminishing budgets, methods for prioritising data for preservation are essential. This talk will detail our application of the Hoffman et al. data rescue framework (DRF) in a recent project to prioritise datasets for rescue, plan workload, anticipate potential obstacles, and approximate resources required. I will detail our novel points-based adaptation of the DRF which facilitated the decision-making process of which dataset to save, accounting for our limited budget. I will also describe how we used this adaptation to quantitatively compare the dataset before and after rescue, taking the FAIR principles into account.

The methodology I will describe is likely applicable to countless similar datasets currently held in inaccessible locations and gives a step-by-step structured process for data curation professionals to follow from prioritising data through to publication. It could greatly improve efficiency and prioritisation of data rescues if adopted by other institutions, particularly those affected by scarcity of budget and resource.