LEADING CORPORATIONS INTRODUCE DATA PROVENANCE STANDARDS

**First cross-industry standards to bring transparency to the origin of data,
enhancing trustworthiness of many data and AI applications**

NEW YORK, NY -- November 30, 2023 -- Based on the work of experts from nineteen leading enterprises, the Data & Trust Alliance (D&TA) announced proposed data provenance standards, believed to be the first with cross-industry applicability. The standards are designed to help companies understand where, when and how data they manage was collected or generated. When implemented, the standards will provide transparency into the origin of the datasets used for both traditional data applications and a rapidly growing number of artificial intelligence (AI) applications, which is expected to enhance AI value and trustworthiness.

Trust in the insights and decisions coming from data-enabled systems is enhanced when companies understand the origin, lineage and any rights associated with the data that feeds them. However, cross-industry provenance standards do not currently exist. This is one reason data scientists spend almost 40% of their time on data preparation and cleansing tasks, according to a 2022 Anaconda report. And 61% of CEOs cite the lack of clarity on data lineage and provenance as a top barrier to adoption of generative AI, according to the 2023 annual IBM Institute for Business Value CEO study.

The proposed standards were developed by data, AI, ethics, compliance and legal experts from Alliance companies including AARP, American Express, Deloitte, Howso, Humana, IBM, Kenvue, Mastercard, Nielsen, Nike, Pfizer, Regions Bank, Transcarent, UPS, Walmart and Warby Parker. All are members of the Data & Trust Alliance, a not-for-profit, cross-industry consortium that develops practices for the responsible use of data and AI.

"As businesses scale and accelerate the impact of AI with trusted data, it is necessary to ensure the technology is developed and deployed responsibly," said Rob Thomas, senior vice president, software and chief commercial officer, IBM and chair of the D&TA Data Provenance initiative. "These practical data provenance standards, co-created by senior practitioners across industry, are designed to help ensure AI workflows are not only compliant with ever-changing government regulations and free of bias, but also developed to generate increased business value. While the standards may not address every application of AI, we believe they fill an important, longstanding need."

Standards for Datasets, Surfaced in Metadata

D&TA's eight proposed data provenance standards surface metadata on source, legal rights, privacy & protection, generation date, data type, generation method, intended use and restrictions and lineage. In addition, the standards call for using a unique provenance metadata ID with each dataset. This essential information about the origin of and any rights associated

with data allows enterprises to make informed choices about the data they source and use. The result can be improved operational efficiency, regulatory compliance, collaboration and value generation.

Of the D&TA's eight data provenance standards, only one—generation date—is consistently surfaced in metadata today. Five standards curate data classification values that currently exist but are not surfaced consistently in metadata. For instance, the privacy classification values PII and PHI are widely understood, but they are not always present in metadata, leading to heightened risk and inefficiencies, as data must be reviewed and cleared multiple times for use. Entirely new are the intended use and restrictions standard, to surface the boundaries of data use for AI; and the provenance metadata unique ID, which will help track lineage over time.

The standards are designed to be used both within a company and with the company's ecosystem of data providers and data partners for use cases across the enterprise. They are less applicable to large language models trained with public, web-scale datasets. By adopting the data provenance standards, businesses will have a more effective way to understand datasets before purchase or use—and have a basis to decline data or request changes from third parties. Meanwhile, data providers will only need to address one set of standards, greatly increasing collaboration across the business ecosystem.

The proposed data provenance standards are currently being tested across the Alliance—in test cases ranging from regulatory compliance and supply chain to procurement and virtual patient healthcare.

"As a leading global provider of business decisioning data and analytics whose responsible AI strategy is anchored on transparency and trust, Dun & Bradstreet is pleased to partner with the Alliance to test the proposed data provenance standards," said Gary Kotovets, chief data & analytics officer, Dun & Bradstreet. "We believe the proposed data provenance standards will help organizations establish trust in solutions and experiences that leverage data and AI technologies through increased transparency, interoperability and compliance insights to support accountability—all of which are essential building blocks in this rapidly evolving space to help everyone achieve better outcomes."

The Data & Trust Alliance is actively soliciting input. Interested practitioners can visit dataandtrustalliance.org to learn more about the standards and contribute to them. "The standards were derived from pain-points that our members shared around data provenance—across a variety of industry use cases," said Saira Jesani, deputy executive director, D&TA. "Now, it's important for us to open it up to the broader business ecosystem. We are inviting practitioners from all industries to give us input and join a community of practice to share new use cases and make the metadata more robust and adoptable for all."

The Alliance currently expects to release Data Provenance Standards V1 in 2024.

Support for the Data Provenance Standards

Neil Blumenthal, co-founder and co-chief executive, Warby Parker: "Transparency and accuracy around the origin of food, water, raw materials and capital are fundamental prerequisites for society, essential to establishing trust and defining quality. At Warby Parker, we've always felt the same standard must apply to data. We are excited by the rapid evolution of AI and believe we are uniquely positioned to bring this innovation to the optical industry. Expanding the use of AI is only as good as the data we have, and we believe these new data provenance standards will lead to better and more accessible products and services for customers, as well as productivity gains throughout the industry."

