May 2025

Report on Day 3 of DSW 2025 - Research Directions in Data Spaces-

1 Abstract

The third day of **DSW2025 in Chennai**, held on **9 April 2025**, focused on the theme "**Research Directions in Data Spaces – A Broad Spectrum.**" The sessions brought together leading scientists, institutional heads, and technology practitioners from across India and Asia to explore current challenges and future directions in building trustworthy, interoperable, and privacy-preserving data ecosystems.

Speakers presented diverse perspectives—from **national data infrastructure** and **digital public goods** to **AI governance**, **semantic interoperability**, and **privacy-enhancing technologies**. Highlights included keynote talks from **CSIR-4PI, Indian Institute of Finance, Indian Institute of Science, Indian Statistical Institute**, **Nanjing University**, and **Siemens Technology and Services Ltd**, all emphasizing the need for decentralized, consent-based, and inclusive data architectures. Special focus was given to real-world implementations such as **Geospatial Data Interface (GDI)**, **Agricultural Data Exchange (ADeX)**, and **P3DX privacy framework**, positioning India as an emerging leader in the global data space movement.

Closing reflections emphasized **international collaboration, standards alignment**, and the importance of actionable frameworks over theoretical constructs—underscoring a shared commitment to advancing practical, trustworthy, and inclusive data ecosystems through sustained dialogue and technical progress.

2 Attendees

Day 3 Participants: 72+ α （in person:72, including students; online: α）

3 Speeches

3.1 G. K. Patra

Prof. G. K. Patra, Chief Scientist and Head, Director's S&T Secretariat, **CSIR Fourth Paradigm Institute (CSIR-4PI)**, delivered a keynote addressing the promise and challenges of data spaces from the perspective of India's national data science infrastructure. Drawing on CSIR-4PI's long-standing expertise in modeling, simulation, and AI, he provided insights into how centralized systems currently dominate data handling within CSIR and why a shift toward **federated** and **decentralized architecture** is increasingly necessary. (see d3-1AM-01-00- G K Patra_Data-space-week-2025-4PI.pdf)

CSIR-4PI is India's only independent institute fully dedicated to **data science**, serving as a **horizontal enabler across 37 CSIR laboratories**. Originally established as the Centre for Mathematical Modelling and Computer Simulation, the institute was restructured in 2012 to support data-intensive research—the "Fourth Paradigm" of science, where knowledge is derived directly from large-scale data.

The institute has led initiatives in **generative AI, synthetic biometric data** for **Aadhaar, foundational model fine-tuning** for Indian use cases, and **domain-agnostic AI frameworks.** With 4.2 petaflops of compute capacity and a 3-petabyte centralized scientific cloud, CSIR-4PI supports cross-domain AI infrastructure and services. However, data within and across CSIR institutions remains siloed. While the institute has developed federated learning frameworks, it has yet to implement federated or decentralized data sharing in practice.

Patra expressed interest in data spaces as a model that could bridge these silos through secure, sovereign, and interoperable exchange. He described how national programs such as **IndiaAI** and **DEPA (Data Empowerment and Protection Architecture)** are advancing centralized architecture, while also acknowledging the need to assess whether such approaches meet the demands of broader inter-institutional collaboration. **A federated and decentralized approac**h may prove more suitable in this context.

He further emphasized the importance of **privacy-by-design**, **inclusive data governance**, and **open, collaborative research frameworks.** Challenges such as infrastructure accessibility, bias in AI models, and sustainability were identified as key issues, especially for building India-specific AI capabilities. Emerging technologies such as green AI, neuromorphic hardware, and quantum computing were noted as future enablers.

In conclusion, Patra advocated for **responsible and sustainable data practices** that prioritize usefulness over scale, and called for new architectures—such as data spaces—that can enable India to unlock its vast and diverse data assets through secure and federated collaboration.

Q&A Highlights

Comment1: Your institute's challenges match very well with the principles and goals of data spaces. I'm encouraged by how deeply you have explored the concept, especially in an Indian context.

Comment2: In Japan as well, cross-border collaboration is necessary for research and innovation. International data sovereignty and utilization must be balanced, and 2025 marks a good starting point for such efforts globally.

Comment3: From an industry perspective, data spaces open up many research and application opportunities. I hope for more international collaboration in this area.

