May 2025

Report on Day 5 of DSW 2025 – Open Data Forum -

1 Abstract

The final day of Data Spaces Week 2025 in Chennai centered on the theme "Open Data: Contexts, Governance, and Global Cooperation." Sessions examined the dual role of open data—as a catalyst for innovation and as a source of ethical, legal, and operational challenges. Presentations highlighted use cases ranging from **public infrastructure (Indian Railways)** and **democratic accountability** (Sansad Ratna Awards) to privacy-preserving analytics and international healthcare data spaces (**Dataspace4Health**). Discussions explored key issues such as **data credibility, AI hallucination, regulatory gaps**, and the need for contextual modeling, metadata standards, and trust mechanisms. Together, the day emphasized that unlocking open data's potential requires **technological safeguards, ethical frameworks, and global cooperation** across academia, industry, and government.

2 Attendees

Day 5 Participants: 79 (in person: 65, including students; online: 14)

3 Welcome Address – Thematic Overview of DSW2025 Final Day

Gopal Tadepalli (Anna University) opened the final day of Data Spaces Week 2025 (April 11, Friday) by welcoming participants and outlining the week's structure. He explained that the first two days were dedicated to the **IOFDS roundtable**, followed by a **day on research directions**. The fourth day focused on the **IEEE Data Trading System Working Group (DTS WG)**, chaired by Dr. Hiroshi Mano, with strong contributions from international experts including Kohtaro Asai and other DSA members.

The fifth day, he noted, was conceptualized around the theme of **Open Data**. He emphasized the growing role of open and synthetic datasets in academic and industrial research, particularly for early-stage analysis and prototyping. Tadepalli highlighted Professor Inder Gopal's keynote on **data as a public good**, stressing that public institutions in India are increasingly committed to transparency through data sharing.

He also previewed key sessions of the day, including:

- A presentation on the **Sansad Ratna Awards** by its founder, Thiru K. Srinivasan.
- A panel discussion moderated by a **Deputy General Manager from the Reserve Bank of India**, featuring insights from the former Director of IRCTC and other experts.

- An important session on the **Luxembourg Testbed**, coordinated through **NTT DATA**, where Japanese collaborators Masaru Dobashi and Miki Kanno would join remotely despite returning to Japan.

Special mention was made of **Kohtaro Asai** (Mitsubishi Electric & DSA), whom Tadepalli praised for his pioneering work in audio/video compression and data storage technologies over a 44-year career. He recounted a personal anecdote to underline Asai's dedication and humility.

Tadepalli further introduced Prof. Jinju Watada, an 80-year-old emeritus professor from Waseda University, now spearheading data science initiatives at Shimonoseki City University. Watada's vision includes engaging undergraduate students in data space research from the early years of their academic journey, reflecting a proactive and intergenerational approach to building future capacity.

Finally, he welcomed **Dr. Banu Kumar**, Founder Director of Ara Associates, as the **Session Chair** and **Chief Guest** for the day. Tadepalli praised his three-decade career in **automatic identification, data analytics, cloud infrastructure**, and his active role in mentoring students and advising academic institutions. Kumar's industry experience and dedication to societal impact were presented as fitting complements to the open data focus of the final day.

4 Speeches

4.1 Bhanu Kumar

Bhanu Kumar, Founder Director of Ara Associates, delivered a keynote address exploring the promise and paradox of Open Data in the context of responsible digital infrastructure. Drawing on his industry experience, he emphasized the balance between openness, privacy, and ethical governance. (see d5-1AM-01-00-Bhanu Kumar_DSW_2025.pdf)

Bhanu Kumar began by offering a reflective overview of the five-day Data Spaces Week, noting its progression from **societal and research perspectives** to standardization, and finally to open data. Framing his talk around the theme of "**the promise and the paradox,**" he explored both the transformative potential and the ethical complexities of open data in today's digital ecosystems.

He articulated the **promises of open data** as fostering **innovation, collaboration, transparency**, and **public good,** enabling faster scientific discovery, cross-sectoral solutions (e.g., in healthcare, mobility, agriculture), and democratization of knowledge. However, he warned against its paradoxes—including privacy risks, ethical dilemmas, misuse, and the potential for "**context collapse**" in the absence of metadata or consent frameworks.

