Beyond Open Data White Paper

Brianna R. (Pagán) Corremonte & Thomas Hervey

This white paper was co-written by many individuals from the CNG Forum 2025 Beyond Open Data breakout discussion group. Credit to the co-authors and group leaders of this effort — Bill Greer, Kevin Bullock, Pete Gadomski, Tyler Erickson, Marc Prioleau, and Sean Gorman — whose expertise and leadership were integral to this work. Additional credit to the numerous additional contributors whose collective efforts supported the development and completion of this report.

Introduction

The world we live in has been shaped by open data. For decades, governments, researchers, nonprofits, and companies have used the Internet to share data to unlock new science, enable civic transparency, and build infrastructure that industries now depend on. We know that great things can happen when data is shared, but openness has come to be treated as a binary checkbox: choose an open license, publish a URL, call it done. In an era of AI agents consuming data at scale and exponential data growth, that bar is no longer sufficient.

Last year, the first-ever Cloud-Native Geospatial Forum convened in Snowbird, Utah. One track, Building Resilient Data Ecosystems, was designed differently: no recordings, held under the Chatham House rule, participants speaking as themselves rather than as representatives of their organizations.

“Beyond Open Data" was one of the sessions in the Building Resilient Data Ecosystems track. This session presented hard questions about what open data actually means, what it fails to capture, and what goals we should be orienting toward instead. The result is a white paper that covers these three theses:

1. Open data is a spectrum, not a checkbox, and we need better tools to evaluate it.
2. We should not treat open data like it is open software.
3. The power of open data comes from a strong ecosystem, not its license.

(1)Usefulness is a better measure of quality than openness

Why consider what lies “beyond” open data? We are currently in an era of exponential data growth and unprecedented accessibility, driven by rapid technological advancements and the rise of automated agents capable of consuming data at scale. While historical efforts to champion "open data" established admirable goals, the colloquial concept has become lackluster and antiquated. Making data broadly accessible, in theory, does not inherently make it *useful* in practice.

Historically, "openness" has often been treated as a binary checkbox, yet defined by varying degrees of adherence to a range of different criteria such as licensing. This is problematic because, in this example, a license alone does not dictate utility. We need a better way to represent the ultimate goal of open data: moving beyond a basic baseline of “openness” to evaluate data on a future-looking *spectrum of usefulness*.

Defining the Usefulness Criteria Spectrum

To address this evaluation gap, our group set out to define what makes open data genuinely useful. We started by examining established openness frameworks, such as the 5-Star Linked Open Data and FAIR principles. Grounding our discussions in the real-world challenges users face in accessing and applying geospatial data, we debated how to objectively assess dataset quality. Our goal was to shift the evaluation from simply asking "how open is this?" to "how easily can this be used and reused for a specific purpose, or by a general audience with diverse needs?"

Through this workshopping process, we developed a report-card-style rubric. We settled on five core criteria for evaluating a dataset's usefulness, each rated on a 1-4 star scale. We then applied this rubric to various datasets. Ultimately, this distilled a core set of criteria that serves as a concrete reference guide for data providers. Below are the proposed criteria:

- **License**
  A license is the legal permission to use and reuse data or a software product. It is traditionally the primary criterion for determining whether the data is open. To minimize the effort required for users to understand and comply with license terms, data providers should use standardized open licenses rather than custom ones. Importantly,

however, just because data has a more open license (i.e., a license with fewer use and reuse restrictions) doesn’t mean it's useful or easy to reuse.

- **Cost**
  Open and free of cost are not synonymous. Open data can incur associated (financial) costs that vary significantly. Importantly, once data is made freely available, it's hard to restrict its usage.

- **Burden**
  Burden measures how long it would take a reasonably skilled user to access and start using the data of interest. Importantly, many datasets could be considered highly “open” under an open data framework (e.g., FAIR principles), yet remain burdensome to access, use, and reuse. This may be due to anti-patterned user interfaces, convoluted search/information retrieval systems, or nested, siloed data management systems that make accessing the data difficult.
  Furthermore, some organizations will email a requested dataset, which can lead to long wait times. In other cases, some organizations use niche data formats that are supported by only a few software programs. Interoperability, which we define as the ability of your data to work with other systems (proprietary or not), can add a significant burden and contribute to the overall burden measurement.