3.2 Aman Agarwal

Prof. Aman Agarwal, Indian Institute of Finance, New Delhi, delivered a keynote on the interplay between **AI, data spaces,** and **knowledge** as a resource for building human capital. Drawing from his background in **finance and economics**, he argued that the value of data lies not in its raw form, but in the **knowledge and economic impact it enables.** (see d3-1AM-02-00-Aman Agarwal_Data Science Anna University Chennai.pdf)

He emphasized that data must be **structured, trusted, and ethically governed** to serve national and societal goals. Reflecting on global inefficiencies witnessed during his time at the **World Bank,** he highlighted the need for data harmonization across institutions to avoid duplication and waste.

Prof. Agarwal introduced the concept of an **Automated National Labor Exchange (NLX)**—a data-driven, AI-enabled platform for real-time labor-market matching. Unlike traditional employment systems, NLX treats labor hours as tradable assets, supporting transparent, skill-based hiring free from institutional or human bias. This model, he argued, could improve employment, promote gender neutrality, and reduce information asymmetries.

He also reflected on the role of frameworks like **Aadhaar** in integrating cross-sector data, while cautioning against overreliance on idealized privacy. In his view, trust—both technical and institutional—is the critical currency in any data-driven system.

In conclusion, he called for the development of data spaces that are secure, interoperable, and aligned with economic, social, and ethical objectives, positioning them as foundational infrastructure for inclusive and sustainable growth.

3.3 Jyotirmoy Dutta

Jyotirmoy Dutta, Principal Scientist at the Indian Institute of Science (IISc), Bengaluru, delivered a keynote on data exchange platforms as critical components of **Digital Public Infrastructure (DPI),** with a focus on the **Geospatial Data Sharing Interface (GDI).** (see d3-1AM-04-00-Jyotirmoy Dutta_Data Exchange Platforms as Digital Public Infrastructure.pdf)

He underscored the pressing need to address fragmentation in public **geospatial data**, which is often scattered across ministries, stored in incompatible formats, and difficult to access. His team builds data exchanges that connect data providers and application developers through modular, privacy-preserving, and federated architecture—where data remains with the provider, but **cataloging, authentication, and consent mechanisms** are managed centrally.

At the core of this work is **GDI,** developed under India's 2022 **National Geospatial Policy,** which mandates that data generated with public funds be made fairly and legally accessible. GDI already aggregates datasets from over **37 providers**, offering APIs, metadata catalogs, and consent

workflows to enable trusted, scalable data use across sectors such as urban planning, agriculture, and environmental monitoring.

He emphasized that **one-size-fits-all architectures do not work,** and that **interoperability, data trust,** and **domain-specific consent** must be embedded from the start. Real-world use cases—such as smart waste routing, satellite image analysis, and data-driven welfare eligibility—demonstrated how these platforms reduce barriers to innovation and public service delivery.

Dutta concluded by positioning data exchanges as India's emerging "third layer" of public digital infrastructure—complementing **Aadhaar (identity)** and **UPI (payments)**—and essential for achieving national goals in digital inclusion and economic growth.

3.4 G. Ravindran

Dr. G. Ravindran, Professor & Head of the SQC and OR Unit, Indian Statistical Institute (**ISI**), Chennai, delivered a comprehensive talk on **the role of statistics in enabling responsible data spaces and data science**. Drawing on ISI's institutional experience, he highlighted how statistical thinking is vital for building trustworthy, scalable, and inclusive data ecosystems. (see d3-1PM-01-00-G. Ravindran_Data Science, Statistics, and Data Spaces.pdf)

Dr. Ravindran emphasized that **data spaces** are not mere data repositories but federated socio-technical infrastructures with shared vocabularies, governance rules, and trust frameworks. He contrasted them with the broader field of data science, which focuses on extracting insights from complex, high-volume datasets.

Key contributions of statistics to data spaces include:

- **Data harmonization and bias correction** across fragmented sources.
- **Uncertainty quantification** using Bayesian methods, bootstrapping, and confidence intervals.
- **Auditing algorithms** for fairness and transparency through statistical testing and causal inference.
- **Privacy-preserving techniques**, such as differential privacy and synthetic data.
- **Interpretability of AI models** using regression, additive models, and PCA.

A notable example was ISI's support during India's **COVID-19 vaccination drive**, where it helped optimize logistics using hierarchical models and spatio-temporal methods.

He introduced ISI's new **BStat in Data Science** program (launched 2024), which combines classical training in probability and inference with modern tools like machine learning and R/Python. The goal is to nurture **ethically grounded, domain-aware data scientists.**

He concluded with a call to view data not as abstract entries, but as reflections of social systems, echoing ISI founder P.C. Mahalanobis: "Statistics must serve the real needs of people."

