

PROGRAMME

DAY 1 – MAY 13

8:30 – 9:30	Registration
9:30 – 11:00	<p>Conference Welcome & Harvard Dataverse Updates <i>Room: Auditori</i> Opening: Teresa Sanchís (Director general for Research, Catalan Government) and Mercè Crosas (BSC CSSH Lab Director) Chair: Danny Ebanks Presenters: Community updates - Philip Durbin Dataverse software update - Gustavo Durand Dataverse project update - Ceilyn Boyd Frontend features update - Ellen Kraffmiller</p>
11:00 – 11:30	Coffee Break (Room: Port Vell)
11:30 – 13:00	<p>Plenary: Dataverse Repositories and Services Showcase <i>Room: Auditori</i> Chair: Ingrid Dillo Presenters:</p> <ul style="list-style-type: none"> ● Laia Garcia: Building a community-driven Dataverse service for the Catalan research system ● Aina Jene: BSC Dataverse: An Institutional Repository for Research Data and High-Performance Computing ● Ceilyn Boyd: "Locally FAIR" Data: Restricting Data Visibility in Dataverse Installations ● Bapon Fakhruddin: Climate and Disaster Risk Reduction in Fiji: Building a FAIR Data Ecosystem ● Billie Hu: Addressing Architectural Challenges in Scaling a National Dataverse: Lessons from Borealis ● Rene Faustino Gabriel Junior: Advancing Open Science in Brazil through DOI Assignment for Research Data: The ConsCiência Consortium ● Taras Zhyhulin: An End-to-End Toolkit Facilitating Data Sharing and Reuse
13:00 – 14:00	Lunch Break (Room: Port Vell)
14:00 – 15:30	<p>Parallel Sessions 1 Topic: AI for Data Discovery and Management <i>Room: A1</i> Chair: Nadia Tonello Presenters:</p> <ul style="list-style-type: none"> ● Adília Alves: ImageSearch – a tool for image semantic retrieval using text or image queries ● Tilo Mathes: AI-Powered dataset integration between Dataverse and RSpace using CDIF variables ● Zhen Meng: Biological Big Data Integration and AI-Driven Applications ● Pedro Principe: Enhancing repositories with interconnected research intelligence ● Slava Tykhonov: AI-Driven Data Annotations with Semantic Croissant and CDIF

14:00 – 15:30

Parallel Sessions 1

Topic: Dataset and Tool Interoperability

Room: A2

Chair: Ellen Kraffmiller

Presenters:

- **Eryk Kulikowski:** Introducing DDI-CDI to Dataverse: Automatic Metadata Generation and Interactive Editing
- **Lukasz Opiola:** Bridging Active Data Management and Research Data Publication: Integrating Onedata with Dataverse
- **Viviane Veiga:** Sociotechnical framework for Dataverse health research data management: Protocols for Scoping Reviews on Stakeholder Perceptions and Practices and Interoperability Requirements
- **Rory Macneil:** Learnings on interoperability and cross-tool persistent identifier sharing through the RSpace and Dataverse use case

Topic: Trusted and Secure Data Environments

Room: A4

Chair: Eva Alloza

Presenters:

- **Philipp Conzett:** From Fragmentation to Federation: A Unified Nordic-Baltic Hub for Trustworthy Dataverse Repositories
- **James Myers:** Advancing Indigenous Data Sovereignty through Dataverse and Local Contexts Integration
- **Zhihong Shen:** CoNet: A Collaborative Platform for Cross-Center Data Analysis
- **Thomas Van De Velde:** FAIRVault; secure storage and preservation for sensitive & restricted datasets

Topic: Showcase: Comparing approaches for Dataverse repositories

Room: A3

Chair: Jonathan Crabtree

Presenters:

- **Bethany Seeger:** Getting started with Dataverse? Tips from how we approached the work at Yale
- **Julian Gautier:** Evaluating progress of Harvard Dataverse deposit form redesign
- **Byomkesh Talukder:** A Dataverse-Integrated Framework for AI-Enabled Climate Health Simulations
- **Rachel Fernandez:** ASU's Dataverse Snapshot: Showcase of technical infrastructure, policy, and services
- **Juan Corrales:** Joint Developments and Shared Services for Dataverse: The CSUC–Consortio Madroño Experience
- **Ryan Perry:** From Policy to Practice: Automating Curation, Preservation and Sensitive Data Access in Dataverse

15:30 – 16:00

Coffee Break (Room: Port Vell)

Organized by

16:00 – 17:30	<p>Keynote Addresses <i>Room: Auditori</i> Chairs: Mercè Crosas and Diego Perino Keynote Speakers:</p> <p>Ian Foster (Argonne National Laboratory and University of Chicago): Repositories as Infrastructure for Machine Reasoning</p> <p>Hilary Hanahoe (Research Data Alliance): Artificial Intelligence and the global research data community: a perfect pairing?</p> <p>Rosie Hicks (Australian Research Data Commons): A connected and ethical AI-ready data ecosystem</p> <p>Panel: Ian Foster, Hilary Hanahoe, Rosie Hicks, and Diego Perino (director of BSC AI Institute)</p>
17:30 – 19:00	<p>Institutional Welcome, Reception and Posters Opening <i>Room: Port Vell</i> Department of Research and Universities (Catalan Government) Cristian Canton (BSC Executive Director)</p>

DAY 2 – MAY 14

9:00-9:30	<p>Registration</p>
9:30 – 11:00	<p>Keynote Addresses <i>Room: Auditori</i> Chair: David Fusté Keynote Speakers:</p> <p>Enrique Bernal-Delgado (Institute for Health Sciences. IACS): Governing federated multi-country observational research: lessons for an extensive reuse of sensitive multimodal data</p> <p>Emily Jefferson (DARE UK and Health Data Research UK): Enhancing TRE Capabilities to Support Federation and AI Projects: A UK Perspective</p> <p>Jano van Hemert (University of Edinburgh): Lessons learnt from ten years of safe haven services: opportunities and challenges for future services</p>
11:00 – 11:30	<p>Coffee Break (Room: Port Vell)</p>
11:30 – 13:00	<p>Parallel Sessions 1 Topic: AI for Better Research Data Repositories <i>Room: A1</i> Chair: Pedro Principe Presenters:</p> <ul style="list-style-type: none"> ● Danny Ebanks: A. I. for Enhancing Repository Experience and Accelerating Research ● Aalap Doshi and Jeanette Jackson: Approaching AI in trusted data repositories: Reflections from ICPSR ● Rachel Fernandez: Combing through the metadata: Looking back at 5 years of research data deposits at Arizona State University ● Balázs Pataki: VibeARP: A Data Steward Assistant for the Hungarian Research Data Repository Platform ● Zhenxin Wu: Transforming digital preservation through artificial intelligence: a comprehensive framework and analysis

Organized by

11:30 – 13:00

Parallel Sessions 1

Topic: Data Infrastructure for Interoperability

Room: A2

Chair: Alexandra Malaga

Presenters:

- **Oliver Bertuch:** Metadata code infrastructure: evolving to fit for the next 10 years
- **Joan Giner-Miguel:** What's new in Croissant? Empowering AI Agent Data Discovery through Machine-Actionable Provenance and Usage Conditions
- **James Myers:** Leveraging the Power of PIDs in Dataverse: 2026 Update
- **Steven McEachern:** Shaping the Future of Dataverse in Europe with the Cross-Domain Interoperability Framework
- **Anna Caellas-Camprubí:** From Sharing to Impact: Tracking Dataverse datasets citations across a decade of indexed publications

Topic: Showcase: Better curation and annotation, better data and metadata

Room: A4

Chair: Marta Curto

Presenters:

- **Tom Emery:** Annotating Administrative Data with the Scientific Literature for enhanced findability
- **Angelica Maria Maineri:** The ODISSEI Portal: leveraging metadata to enhance data discovery
- **Vaidas Morkevicius:** Better Data Files for Survey Data? A Case of LiDA
- **Paula Moura:** Engaging users in improving data curation workflows: DataRepositóriUM use case
- **Lok Hei Lui:** Streamlining data curation - a prototype of a checklist-based data curation tool
- **Bethany Seeger:** Yale Application for Research Data (YARD): Curation Tool Integration

