# Lightning Talks

**10 Years Of Research Data Management At The University of St Andrews: Reflecting On Progress, Challenges And Future Goals by Federica Fina**

In this honest and open talk, we reflect on our successes and our setbacks over the past 10 years or research data management. We will touch on what we have learnt from our mistakes and how we will make sure to use those lessons to try and achieve our goals for the next 10 years.

**A Digital-First Approach for the Kerby A. Miller Collection of Irish emigrant letters and memoirs by Marie-Louise Nathalie Brigitte Rouget**

In 2021, the historian Kerby A. Miller donated his collection of research material about Irish immigration to North America and Irish diaspora identities to the University of Galway Library. In March 2024, the first tranche of curated material was released online via a dedicated digital repository, titled Imirce. To make the materials accessible to academic users and the public alike with minimal delay, the Library opted for a digital-first approach. To define digital-first, we mean a methodology and a process that prepares materials for and releases materials to digital channels as our first priority. The foundations for this principle are accessibility and collaboration, in line with the Library’s strategic goals of being digital, driving research and opening scholarship.

In 2023, the entire collection (150,000+ pages) was scanned to develop and test a new digital procedure for the curation, description, ingestion and publication of an archives collection. Since the 1970s, hundreds of letters were shared with Miller which he and his research assistants then photocopied, transcribed and returned to emigrant descendants in Ireland and the USA. Miller understood that these records held significant sentimental value to the individuals he corresponded with, but with the tendency of heirlooms to go astray in the generations that have followed, Miller’s carefully collected derivatives may now represent the last existing versions of these records.

The first phase of the project has focused on Miller’s collection of Irish emigrant letters, and the curation of related pages of:

- digitised typed transcripts created by Miller and team,

- his collected original manuscripts; and

- his collected reproductions (copies) of manuscripts used to create typed transcripts for his research.

An example of a mixed material item can be found here: <https://t.ly/ccKax>

The proposed 10-min lightning talk will provide insights into the evolving digital curation workflow, including highlights of the following aspects:

- The opportunities and challenges of curating mixed material items (combining pages of Transcripts, Reproductions and/ or Manuscripts into a single digital object)

- Defining the quality methodology for digitally assessing duplicate transcript and reproduction pages for publication to Imirce

- Outlining the approach to Imirce’s search functionality and data visualisations, built on a MODS approach that is archives and DAMS-friendly, and supplemented by OCR transcripts for full text search

- Showing some of the benefits of working on a digital archives collection in an iterative way, with regular releases of newly curated material

Forthcoming: Miller’s collection of emigrant memoirs will be curated and released online in late 2024. The launch of the digital repository also included an open call to the public to contribute new material to ensure the collection's continued growth and relevance. Areas of particular interest are letters written in Irish in North America, and letters and memoirs produced in any language by emigrants from Irish-speaking districts. So far, 35 new donors have come forward, pledging an additional 1,600+ new letters to the corpus.

Imirce is more than a digital archive—it's a bridge connecting scholars, descendants, and curious minds across time and the Atlantic Ocean.

Visit Imirce here: imirce.universityofgalway.ie

**Associated Data Stewards as Means to Broaden Domain-Specific Expertise by Marcus Schmidt, Joost Albers, Eri van Heijnsbergen, Kerstin Belin, Hanna Lindroos, Shauna Ní Fhlaithearta and Nikolai Svoboda**

Although well-trained staff at the facility can help with research data management requests, they are generally not able to handle all cases or are not provided for in the facilities. This is where domain-specific helpdesks come into play, providing additional expertise. However, even these staff soon reach their limits in tricky scenarios resulting from variety of data formats, workflows and available repositories. Agricultural data management requests for example may range from large geodata to growth parameters all the way to genetic information and the proper handling of qualitative data. In these cases, it is paramount to be able to rely on a functioning network of associated data stewards who are experts in their (sub)field of research to engage with expertise requested. In this talk, we will present our concept of associated data stewards: When a request reaches a helpdesk and goes beyond its respective expertise, it is forwarded to a committed data steward, researcher or pool thereof, from another - in this case European - institution to which the request fits thematically, methodologically or locally. The associated data stewards may answer the request and return it, or several data stewards from interlinking fields may discuss and iterate the request. Currently, the Leibniz Centre for Agricultural Landscape Research (ZALF, Germany), the Wageningen Data Competence Center (WDCC, The Netherlands) and the Swedish University of Agricultural Sciences (SLU, Sweden) are piloting a domain-specific associated data steward network for agricultural research data management. In this talk, we will highlight the proposed setup and advantages of this network which has three important prerequisites: a) a clear definition of the agricultural sub-domains covered by each helpdesk including expertise of each data steward, b) a mutual understanding of each other's workflows and technical solutions and c) a structured workflow used to distribute requests which also re-informs the original receiving helpdesk of the result of the request (monitoring and feedback) - all the while complying with data protection regulations. The presentation uses agricultural data as a real-life example but reflects on issues and possible solutions that apply to a variety of research domains.

**Automating Digital Preservation with Archivematica’s API and Python by Guanwen Zhang, Kenton Good and Sharon Farnel**

Preserving large datasets by Archivematice becomes increasingly less effective using its web-based interface. In this study, we would like to highlight an automation application tool implemented by us in Python by leveraging the RESTful API coming with Archivematica. The automated workflow encompasses pre-processing of digital objects to ensure consistency and accuracy of digital objects to be preserved, batch uploading, unpacking, checksum verification and/or generation, format identification, normalization, metadata extraction and creation, and creation of Submission Information Packages (SIPs), Archival Information Packages (AIPs) and Dissemination Information Packages (DIPs). It also enable us to monitor the progress, status and resource usages of batch submitted digital preservation jobs, allowing real-time computer resource budgeting and cleaning and job scheduling. This approach was applied to preserve the Sir Samuel Steele Collection curated and digitized by the University of Alberta Library (UAL). It is evident that this automation approach not only enhances operational efficiency but also ensures consistency and accuracy through the reduction of frequent human intervention, which is both time-consuming and error prone.

**Barriers to Data Sharing in Infrastructure Systems Engineering by Brian Matthews, Katie Cartmell, Adrian Hines, Catherine Jones and Elizabeth Newbold**

Data Sharing is recognised by policy makers as a key enabler for research. However, different research communities are at different stages in adopting the cultural and technological changes required for practical research data sharing. Engineering sciences are seen as relatively low adopters of data sharing in comparison with other research disciplines. Further advocacy and practical development need to be undertaken for engineering researchers to further adopt data sharing practices in the future.

The Data and Analytics Facility for National Infrastructure (DAFNI) is leading a study, Data Infrastructure for National Infrastructure (DINI), into data sharing practices within infrastructure systems engineering. In this domain, we are interested in the engineering of systems that involve the interaction of physical components, machines, or buildings within a common ecosystem such as water or energy supply, transport networks or telecommunication links, and the impact of systems within the natural and human (economic and social) environments. The impacts of this research are less on the construction and maintenance of assets, but rather on the wider commercial and government policy landscape.

Data sources within Infrastructure Systems Research are often the operational data of commercial enterprises or local or national government rather than the “facts-of-nature"" from prior research. Obtaining such data requires negotiating with commercial and government entities, and commercial confidentiality and IP are of high concern. Some data can fall under Critical National Infrastructure as it may expose vulnerabilities to hostile actors, while other data may be personal sensitive, for example energy usage information at a household level. Further, for impact, researchers need to share results with stakeholders in government and industry, and data providers may be willing to share data with researchers, but they may be reluctant to share with impact partners.

Data sources are also highly diverse, from many organisations and systems, and the data needs to be interoperable with data from other domains, such as environmental sciences, social and economic data, health data, and within an underlying geospatial context. This increases the diversity of data and its supporting tools and standards.