- **Access**
  Similar to the “Accessible” definitions in the FAIR principles framework, access is the degree to which all audiences can use data. However, in practice, data can be available under an open license yet be difficult or impossible to access. For example, putting open data behind a registration or paywall reduces access. Additionally, if the data is in a proprietary format that requires downloading specialized software, access is further limited. If the data can't be directly shared, you’re limiting downstream access for other users. Importantly, we had initially listed searchability as a separate criterion, but decided to lump it in with Access for simplicity.

- **Provenance**
  How can you trust your data? What is its lineage? Provenance as a criterion is the completeness of a data lineage. Data can be open and either of very poor quality or poor provenance. For example, consider malicious edits to Wikipedia articles that introduce false information. It's open, but being able to track the provenance and trustworthiness of the data is key to actually using it openly.
  The key to strong data provenance is to ensure you can trace the data back to its source

and see any manipulations that occurred along the way. Importantly, AI can generate trustworthy-looking data, but quality typically falls apart when you try to trace back to the original data source.

- ***Add your own***
  The previous criteria were identified as being most common and impactful to the usefulness of *any* data. However, data's usefulness ultimately depends on the guiding applications or desired outcome. For example, for medical applications, data compliance may be a primary criterion. Therefore, data providers should convene with experts in their field to help determine which, if any, additional criteria are most important for promoting useful and open data.

During the definition process, the following additional criteria were considered but ultimately removed or lumped into the criteria previously stated:

- Interoperability
- Intent
- Data domain coverage and granularity
- Data structure
- Integration points (APIs)
- Documentation and user support
- Compliance
- Schema completeness

Visualizing the Spectrum

To quantify the relative usefulness grade of open data or an open dataset, we create a scorecard with five primary criteria and assign each a score of 1-4 stars. Users can customize it by adding data-specific criteria tailored to their needs. The figures below show how we defined the star scores (Figure 1) and an example scorecard for FEMA's open data (Figure 2). We fully recognize that these evaluations were subjective and that collaboration with a larger group should necessarily evolve and harden the criteria spectrum.

Criteria	⭐	⭐⭐	⭐⭐⭐	⭐⭐⭐⭐
License	Custom license (need a lawyer)	Open license with burdensome terms (copyleft, share alike) (ODBL)	Commonly used open license with reasonable requirements (such as attribution) (CC4.0 versions)	Public Open Domain, no restrictions. (CC0)
Cost	Pricing is too high to justify	Nominal costs	Free with other indirect costs	Free
Burden (includes Interoperability)	5+ hours	2-5 hours	1-2 hours	Less than an hour
Accessibility	Custom format, clunky website, no APIs	Custom format, but the website or platform is functional	Standard open format, decent website with APIs	Standard open format, performant and easy website, APIs available.
Provenance	Unknown source and methods	No documentation, but it has metadata	Some documentation and metadata	Fully documented with metadata

Figure 1. Definitions of the criteria's four-star scores.

Criteria	Score	Additional Details
License	⭐⭐⭐	Public domain with several terms and conditions: https://www.fema.gov/about/openfema/terms-conditions
Cost	⭐⭐⭐⭐	No cost
Burden	⭐	Lots of 404s, dead ends, and it's very hard to search for data.
Accessibility	⭐	No APIs for NFHL (some exist for other FEMA info, but very fragile)
Provenance	⭐⭐⭐	Some documentation was available, and metadata exists.
Total	12/20	The dataset scores 60% useful based on our criteria.

Figure 2. An example of a scorecard for FEMA open data derived by a group participant.

(2) We shouldn’t treat open data just like it is open software

Although open data and open source software (OSS) share a common rhetoric of openness, they operate with distinct challenges, legal frameworks, and sustainability models. This distinction is important. OSS benefits from mature licensing infrastructure, clear dependency tracking, and established community governance patterns; open data exists in a complex matrix where the openness of raw data sources, processing tools, and resulting datasets can vary independently. This creates complex scenarios, such as when OSS processes proprietary data and outputs openly licensed data.

Geospatial data exemplifies these unique challenges. Privacy risks extend to spatial relationships that can reveal sensitive patterns even in anonymized data. Stakeholder interests (especially when it comes to intelligence and defense) reflect legitimate concerns about real-world harm rather than just intellectual property protection. Corporate actors also play an outsized role in shaping data governance through private contractual arrangements and business models, contrasting sharply with OSS's community-driven ecosystem.