13:00 – 14:00

Lunch Break & Poster Presentations (Room: Port Vell)

14:00 – 15:30

Parallel Sessions 2

Discussion session 1: Curation

Room: A1

Chair: Ceilyn Boyd

Presenters:

- **Dieuwertje Bloemen:** Automation-assisted curation at KU Leuven
- **Rachel Fernandez:** In the time of AI and large data: reshaping curation workflows for the ASU Research Data Repository
- **Thu-Mai Lewis:** Policy and Pipelines for Data Curation at Scale

Discussion session 2: Development

Room: A2

Chair: Philip Durbin

Presenters:

- **Gustavo Durand:** Dataverse Marketplace Updates
- **Ellen Kraffmiller:** Building the Dataverse SPA: Recent Development Highlights
- **James Myers:** Scaling Dataverse with Data Size: Configuration Strategies and Recent Performance Improvements

Organized by

14:00 – 15:30

Parallel Sessions 2

Tech demo

Room: A3

Presenters:

- **Oliver Bertuch:** Let's Play Upgrade Monopoly: go directly to latest!
- **Oliver Bertuch:** Continuous Delivery for Dataverse on K8s: "vlûch serviert"
- **Eryk Kulikowski:** Creating New UI components that can be used in Dataverse JSF & SPA: A DualMode Architecture for Dataverse UI Extensions

15:30 – 16:00

Coffee Break (Room: Port Vell)

16:00 – 17:00

Parallel Sessions 3

Topic: AI in the Humanities

Room: A1

Chair: Mariona Coll

Presenters:

- **Yin Shenqin:** Building AI-Ready Datasets in Archaeology: A Large-Scale FAIR Workflow Based on LLMs
- **Juan Antonio Barceló:** A Linked Data Framework for 3D Cultural Heritage Repositories: Integrating Geometry, Visuals, and AI-Driven Semantic Annotation
- **Miquel Centelles Velilla:** Semantic interoperability and metadata crosswalks for epistemic justice in digital humanities
- **Slava Tykhonov:** The minority report - AI arbitrage translator for vocabularies and ontologies

Topic: Dataset Interoperability and Linkage

Room: A2

Chair: Eudald Lerga

Presenters:

- **Changfa Lu:** CSTR support dataset interoperability and linkage
- **Tilo Mathes:** Integrating RAiD into dataverse: university perspectives and workflow integration
- **Slava Tykhonov:** Nectar Publisher as a Human-in-the-Loop for Dataverse

Topic: Trustworthiness and Sensitive Data

Room: A4

Chair: Tatiana Vagramenko

Presenters:

- **Ceilyn Boyd:** Data Trustworthiness in Dataverse
- **Julian Gautier:** What repository characteristics affect a potential data reuser's opinion of data trustworthiness?
- **Ahmad Hesam:** Data Access Broker: getting sensitive data safely and conveniently from repository to secure compute environments
- **Pablo Ignacio Martinez Diaz:** IMPaCT-Data 2 - deployment of the federated infrastructure IMPaCT-Data to serve personalised medicine use-case

Organized by

16:00 – 17:00	<p>Parallel Sessions 3 Topic: Showcase: FAIR data publication, curation, and assessment <i>Room: A3</i> Chair: Paula Checchia Presenters:</p> <ul style="list-style-type: none"> ● Andre Vieira: FAIR assessment and DataRepositóriUM: OStrails pilot use case ● Dieuwertje Bloemen: Automating curation to support FAIR data publication: the KU Leuven RDR review dashboard ● Björn Selent: Curation at scale: Automation of publication workflows for organizational repositories ● Rory Macneil: Born-Managed Chemistry Data: Automating FAIR Compliance from Instrument to Repository
17:00 – 17:30	<p>Updates GDCC meeting <i>Room: Auditori</i></p>