Kumar emphasized the need for a "responsible openness by design", which entails:

- **Thoughtful curation,** stakeholder input, and ethical review prior to data release
- Establishing **tiered access models,** use-case-specific limitations, and data stewardship
- Ensuring openness is **purpose-driven,** not blanket or indiscriminate

He introduced the idea of "**trading in**" instead of "trading off" between openness and privacy—promoting openness **through** privacy by using technological safeguards such as:

- **Differential privacy, federated data architectures**, and **secure multiparty computation**
- **Homomorphic encryption** for privacy-preserving analytics
- **Blockchain, metadata standards,** and **decentralized identifiers (DIDs)** for auditability and traceability

He also emphasized the importance of **building trust through a triad of security, transparency, and ethics,** describing these as intertwined pillars that support responsible data ecosystems. With security as the backbone, transparency as the operational ethos, and ethics as the societal compass, organizations can avoid "analysis paralysis" and instead deliver targeted, trustworthy insights.

A strong case was made for the role of **universities as neutral, credible custodians of data,** capable of acting as:

- **Trust anchors** for the public
- **Bridges** between policy and practice
- **Catalysts** for embedding ethics into research, education, and data innovation

Kumar concluded by advocating for a **culture shift:** embedding ethics into curriculum, championing citizen data literacy, supporting community data hubs, and co-creating frameworks for inclusive and sustainable open data governance. His final message: "Open does not mean unrestricted. Every stakeholder has a role in shaping the future of data with dignity, security, and responsibility."

4.2 Junzo Watada

Prof. Junzo Watada, Emeritus Professor, Shimoneseki City University, delivered a lecture introducing a fuzzy logic-based method for interpreting expert judgments in uncertain environments, using earthquake damage assessment as an illustrative case. (see d5-1AM-02-00-Junzo Watada_LRM_Open-Data-Forum2025_beamerPPT.pdf)

Prof. Watada presented a method to **mathematically model human intuition and subjective assessments**, especially in cases where data is vague or uncertain. Drawing from real-world scenarios like earthquake damage inspections—where engineers often rely on **linguistic terms** ("severely damaged," "moderately affected") rather than precise numbers—he proposed a system that can convert these qualitative expressions into structured data and back.

At the heart of his approach is a concept called **fuzzy logic,** which allows for shades of meaning rather than rigid yes/no answers. By using a more advanced form known as **type-2 fuzzy logic,** he accounts for multiple layers of uncertainty—both in how people describe what they see and in how different experts might interpret the same situation.

His model enables:

- Standardizing expert opinions into computable values
- Making predictions or evaluations even when input data is ambiguous
- Converting the output back into easy-to-understand language (e.g., "low risk," "very bad")

While technical in foundation, the idea is **to bridge human intuition and machine computation**, especially useful in areas like **disaster evaluation, healthcare, or human-centered AI,** where precise data is hard to obtain.

In short, Prof. Watada's work demonstrates how **soft, fuzzy information from human experts can still be made usable in rigorous decision-making systems**, highlighting the future of explainable and interpretable AI.

4.3 Dipti Kumar

Dipti Kumar, Honorary Advisor of the Sansad Ratna Awards, delivered an informative presentation on the Indian parliamentary system and how structured parliamentary data is used to recognize exemplary Members of Parliament. She was joined by Primepoint K. Srinivasan (Founder), Priyadarshini Rahul (Chair), and V. M. Charumathy (Secretary). (see d5-1AM-03-00-Dipti Kumar_Sansad Ratna Award final ppt.pdf)

Dipti Kumar began with an overview of **India's parliamentary structure,** explaining the roles of the **Lok Sabha** (lower house), **Rajya Sabha** (upper house), and the President of India. She emphasized that India's governance rests on **four democratic pillars**: the legislature, executive, judiciary, and the press (fourth estate), highlighting their respective roles in enabling transparency and accountability.