Recognizing these fundamental differences is essential for building governance models, incentive structures, and licensing frameworks that enable sustainable, equitable, and trustworthy open data ecosystems. Below, we detail a few important cases of distinction and suggestions for moving towards a new future for open data.

Licensing Frameworks and Creation Pipelines

OSS is supported by a well-established licensing infrastructure. The Open Source Initiative (OSI) maintains a list of approved licenses (e.g., MIT, GPL, Apache) that have clear legal precedents. Compliance can be automated using tools such as FOSSology or Snyk.

Open data licensing is comparatively fragmented. While several frameworks exist, including Creative Commons (CC0, CC-BY, CC-BY-SA), Open Data Commons (ODbL, PDDL), and the Community Data License Agreements (CDLA), no single body plays a role comparable to the Open Source Initiative for OSS. Tooling for license verification is limited, complicating compliance and reuse.

Open data and OSS exist in a complex, intertwined ecosystem. Data sources can be private or public. Similarly, the software used to transform this data can be proprietary (ArcGIS) or open source (GDAL or scikit-learn). However, unlike OSS, where dependencies are traceable in source code, data provenance often spans multiple licensing domains with unclear inheritance rules. This "licensing arbitrage" requires sustained relationships between stakeholders with different openness philosophies and creates ongoing compliance challenges not present in traditional software development. This fundamental difference in traceability makes data ecosystems inherently more difficult to govern than software ecosystems.

Data Volatility

Like most software, OSS typically follows a release cycle that starts with a plan or specification and ends with automated testing, manual review by multiple engineers, and a documented versioned release. Alternatively, open data represents unplanned real-world messiness and volatility. This gives rise to unique challenges:

- **Restricting Data from Being Open**
  Unlike software, data can represent sensitive real-world assets, cultural heritage, or economic resources, where unrestricted openness could cause measurable harm to the originating communities. Stakeholders may have strong incentives to restrict data openness for economic, cultural, legal, or competitive reasons.
  For example, farmers may withhold agricultural data to protect trade secrets related to crop yields, soil conditions, or pest management strategies that confer competitive advantages. Archaeologists often restrict site location data to prevent looting and unauthorized excavation. Indigenous communities assert data sovereignty through frameworks such as the CARE Principles, which require collective benefit and community control over data about their lands and cultural practices. Researchers may restrict access to datasets during publication embargoes or to maintain first-mover advantages in grant funding, creating tension between open science ideals and career incentives.
- **Privacy and Ethics**
  Location data can reveal sensitive patterns such as individuals' movements, home addresses, and behavioral routines, even when anonymized. For example, environmental sensor networks that collect data on air quality, water levels, or wildlife movement often inadvertently capture human activity patterns in the vicinity. Another example is the 2018 Strava heatmap incident, which inadvertently revealed the locations and patrol patterns of military personnel at secret bases through aggregated fitness tracking data, demonstrating how seemingly anonymized geospatial data can expose sensitive information when analyzed at scale.
  Geospatial data inherently contains spatial relationships that can be cross-referenced

with other datasets to re-identify supposedly anonymous subjects or reveal commercially sensitive information about land use, resource extraction, or infrastructure vulnerabilities.

- **Temporal Value**
  Unlike software, where newer versions typically supersede older ones, geospatial data often exhibits compound value, in which historical observations provide longer (and richer) temporal baselines for understanding change over time. As a result, data retention decisions are fundamentally different from those in software lifecycle management. The relationship between data recency and value varies dramatically across geospatial domains, creating complex temporal economics unlike software versioning.
  For example, real-time traffic data loses operational value within minutes but retains analytical value for transportation planning over the years. Weather data follows an inverse pattern—current observations are essential for immediate forecasting, while historical climate records become increasingly valuable for long-term modeling as they age. Emergency response data (earthquake sensors, flood monitoring) has an immediate life-safety value that decays rapidly, yet the same data gains scientific value for hazard modeling over time.
- **Provenance and Authority**
  Determining provenance and authoritativeness is, and will likely remain, an ongoing governance challenge for geospatial data. Unlike software version control, the same (or similar) geospatial datasets often exist simultaneously across multiple organizations with different update cycles, coordinate systems, accuracy standards, etc. This challenge compounds when tracking the lineage of data merged from crowdsourced contributions with government or commercial data.
  For example, cadastral data may conflict between OpenStreetMap's and the government's data. Similarly, commercial satellite providers offer different interpretations of land cover classification for the same geographic areas. Sensor networks operated by universities, government agencies, and private companies may provide contradictory readings for environmental conditions, raising questions about which source should be considered definitive.