Q&A Highlights

Q: How is ISI supporting government data initiatives?

A: Through audits, survey design, and statistical modeling for programs in Tamil Nadu and national ministries.

Q: What about algorithmic fairness?

A: ISI applies hypothesis testing and causal inference to assess AI fairness and runs ethical AI workshops in collaboration with IISc and others.

3.5 Vasant Rajaraman

Vasant Rajaraman, Center for Data for Public Good, Indian Institute of Science (IISc), Bengaluru, presented a detailed case study on **ADeX: the Agricultural Data Exchange initiative** developed in collaboration with the Government of Telangana. Drawing from work under the **Datakaveri project**, he explained how data exchanges can function as **Digital Public Infrastructure (DPI)** to unlock high-impact, citizen-centric use cases.

He traced the evolution from India's urban data exchange pilot (2019) to **ADeX Telangana,** an operational platform enabling **secure, consent-based, and standards-driven data sharing** among government departments, startups, and financial institutions. Unlike open data portals, ADeX enforces **policy-based access control,** differentiating between **public, private, and personal data**, and incorporates modules for consent collection, metadata catalogs, and immutable logging.

Key outcomes and features:

- **Credit & finance use case:** Integration with HDFC[1] and SBI[2] allowed banks to access verified farmer data, reducing loan processing time from two weeks to **one day**, benefiting over **25,000 farmers**.
- **Agri-advisory applications:** Enabled personalized services for pest alerts, market locations, and soil health, improving **decision-making at the last mile.**
- **Interoperability:** Based on **open APIs, JSON schema validation,** and aligned with standards like OGC and NGSI-LD, making it portable across sectors and states.
- **Governance-by-design:** Consent artifacts, data-sharing agreements, and fine-grained policy control ensure trust and traceability.

He also emphasized the **scalability and flexibility** of the platform—deployed across 50+ virtual machines in a federated architecture—and its open-source foundation, enabling transparent, reproducible deployment across domains like geospatial and health.

---

[1] HDFC Bank（Housing Development Finance Corporation Bank）

[2] SBI（State Bank of India）

Q&A Highlights

Q: Can farmers directly access ADeX?

A: No, ADeX is used by institutional actors like data officers or application developers. Farmers benefit indirectly through services built on ADeX.

Q: How is personal data handled?

A: Via explicit consent mechanisms, based on standards proposed by MeitY[3], ensuring privacy-compliant data flows through consent artifacts and audit logs.

3.6 Gautam Ganesh, Vinay Sharma, and Kaushlendra Pratap Singh

Gautam Ganesh, Vinay Sharma, and **Kaushlendra Pratap Singh**, from Siemens Technology and Services Pvt. Ltd., Bengaluru, delivered a rich and layered presentation titled "**Semantics-Driven Data Strategy and AI Governance**", introducing practical applications of **knowledge graphs, semantic ontologies, and polyglot data models** to enable trusted and interoperable industrial data exchange. Drawing on Siemens' applied research in manufacturing and infrastructure, they presented a unified framework linking data interoperability, digital twin integration, and AI governance.

Gautam Ganesh opened by outlining the growing complexity of industrial data environments, where vast heterogeneity, lack of shared semantics, and low data veracity create barriers to effective reuse and AI deployment. He introduced Siemens' internal **ontology-driven strategy**, based on **polyglot modeling**, to unify disconnected data from ERP, PLM, sensor, and document systems. Siemens developed a curated ontology library, enforced through Git-based version control, to encode and preserve domain expertise and enable **graph-based querying, semantic search**, and **explainable AI**.

Vinay Sharma then expanded on how these semantic tools are applied in the context of the **Industrial Metaverse**—a persistent, immersive virtual replica of real-world industrial systems. He described how Siemens leverages **digital twins, IoT, AI/ML,** and **AR/VR** to support collaborative design, simulation, monitoring, and remote operations. A key challenge is the **lack of interoperability** among legacy machines and diverse toolchains. To address this, Siemens built a **semantic interoperability layer**, using ontology mapping and knowledge graph synthesis to allow applications to communicate without exposing proprietary models—enabling **data trust, system-level integration, and real-time collaboration.**

Kaushlendra Pratap Singh presented a second real-world application: **smart infrastructure and building management systems (BMS)**. Siemens' BMS tools aggregate data from

[3] Ministry of Electronics and Information Technology, India

heterogeneous formats (CSV, TTL, ZIP) and various vendors (Honeywell, Tridium). Manual mapping of data structures to client applications previously required 20–30 days. By introducing a BT-DOM ontology and graph-based enrichment pipeline, Siemens automated the onboarding process, **reducing processing time to 30 minutes**. This transition from tree-based to **graph-based modeling** supports scalable, standard-aligned, and domain-aware interoperability across smart buildings, grids, and cities.