These characteristics form significant barriers to data sharing; however, the potential impact of enhanced data analysis on society could be great by improving decision making and investment in infrastructure. DINI is consulting across the community, focusing on a representative sample of researchers and stakeholder organizations from the energy supply, water supply, and transport (road, rail, air) sectors to explore the barriers and opportunities for data sharing, and identify best practices, supported by use cases. The results of this engagement will be analysed to develop recommendations for further actions to support and enhance data sharing within this sector. In this presentation, we will discuss the challenges of data sharing in this sector, present the results from the community consultation, and propose recommendations for further action to support data sharing and reuse.

**Base4NFDI: Fostering A Cross-Disciplinary Service Landscape For The German National Research Data Infrastructure**

Since March 2023, the NFDI (German National Research Data Infrastructure), funded by the German Research Foundation (DFG), comprises the final 26 discipline-specific consortia. These consortia cover a wide range of scientific disciplines, from the humanities and social sciences to engineering and the life and natural sciences. They focus on developing or enhancing tools, services, and workflows to enable FAIR (Findable, Accessible, Interoperable, and Reusable) data sharing in their respective disciplines. They also drive cultural change towards an open data culture. This will professionalise RDM in a very diverse landscape of technical and organisational solutions.

Base4NFDI is a joint initiative of all the 26 discipline-specific consortia to develop essential basic services through a proposal-driven, bottom-up process. It aims to ensure overall coherence of emerging basic services from a sociotechnical perspective and streamline their development. In this talk we introduce the project Base4NFDI and examples of the basic services funded. We outline the decision making process and how the development of the services is supported, e.g. with consulting and training. We will discuss the challenges faced in the first two years and provide an outlook on questions regarding interoperability at an international level. Proposals for basic services arise from so-called NFDI sections, where cross-domain topics are addressed beyond disciplinary boundaries of the consortia. The sections combine infrastructural and technological expertise with domain knowledge on topics such as Common Infrastructures, (Meta)data, Terminologies and Provenance, Training and Education, Ethical, Legal and Social Aspects, and Industry Engagement.

Basic services encompass technical-organisational solutions, including storage, computing services, software, processes, workflows, and necessary personnel support, provided. They typically bundle or scale existing services and are characterized by scalability and sustainable operating models, developed during the funding period in Base4NFDI. Basic services have to pass three phases: initialisation, integration, and ramp-up for operation. Each phase requires an application and can only be funded after a rigorous review procedure, assessing whether user needs are met and technical quality and coherence criteria are fulfilled. The development of basic services is supported across all phases by Base4NFDI staff. For example, Section Liaison Officers support the NFDI section in preparing basic service development strategies and Service Stewards ensure smooth interplay between sections, NFDI consortia, and partner organisations involved in service development. A NFDI-wide basic service aims to benefit most or all consortia and their users, significantly enhancing the efficiency of the German research community. The first service candidates under development include an authentication framework, a terminology service, a PID and a Jupyter service. After successfully passing the three phases, basic service candidates will be added to NFDI’s service portfolio. As services mature, questions about KPIs, service quality, models for long-term operation and interoperability will need to be addressed. Particularly at the international level, integration with the European Open Science Cloud (EOSC) will be a key focus.

**Caring Types: How Levels Of Retention, Curation And Preservation Drive Transparent Repository Metadata by Hervé L'Hours, Oliver Parkes, Jonas Recker, Mari Kleemola and Olivier Rouchon**

Clearer, common descriptors of repository types and the levels of care they provide are needed to enable the design, planning and implementation of research infrastructures. Explicit and transparent metadata support the range of necessary retention, curation and preservation activities and functions across communities.

As plans to deliver more interoperable and federated scientific infrastructures (including EOSC) have evolved and digital object-focussed criteria (such as FAIR) have emerged, a growing range of actors are involved in repository-related fields. Researchers, funders and other stakeholders need ways to better understand repositories’ services, reliability and sustainability. Clear, common descriptors of repository types and the levels of care they provide, supported by explicit and transparent metadata, can provide orientation around the range of retention, curation, and preservation activities and functions provided by research infrastructures.

The notion of repository criteria, assessment and certification evolved from reference modelling through to the development of standards and processes related to trustworthy digital repositories. The common assumption amongst these approaches was for organisations hosting or sharing digital objects to take responsibility for preservation over time. Both inside and outside the libraries and archives specialist communities the definitions and assumptions around preservation have broadened as a wider number of institutions self-label as repositories and as trustworthy. This beneficial growth and spread has also lead to some blurred concepts and gaps in communication.

More recent applications for the CoreTrustSeal are from a broader range of services, including those not in scope for certification due to the absence of active long term preservation services. For internal and wider community benefit they developed a community driven position paper on the levels of care provided across different repositories. This led to a follow up discussion paper on using the levels to define repositories, including beyond those providing active preservation. The authors (including repository staff and CoreTrustSeal Board members) worked through the implications of these levels and types to consider what structured metadata their adoption would imply.

Taken together, this approach places transparent retention information at the foundation of increasing tiers of care and responsibility from the deposit compliance criteria during appraisal and selection, through to initial curation measures to meet standards and quality expectations. These are all aligned with a more explicit statement about what it means to provide active long term preservation

This talk will examine the drivers for a new perspective and the potential for a future that more clearly defines the responsibilities undertaken by those caring for digital objects in interoperable and federated infrastructures. "

**Celebrating 15 Years of 4TU.ResearchData by Daniel Bangert and Madeleine de Smaele**

2025 marks the 15th anniversary of 4TU.ResearchData, the international data repository service for science, engineering and design. 4TU.ResearchData started as an initiative of the then three technical universities in the Netherlands, later expanding to the four technical universities as part of the 4TU.Federation.

Beginning in 2010 with a collection from the Darelux project, consisting of hydrological measurements collected in a river catchment area in Luxembourg, the repository has grown to 10,000+ datasets coming from various disciplines within the science, engineering and design domains. In addition to offering curation, sharing, long-term access and preservation of research datasets and software, 4TU.ResearchData has established an active community around it to enable members to connect and exchange knowledge about best practices for the creation and reuse of FAIR data and software within the technical sciences. This community of researchers and data professionals form topical Working Groups and stimulate and broaden the impact of FAIR data.

One of the characteristics that set 4TU.ResearchData apart from other, more generic repositories, is the support of data in NetCDF format via the OPeNDAP protocol. One of the advantages of the OPeNDAP protocol is that you can query a dataset, retrieve subsets of files, without the need to download the complete dataset first. Building on open standards and protocols to ensure operability between systems, in recent years we have also developed our own free and open source repository software as part of our strategy to support open science.

In this lightning talk, we will reflect on the challenges we faced during the development of the repository service, discuss changes and trends we have observed over time, and look forward to future, innovative developments we consider necessary to best serve the community. We will include not only aspects related to the technical infrastructure on which the repository runs, but also discuss human expertise, capacity building and community involvement as vital aspects of a constantly innovating, high-quality service."

**Combining Algorithms and Human Expertise: OpenAIRE's Entity Disambiguation Method by Miriam Baglioni, Claudio Atzori, Ivana Končić, Stefania Amodeo, Alessia Bardi, Gina Pavone, Bojan Macan and Martina Žugaj**

The OpenAIRE's solution to the “entity disambiguation” problem: a hybrid data curation method that combines deduplication algorithms with the expertise of human curators to ensure high-quality, interoperable scholarly information.

**Curating Archive and Library Special Collections for Large-Scale Environmental Data Collection: A Case Study by Nicole Elizabeth Wood**

This lightning talk will discuss preliminary findings from a case study conducted at the Huntington Library, Art Museum, and Botanical Gardens, which sought to identify current practices for organizing, digitizing, and describing materials suited for the study of environmental history and to ascertain best practices that facilitate easy, large-scale retrieval of data from sources traditionally siloed for qualitative use. The case study specifically attempted to develop a long-term spatiotemporal record of rainfall in California prior to the implementation of rain gauge measurements using print, digitized, and born-digital collections. In addition to sharing findings on the curation practices that may support or impede the large-scale data collection of environmental observations, this lightening talk will discuss the ethical considerations of curating digital collections for large-scale data retrieval.