Corporate and Legal Dynamics

In software, licensing disputes and corporate practices have established norms over the course of decades. By contrast, for data, legal frameworks are still developing across multiple jurisdictions with conflicting approaches. Below are some specific notable differences:

- **Fragmented Legal Frameworks**
  Unlike software's relatively mature copyright system, data protection operates under a patchwork of overlapping rights. For example, the European Union's sui generis database rights[1] provide 15-year protection for substantial investment in data compilation, while the United States explicitly rejects such protection under the Feist doctrine[2]. This jurisdictional divide creates regulatory arbitrage, allowing companies to forum-shop for favorable legal treatment, whereas in software, international copyright treaties provide relative harmonization.

- **Corporate Business Models**
  Data-driven organizations struggle with revenue generation in ways OSS companies have largely resolved. While software companies can leverage open-core models, support services, or dual licensing, data companies face unique challenges where the "product" may lose value through sharing, cannot be easily versioned, and often requires ongoing collection costs. Corporate actors have developed novel frameworks, such as "Contributors Clubs" (exemplified by the Overture Maps Foundation), that provide exclusive insights in exchange for data contributions and tiered licensing based on organizational size or commercial use.

- **Limited Defensive Mechanisms**
  As an example, the ongoing lawsuit by the New York Times against Microsoft and OpenAI highlights nuanced questions about AI training data, derivative works, and fair use[3]. Given the potential copyright infringement damages of $150,000 per willful violation, the usage of millions of potentially copied works could be a serious threat to many companies.
  While copyleft[4] data licenses exist in principle, they provide weaker protection than software copyleft due to definitional ambiguities around "derivative works" and limited enforcement mechanisms. Corporate actors, therefore, still play an outsized role in

---

[1]European Union, "Database Protection," Your Europe, accessed August 2025, https://europa.eu/youreurope/business/running-business/intellectual-property/database-protection/index_en.htm
[2]*Feist Publications, Inc. v. Rural Telephone Service Company, Inc.*, 499 U.S. 340 (1991). For an accessible overview of the case and its impact on copyright law, see the Wikipedia entry: https://en.wikipedia.org/wiki/Feist_Publications,_Inc._v._Rural_Telephone_Service_Co.
[3]*The New York Times Co. v. Microsoft Corp.*, No. 1:23-cv-11195 (S.D.N.Y. filed Dec. 27, 2023). The plaintiffs include The New York Times Company; the defendants include Microsoft Corporation and multiple OpenAI entities.
[4]Luis Villa, "Copyleft, attribution, and data: other considerations," lu.is (blog), September 21, 2016, https://lu.is/2016/09/copyleft-attribution-and-data-other-considerations/.

shaping licensing precedents and sustainability models, often preferring permissive licenses or private contractual arrangements over community-driven copyleft standards.

There are some copyleft-style licenses for data, meaning that derivatives retain the same rights and freedoms, provided they are under the same copyleft license, such as the Open Database License (ODbL). However, they face adoption and enforcement challenges that differ significantly from software copyleft. For example, ODbL requires *derivative databases* to be shared under the same license, but not applications that use the data. This creates loopholes that allow corporations to extract value from ODbL-licensed datasets (such as OpenStreetMap) in *their applications* without triggering reciprocal obligations, provided they don't redistribute modified versions of the database itself. Additionally, ODbL enforcement relies heavily on community vigilance rather than automated compliance tools available for software licenses.

Looking to the Future

Applying OSS frameworks directly to open data is insufficient. Sustainable open data ecosystems must incorporate the following:

- **Governance structures that address provenance, authority, and ethics**
  Unlike software repositories with clear maintainer hierarchies, data governance must establish authoritative sources across distributed collection networks, implement ethical review processes for sensitive datasets, and maintain detailed provenance tracking that spans organizational boundaries. This requires multi-stakeholder governance models that balance the interests of data collectors, processors, and users while addressing concerns from affected communities. Frameworks such as the CARE Principles for Indigenous Data Governance demonstrate that ethical considerations must be embedded in governance structures from inception, not retrofitted as compliance requirements.
- **Incentive models for proprietary data holders**
  Traditional OSS relies on intrinsic developer motivation and career benefits, but data contributors often face direct costs (collection infrastructure, processing resources, legal compliance) without equivalent professional rewards. Successful models must provide tangible value exchanges, such as exclusive access to aggregated insights, preferred API access, co-marketing opportunities, or revenue-sharing arrangements. The challenge lies in creating sustainable economic incentives that don't undermine the openness goals, requiring careful balance between contributor benefits and public

access. The Overture Maps Foundation is trying this through their "Contributors Club,"[5] which provides access to exclusive insights in exchange for contributions.