The team concluded by reaffirming the importance of **open semantic standards, reusability, and modular data translation services,** with polyglot models serving as a bridge to industry-wide efforts like **Manufacturing-X.** They emphasized that true data spaces require not only technical alignment, but **formalized governance and continuous feedback,** grounded in practical use cases and institutional knowledge.

Q&A & Closing Comments

Gautam Ganesh summarized the team's philosophy: polyglot ontologies are not only about structure, but about **bridging global and local truths**, reducing hallucination in AI systems, and making complex domains tractable through semantic abstraction.

Gopal Tadepalli acknowledged the Siemens team's sustained collaboration and thanked them for translating conceptual frameworks into practical implementations across multiple industrial domains. **Hiroshi Mano** presented a commemorative memento on behalf of the organizing committee.

3.7 Xue Wang

Xue Wang, on behalf of **Yikun Xia** and the **Data Management Innovation Research Center, Nanjing University (China)**, presented a study titled "**Analysis of the Characteristics, Challenges, and Strategies for Constructing Subject Value Networks in Industry Data Spaces.**" The presentation explored how multi-stakeholder collaboration can be structured and sustained in industrial data spaces through value networks.

Drawing on ecological synergy theory and four case studies (in mobility, agriculture, energy, and green sectors), the study examined:

1. **Stakeholder roles** — including governments, operators, service providers, standards bodies, regulators, data providers, and users.
2. **Structural features** — such as regulatory alignment (e.g., GDPR, DGA), cross-sector connectivity, maturity-based progression, and value co-creation.
3. **Practical challenges** — including:
   - **Uneven participation**, with SMEs at a disadvantage
   - **Fragmented interoperability** and lack of shared data models

- **Rigid governance structures,** unable to adapt to change
- **Unclear benefit-sharing,** limiting sustainability

To address these, the team proposed:

- **Capability building** for SMEs via training, toolkits, and incentives
- **Multi-party governance** led by government–industry–academia alliances
- **Dynamic frameworks** using AI, sandboxes, and policy automation
- **Innovative business models** with blockchain-based revenue sharing and SME-friendly pricing

The study concludes that while international examples show progress, **sustainable, interoperable, and inclusive value networks** remain a key challenge—and opportunity—for data space development globally.

3.8 Inder Gopal,

Inder Gopal, Director of the **Center for Data for Public Good (CDPG)**, IISc Bengaluru, delivered a keynote titled "**How Data Spaces Improve the Lives of Indian Citizens.**" Grounded in practical experience, his presentation focused on real-world implementations of data spaces as digital public infrastructure.

He began by explaining the **mission of CDPG**: to combine academic research with tangible societal impact through the construction and deployment of **live data exchange platforms**. Rather than theorizing data architectures, the center builds and operates systems—such as the **India Urban Data Exchange (IUDX)** and agriculture/benefit eligibility platforms in states like Telangana and Chhattisgarh.

Dr. Gopal defined a data space as a socio-technical ecosystem built around a data exchange platform, which serves as a federated middleware layer that supports:

- **Standardized APIs and data models**
- **Decentralized data control** (data sovereignty preserved by providers)
- **Secure data sharing** via catalogs, consent, authorization, and analytics services

He emphasized the importance of:

- **Differentiating data types:** public, private, personal/sensitive, each requiring different privacy and governance tools
- **Data policy frameworks,** such as India's DPDP Act and National Geospatial Policy
- **Reusable standards,** such as NGSI-LD and OGC, to ensure interoperability
- **App ecosystems** that drive citizen-facing services across sectors (e.g., transport ETA predictions, benefit eligibility computation)

The talk also addressed hierarchical data governance, including privacy-preserving technologies like:

- **Differential privacy**
- **Secure multiparty computation**
- **Homomorphic encryption**

Rather than classifying data themselves, CDPG builds tools that **enable data providers to enforce their own access, privacy, and monetization policies.** Dr. Gopal closed with the message that **real impact requires implementation, not just architecture**—and that **trust, interoperability, and scalability are key** to sustainable public data ecosystems.