DAY 3 – MAY 15

9:00-9:30	Registration
9:30 – 11:00	<p>Keynote Addresses <i>Room: Auditori</i> Chair: Simon Hodson Keynote Speakers:</p> <p>Margaret Gold (Leiden University and Citizen Science Lab): Four waves of development in Citizen Science: a global to local view of a field in motion</p> <p>Josep Perelló (Universitat de Barcelona): Co-Producing Data and Knowledge Through Citizen Social Science: Social Compromise with Communities and Civil Society Organisations</p> <p>Stefaan Verhulst (The Data Tank; The GovLab): The Best of Times, and the Worst of Times? Data Access in the Age of Generative AI</p>
11:00 – 11:30	Coffee Break (Room: Port Vell)
11:30 – 13:00	<p>Parallel Sessions 1 Topic: Interoperability Use Cases <i>Room: A2</i> Chair: Philipp Conzett Presenters:</p> <ul style="list-style-type: none"> ● Felipe Leonardo Gómez Cortés: Implementing DDI-CDI for Machine-Actionable Qualitative Data Description: A Case Study of Medieval Notary Manuscripts in Dataverse ● James Myers: Making Coded Qualitative Data Discoverable: Previewers for REFI-QDA Files in Dataverse ● Marc Ter Horst: Metadata for Chemists ● Johanna Havemann: Scaling Open Data Repositories in Eastern and Southern Africa: A Networked Dataverse–DSpace Approach ● Bapon Fakhruddin: Harnessing Ocean Data for Coastal Resilience and Blue Economies

Organized by

11:30 – 13:00

Parallel Sessions 1

Topic: Resilient and trusted data systems

Room: A4

Chair: Eudald Lerga

Presenters:

- **Deirdre Lungley:** Powering curation tooling, including statistical disclosure risk analysis, with Dataverse
- **Hanyu Li:** Digital academic research paradigm transformation and trust reconstruction from institutional trust to systemic trust in decentralized science
- **Carlos Perez Dengra:** Automating trust: A Lagrangian Deep Reinforcement Learning framework to optimize agnostic sensitive data sharing
- **Guy Collins:** From metadata to seeds designing data systems that survive institutional forgetting

Topic: Showcase: Sustainability and Systems Integrations

Room: A3

Chair: Marta Curto

Presenters:

- **Domingo Scisci:** Embedding Dataverse in a modular national data infrastructure: architecture and workflows of the DASSI services
- **Eryk Kulikowski:** Implementing & Improving Globus in Dataverse: Uploads, Downloads & Preview URLs
- **Alyssa Conlon:** The Internationalization of Borealis – A Focus on Sustainability
- **Santi Thompson and Xuan Zhou:** Advancing data retention in the Texas Data Repository
- **Pedro Filipe Sobral:** Fantastic integrations and where to find them: interoperability in the POLEN DataHub research data repository

13:05 – 13:15

Next DCM Announcement & Closing

Room: Auditori

13:15 – 14:15

Lunch Break (Room: Port Vell)

Note: You can find the list of Social events in this [link](#).

Organized by

Poster Presentations

Sana Abdelkader: Establishment of a new curation process in Edmond, the Max Planck Society open research data repository

Lisa Bagdenand: Integrating dataverse datasets into knowledge graphs for research

Ceilyn Boyd: Differentially Private Data: Using DP Wizard to Share Sensitive Data

Oliver Bertuch: Dataverse plugin ecosystem: towards stability, compatibility and sustainability

Oliver Bertuch: Restructuring Dataverse exporters: unifying export and citation for both datasets and files, allowing citation exports in bulk

Dieuwertje Bloemen: Fantastic Docs and Where to Find Them- Field Notes from the Dataverse Documentation Working Group

Dieuwertje Bloemen: Automating curation to support FAIR data publication: the KU Leuven RDR review dashboard

Emilio Cammarata: FAIRness in Reference Managers usage: utopia or reality?