**Data Management And Archives – Where Do We Go From Here? By Alexandra Stam and Auriane Marmier**

Notions of the importance of openness, transparency, and data sharing for the betterment of science have been around for decades, and data archives have been significant players in promoting such notions within research institutions and different scientific communities. To fulfil their mission of preserving and disseminating high-quality data, and in the absence of the needed skills for data sharing among researchers, archives often took on a lead role in training researchers in relevant aspects of data management. For instance, at the Swiss Centre of Expertise in the Social Sciences, FORS, we have since our founding in 2008 maintained our national data archive as well as provided trainings, guides, and consultations in data management for the community of social science researchers in Switzerland.

However, with the rapid acceleration of the open data movement in the last decade or so, the role of archives regarding data management has begun to change. Policy changes among funders and journals have led to new pressures on researchers to comply with data sharing requirements. This has meant for archives an increase in submissions and workload, often with new and challenging data types (e.g., sensitive data). So, the related intensified needs of researchers in data management present challenges to archives with limited resources in this regard. At FORS, we have recognised that the current demand far outweighs our capacity to provide the needed trainings and advice. This may only amplify in the future with new initiatives whereby universities may move towards giving formal credit to researchers for data sharing for hiring and promotion purposes.

Fortunately, in recent years research institutions have begun to step up efforts to support researchers in relation to the data sharing requirements, specifically by way of data stewards and data librarians within their institutions, as well as intensified and broadened trainings in data management. Also, it is not inconceivable that data management will gradually become part of the core curriculum in graduate programs for researchers across disciplines.

For this lightning talk, we will briefly present how FORS has been involved in the provision of data management training over the last years according to researchers needs and the evolving open research data landscape. We will then look ahead, and address the question of the role that data archives should play in advancing open research data in the future, specifically with respect to data management. If institutions continue to invest in data stewards, data management trainings, and support services, what should be the nature of the relationship between archives and these institutional services? If data management is at some point fully covered within research institutions, how should resources be reallocated within archives? What should archives do in addition to facilitate the acquisition and sharing of high-quality data? We have some ideas on this and will open discussion on the question “Where do we go from here?”.

**Dataset DOIs for Improved Findability, Acknowledgement of Data Creators, and Tracking of Data Reuse by Kalynn E Kennon and Jennifer H Lee** Launched in 2016, The Infectious Diseases Data Observatory (IDDO) is a scientifically independent, multi-disciplinary coalition of the global infectious disease community. IDDO aims to expand data reuse and collaboration within the global health community to deliver better treatment and control of poverty-related infectious disease and emerging infections, expanding its proven model of evidence generation and building a unique global data library. It provides the methods, governance, and infrastructure to translate data into evidence that improves outcomes for patients worldwide. IDDO builds on and incorporates the pioneering work of the WorldWide Antimalarial Resistance Network (WWARN), a unique, decade-long collaborative data-sharing framework that proved it was possible to produce policy-changing scientific evidence from historical data, benefitting the health of millions.

Ethical data reuse is at the heart of IDDO’s mission. Barriers to data sharing and data reuse such as lack of incentives or concern about ownership or receiving credit can be overcome by tracking metrics for data reuse in a way that credits the original creators of the data assets. Acknowledgement of the original creators of data resources is an important part of that work, especially as much of the data held by IDDO is generated in low and middle-income countries (LMICs).

We mint a Digital Object Identifier (DOI) for every dataset contributed to the IDDO platform. This unique, persistent identifier (PID) contains rich metadata describing the data asset as well as information about the creators of that resource. Linking these creator names with their own PIDs, such as ORCID IDs or ROR IDs, allows for the tracking and linkage of data assets through the reuse lifecycle. This provides a clear, traceable record of data requests for reuse that can generate reuse metrics for a data asset and measure the impact of that resource beyond its original creation and use.

The minting of DOIs also supports FAIR principles by the creation of rich, descriptive metadata about each dataset contributed to the IDDO platform. This increases the visibility and findability of datasets which in turn facilitates further reuse and generation of new evidence. Linking personal PIDs to these DOIs enables the tracking of data reuse through citations and publications that reference the study DOI (as required by IDDO’s Data Sharing/Use Agreement). These metrics can then be linked to institutions, investigators, and others involved in creating that data asset and can be tracked and visualized via resources like DataCite.

IDDO’s practice of requiring those who request data from the platform for reuse to reference the individual dataset DOIs ensures proper attribution and recognition of the data creators and ensures researchers receive due credit every time their data is reused. These quantitative metrics allow for a transparent and traceable method of measuring impact of individual data assets over time as they are used again and again to generate new evidence by the research community."

**Development of CI PAOS and DCL on RDM at the U.S. National Science Foundation (NSF) by Plato Smith, April Bernard and Marie Coppola**

The Directorate for Computer and Information Science and Engineering Office of Advanced Cyberinfrastructure at the U.S. National Science Foundation make strategic investments to support public access and open science. One such strategic investment includes the new Cyberinfrastructure for Public Access and Open Science Program (PD 24-7414), a funding opportunity made available to the public on April 8, 2024. The Cyberinfrastructure for Public Access and Open Science program supports the Public Access and Open Science working group at the U.S. National Science Foundation that leads internal and external public access and open science efforts and collaborative initiatives. The innovative program also launched the first funding opportunity on Leveraging Cyberinfrastructure for Research Data Management (RDM) through a Dear Colleague Letter on May 1, 2024, (NSF 24-085). The lightning talk will discuss development of the Cyberinfrastructure for Public Access and Open Science program, innovation on Leveraging Cyberinfrastructure Research Data Management, and the program’s contributions to the agency’s public access and open science working group broader efforts.

**Development of DMP Services since 2015 by Mari Elisa Kuusniemi and Soile Manninen**

We regularly provide feedback on DMPs and now is time a review of our own performance. To achieve this, we analyse DMPs stored since 2015 by conducting a content analysis of the feedback provided by the data team. In addition, we interview the people who have supported data management planning, including a long-standing members of the data team, more recent recruits and experts outside the data team such as IT specialists and legal experts. This analysis will identify changes in the focus of our feedback, shifts in comment topics, and the evolution of DMPs over time.

Our study has two main objectives: firstly, to document our work and identify potential areas for improvement, and secondly, to improve our feedback guidelines and DMP instructions for researchers. The findings of this study will also support the development of the university’s machine-actionable DMP (maDMP) template.

Our aim is to improve our data management planning support to ensure it effectively meets the needs of researchers.”

**Digital Preservation Resources For Higher Education** **by Paul Stokes and Karen Colbron**

Those of us that are already active in digital preservation know that there are many excellent resources for beginners (and experts) out there, but, as is often the case with people starting in a new relatively strange activity, they can be difficult to find for newcomers. If you’re not in the know, if you don’t know how to describe or label what you’re looking for then it can be very frustrating.

This frustration can be exacerbated when terminology that has multiple (often subtly) different definitions across domains is employed. Every community has its own glossary and jargon and there will often be a need for a ‘bi-lingual’third party to translate from one domain to another.

And that’s where Jisc comes in.

We at Jisc are a membership organisation. Our core membership lies within the higher education and further education sectors. So, although we do undertake a wide range of activities in many different sectors, we are (unsurprisingly) very much focused on the needs of our core education members. Part of the remit of the Content and Discovery group within Jisc is to help our members with advice and guidance around digital preservation. With that in mind we are acting as the “bi-lingual third party”and are currently compiling a list of available resources that are of particular interest to /useful for the education sector. We speak education/research AND preservation.