She then introduced the **Sansad Ratna Awards,** a data-driven initiative honoring high-performing Members of Parliament (MPs), based on publicly available parliamentary records. Dipti emphasized that this project bridges **governance and open data** and is not based on subjective perception or political bias.

Primepoint K. Srinivasan expanded on the origin and motivation of the awards. The initiative was launched in 2010 on the suggestion of **Dr. A.P.J. Abdul Kalam**, with the goal of highlighting MPs who contribute constructively but are often overlooked in the media. Using **quantifiable indicators** such as:

- Debates participated

- Questions asked
- Private members' bills introduced
- Attendance
- Committee involvement

The team has created a **custom AI-powered chatbot** (based on Perplexity AI) that allows users to retrieve detailed performance data of MPs across the 14th to 18th Lok Sabha. The chatbot is carefully trained on curated sources like PRS Legislative Research and Prime Point Foundation data archives to ensure accuracy and factual consistency.

V. M. Charumathy demonstrated how the chatbot can generate **comparative analyses**, such as the performance of first-time MPs or gender-based statistics, and showed how individual MP profiles like that of Supriya Sule (a top performer and jury member) can be retrieved.

Srinivasan also addressed the **importance of parliamentary committees** (including Joint Parliamentary Committees and the Public Accounts Committee) in India's democratic process, stressing that data from these committees is rich but underutilized.

Finally, the speakers emphasized that **open data should be trustworthy and verifiable**, and that through this initiative, **universities, technologists, and civic actors** can play a role in shaping **data-informed public discourse.**

4.4 Sriram Venkatachalam

Sriram Venkatachalam, formerly Director of **Indian Railway Catering and Tourism Corporation (IRCTC),** delivered a wide-ranging lecture on the Indian Railways' digital transformation and the immense scale of its data generation, integration, and utilization. Drawing on decades of experience, he traced the evolution of key ICT systems, from early computerization in the 1970s to today's cloud-based, API-enabled infrastructure. (see d5-1PM-01-00-Sriram Venkatachalam_few slides for COE Guindy.pdf)

Indian Railways, operating 24,000 trains daily across 7,461 stations and employing 1.2 million people, is among the world's largest data-generating organizations. Passenger-focused systems like the Passenger Reservation System (PRS), the Next Generation e-Ticketing (NGeT), and the Unreserved Ticketing System (UTS) handle tens of millions of transactions each day. These are complemented by systems for freight movement (FOIS), real-time train tracking (RTIS), and data analytics (including data lakes and dashboards like e-Drishti).

He highlighted the **IR Cloud** and **Enterprise Service Bus** as foundational elements for application interoperability across more than 70 ICT systems. The Indian Railways' backend now supports **route optimization, coach maintenance, crew scheduling, freight forecasting,** and **catering logistics,** among others. Specific applications such as Food Track for meal services and

API gateways for integration with third-party services (e.g., MakeMyTrip) exemplify how data enables new business models and improved passenger experience.

Yet, Venkatachalam emphasized that **data utilization** still lags behind data availability. Institutional barriers, legacy systems, and limited cross-sector integration have hindered full exploitation. He advocated for greater collaboration with data scientists, universities, and government IT agencies (e.g., MeitY[1], NIC[2]), noting that Indian Railways has started engaging external experts to unlock this potential.

He concluded by underscoring data as a **renewable, ever-enriching resource**, not merely "the new oil." With continued investment in analytics and cross-functional expertise, Indian Railways can serve as a model for national-scale, **mission-critical data spaces**.

4.5 Niel Gordon

Prof. Niel Gordon, Professor of Computer Science at the University of Hull (UK), delivered a comprehensive presentation on the ethical, social, and global dimensions of open data in the modern world. Drawing on his experience with the British Computer Society and the BCS Working Group on Social Accountability, he explored how data generated from everyday digital interactions can be harnessed responsibly for the public good. (see d5-1PM-02-00-Niel Gordon_OpenData.pdf)

Prof. Gordon began by tracing the evolution of computing from centralized mainframes to ubiquitous, interconnected devices, highlighting how the emergence of smart homes, wearables, and IoT systems has transformed personal environments into **domestic data spaces**. He emphasized that while such data-driven systems offer benefits in healthcare, logistics, and convenience, they also raise profound questions about **privacy, surveillance, and ownership**.