Paula Checchia: BSC Dataverse: An Institutional Repository for Research Data and High-Performance Computing

Anne Clinio: Is Dataverse a potential infrastructure for citizen science? An exploratory analysis

Alyssa Conlon: The Internationalization of Borealis – A Focus on Sustainability

Tom Emery: Annotating Administrative Data with the Scientific Literature for enhanced findability

Rachel Fernandez: ASU's Dataverse Snapshot: Showcase of technical infrastructure, policy, and services

Maria Paola Ferri: From openVRE to safeVRE: An Interoperable Virtual Research Environment for Secure Data Access, Scalable Computation and Collaboration

Bryan Gee: Development of a programmatic approach for large-scale metadata re-curation in the Texas Data Repository

Joan Giner-Miguel: What's new with Croissant? Bridging Repository Standards for AI Discovery

Marc Ter Horst: Managed data in Chemistry

Dorothea Iglezakis: Metadata2Dataverse – a metadata mapping tool

Daan Janssen: Advanced Editing Toolkit: Extending Dataverse with integrated metadata and data processing workflows

Agate Jarmakovica: Building Trustworthy Repositories: A Coordinated National–Institutional Model for DataverseLV and RSU Dataverse

Rene Faustino Gabriel Junior: Disciplinary Metadata Management in the Face of Sensitive Data: Tensions, Limits, and Solutions in Research Data Repositories

Changfa Lu: Scientific data center software stack

Changfa Lu: DataSpace: A Collaborative Management Tool for Scientific Data

Rory Macneil: Born-Managed Chemistry Data: Automating FAIR Compliance from Instrument to Repository

Angelica Maria Maineri: The ODISSEI Portal: leveraging metadata to enhance data discovery

Paloma Marín-Arraiza: Supporting enhancements in the Dataverse-ORCID integration with ORCID's Global Participation Fund

Vaidas Morkevicius: Better Data Files for Survey Data? A Case of LiDA

Paula Moura: Engaging users in improving data curation workflows: DataRepositóriUM use case

Constanza Pardo: Research Collection: a dataverse research data repository at Institut de physique de globe de Paris

Ryan Perry: From Policy to Practice: Automating Curation, Preservation and Sensitive Data Access in Dataverse

Pedro Principe: Linking datasets and publications with COAR Notify: the case of University of Minho repositories

Laura Vilela Rodrigues Rezende: Brazilian case study of climate scientific data workflow, including a Dataverse institutional research data repository

Bethany Seeger: Yale Application for Research Data (YARD): Curation Tool Integration

Domingo Scisci: Embedding Dataverse in a modular national data infrastructure: architecture and workflows of the DASSI services

Chandra Shekhar Roy: AI Powered Historical Data Repositories A Case Study in Managing Big data and Enhancing Discoverability for National Statistical Archives

Don Sizemore: A Path to Sustainability: UNC Dataverse's Role in the Research Data Management Core

Pedro Filipe Sobral: Fantastic integrations and where to find them: interoperability in the POLEN DataHub research data repository

Yin Shenqin: Data Publishing in the AI Era: Practices and Challenges from Library Journal (China)

Yin Shenqin: Research Data Services in the AI Era: A Four-Tiered Framework at Fudan University Social Science Data Platform

Tsaone Swaabow Thapelo: Event-Driven Agricultural Evidence for Crisis-Aware e-Governance

Martin Vávra: Understanding user needs: a user-centered approach to improving the Dataverse experience at the Czech Social Science Data Archive

Organized by

Andre Vieira: FAIR assessment and DataRepositóriUM: OSTRails pilot use case

Emma Louise Walton: Unifying training and outreach efforts for DataverseNO: a cross-institutional resource package

Enrique Wulff: The Metaverse as a Cultural Continuum: Insights from the EOSC projects with CSIC participation

Chase Yakaboski: Harmonizing consortium-scale rare disease data assets with a provenance-preserving semantic model

Xuan Zhou: Activating Open Data on Campus: A Datathon Use Case with Dataverse

Gregor Žibert: Rebuilding National Data Services with Dataverse: The Slovenian Social Science Data Archives' Modernisation Journey