Bruce D. Broussard, president and chief executive officer, Humana Inc: "The reliance on high-quality trusted data is critical to ensure the value of AI, and as businesses increasingly use the technology to better serve customers, members, and patients, it's vital we take proactive steps to preserve their trust and make certain AI works as intended. The need goes beyond an effort for just one company. One day, regulation may help address this need, but there's a significant opportunity in AI today, and business must act swiftly."

Mike Capps, chief executive officer and co-founder, Howso: "Garbage in, garbage out: that's the problem with AI today. Most AI systems are black boxes: data goes in and an answer comes out, but we have no idea what data was used, where it was sourced, or how the AI interpreted it. The new cross-industry standards from the D&TA are a huge leap forward in increasing trust and transparency in AI because they will ensure models are trained on reliable data from a traceable provenance."

Jason Girzadas, chief executive officer, Deloitte US: "We believe creating responsible technology is everyone's responsibility. That's why Deloitte is proud to collaborate with leading organizations promoting transparency in the datasets that power AI. The development of the first cross-industry standards around data provenance is an important step forward to help businesses more confidently take advantage of evolving AI technologies."

Jo Ann Jenkins, chief executive officer, AARP: "As a trusted source for critical information impacting those over the age of fifty, AARP applauds the data provenance standards proposed by D&TA. These standards align with AARP's mission to provide clear, simple and transparent information on matters of importance to those 50+, including the trustworthiness of AI and the data that powers it."

David Kenny, executive chairman, Nielsen: "Trust and transparency in the data that fuels media industry economics are critical. Leaders in the private and public sectors have a deep responsibility to build a thoughtful framework around the use of AI that enables its benefits while sternly mitigating its risks. Central to this work must be our ability to validate data

sources and protect and credit intellectual property across the vast communities of creators and technology innovators. The adoption of these data provenance standards will be a key step towards ensuring data integrity throughout the content and advertising ecosystems."

Arvind Krishna, chairman and chief executive officer, IBM: "In this era of generative AI and rapid technological advancement, open innovation is key to driving effective outcomes. By adopting and amplifying these data provenance standards across industries, enterprises can create an ecosystem that fosters greater transparency and accountability in service of the safe and responsible deployment of technology."

Glen Tullman, chief executive officer, Transcarent: "Healthcare is an information business. For Transcarent, and an increasing number of healthcare companies, information based on high-caliber data is foundational to everything we do. Thoughtful and practical data provenance standards will be key to enabling physicians and other health and care professionals to deliver high-quality, cutting-edge care with confidence, so they know where, when, and how the data they are using to make treatment decisions was collected and generated. Data quality is a matter of safety for people receiving care and is critical to the well-being of our industry. We applaud the Data & Trust Alliance for being a cross-industry convener committed to developing practical resources."

Ken Finnerty, president; IT & data analytics, UPS: "The creators of AI platforms are not the only players in this inflection point. Enterprises in every industry are deploying data and intelligent systems that are core to their business. Companies like ours feel a deep responsibility to ensure new value creation, as well as trust and transparency of data with all of our customers and stakeholders. Data provenance is critical to those efforts."

Nuala O'Connor, SVP and chief counsel, digital citizenship, Walmart Inc.: "As the pace of innovation increases and more sophisticated data assets and AI models are integrated into our customer experiences and business operations, it's important those we serve feel confident and comfortable with the ways we use data and technology. The D&TA's proposed data provenance standards will help businesses understand and manage data accordingly to safeguard its integrity."

JoAnn Stonier, Mastercard Fellow of Data and AI: "As AI advances rapidly and opportunities grow, so do data risks. Mitigating these risks requires transparency, accountability and privacy. Data provenance is a crucial discipline for ensuring data integrity and ethical AI development to build trust between organizations. The D&TA's cross-industry provenance standards are a helpful guide for a future of responsible AI practices to reinforce trust in new products and business applications."

Bernardo Tavares, chief technology & data officer, Kenvue: "As a digital-first company, at Kenvue we are focused on building trust with science, and that includes consumer data and AI.

We are proud to support and partner with the D&TA on the proposed data provenance standards. Consumers trust us with their information every day, and by working with experts across industries we are creating transparency and building trust in new technologies. These data provenance standards are a step in the right direction to ensure data across a variety of platforms is used in an honest and ethical way."

"The Data & Trust Alliance came together because of a shared belief that data and AI would be critical to our member companies' future," said Ken Chenault, D&TA co-chair, General Catalyst chairman and managing director, and American Express former chairman and CEO. "Given the speed of AI's adoption and the potential impact to business and society, companies need to put in place the infrastructure for data transparency to establish trust across stakeholders." Added Sam Palmisano, D&TA co-chair, chairman of the Center for Global Enterprise and former chairman of IBM, "D&TA is focused on real-world implementation of effective and responsible AI, and our new data provenance standards bring that pragmatic approach to one of the critical dependencies for both business and society."

Data & Trust Alliance's 26 members span 15 industries, operate in more than 175 countries, and generate more than $1.6 trillion in annual revenues. Additional information on the Alliance and the Data Provenance initiative is available at dataandtrustalliance.org.

# # #