In framing **open data** as both an opportunity and a responsibility, he presented it as a potential engine for **social accountability, policy transparency,** and **international cooperation.** Drawing on case studies from the UK, the EU, and beyond, he showcased public open data portals—such as the UK Government Data Store and the EU's data.europa.eu—as examples of initiatives promoting **data accessibility for research and civic engagement.**

The lecture then shifted focus to the **ethical dilemmas** posed by AI, cloud services, and cross-border data flows. Prof. Gordon warned against overreliance on opaque AI models trained on questionable datasets and stressed the importance of **governance frameworks** that prioritize

[1] Ministry of Electronics and Information Technology, Government of India. The "Y" intentionally derived from the last letter of "Technology" to emphasize the ministry's technological focus.

[2] National Informatics Centre (NIC) is an attached office of MeitY

fairness, data protection, and user agency. He highlighted the potential of open data in addressing global challenges—such as climate change, public health, and demographic shifts—especially when aligned with **UN SDGs.**

In closing, Prof. Gordon advocated for data practices rooted in **transparency, inclusiveness, and ethical responsibility**, concluding with the reminder that open data must be accompanied by open minds and a shared purpose for the common good.

4.6 Miki Kanno (NTT DATA, Japan), Michael Mossal and Katharina Hrom (NTT Data, Luxembourg)

Miki Kanno (NTT DATA, Japan), **Michael Mossal and Katharina Hrom** (NTT DATA Luxembourg) delivered a joint presentation on cross-regional data space collaborations and the development of **Dataspace4Health.** Drawing on their operational and strategic work in Europe and Japan, they shared insights into international testbeds, trusted data exchange, and healthcare data infrastructure. (see d5-1PM-03-00-MikiKanno_OpenData_NTT_Internationalcollaborativeinitiatives_day5.pdf)

The presentation opened with **Miki Kanno** outlining NTT DATA's global efforts to enable **inter-company and cross-border data collaboration**. She emphasized the growing challenges of managing data sovereignty, compliance, and interoperability across diverse industrial and regulatory environments. NTT DATA has established a **Global One Team** to coordinate data space-related activities across Japan, Europe, and other regions. Their active involvement includes testbeds with the **University of Tokyo**, participation in the IDSA and **Gaia-X community**, and implementation of trusted data exchange solutions based on Gaia-X's clearing house and connectors referencing Japanese and European ecosystems. A highlight was the **technical linkage established between testbeds in Japan and Luxembourg,** enabling secure, cross-border data transactions—an important milestone for trusted international interoperability.

Following this, **Michael Mossal** traced the origins of the **Dataspace4Health** project in Luxembourg. It arose from two earlier healthcare initiatives: one focused on cardiovascular treatment using cloud-based data lakes, and another centered on managing pandemic-related mass testing logistics. Despite technical success, both faced regulatory hurdles—particularly around public cloud storage of sensitive health data. This spurred the development of a sovereign, regulation-compliant data space for healthcare, co-financed by the Luxembourg Ministry of Economy and led by NTT DATA Luxembourg.

Katharina Hrom elaborated on the project structure and use cases. **Dataspace4Health** is a national initiative involving hospitals, research institutes, public authorities, and compliance partners. The

project has four strategic pillars: 1) establishing a national health data strategy; 2) implementing a minimum viable health data space; 3) developing decision support systems in diabetes and oncology; and 4) preparing for long-term operational sustainability. Two pilot use cases focus on:

- **AI-assisted diabetes management,** using anonymized hospital data to predict complications;
- **A distributed oncology clinical study** infrastructure, overcoming Luxembourg's limitations in centralized university hospitals by enabling secure remote collaboration.

The team emphasized alignment with the **European Health Data Space (EHDS)** regulations, while also demonstrating international extensions—such as a planned use case on Parkinson's disease in collaboration with Japanese institutions.

