RESEARCH DATA ALLIANCE

From Discovery to Action: Tackling AI-Driven Data Challenges Through Global Collaboration

eResearch Australasia 2025
Wednesday 22nd October, 2025

Presenter: Hilary Hanahoe, Secretary General, Research Data Alliance
Contributors: Trish Radotic, Connie Clare, Shalini Kurapati - RDA Secretariat

The AI Transformation in Research

- AI is rapidly transforming research across all disciplines
- 50% of researchers expect AI to play a major role in discovery within 5 years (Nature 2025 survey)
- High quality, well-managed data is crucial for accurate, reproducible AI-driven insights
- The Challenge: Traditional bottom-up RDA approach needs acceleration for AI's urgent pace

RDA
RESEARCH DATA ALLIANCE

What is the RDA?

The RDA is an **international member-based organisation** focused on the development of **infrastructure** and **community activities** that reduce barriers to **data sharing** and **accelerate data-driven innovation** worldwide.

Vision

Researchers and innovators openly share and reuse data across technologies, disciplines, and countries to address the grand challenges of society.

Mission

The RDA builds the social and technical bridges that enable open sharing and reuse of data.

16,000+ members

150+ countries

Researchers, scientists, data professionals

Different disciplines, domains, thematic fields

Guiding Principles

- Openness
- Consensus
- Inclusivity
- Harmonisation
- Community-driven
- Non-profit & technology-neutral

How does the community collaborate?

110 ACTIVE GROUPS

WORKING GROUPS (WG)
46

Develop tools, policies, practices and products for data that are adopted by projects, organisations and communities. Lifespan: 18-24 months.

RECOMMENDATIONS:
Concrete deliverables (e.g., running code, tools, standards)

All Working Groups

INTEREST GROUPS (IG)
62

Focus on solving a specific data sharing problem and identifying what kind of infrastructure needs to be built. Lifespan: As long as group is active.

OUTPUTS:
Best practices, guidelines, new WGs.

All Interest Groups

COMMUNITIES OF PRACTICE (CoP)
2

'Umbrella groups' that focus on a domain / discipline and have a coordination and awareness raising role. Lifespan: As long as CoP is active (review every 18 months).

OUTPUTS:
New WGs and IGs, bridge building across the RDA and externally.

All Communities of Practice

Why a Top-Down Approach for AI?

RDA's typical process

- Community identifies needs
- Groups form organically
- Solutions develop over 18-24 months

Why AI required a different approach

- Speed and scale of AI impact unprecedented
- Immediate gaps in reproducibility, governance, ethics
- Risk of fragmented solutions
- Need for coordinated global response

Strategic decision to launch the Discovery Initiative to identify priorities, then mobilise the community

The Discovery Journey

Our Three-Phase Engagement Strategy:

Phase 1:
State of AI Pilot Workshops (September 2024)

- Engaged key stakeholders
- Identified pressing challenges

Phase 2:
AI in Action Workshops (February 2025)

- 266 participants from 36 countries
- Showcasing of practical implementations
- 2 sessions to cover global time zones

Phase 3:
RDA-Microsoft Roundtables (May 2025)

- 135+ stakeholders from 20+ countries
- Deep dive on data readiness, AI in research, AI governance

What we heard: Key challenges

Community insights on the pain points

RDA
RESEARCH DATA ALLIANCE

	Data Readiness	AI in Research	AI Governance
Problem	Data lacks complete metadata, provenance, standardisation	Reproducibility crisis - variable quality of datasets and models	Fragmented implementation across regions
Impact	“If your AI strategy doesn’t include your data strategy, it’s incomplete”	Undermining reliability and limiting reuse	Regulatory development not keeping pace with innovation
Gap	Lack of tools for automated data prep across domains	Lack of institutional and journal guidelines on AI usage	Governance literacy among researchers and policymakers

Expert Perspectives

"Repositories should curate data not just for access, but to make it machine-actionable, so that it's AI ready."

Reyna Jenkins
World Data System

"We don't curate just for access, we curate to ensure data trust and AI-readiness at scale."

Mukkesh Kumar
A*STAR, Singapore

"We are not just using AI to assist science. We are training AI to conduct science... grounded in trust, explainability, and shared purpose."

Kyon-Ha Lee
KISTI, South Korea

Community Solutions

Participants were divided into small breakout groups to explore the positive aspects (Roses), challenges (Thorns), and opportunities (Buds) of three key thematic topics.

Data Readiness

- Broad uptake of **FAIR data principles** and understanding of importance of **data quality, transparency, and openness** in making data AI-ready.
- Researchers more aware of **bias**, **transparency**, and **machine-actionable formats**.