We’re currently at the literature review / gap analysis stage and this lightning talk is intended to provide an overview of the resources identified to date, how they are of particular use to the education sector, describe any perceived gaps, and outline what might be done about those gaps (and by whom). We intend to use this information to present about more rounded education focused resource at a future conference.

**DMP Assistant: Stabilizing for the Future by Marcus Closen**

DMP Assistant is the Canadian solution for data management planning. It provides a national, bi-lingual platform for Canadian researchers, based on DCC's DMPOnline. In the two years we have undertaken to stabilize our platform, as we think about the future of data management planning in the Canadian context.

**DMPs and Sensitive Data by Marcus Closen and** **Victoria Smith**

We evaluate cases for data management plans (DMPs) as public-facing outputs for researchers to use in open science contexts when data preclude sharing or reuse through our work in research data management in Canada.

**Elevating Community-Led Digital Curation Resources to an Institutional Level by Lena Karvovskaya, Chris Hartgerink and Elisa Rodenburg**

This presentation explores a case study from Digital Curation in the context of a research organisation. We present the ongoing work on creating and maintaining a community-led university resource with guidelines and best practices.

Digital Curation is an important part of research and Open Science as it ensures long term preservation of research data, replicability of research results, and trustworthiness of research in general. Many open science resources are community-led, creating a sense of shared ownership and engagement.

Inspired by community-led resources like the Turing Way, the VU Amsterdam Research Support team started working on a collaborative handbook that is community-driven and -maintained. This enables VU Amsterdam colleagues to co-create guidelines that are relevant and useful to them. The handbook is meant to complement official university webpages which are managed by the university-wide communications department. Such institutional pages are challenging to update according to the user needs - they risk to become outdated and disconnected from the daily life of researchers and their needs.

The field of research data management is dynamic and changes rapidly. Finding up-to-date and trustworthy information that is immediately relevant for one’s work is critical to successfully fulfilling one’s role. There is an abundance of resources that focus primarily on the national and international level. As a result, current resources are not always applicable to the local context. The user experience research conducted for VU Amsterdam in 2020 and 2023 highlighted that official institutional pages mostly fail to guide the users (in our case researchers) to the information they need. They often turn out to be outdated, fragmented and not interlinked in a proper way. Working on a community-led resource offers flexibility and makes it easy for the community to update resources, signal problems, and suggest improvements.

At the same time, we observe that the community approach does not fit well into the university way of working. Various departments struggle to endorse it without the certainty of what they are endorsing. Yet, community-led means that the outcome is not predetermined. In our efforts to create community resources and having them accepted as institutional resources, we encounter an institution that struggles to accept working openly, despite all our efforts to change the culture at our institutions. It highlights how the wider culture change towards working openly should not overlook the departments that are not present within the open science movement. We want to share some of our success and highlight the challenges we encounter and look forward to hearing from the audience about similar resources they’ve been developing.

**Emerging Challenges in Sensitive Research Data Management in Québec**  **by Alisa Beth Rod, Caroline Baril, Stéphanie Pham-Dang and Teresa Bascik**

This lightning talk will provide an overview of a study involving academic researchers in Québec who collect, manage, and share sensitive research data containing PII. The study will focus on their perspectives regarding the impact of Law 25 on their research and their ability to align with RDM best practices (e.g., the FAIR principles). In addition, we aim to evaluate the role of librarians and data curators in supporting their work, and their experiences in using compliant infrastructure platforms or services. This novel study is designed as descriptive in nature, aiming to highlight practical needs and situational experiences around the topic of sensitive research data in the context of Québec’s regulatory environment. Data collection will rely on a semi-structured interview methodological approach to be conducted in Fall 2024. We aim to interview approximately 10 participants consisting of faculty researchers with appointments at an academic institution or affiliated research hospital in Québec. We will present preliminary findings from an analysis of qualitative interviews with Québec-based researchers working with sensitive research data.

**European Perspectives on Training Data Stewards by Nida van Leersum and Saba Sharma**

Training for Data Stewards is necessary given that there is a need to professionalise Data Stewards within research performing organisations to advance Open Science skills. Training should be based on the identified skills and competencies that are required by Data Stewards. Given that Data Stewards work within different institutional and country context, this lightning talk will focus on how training can be developed that can be tailored for different contexts while still addressing the knowledge and skills required by entry level data stewards.

**FAIR Principles Implementation in ML/AI - Findings from Skills4EOSC by Delphi Study**

Implementing FAIR principles in ML/AI data and models could have several benefits. These include facilitating education in data science and machine learning, ensuring sustainability of fields reliant on AI and ML, enabling of collaborative research and development and setting of benchmarks for interpretability of models, etc. However, researchers and curators currently do not have clear, practical guidelines or consistent metrics to support, evaluate and improve FAIRness of ML/AI data and models. Skills4EOSC Work Package 6 Professional Networks for Lifelong Learning is running a Delphi Study towards gathering consensus among experts on the implementation of FAIR principles in the development of ML/AI models.

The Delphi method involves two rounds of surveys and a final round of focused and detailed discussion. In the first round, participants rate suggested practices related to FAIR principles implementation in ML/AI, and they also have the options to motivate their rating and suggest additional practices they feel are important and/or feasible. In this approach, the initial selection of suggested practices, obtained through literature review and consulting a few researchers in the field, will be refined and augmented by the participating experts. In the second round, the aggregated group opinion will be fed back to participants, and they will be asked to vote on practices a second time. Further, they will be asked to vote on any additional practices suggested during the first round. A total of 70 ML/AI experts from across Europe and beyond are participating in rounds one and two of the study. In the third round, selected participants will be invited to discuss and review the results, analyses and interpretations. The final output will be a “Top 10 practices for FAIR principles implementation in ML/AI”. The first round closes in August, 2024 and the second will run in October, with final discussions and reviews taking place in December.

The approach taken by Skills4EOSC recognises that while researchers are extremely busy and cannot be expected to produce these guidelines themselves, if practices are to be implemented, it is important that these be scrutinised, vetted and approved by the community. We hope that the “Top 10” will serve as a basis for clear guidelines that allow researchers, repositories and curators to work harmoniously in implementing FAIR principles in ML/AI, with the related benefits to science and society to follow.

**Fostering the Next Generation of Open Scientists: Insights from a Project on Encouraging Data Reuse in Students' Theses by Anna Daudrich, Elena Šimukovič and Simon van Rekum**

One of the primary objectives of data curation and Open Science is to ensure the reusability of research data. Currently, data reuse is predominantly centred on advanced academic research. We propose that, in order to establish data reuse as a common practice, it is crucial to begin teaching data literacy as early as possible in academic life, starting at the bachelor's level. This approach can ensure that the necessary practical skills and awareness are developed early on and are firmly anchored for the rest of young researchers’ careers.

In view of this, we launched the ROADS project (Reusing Openly Available Data for Student theses) in December 2023 at the ZHAW Zurich University of Applied Sciences, which will run until June 2025. Our objective is for students from ZHAW at various levels (Bachelor’s, Master’s, and PhD) across disciplines (ranging from media linguistics to photovoltaics) to reuse existing datasets in different use cases for their theses, term papers, and other assignments. The process of data reuse occurs under the supervision of researchers and lecturers from various ZHAW departments. Additionally, it is facilitated by the data stewards from the university library and a central service unit ZHAW Services Research Data.

With this project, our aim has been to expand the scope of data reuse to include not only research but also educational purposes, helping students (and, partially, their supervisors) to acquire the necessary competencies in Research Data Management. This will enable them to understand how to reuse data and encourage their engagement in such practices. Our initial findings suggest that limited data literacy is indeed one of the major hurdles to data reuse. Without sufficient competencies in searching for datasets, using specific software, or understanding what data quality means in each case, data cannot be effectively reused. In other words, datasets can remain unusable even if all necessary digital curation measures have been taken.