3.9 Yikun Xia and Cong Tian

Yikun Xia and Cong Tian (Nanjing University) presented a conceptual study titled "A Research Review of Data Trust from a Systems Theory Perspective." Their talk tackled the growing mistrust in data—even as data becomes central to innovation and governance. The erosion of trust, they noted, is visible among individuals, institutions, and across economic systems.

They argued that current literature treats data trust in fragmented ways: sometimes as **technical reliability,** sometimes as **institutional governance**, or as **individual perceptions.** However, these perspectives miss the big picture: **trust is not only emotional or legal**—it is a **system-level mechanism that allows data ecosystems to function**.

To respond, thet offered a **new, structured definition**: Data trust is the **positive expectation** that people or institutions hold toward data, platforms, and governance mechanisms—shaped by their context and interaction. This involves **not just trusting the data itself**, but also the **technologies, institutions, and people** behind it.

They emphasized that data trust has three interrelated dimensions:

(1) **Who** is trusting (users, providers, regulators);
(2) **What** is being trusted (data, platforms, laws, people);
(3) **Under what conditions** trust is built (technical security, legal clarity, social perception).

In practical terms, they proposed that **building trust requires simultaneous effort on three fronts:**

- **Technical infrastructure** that ensures security and control (e.g., encryption, privacy tools);
- **Institutional frameworks** that are clear, enforceable, and transparent;
- **Public engagement and shared values** that make people feel included and protected.

Their core message: **Data trust is not automatic—it must be co-created** by aligning tools, rules, and values. Only then can we build sustainable, trustworthy digital ecosystems.

3.10 Bharath Ramesh

Bharath Ramesh, Solutions Architect at the **Center for Data for Public Good (CDPG),** Indian Institute of Science, Bengaluru, delivered a practical and insightful talk on **privacy-enhancing technologies (PETs)** for sensitive data processing. Drawing from CDPG's field implementations, he introduced the concept of **P3DX (Privacy-Preserving Processing and Data Exchange)**—a secure computing framework designed to enable data collaboration without compromising privacy.

Starting with real-world challenges like inter-departmental beneficiary identification without revealing PII, he explained how PETs—**Secure Multiparty Computation (SMPC)**, **Federated Learning**, **Trusted Execution Environments (TEs)**, and **Differential Privacy**—can empower public data infrastructure to process personal and sensitive data securely.

Bharath showcased use cases such as:

- **Confidential benchmarking** across hospitals,
- **Privacy-preserving legal complaint resolution**,
- **Federated AI training** across healthcare centers,
- And l**oan disbursement for farmers** (e.g. in Telangana) using TEs to compute eligibility without exposing individual records.

He emphasized the maturity of federated learning and TEs in CDPG's implementations and demonstrated a secure farm loan estimation tool using a TE. His call to action: PETs must evolve alongside legal and institutional mechanisms to support trusted public digital infrastructure.

4 Wrap-Up Summary of Day 3 – DSW2025, Chennai

The third day of DSW2025 concluded with informal remarks and reflections by **Dr. Gopal Tadepalli** and **Dr. Hiroshi Mano,** offering a thoughtful closure to a day filled with rich presentations on privacy, data trust, and data space applications.

Dr. Gopal emphasized the depth of interest in **cybersecurity** and **privacy-preserving technologies** across Indian academic and research institutions. He acknowledged the contributions made by remote speakers, including Prof. Inder Gopal and Bharath Ramesh, noting the high value of their talks, all of which were recorded for public access.

Dr. Mano expressed admiration for the **breadth of ideas and practical efforts** showcased by Indian and Asian contributors over the past few days. He reaffirmed that data spaces are inherently borderless and encouraged further alignment with international initiatives such as **IOFDS** and **IEEE** standards development. His comments stressed the importance of privacy protection and global interoperability, urging continued collaboration across regions and disciplines.

In closing, Dr. Gopal reflected on the guiding principle of "**Progress through Knowledge,**" a motto upheld by Anna University. He emphasized the need for knowledge-led innovation in developing trustworthy data ecosystems and signaled that the discussions would continue with more focused technical work in the upcoming **IEEE DTS Working Group meetings**.

The day ended not with a formal panel, but with shared appreciation for the collective progress made and a clear commitment to sustaining momentum through standards and open dialogue.

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.