- Data often **lacks complete metadata**, **provenance, and standardisation**.
- **Lack of tools** for **automated data** prep across domains.
- **Infrastructure costs** and **skills gaps**

- Use AI to support **automated schema generation**, **metadata enrichment**, and **interoperability solutions**
- Invest in **domain-specific standards**, **curated small models**, and **FAIR implementation** profiles to ensure **systemic readiness**

AI in Research

- AI enabling researchers to work with increasingly complex, multimodal datasets, and enabling breakthroughs.
- Use of **Large Language Models (LLMs)** reducing time and costs in research workflows.

- **Reproducibility and research integrity; quality** and **documentation** of datasets and models are variable, undermining reliability and limiting reuse
- Lack of institutional and journal **guidelines** on **AI usage in research**.

- Adopt **"Model Openness Framework"** to improve **model sharing**, **documentation** and **reporting.**
- Wider adoption of **reproducibility-first protocols**, **AI-specific metadata standards**, and **interdisciplinary data sharing** to strengthen **accountability** and **collaboration**.

AI Governance

- Emergence of new regulatory frameworks**: EU AI Act**, **UNESCO AI recommendations**, and **national policy** roadmaps.
- Many institutions beginning to **embed responsible AI policies** internally.

- **Implementation is fragmented** across regions, **lack of governance literacy** among researchers and policymakers.
- **Trust vs. speed**: regulatory development is not keeping pace with technological innovation..
- **Privacy vs data sharing** with **sensitive datasets**.

- Development of **best practice repositories** and **university-level AI policies** to institutionalise governance processes.
- **International standards** (e.g from the RDA) to **promote globally harmonised governance frameworks.**
- **Shared standards** and **increased awareness of societal impacts** = **responsible deployment.**

8 Thematic Priorities Identified

1. Skills development: Prompting, agentic AI, data literacy, ethics
2. Human-AI collaboration: Empowering researchers, not replacing them
3. Use case guidance: Writing, coding assistance, productivity, clear guidelines
4. Data foundations: AI-readiness imperative
5. Efficient AI: Small models, good data for democratisation
6. Responsible AI: Reproducibility and data ethics challenges
7. Secure infrastructure: Trusted and secure environments
8. Harmonised governance: Global regulation, agreement on need

Key Recommendations

Eight actions for the Community:

1. Upskilling: Prompting, agentic AI, data-for-AI training with ethics focus
2. Empower Researchers: AI as augmentation, not replacement
3. Usage Guidance: Clear institutional policies on AI in research workflows
4. Data Infrastructure: Invest in automated schema generation, metadata enrichment
5. Democratise AI: Small models with quality data vs large model dependency
6. Reproducibility Standards: Adopt Model Openness Framework, metadata standards
7. Secure Environments: Sandbox infrastructure, governance frameworks
8. Global Standards: RDA-led harmonisation of governance frameworks

RDA-Microsoft White Papers

- **Open white papers**
- **Practical frameworks** and **best practices**
- **Community dialogue** and **valuable resources**

https://bit.ly/RDAandAI

- How researchers **prepare data for AI** and **develop skills** for AI-driven research
- How **technologies** like **HPC** and **QC transform research** capabilities

Agentic AI in support of Science and Research

- New community consultation on **agentic AI in research**
  - Autonomously assist with research tasks
  - Accelerate research, lower costs and democratise access to tools
- **November 2025: Survey** and **online sessions** - *watch this space!*

https://bit.ly/RDAandAI

Conclusion

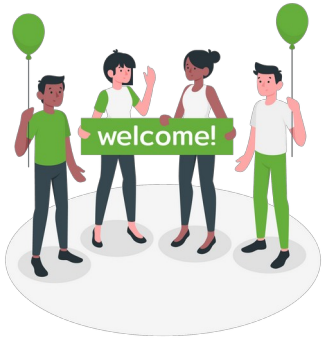
WHAT WE'VE LEARNED

- ✓ AIs impact on research data is immediate and profound
- ✓ Community expertise exists - and needs coordination
- ✓ Bottom-up AND top-down approaches both necessary

THE CHALLENGE

- ✓ Ensure AI-driven research remains open, trustworthy, impactful
- ✓ Build infrastructure for responsible AI at scale
- ✓ Bridge gaps between technology, policy, and practice

Get involved



- Join RDA (free individual membership)
- Propose or join Working and Interest Groups
- Follow-up AI activities (Q4 2025)
- Raise/participate in BoFs at future plenaries
- Share your expertise and use cases

Resources:

- All RDA outputs curated at: rd-alliance.org/value_rda
- RDA AI community expert directory

Hilary Hanahoe
Secretary General
hilary.hanahoe@rda-foundation.org

Trish Radotic
Community Development Manager Oceania & East Asia
trish.radotic@rda-foundation.org