In this contribution, we will particularly focus on the educational strategies relied upon within the ROADS project, considering the perspectives of both students and their supervisors. Specifically, we will address the following questions: Which skills do all participants need to reuse data effectively? Where are the major gaps? What strategies have been implemented to overcome these gaps? What are the main challenges regarding the integration of best practices in Research Data Management into academic curricula? By sharing our findings, we hope to contribute to making data reuse and Research Data Management an integral part of the academic process, and to help sustainably foster a new generation of open scientists.

**Frameworks for Evaluating Digital Sustainability in Distributed Research Data Ecosystems by Heather Lynn Barnes, Kira Bradford, Allie Gartland-Gray and Alison Banger**

Research institutions, funders, and policymakers are wrestling with the exponential growth of biomedical research data and the need to ensure it endures over time and remains sustainable. A distributed research data ecosystem, an increasingly common model for data sharing, provides a central discovery and search portal but depends on externally managed digital repositories such as ICPSR to host research datasets. Repositories within a distributed research system maintain their own governance, data archiving, and long-term preservation protocols, and it is currently difficult to assess the degree to which all repositories vetted for use in a distributed research ecosystem coordinate around a set of data sustainability standards. In addition, while sustainability is often considered an important element in planning and operating research data ecosystems, there is not yet a standardized set of metrics for assessing their long-term sustainability. In this presentation, we review common digital preservation frameworks and explore their potential for contributing to a new framework for evaluating sustainability within distributed research data ecosystems. In addition, we call for the digital preservation community to contribute to a conversation around designing new models for evaluating sustainability across a distributed research data ecosystem.

**From FAIR Research Data to FAIR Government Information by Deborah Yun Caldwell, Lynda Kellam and Shari Laster**

As one of the largest publishers of information and data in the world, the United States government has the potential to influence the widespread adoption of the FAIR Principles as a standard for curating information resources. Over the past decade, U.S. federal agencies have been increasingly using the FAIR Principles as a standard metric to ensure that research data funded by the government is curated and made available. For example, the National Institutes of Health (NIH)’s Data Management and Sharing Policy explicitly encourages federally-funded researchers to align their data management practices with the FAIR Principles. However, research data is a relatively small segment of federal government information dissemination. A breadth of products and published formats make up the larger digital government information ecosystem, ranging from ephemera such as pamphlets and PowerPoint slides to complex digital objects such as databases and interactive applications. Moreover, digital government information and its metadata are themselves data, making the ecosystem particularly amenable to a digital curation framework.

This presentation provides a starting point for considering the wider U.S. digital government information ecosystem within the framework of FAIR. We first take a look at how U.S. government information dissemination has evolved over the past several decades as the federal government has transitioned from print to digital dissemination models. In doing so, we will highlight the current gaps in what is collected, described, made available, and preserved for future discovery and reuse.

Next, we will explore how FAIR Principles can be applicable to the complex ecosystem of federal government information. We will present several examples and specific cases, analyzing them in relation to the FAIR Principles. Our goal for this project is to craft a model for applying FAIR Principles to our wider understanding of digital government information. We will close with a brief overview of a “possible future” in which curated U.S. government information is commonplace and contributes to social good worldwide and identify potential mechanisms to move us toward this vision.

**From Idea to Implementation: An Institutional Catalyst Helps Bring About the Yale Dataverse by Limor Peer and Barbara Esty**

This paper describes key ingredients in the evolution of the Yale Dataverse from idea to implementation: the establishment of a new organization and the involvement of domain expert curators.

**Fundamentals of Scientific Metadata – A Hands-on Training Course on FAIR Data Handling for Researchers and Data Stewards by Silke Christine Gerlich, Abril Azocar Guzman, Volker Hofmann and Stefan Sandfeld**

Scientific research has been subject to the fast-progressing digitalisation with impacts how research is conducted today. Generation and sharing of data according to best practices that support the digital change, bears numerous challenges for the scientists.

Implementation of data documentation recommendations like the FAIR Principles [1] require profound knowledge and technical skills and thus well-trained, data-competent, and technically skilled researchers. The education for this is, however, not part of typical academic curricula: More than 45 % of scientific staff state to have little to know prior knowledge about the FAIR principles and metadata handling [2]. However, general interest in training formats on these topics is high (91.7 %). The realization of FAIR research data thus depends on the profound and high-quality training material and training of scientific staff.

With our training course „Fundamentals of Scientific Metadata”, we established training material that covers the fundamental elements of (meta)data annotation and addresses early-career researchers of any scientific domain. The didactic concept of the course encourages and motivates the participants to begin and sustainably proceed with the structured, schema-conform documentation of their scientific (meta)data. The material covers the fundamentals of (semi-)structured metadata, schemas and standards, as well as persistent identifiers (PIDs).

(Meta)data annotation is a predominantly practical skill that should be acquired and consolidated in a hands-on manner. Therefore, our course makes use of familiar problems and interrelated exercises to encourage the participants to practically test and consolidate the newly acquired skills and concepts. An initially unfamiliar data object is annotated with increasingly structured metadata throughout the course, complying with the FAIR principles. We set our focus on the confident handling on JSON and the development and understanding of JSON schemas.

The training material was created in a modular manner that effortlessly enables the adaption of the material to various target groups in skill level as well as scientific domains: we have realized adaptions in the domains of Materials Science and Engineering and Particle Physics as well as for the target groups Researchers and Data Stewards. The domain-agnostic version of the training material has been published comprehensively via The Carpentries Incubator [3]. Publications of the adaptions are in preparation.

To date, 11 instances of the domain-agnostic course and its individual adaptions have been conducted. Each instance was met with overwhelming interest in participation. We conclude every course instance with a comprehensive participants’ evaluation. The evaluation results confirm the target group-oriented accuracy of the course contents as well as the high quality of our material.

**Implementing Reuse Practices for Research Data in the National Institute for Public Health and the Environment (RIVM) of the Netherlands by Andrés Felipe Ramos Padilla, Marjolein Kooijman, Ellen Leenarts, Jente Houweling, Nikée Groot and Inez Joung**

As part of the data strategy at the RIVM, we aim to improve the conditions and usual practices in which research outputs are prepared and shared for reuse, not only for researchers but also for the general public. We have been tackling the reproducibility and reuse issues via four strategies in the RIVM: education, implementation, evaluation, and improvements, to take a step forward in implementing research data reuse practices. In the lightning talk we present the most important findings and challenges of the four strategies.

**Introducing Doctoral Students to Data Management Plans. Insights from a Transversal Skills in RDM course by João Aguiar Castro**

This lightning talk will present the "Transversal Skills in Research Data Management" course provided by the Faculty of Engineering at the University of Porto. This bi-annual course is aimed at doctoral students,with a focus on creating comprehensive DMPs, following FAIR data principles. The talk will cover the course structure, students results and key lessons from seven editions, concluding with insights on the importance of training researchers in RDM.

**Meeting Domain-Specific Requirements With The 4TU.ResearchData Repository Madeleine de Smaele, Roel Janssen, Daniel Bangert and Jan van der Heul**

4TU.ResearchData is an international data repository service for the science, engineering and design domains, led by the four technical universities in the Netherlands.

Since 2023, the repository service runs on in-house developed free and open source software called Djehuty. Through Djehuty, we are facilitating contributions from the 4TU.ResearchData community and are collaborating on its use as a multi-organisational repository system. For example, Nikhef, the Dutch Institute for Subatomic Physics, recently launched their data repository based on Djehuty.

To meet domain-specific needs, 4TU.ResearchData supports access to data using open, data-optimal protocols like OPeNDAP. The main advantage of OPeNDAP is the ability to retrieve subsets of files without the need to download the complete dataset first. Through our OPeNDAP service we enhance access to specifically netCDF datasets; a data format and model that, although generic, is mainly and widely used in climate and atmospheric sciences and oceanography. Presently, about 50% of the data stored in our repository is atmospheric and environmental research data coded in netCDF.