Oussama Zoubia: Co-Designing FAIRdata Cologne: A Researcher-Centered Metadata Catalogue for Experimental Research

Ivanete Presot: New Opportunities in the Application of the Electronic Laboratory Notebook (ElabFTW) and Its Contribution to Open Science

Kristine Racina: Ten years of DataverseNL: lessons learned and future directions

Organized by

KEYNOTE ADDRESSES

ENRIQUE BERNAL-DELGADO

Title: Governing federated multi-country observational research: lessons for an extensive reuse of sensitive multimodal data

Abstract: Our research often requires pseudonymised, multimodal, and longitudinal individual data linked from multiple sources stored in multiple locations. Generally, our research requires a large amount of data representing the universe of potential observations. In some cases, the large amount of data and its dimensionality may require computational partitioning. These are the usual arguments supporting the need for a federated approach; however, intrinsic to the federated approach is the need to address a number of legal, organisational, semantic and technological challenges. Based on several exemplary case studies, we will discuss how to overcome these challenges while complying with the new European Health Data Space regulation for secondary use and the principles of Open Science.

IAN FOSTER

Title: Repositories as Infrastructure for Machine Reasoning

Abstract: Research data repositories were built for preservation and human reuse. Increasingly, however, their primary consumers will be AI systems—agents that integrate data at scale and drive downstream computation across institutional boundaries. Control over scientific reasoning will follow the infrastructure that enables this integration. If repositories remain optimized for download and documentation, proprietary AI platforms will absorb public data and expose reasoning through closed systems. Retaining public control requires repositories to interoperate at the level of semantics and execution, supporting policy-bound computation, machine-readable usage constraints, and verifiable provenance across institutions. In doing so, they become not merely archives, but a federated substrate for machine-mediated scientific reasoning.

MARGARET J GOLD

Title: Four waves of development in Citizen Science: a global to local view of a field in motion

Abstract: In this keynote talk, Margaret will open with a brief introduction to Citizen Science and what is particular about CS data, followed by an overview of a field in rapid development, from Global SDGs and national research policies to institutional Open Science programmes and grassroots collaborations with citizen and societal actors.

HILARY HANAHOE

Title: Artificial Intelligence and the global research data community: a perfect pairing?

Abstract: Since 2025, the Research Data Alliance (RDA), in collaboration with Microsoft, has been conducting a global community consultation to explore how researchers are currently using agentic AI — defined as AI systems capable of autonomous operation with minimal human oversight — and to assess its value across the research lifecycle.

Grounded in the FAIR and CARE principles, the consultation was open to all, regardless of AI expertise or location, and built upon earlier RDA-Microsoft work that had identified automated data preparation as a critical researcher need.

Participants evaluated eleven proposed agentic AI tools spanning the full research lifecycle, from planning and funding through to publication and impact reporting. Three emerged as clear community priorities: the Literature Librarian, supporting natural language literature searches integrated with library subscriptions; the Data Director, assisting with FAIR-compliant data preparation and sharing; and the Funding Finder, helping researchers identify relevant opportunities and navigate application processes. These rankings were consistent across both regional and stakeholder analyses.

Beyond preferences, respondents also shared insights on desirable tool features and anticipated challenges. The presentation will summarise the consultation findings and provide an update on developing a blueprint for global community use of one of the top three ranked tools.

ROSIE HICKS

Title: A connected and ethical AI-ready data ecosystem

Abstract: Capitalising on advances in AI/ML requires a connected, ethical and AI-ready data ecosystem that enables data access, interoperability and reuse at scale. Connectivity addresses the fragmented nature of Australia's repository and data storage ecosystem, which makes it difficult for users to navigate, inefficient to implement FAIR and CARE, and impossible to scale data engineering services. Ethical considerations need to be encoded through provenance and cultural/intellectual property rights to ensure AI models respect researchers' and First Nations' rights, secure trust and address AI bias. These foundations enable the creation of trusted AI models and sensitive data assets that can be shared, analysed and reused with confidence. This presentation will provide an overview of Australia's current progress towards an AI-ready data ecosystem.