- **Licensing frameworks tailored specifically to the realities of data, rather than inherited wholesale from OSS**
  Data licenses must address temporal considerations (data currency and refresh rights), attribution complexities (multi-source aggregation), jurisdiction-specific privacy requirements, and definitions of derivative work that account for algorithmic processing. While Creative Commons and Open Data Commons provide starting points, emerging challenges like AI training rights, synthetic data generation, and cross-border data transfers require new licensing approaches that software licenses never anticipated. These frameworks must be both legally robust and practically implementable across diverse technical and organizational contexts.

[5] Marc Prioleau, "The Unique Challenges of Open Data Projects: Lessons from Overture Maps Foundation," Linux Foundation Blog, January 13, 2025, https://linuxfoundation.org/blog/the-unique-challenges-of-open-data-projects-lessons-from-overture-maps-foundation.

(3) The power of open data comes from a strong ecosystem, not its license

Open Data Ecosystem Obstacles

Like a biological ecosystem, where organisms interact to drive energy and nutrient cycles, ideally, a data ecosystem, composed of data, software, people, and systems, symbiotically interacts to drive information and knowledge flow. Therefore, our approach to organizing geospatial data should move beyond open data to a more robust, useful ecosystem that ultimately helps people achieve their goals more easily.

Unfortunately, open data often appears as loosely organized collections of publicly available datasets, grouped into heterogeneous catalogs with varying quality levels. Many industries (including the geospatial industry) struggle to keep their ecosystem running efficiently and effectively, often due to their size and complexity.

So what is needed? In this work, we introduce several problem areas that prevent data ecosystems from working effectively and efficiently, what we can improve, and the benefits of doing so. During our in-person session at CNG Forum 2025, our conversations often returned to identifying persistent problems and new problems that have emerged from the rapid integration between geospatial data and AI.

The Governance Challenge

To begin, there is a proliferation of data formats and data structures with inconsistent adherence to a patchwork of loosely enforced open data principles. It's too easy for a commercial entity to call geospatial data "open" simply by granting it a flexible re-use license or publishing it publicly for free. There's no requirement to maintain source or derivative datasets.

Continued refinement of open data criteria (see the previous section) and standards, updating hosting and transfer protocols, and assigning ownership of released geospatial open data products will be essential for an ecosystem to run itself. This means a heavier hand from standards organizations. This would be beneficial because it encourages public and private development of tools that adhere to best practices and web standards, perhaps incentivising with a new form of accreditation.

The Interoperability Challenge

Integrating disparate datasets that refer to the same physical features but are captured in different formats, at varying granularities, or by various authorities is a persistent challenge. This often results in time-consuming and costly processes known as “data conflation,” where similar features must be matched across datasets using spatial proximity, textual similarity, or domain-specific rules. Compounding the issue is the fact that as more organizations (i.e., government agencies, private logistics firms) publish or consume location-based datasets, the volume and diversity of spatial information have grown dramatically, introducing a "conflation tax." The ecosystem would benefit from streamlined interoperability protocols because people would spend less time pre-processing and post-processing data. One approach is to use globally unique and persistent identifiers for geospatial entities such as roads, buildings, and places. This concept is not new, but successes to date have been limited. A recent example of this approach is the Global Entity Reference System (GERS)[6], introduced by the Overture Maps Foundation. Conceptually, entity-centric spatial referencing is applicable across platforms. A shared ID system maintained through open, collaborative governance can significantly reduce the cost and complexity of using geospatial data.