Another area in which we are exploring options for 4TU.ResearchData to expand its services is related to the sharing and reuse of research software. Software is increasingly recognized as a key research output, requiring the same level of transparency, accessibility, and disclosure as data. As a result, we have observed a growing number of software publications in our repository and are developing methods to support software reproducibility, discovery and transparency. For example, the 4TU.ResearchData approach of protocols-first has resulted in the implementation of the Git protocol to enable a smooth integration of any git repository and 4TU.ResearchData. This allows users to easily publish their code associated with a research output, get a DOI (Digital Object Identifier) for their code/software and ensures that the code/software will be available in the future. Further, as part of our collaboration with the Netherlands eScience Center, we are working on an integration with the Research Software Directory, to enhance the discovery, reuse, impact and citation of research software.

In this lightning talk we will address and showcase the value of community contributions and collaborations to meet domain-specific needs and share future plans to support these further.

**Navigating the Future of Digital Curation: The Transformative Role of Persistent Identifiers by Steffi Genderjahn, Andreas Czerniak, Marc Lange and Heinz Pampel**

Persistent identifiers (PIDs) are crucial for the development of a common infrastructure for open science and contribute to scientific integrity. They enable the unique and persistent identification of research objects and scientific resources. At the same time, the associated descriptive metadata contain information about the resource and are essential for a digital, interconnected scientific landscape. Globally, the development and adoption of PIDs for various institutions has gained momentum as stakeholders have recognized their importance in improving research visibility, attribution and data sharing. This progression of PIDs has fundamentally changed the way researchers and resources are identified and connected in the global research ecosystem.

Targeted surveys and workshops organized for the German research landscape by the projects PID Network Germany and PID4NFDI help to identify the use of PIDs as well as the needs and challenges in different research and infrastructure institutions and promote the networking of different open PID systems on a national and international level. The latest survey results provide information on the current status of the use and implementation of PIDs at research and higher education institutions in Germany. In the lightning talk, we will present the results and place them in the context of existing findings.

The PID Network Germany project builds on the successful efforts of ORCID-DE projects, two forerunner projects that ran from 2016 to 2019 and 2019 to 2023 respectively. It aims to further integrate PIDs into the research landscape and address existing challenges. In this context, aligning efforts with principles such as ACTION - Advocacy, Collaboration, Training, Implementation, Outreach and Networking - has proven effective in overcoming reservations and in promoting the widespread use of PIDs in different contexts.

The presentation will address the extent to which awareness of different PID systems has changed over the last decade and what is being done to promote their establishment. By examining progress in the research, cultural, and funding institutions landscape, we aim to provide insights into best practices for the coming decades of digital curation and propose actionable strategies, such as the development of a PID roadmap.

**Novel SDTM Implementation to Maximize Benefits of Sharing Legacy Data by Jennifer Hanbyul Lee and Kalynn Kennon**

The Infectious Diseases Data Observatory (IDDO) is a scientifically independent multi-disciplinary coalition of the global infectious disease community. It is a data sharing platform which hosts poverty-related infectious disease data and promotes data sharing and re-use. IDDO is unique in that we curate individual participant data (IPD) in-house from many different sources, to produce freely available harmonised datasets, enabling scientists to answer new research questions from existing data.

IDDO is built on and incorporates the pioneering work of the WorldWide Antimalarial Resistance Network (WWARN). WWARN is a collaborative data-sharing platform begun in 2009 that proved it was possible to produce policy-changing scientific evidence from pooling existing data. Pivoting from WWARN’s success, IDDO developed to fulfil the need of various infectious disease research communities to create a similar IPD sharing platform. The IDDO data sharing platform hosts IPD for over one million participants across a wide range of poverty-related infectious diseases data including malaria, Covid-19, Ebola virus, Visceral leishmaniasis, Chagas disease, schistosomiasis and soil-transmitted helminthiases, and antimicrobial resistance with more being scoped for feasibility.

Due to the disparate nature of the IPD contributed to IDDO, transformation to a single data standard was key – to alleviate the burden of having to create a unique standard and repository for each disease type and, to harmonise the data allowing pooling for secondary analysis of IPD across disease types, interventions, and populations. IDDO chose the Study Data Tabulation Model (SDTM) developed by the Clinical Data Interchange Standards Consortium (CDISC), which provides powerful tools for collection, storage, and analysis of clinical trial data. The comprehensiveness of CDISC SDTM allowed for aggregation and storage of standardised data while providing the flexibility needed to accommodate the intricacies of the real-world data IDDO receives from various clinical trials and observational studies.

The SDTM model and implementation was originally created to organize clinical trial data for submission to drug regulatory authorities. There were instances where standard implementation of the SDTM model was not appropriate for the legacy and real-world data contributed to IDDO. Therefore, while embracing the core SDTM model and adhering to the rules and original intent for the standard, IDDO has made use of a bespoke implementation of SDTM. We have made specific adaptations with the aim to maximise usefulness and usability of the data for pooled analysis, especially to those unfamiliar with CDISC standards, to encourage reuse of this limited resource and enable the generation of new evidence to impact public policy and treatment guidelines to improve the lives of those suffering from these diseases.

This talk will provide an overview of our data reuse platform, a single repository that allows for pooling of existing data across all infectious diseases data held by IDDO. We will discuss the challenges inherent in this process as well as our solutions, including customised adaptations of CDISC SDTM to promote findability, accessibility, and interoperability of legacy and real-world patient data, allowing for the generation of new evidence and extending the life of existing data resources beyond a single clinical trial."

**Open Science and Open Government at RIVM: A Balancing Act between Data Privacy and Utility by Johanne Margrieta Houweling, Ellen Leenarts, Marjolein Kooijman, Andrés Felipe Ramos Padilla, Sandy Chung-A-Hing, Inez Joung**

The Dutch National Institute for Public Health and the Environment (RIVM) addresses the challenges of Open Data, Open Government, and Open Science through the use of Statistical Disclosure Control (SDC). SDC minimizes re-identification risks in data sharing while preserving data utility, a balance that involves assessing risks, applying anonymization techniques, and measuring information loss. Given the complexity of these tasks and the diverse nature of the data handled by RIVM, simple, universal rules are insufficient. Instead, RIVM has formed a multidisciplinary SDC working group consisting of data stewards, statisticians, and privacy experts. This team not only advises on but also creates guidelines for SDC, enhancing organizational knowledge through training and presentations. They use the open-source tool sdcMicro to aid their operations.

This lightning talk will delve into how RIVM implements SDC, with a particular focus on examples from the COVID-19 pandemic, illustrating the shifting landscape of data sharing expectations. The talk demonstrates RIVM's strategies for maintaining public trust and meeting ethical and legal standards while promoting transparency and advancing Open Science and Government initiatives."

**Piloting maDMPs for Streamlined Research Data Management Workflows by Maria Praetzellis**

In response to the growing federal mandates for data management and sharing, the California Digital Library (CDL) and the Association of Research Libraries (ARL) have launched a pilot project to test new workflows at 10 U.S. universities. The project utilizes machine-actionable DMPs (maDMPs) to streamline processes, improve communication, and ensure compliance with federal requirements. This lightning talk will discuss the pilot by exploring three case studies that showcase the use of generative AI for crafting compliant data management plans, maDMPs for deploying computing and storage resources, and tracking research outputs through integration with internal data management platforms.

**Promoting Fair Principles And Research Integrity Through Manual Curation Of Manuscripts at the Point of Submission by Mary Ann Tuli, Yannan Fan, Bastien Molcrette and Chris Hunter**

GigaScience Press publishes award winning Open Science journals, which have the goal of making scientific communication reach researchers and communities around the globe. GigaScience focuses on ‘big data’ research from the life and biomedical sciences, and GigaByte focuses on short, data-driven articles, promoting a rapid exchange of scientific information. Both journals have implemented rigorous procedures to promote FAIR (Findable, Accessible, Interoperable, Reusable) data principles and research integrity starting at the point of manuscript submission.

