



Rajashree Das
Expert in Residence

Data Sharing is Caring (But Take Care!)

Collaborating in data ecosystems gets more value out of data, both internally and externally, as long as it's done with care

Collaborative data ecosystems are transforming the way organizations share, innovate, and thrive — even alongside competitors. Think of it as everyone bringing their best to the table: data that's secure, consumable, and shared in real-time. What works externally can also boost collaboration within. But with generative AI raising the stakes, there's no room for shortcuts. Data integrity, privacy, and authenticity are non-negotiable. Both providers and consumers of data products must take responsibility to avoid missteps and ensure compliance. The result? A smarter, faster, and more connected ecosystem — where everyone has a seat at the table. Can we get a Like for that?

What

- Collaborative data ecosystems, different organizations sharing data under applicable regulations to create new value for all participants, take many different forms relying on foundational capabilities such as privacy, ethics, ownership, trust, compliance, and accessibility.
- Federative approaches — data mesh — are emerging, pushing business domains to truly own and manage their data and its uses, to actively collaborate with internal and external partners. Treating data as a first-class product, not only providing trusted and fresh data but also making data products available to internal and external consumers through a compelling, self-service experience.
- Interoperability and open data protocols are required to allow systems to communicate and exchange information timely and securely.
- Organizations need to comply with the Data Privacy Act. Therefore it is important to have data sharing agreements with the organizations involved, and have the data privacy officer sign-off on certified and trusted data. They also need to make sure that certified LLMs are being used within enterprise.
- Data governance platforms and market places are enablers for data monetization, unlocking data value for competitive advantage, operational efficiency, customer experience, product enhancements, while taking care of potentially sensitive data, with anonymization.

Use

- **Researchers at Asan Medical Center** in South Korea have successfully applied [homomorphic encryption \(HE\)](#) to big data from multiple institutions, demonstrating its effectiveness in securely combining diverse data sources for predictive AI models while maintaining privacy.
- **The WISDOM Project**, [comprising eight European universities](#), is using AI to address chronic immune-mediated diseases while integrating VEIL.AI's (a Finland-based startup) anonymization solutions to ensure compliance with GDPR and EHDS in handling sensitive health data.
- **The US Navy** collaborated with Anjuna Security, a US-based computer and network security company, and NVIDIA to deploy [Llama3 LLM models](#) on confidential NVIDIA H100 Tensor Core GPUs within the NVIDIA LaunchPad environment, using Anjuna Seaglass platform. It provided a virtualized, secure environment that was easy to deploy and operate.
- **Netflix** has partnered with clean room suppliers Snowflake, InfoSum, and LiveRamp to create [data clean rooms](#) that will allow advertisers to share data and build audiences in a secure and privacy-compliant way.

Impact

- Capgemini's [past research](#) has shown the benefits of data ecosystems for the private sector. These included a 15% improvement in customer satisfaction, 14% improvement in productivity/efficiency, and an 11% reduction in costs annually over 2–3 years.
- Collaborative data ecosystems are key to address societal challenges and organizational purposes, for example in health, public and citizen services, energy consumption, agriculture, and sustainability.
- In collaborative data ecosystems, organizations are likely to find unexpected new partners — fueling new, data-powered value streams, data monetization, and even breakthrough innovative business models.
- Embracing ownership and data-product management by business domains is vital to create an organization-wide, data-powered culture. Sensitive data can be anonymized through generation of synthetic data, which again can be fed into AI models as data-products.

Tech

- **Data Exchanges and Marketplaces:** [AWS Data Exchange](#), [Snowflake Data Marketplace](#), [Dawex](#), [Human Data Income \(HUDI\)](#) [Defi token-driven data monetization](#), [Harbr Data Marketplace](#), [Nokia Data Marketplace](#), [Collibra Marketplace](#), [Microsoft Azure Marketplace](#)
- **Data-sharing Platforms:** [Amazon Redshift Data Sharing](#), [Microsoft Azure Data Share](#), [Snowflake For Collaboration and Data Sharing](#), [Databricks Delta Sharing](#), [Google Analytics Hub](#), [IBM Data Product Hub](#), [eightwire](#), [Immuta](#), [baffle](#), [bobsbled](#)
- **Data-collaboration Platforms:** [Harbr](#), [Snowflake Data Cloud](#), [InfoSum Data Collaboration Platform](#), [Duality Data Collaboration](#), [Hex](#), [LiveRamp](#), [Walmart Luminate](#)
- **Federated Learning:** [IBM Federated Learning](#), [TensorFlow Federated](#), [Owkin Open Source FL](#), [Microsoft FLUTE](#), [Sherpa.ai](#), [Nvidia Flare](#), [FedML](#), [Gemmo.ai](#), [Flower](#), [Duality](#), [acuratio](#)
- **Data Clean Rooms:** [Snowflake Data Clean Rooms](#), [InfoSum](#), [LiveRamp](#), [AWS Clean Rooms](#), [AppsFlyer](#), [Epsilon](#), [Google BigQuery](#), [Claravine](#), [acxiom](#), [Optable](#)
- **Differential Privacy and Cryptography:** [Microsoft Differential Privacy](#), [Google BigQuery Differential Privacy](#), [Sarus Differential Privacy](#), [Cosmian Confidential AI for the Cloud](#), [Zama Fully Homomorphic Encryption](#), [Cryptomathic](#)
- **Data Collaboration/Data Mesh Enablers:** [DataPlex](#), [Cinchy](#), [IBM Data Fabric](#), [Talend](#), [Dremio](#), [Snowflake](#), [Omnisient](#), [Starburst](#), [Dataiku LLM Mesh](#), [Splunk](#), [Datastreams](#), [Databricks](#)