EMILY JEFFERSON

Title: Enhancing TRE Capabilities to Support Federation and AI Projects: A UK Perspective

Abstract: Trusted Research Environments (TREs)—also known as Secure Data Environments or Safe Havens—are secure computing infrastructures that enable the analysis of sensitive data (e.g. health, administrative, genomic, and financial data) without that data leaving the secure environment. They operate in accordance with the Five Safes principles.

More than 60 TREs are currently operating across the UK, collectively supporting around 10,000 research projects. However, by design, TREs can create secure data silos, making it challenging to link and query data across different environments. In addition, supporting artificial intelligence (AI) projects within TREs presents technical, procedural, cultural, governance, and regulatory challenges.

This talk will provide an overview of the UK TRE landscape and its relationship to developments across Europe. It will highlight the work of DARE UK and HDR UK in convening the UK TRE ecosystem to develop standardised, interoperable models of federation. The session will also explore ongoing efforts to address the specific challenges of enabling AI research securely within TREs.

Organized by

JOSEP PERELLÓ

Title: Co-Producing Data and Knowledge Through Citizen Social Science: Social Compromise with Communities and Civil Society Organisations

Abstract: Computational social science and complex systems research increasingly call for a commitment to align scientific inquiry with the needs, priorities, and values of the people and civil society organisations directly concerned with social and urban challenges. This talk presents how our group, OpenSystems at the Universitat de Barcelona, develops citizen science methodologies in which data is co-produced with communities—often in vulnerable situations—and civic actors, ensuring that research outcomes remain both scientifically robust and socially meaningful. Drawing from recent interdisciplinary work, I will describe cocreated initiatives on mental health, where shared data and narratives uncover multiple dimensions of community health-care provision and social support networks, informing interventions developed together with civil society organisations. I will also present our thermal walks, where groups of citizens contribute experiential input and low-cost sensor data while exploring their neighbourhoods, helping guide climate adaptation efforts in collaboration with local civil society organisations. Additionally, research on walkability shows how integrating movement traces, embodied experience, and civic dialogue reveals new insights into urban accessibility and community placemaking. Across these examples, I will emphasize how citizen social science produces knowledge capable of informing public debate and institutional decision-making.

JANO VAN HEMERT

Title: Lessons learnt from ten years of safe haven services: opportunities and challenges for future services

Abstract: Since 2015, EPCC has operated Safe Haven Services where it hosts multiple safe havens. The first and largest is the Scottish National Safe Haven, which enables researchers to process health, social and administrative data about the 5.5 million people in Scotland. For instance, it holds copies of all prescribed medicine in Scotland since 2010 and all radiology images between 2010 and 2018. I will share learnings of the operation and development of these services that collectively have supported over 1000 studies. In addition, we have undertaken a study to understand what barriers researchers face in studies that require microdata to be integrated from two safe havens. From these learnings I will highlight what opportunities and challenges exist for safe havens.

STEFAN VERHULST

Title: The Best of Times, and the Worst of Times? Data Access in the Age of Generative AI

Abstract: In the age of generative AI, we are living through what Charles Dickens might call both “the best of times and the worst of times” (A Tale of Two Cities) for data: never has data re-use held such promise for addressing societal challenges yet never has access to that data felt so fragile. We are witnessing a growing “data winter,” marked by shrinking access to non-traditional data, the privatization of climate and other high-value research datasets, legal anxieties around AI training, and a slowdown in open government data - all which risk concentrating insight in the hands of a few. And yet, alongside this contraction, a spring of hope is emerging AI-enabled tools and interfaces, privacy-enhancing technologies, and new kinds of data commons are creating new pathways for responsible access and collaboration. The question before us is whether we allow extraction, weaponisations, fragmentation, and control to define the future of data, or whether we intentionally build a Fourth Wave of Open Data — one that is equitable, responsible and designed for meaningful reuse. The trajectory is not yet predetermined; it depends on the governance and investment choices we make today, and on whether we can turn the paradox of abundance without access into a new era of shared insight and collective progress.

Organized by