Once it has been established that a new submission meets the scope of the journal, a thorough manual review is conducted to ensure compliance with FAIR guidelines and to verify research integrity.

Data curators confirm that all data underlying the study’s findings are made available in subject-specific public repositories (e.g. EMBL, ArrayExpress) or are provided as data files to be hosted in GigaDB, the database that accompanies GigaScience and GigaByte. Furthermore, they will verify that stable accession numbers and DOIs are assigned for deposited datasets and that the appropriate license has been obtained. Additionally, the metadata accompanying any datasets are assessed for completeness and quality to enable future reuse and reproducibility. Where a public repository is not available, authors will submit their data or tools to GigaDB. These files are manually inspected to verify they are in non-proprietary, machine-readable formats that align with community standards. Figures and images are also examined using specialised software tools which can identify signs of manipulation or duplication.

Data curators also inspect custom code and software developed for the study to ensure it is shared in a suitable public repository (e.g. GitHub) with the appropriate documentation and licensing.

A recent development has seen us work with the DOME consortium. The DOME-ML (Data, Optimization, Model and Evaluation in Machine Learning) recommendations establish standards of supervised machine learning validation in biology. For studies that use machine or deep learning methods, authors are encouraged to submit supporting annotations to the DOME-ML registry.

For studies involving human subjects or animals, data curators take great care to confirm that all ethical approvals are documented, and that data sharing complies with consent and privacy regulations.

In addition to the above assessment of the scientific data, the editorial team will verify the roles of all authors and assess any potential conflicts of interest and will screen all manuscripts for plagiarism or text recycling using similarity detection software.

This rigorous and comprehensive manual inspection of manuscripts ensures that issues related to data sharing, reproducibility, and research integrity can be identified and addressed prior to peer review which not only ensures higher quality submissions but also educates authors on best practices for open and reproducible science. Furthermore, reviewers can focus on evaluating the manuscript rather than spend valuable time ensuring data integrity.

The Press’ commitment to reproducibility through open access data sharing and digital curation even before the FAIR guidelines were introduced has put us at the forefront of Big Data publishing in the life sciences, a position we are committed to maintaining over the next 20 years and beyond.

**RDA TIGER: Soft Infrastructure to Facilitate and Shape Community-Driven Standards for Research Data Management by Liise Lehtsalu, Ryan O'Connor and Athina Papadopoulou**

RDA TIGER is a Horizon Europe funded project, led by RDA Europe, that develops and provides services to RDA Working Groups throughout the working group lifecycle. One of the key services developed is the RDA TIGER Facilitation Service. The Facilitation Service is soft infrastructure that supports RDA WGs in the production of community-based standards and best practice development in research data management. The success of the RDA TIGER Facilitation Service evidences the growing importance of soft infrastructures in digital curation and research data management. By providing building blocks for social bridges through a proactive community liaison approach, the RDA TIGER Facilitation Service creates human capital for the Working Groups (WGs) and supports them in confronting technical and semantic interoperability challenges and bridging domain-silos.

RDA is an international, community-driven initiative that aims to build the social and technical infrastructure that enables data sharing, reuse and data-driven innovation. Within the RDA, members initiate WGs that focus on co-developing practices, tools, policies or products in support of FAIR data sharing and reuse. While over the years the RDA WGs have proven extremely efficient in co-developing internationally accepted recommendations and best practices, they are also volunteer-led efforts with related challenges. In particular, ensuring effective and sustained engagement of relevant stakeholders, developing and keeping to a work plan, administrating the regular meetings and record keeping of the WGs, and ensuring timely output finalization and broad-based dissemination can prove challenging to volunteer-led WGs and limit the impact of their outputs. The RDA TIGER project, coordinated by RDA Europe and involving CODATA, the Netherlands eScience Center, and DANS as partners, responds to these challenges and provides full-lifecycle support services to WGs, ensuring that the WGs effectively move from ideation to delivering FAIR-enabling results within defined time frames.

In this lightning talk, we introduce good practices in community facilitation that the Facilitation Service is developing and evidence the value of such soft infrastructure in furthering community-based standards and best practice development in research data management. RDA TIGER project currently supports 18 RDA WGs in various stages of their lifecycle; thanks to the support of the Facilitation Service, WGs are effectively progressing, reaching out across disciplinary and geographical silos, and delivering outputs that are of cross-domain relevance and contribute to resolving major technical and semantic interoperability challenges faced by the data curation and research data management communities. The RDA TIGER facilitation model is being tested and validated by the RDA TIGRUS project led by RDA-US; the TIGER and TIGRUS projects are in close contact with each other, ensuring that both services are developed for the good of the community. We argue that soft infrastructures like the RDA TIGER Facilitation Service are key to ensuring the long-term success and sustainability of technical infrastructures and their interoperability.

**Research Software Metadata Curation Roadmap by Morane Gruenpeter, Sabrina Granger, Alain Monteil, Jozefina Sadowska and Estelle Nivault**

Software is essential for academia, playing a crucial role in research, yet it is not fully recognized as a first-class output. Quality metadata and standardized identifiers can only be achieved if infrastructures provide users with the necessary functionalities and institutions offer the incentives to encourage researchers to archive and describe their software. In this talk, we outline the current status of research software metadata curation, the ongoing implementation within FAIRCORE4EOSC project, and a proposed roadmap for the future.

The FAIRCORE4EOSC project focuses on the development and realization of core components for the European Open Science Cloud (EOSC), launched in June 2022. Software Heritage, the INRIA research center and other partners are collaborating on the workpackage ""Services and tools to archive, reference, describe, and cite research software."" In this short lightning talk, we will share the implementation mechanisms regarding the curation of research software. These curation mechanisms for research software rely on the use of the CodeMeta vocabulary and crosswalk table, allowing the infrastructures to strengthen interoperability and provide researchers with adequate tools to better describe and cite software. The academic community has continuously contributed to this effort through engagement with FORCE11, RDA, and the SciCodes consortium.

The proposed roadmap outlines the future development and integration of these tools within the broader EOSC ecosystem to reinforce software as a first-class output. This plan includes strategies for enhancing interoperability between existing infrastructures while relying on the universal source code archive, Software Heritage, improving metadata standards, and enabling greater adoption of best practices in software curation. Additionally, we will discuss the collaborative efforts and community engagement necessary to ensure these tools effectively support the FAIR (Findable, Accessible, Interoperable, and Reusable) principles and the RSMD (Research Software MetaData) guidelines.

Although librarians have been accustomed to describing various types of documents, especially with the rise of data-driven research, software-specific nuances need to be explained to the teams in charge of metadata curation. The aim is to enable moderators, archivists, and curators to understand this distinct academic output. We need to develop comprehensive training programs to equip data curators, researchers, and supporting staff with the necessary knowledge and skills for effective research software management. Continuous training and support from software archival experts will ensure that these professionals are well-prepared to handle current and future challenges in digital curation.

Infrastructures and institutions must collaborate to develop innovative tools and systems while addressing the education and training needs of curators and researchers within the research landscape. This collaboration is essential for improving the archival, referencing, description, and citation of software, ensuring that these processes are aligned with best practices and evolving standards."