The AI Integration Challenge

At the time of this writing, the immense attention and rapidly evolving technologies in the (generative) AI industry feel as though existing technological problems are disappearing and new ones are appearing faster than we can comprehend the change. For example, the Model Context Protocol (MCP) is benefiting the open data ecosystem by enabling AI tools to quickly interpret and interact with other tools in a structured way, such as querying geospatial databases without requiring building bespoke ETL pipelines. However, this has shed light on the fact that quickly connecting AI tools is risky when they return ungrounded, incomplete, or nonfactual results. So, we see two primary ways in which generative AI is leveraging geospatial data but needs improvement to make the open data ecosystem more robust: 1. grounding to prevent hallucinations, and 2. embeddings to facilitate geospatial reasoning.

Grounding LLMs with Geospatial Data

Grounding in this case means ground-truthing generative AI responses based on the facts and values of things in the real world, much like a retrieval-augmented generation (RAG) system

---

[6] Ed Parsons, "The Quest for a Universal Geospatial ID: From the One Ring to the Holy Grail?," edparsons.com (blog), July 7, 2025, https://www.edparsons.com/2025/07/the-quest-for-a-universal-geospatial-id-from-the-one-ring-to-the-holy-grail/. See also Overture Maps Foundation, "Global Entity Reference System (GERS)," documentation, https://docs.overturemaps.org/gers/.

would, rather than on stochastically predicted values. In other words, grounding concretizes conversational intent using the structured reality that software needs to do something useful.

Unfortunately, at this time, the grounding process is immature. Many groups in industry and academia are working to improve models' understanding of geography, as evidenced by working groups and workshops such as the upcoming AGILE 2026 workshop, [Geography According to Foundation Models](). This needs improvement because the accuracy and quality of results heavily depend on the quality of the protocols, semantics, and breadth and quality of the (typically text) data used to ground models, including both foundational models and models built on top of them. Specifically, reconciliation matters because language models alone often lack sufficient semantic and geometric precision. For example, geospatial identifiers and contexts commonly change and/or become overloaded and overlapping, such as businesses that share similar names, relocate, or modify attributes over time. Furthermore, geospatial semantics and reasoning are generally lost on LLMs because they train on text representations of space, not geospatial relationships, geometry, or topology. Improved groundings will improve the ecosystem because people, tools, and systems won't need to worry as much about the robustness of LLMs' results.

Geospatial Embeddings Must Mature

Like in other fields, there is a growing interest in richer multi-dimensional representations of geospatial data. Unfortunately, while the current models accelerate at many tasks, they struggle with geospatial semantics and geometric representations. We see generative AI models excelling at linguistic cognition and approximating spatial descriptions, but they don't natively handle metric spatial reasoning. Instead, foundational Models should provide us with a more comprehensive, grounded, and truthful representation of our planet based on 1. more precise measurements, and 2. human cognitive and linguistic understandings of our world. If achieved, strong models would benefit the ecosystem because the tools and systems would spend less time validating, conflating, and running operations on open data and instead rely on the model's representations. Some examples of foundational geospatial models are attempting to embed natural representations of the planet, including LGND.ai's model, DevelopmentSeed and Clay's v1, Earth Genome's Earth Index, and Google DeepMind's AlphaEarth. However, as mentioned previously, these models still lack sufficient precision for many tasks. On the human-geography side, we should consider that the human brain is complex and uses distinct circuits for linguistic (symbolic, sequential) vs. spatial (continuous, metric) cognition. This makes it difficult for us to evaluate a model's ability to reason spatially, such as topics like topology, cardinality, vague geographies (e.g., "Northern California")

Conclusion

The world we live in has been shaped by open data, but the framework built to support it–binary licenses, inherited software governance models, loosely organized data catalogs–was never designed for what the geospatial data ecosystem has become. The three theses in this paper are not independent critiques. They are the same problem seen from three different angles.

Treating openness as a checkbox ignores whether data is actually usable or useful. Applying open source software frameworks to open data ignores the fundamental differences in how data is created, governed, and valued over time. And publishing datasets without maintaining the relationships between them–the governance, the identifiers, the accountability–does not produce an effective ecosystem. Each failure compounds the others.

The path forward requires the geospatial community to do three things it has largely avoided: evaluate data on usefulness, not just openness; build licensing and governance frameworks tailored to data's specific realities; and treat the ecosystem, not the dataset, as the thing worth maintaining. None of this is technically hard. It requires sustained commitment from standards organizations, data providers, and the community that depends on this infrastructure.

The goal was never openness for its own sake. It was to make data useful. That goal is still worth pursuing, with better tools, better frameworks, and clearer eyes about what open data actually demands.