The session concluded with a future outlook on broader global cooperation. The speakers shared their vision for building **interoperable, secure data spaces** between Luxembourg, Japan, and additional partners like Spain, framed within the **DataSpace One** initiative. They stressed that data exchange mechanisms must be both technologically sound and socially responsible, aligning with ethical frameworks, data governance standards, and public trust imperatives.

Q&A Session Highlights

No formal Q&A was included due to time constraints, but closing remarks from Prof. Gopal Tadepalli commended the collaborative depth between NTT DATA and the Luxembourg consortium, especially in establishing internationally connected, regulation-compliant data infrastructures. He highlighted the initiative's significance in improving diagnosis, treatment, and preventive healthcare on a global scale.

5 Pannel Discussion: Contextualizing Open Data

Moderator: Ms. Kalaranjani Mahavir, Deputy General Manager & Head, Cyber Security and Information Technology Examination (CSITE), Reserve Bank of India, Chennai

Panellists:

Dr. Parthajit Kayal (Madras School of Economics)
Mr. Rahul Krishnan (MathWorks, Bengaluru)
Advocate Manoaj (Distinguished lawyer with expertise in civil, criminal, and corporate law)

This multidisciplinary panel explored the opportunities, responsibilities, and challenges in leveraging **open data** within legal, academic, and technological domains. Drawing on their diverse expertise, the speakers emphasized the importance of **data credibility, ethical usage, contextual modeling,** and **regulatory preparedness** in an era of accelerating AI and data integration.

Key Themes:

Risks and Responsibilities of Open Data:

- Rahul Krishnan emphasized that open data is a "double-edged sword" — its usefulness depends on source verification, contextual integration, and responsible modeling frameworks.
- Advocate Manoaj cautioned against over-reliance on AI-generated outputs without checks, citing legal cases of fabricated judgments. He stressed the need for guardrails, professional accountability, and a regulatory framework.
- Dr. Parthajit Kayal highlighted the ethical dilemmas faced by researchers when the same data, modeled differently, leads to diverging conclusions, sometimes with reputational consequences.

Technical & Academic Approaches to Contextualization:

- Technologies such as **graph search**, **transformers**, and **AI governance frameworks** were presented as tools to enhance metadata integration and ensure traceability in modeling pipelines (Rahul Krishnan).
- Academic training at the Madras School of Economics involves exposing students to open datasets, guided research with data dashboards, and applications of **ANOVA, logistic regression**, and nowcasting for economic indicators (Dr. Kayal).
- Students are taught to replicate published results using openly referenced datasets and to develop independent interpretations within resource constraints.

Legal Considerations:

- Advocate Manoaj underscored the lack of formal legal frameworks in India for open data contextualization and AI accountability.
- He referenced European AI legislation that includes liability, IP protection, and insurance provisions as necessary precedents.
- He stressed the need for certification systems, professional ethics enforcement, and user education to prevent misuse.

Data Integration & Bias Mitigation:

- The challenge of combining numerical, textual, and categorical data in models like IPO price forecasts was shared (Dr. Kayal). He noted the black-box nature of ML models, advocating for explainability tools like SHAP values to make outputs more interpretable.
- Both panelists acknowledged the presence of algorithmic bias and called for global certification mechanisms to ensure fairness.

Security, Privacy, and Infrastructure Gaps:

- The moderator and audience raised concerns about the digital infrastructure limitations in academia for secure data handling.

- Rahul noted the difficulty of applying privacy protections retroactively to open data and advocated for secure internal processing environments from the outset.

Summary Statement by Moderator:

Ms. Mahavir concluded that open data should be:

- Accessible, machine-readable, and free under an open license
- Designed to promote democracy, transparency, and innovation
- Accompanied by a responsible regulatory framework for ethical and secure use

The discussion was praised for its interdisciplinary perspective, with academia, legal systems, and technology sectors offering complementary insights on a shared societal challenge.

Acknowledgments and Editorial Note

Isamu Yamada of the Data Society Alliance (DSA), serving as the secretariat of the International Open Forum on Data Society (IOFDS), compiled this report based on the speaker's presentations and discussions, including participants. Each speaker has made efforts to review the content. We thank all contributors for their valuable input and cooperation.