**Software Curation and Archiving: A Data Curator’s Perspective of the Last 10 Years by Fernando Rios**

It can be argued that research software, that is, software that is used or created in support of scholarly research activities, is as important to supporting the understanding, verification, and reproduction of research results as journal articles and associated research data. However, it has often been relegated to a second-class research output. This has led to a wide variety of initiatives to raise it to a level that gives it the recognition it needs. They have ranged from work towards raising its profile as a first-class research product in the eyes of the research community, to the development of tools and resources for data managers and curators to help improve the reproducibility and reuse of potential of research products. The aim of this talk is to briefly review some of the software curation and archiving work of the past 10 years from a data curator’s perspective. At the end of the presentation, my hope is that the audience will have gained a glimpse of the software curation field from one practitioner’s point of view, learned about projects or tools they might not have heard (or forgotten) about, and leave with some thoughts on how they might continue to evolve in the field from wherever they are situated professionally.

**The Copyrights Chain In The Digital Curation Process: “Which Copyrights” Project At The Nazarian Library, University Of Haifa / Keren Barner Digitization, Curation, Preservation: Digital Projects and Special Collections, Younes & Soraya Nazarian Library, University of Haifa by Keren Barner**

Open access, image sharing, and mostly the rise of AI emphasize the importance of providing clear licenses for the reuse of collections by memory institutes as a key step in the digital curation process.

Trusting individuals to control their own data and its (re)use was an important concern since the early establishment of our digital collections in 1998 with unsatisfied solutions.

This talk will describe our experience in a recent ongoing project to enrich the University of Haifa digital collections with clear copyright status, to benefit the user and owner’s experiences, inspired by WorldFAIR Recommendations Report for Cultural Heritage Image Sharing (Knazook and Murphy, 2023)

We will explore “Which copyrights” project 2024-2025. Scope of 80,000 bibliographic records contains 200,000 digital representations covering local cultural heritage data. 50% have already processed and available with the copyright’s representation notice online.

Our next step - Use a public platform to make the dataset available for the public. Sharing our collections, expand visibility and usage to a broader public, via non-academic platforms such as Wikipedia/Wikimedia.

We hope our library experience will be useful to other institutions working with digital heritage collections, for the benefit not only of researchers but of the broader public as well.

**The Helmholtz Metadata Collaboration – Paving the road for a Helmholtz FAIR data space by Constanze Curdt, Christine Lemster and Sören Lorenz**

The Helmholtz Association of German Research Centres set up the Helmholtz Metadata Collaboration (HMC) platform in 2019 to translate global metadata concepts into standardized formats and harmonize scientific practice for the Association. HMC has the mission to leverage visibility and reusability of data across the Helmholtz Association and beyond, and to turn FAIR data into reality. Our long-term goal is to create a sustainable, distributed, semantically enriched FAIR data space that spans across all 18 autonomous Helmholtz centers and the six Helmholtz research fields (Aeronautics, Space & Transport, Earth & Environment, Energy, Health, Information, and Matter).

Our work in HMC is focussed on activities in three strategic areas: (1) Assessing and monitoring the state of FAIR data across Helmholtz, (2) Facilitating connectivity of Helmholtz research data and (3) Transforming (meta)data recommendations into implementations. With our activities we address our multiple stakeholders (e.g. scientific community producing data, data professionals, research data infrastructures, technicians and administration responsible for research infrastructure) to increase the coherence and connectivity of metadata in Helmholtz.

Over the last years, we analyzed the state of research data management and data FAIRness within Helmholtz through surveys and FAIR metrics. Awareness was raised about the importance of metadata through outreach events, training and consulting. We developed the technical backbone necessary for connecting FAIR data in Helmholtz and establishing the Helmholtz FAIR data space. Together with our communities, we worked towards the implementation of aligned FAIR metadata practices and recommendations. In addition, we funded 28 community projects across Helmholtz to implement solutions for practical metadata challenges. Through our work in various panels and committees we are closely intertwined with national (e.g. NFDI) and international (e.g. EOSC, RDA) initiatives in research data management.

In this lightning talk, we will present details of our approach, some results and lessons learnt on our road to a Helmholtz FAIR data space. By sharing our experiences, we engage in active discussions on (meta)data, relevant stakeholders and how to shape our path to a FAIR data ecosystem together."

**The OSF Genes Within The DNA Of Research Data Management Infrastructure by Kirianne Goossen**

The Research Data Management (RDM) infrastructure at the Vrije Universiteit (VU) Amsterdam consists out of a variety of applications. Most applications support only one part of the research life cycle, while there is one application that can be used for the entire research life cycle – the Open Science Framework (OSF). During this lightning talk the VU RDM infrastructure will be discussed and we will dive into the world of OSF.

**Thoth Open Archiving Network (TOAN) Update: Lessons Learned & New Directions by Miranda Barnes**

We wish to share our progress on the Thoth Open Archiving Network (TOAN) with the IDCC, including the lessons we have learned in our developmental research and how these have led to new directions for the Network. These include expanding our view of repository types, the challenges of sustainability, technical development time, metadata quandaries, and repository collection policies. Join us for a rapid-fire tour of our next steps, and how we envision TOAN as an open infrastructure, now and into the future.

**Twenty Years Back, Five Years Forward: The Past, Present And Future Of DANS by Ingrid Dillo**

The first IDCC took place in 2005. In that same year, the Dutch national research funder and the Royal Academy of Arts and Sciences created a new organization called DANS with a mission to preserve and guarantee continued access to research data in the humanities and social sciences.

Twenty years later DANS has become the Dutch national centre of expertise and repository for research data, with a mission to enhance the reusability of research data and thus the quality of scientific research. And what will the future bring for DANS? We are currently working on our strategy for the coming years.

In this presentation I will give a concise overview of the DANS organization and services as well as the main developments in its environment, in the past (2005), present (2025) and future (2030). This will show that the environment in which DANS is positioned changed dramatically over the last twenty years, with the rapid development of the digital world, the arrival of the Open Science and FAIR concepts and the unprecedented impact of AI. On the other hand we will see that a lot of the challenges that were identified in 2005 still exist.

**What Does Your Data Curation Look Like? Better Collaboration Between Research Tool Providers and Curators by Vaida Plankytė**

We wish to present the previous and ongoing work by our upcoming RDA Interest Group, centered around the intersection between research tools and data curation workflows.

We have held a Birds of a Feather session entitled “Research Tool Usage in Data Curation Workflows: Responsibilities & Collaboration” at the RDA Virtual Plenary 22, where we highlighted that “data curation” is interpreted and approached in multiple different ways across institutions, as an activity and as a role, as well as having different ways of existing within an institutional RDM frame: as a separate department, part of the Library services, or embedded. This results in each institution having both subtly and significantly different requirements when it comes to what, how, and which research tools can best assist with the kind of data curation they perform.

Presentations by RDM managers and research tool providers reflected on how collaboration between institutional RDM and research tools can be streamlined through an explicit definition of what data curation involves for that institution. The development of a shared language between parties and clarification of the tasks involved that require support then enables greater understanding of how each party’s role in the successful adoption and usage of tools for data curation.

Looking forward, the RDA Working Group “Mapping the landscape of digital research tools” has carried out a comprehensive review of existing research data lifecycle models, and is publishing their harmonized MaLDReTH model in August 2024. Crucially, their model has also mapped existing research tools, based on their feature set, onto each stage of the MaLDReTH model.

At IDCC25, we would like to summarize the learnings from the Birds of a Feather session, as well as present the outcome of reviewing the MaLDReTH model in the context of our interest group: we aim to develop guidelines and facilitate a shared language between research data managers and research tool providers around data curation, and by grounding ourselves in a universal lifecycle model, we hope to formalize institutional RDM approaches and their resulting research tool needs and requirements. We aim for this work to help inform the research tool selection and adoption process through a well articulated statement of purpose.

We believe our work addresses concrete practical issues that cause points of friction in institutional research infrastructures, which are constantly increasing in complexity and becoming essential to all research disciplines, requiring great understanding of researcher and curator needs. We trust IDCC25 would be the ideal space to present our findings, and discuss the intersection between research tools and institutional data curation in terms of communication, onboarding, adoption, and workflows."