



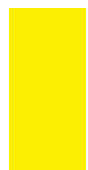
# SOVEREIGN DATA COMMONS AND PUBLIC DATA INFRASTRUCTURE

*Policy Building Blocks for the Digital Commons #3*

**OPEN FUTURE  
POLICY BRIEF**

Authors: Paul Keller & Aditya Singh

**2 APRIL 2026**



## ABOUT THIS POLICY BRIEF

NGI Commons (Open Source and Internet Commons for Europe’s Digital Sovereignty) project is funded by the European Union’s Horizon Europe research and innovation programme under Grant Agreement number 101135279. This work has received funding from the Swiss State Secretariat for Education, Research, and Innovation (SERI).

Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

The text of this document is based on a report–Deliverable 3.4–submitted to the European Commission as a deliverable of the NGI Commons project. It is part of a [series of policy briefs](#) to inform advocacy on Digital Commons and digital sovereignty in the context of the upcoming Multiannual Financial Framework (2028–2034). All NGI Commons deliverables are available on the [project website](#).

[Open Future](#) is a European think tank that develops new approaches to an open internet that maximize societal benefits of shared data, knowledge and culture. Open Future advocates for Digital Commons–characterized by distributed production, collective governance, and shared stewardship–as offering the most viable path towards a resilient digital ecosystem.

[Paul Keller](#) is a co-founder and director of policy at Open Future. His work focuses on the intersection of copyright policy and emerging technologies. He works on policies and systems that improve access to knowledge and culture and protect the digital public sphere.

[Aditya Singh](#) is a Senior Policy Analyst at Open Future with expertise in digital rights, data governance, and technology ethics. His doctoral research on agricultural data governance informs his interests in commons, food systems, and knowledge infrastructures.



This policy brief is published under the terms of the [Creative Commons Attribution License](#).

# INTRODUCTION

Digital sovereignty, understood as the ability of states, organizations, and individuals to independently shape and control their digital environments, has become a central concern for the European Union. Achieving it requires more than regulatory oversight of dominant technology providers: it requires building credible alternatives. This includes building alternatives to the dominant paradigm relating to data and AI.

Cultural heritage institutions, research organizations, public administrations steward the data layer that can power alternatives that foreground public interest. However, they remain largely dependent on commercial infrastructure, undermining their ability to govern a collectively produced resource in line with public interest and sovereignty.

The EU's Data Union Strategy sets ambitious goals for leveraging public data assets for AI development. This brief proposes the infrastructure layer necessary to realize these ambitions: dedicated public data storage, governed as a commons. The proposed infrastructure would provide long-term storage and access for public interest data—cultural heritage, research outputs, linguistic resources, administrative data sets—under a governance model that ensures public interest orientation, prevents commercial capture, and enables collective decision-making by contributing institutions.

## PROBLEM DEFINITION: THE MISSING INFRASTRUCTURE LAYER

The European Commission's Data Union Strategy, published in November 2025, sets ambitious goals for leveraging Europe's public data assets—scientific, cultural, and linguistic resources—for AI development.<sup>1</sup> The strategy positions common European data spaces<sup>2</sup> as a central infrastructure for data sharing and introduces Data Labs to connect these spaces with the AI ecosystem.

However, the strategy's implementation depends on an assumption that is not borne out by the current state of European public data infrastructure: that the underlying storage and long-term stewardship capacity for public data already exists or will emerge organically. Without dedicated public infrastructure for storing and governing public data assets, Europe risks undermining its capacity to act as a sovereign actor in the data economy—dependent on commercial providers for access to its own publicly-funded digital resources.

---

<sup>1</sup> European Commission (2025). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions – A Data Union Strategy: Unlocking Data for AI and Innovation. COM(2025) 835 final. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52025DC0835>.

<sup>2</sup> European Commission (2025). European Data Spaces. Directorate-General for Communications Networks, Content and Technology (DG CONNECT). Available at: <https://digital-strategy.ec.europa.eu/en/policies/data-spaces> (accessed 18 February 2026).

In practice, decades of project-based funding for digitization and data initiatives have created a fragmented landscape. Many public data sets are generated through time-limited projects, but when funding ends, the infrastructure to maintain, update, and provide access to these data sets often disappears. The result is a pattern familiar across European digital policy: substantial investment in creating digital assets, followed by their gradual degradation, loss of accessibility, or capture by private actors better positioned to provide long-term hosting.

The Data Labs introduced by the Data Union Strategy address a different problem. They focus on processing, curation, pseudonymization, and connecting data spaces with AI developers. As currently conceived, they do not provide the durable storage infrastructure that would ensure public data remains accessible, governed, and usable over time. The Data Spaces themselves define governance frameworks and interoperability standards, but they similarly assume that participating institutions have the capacity to maintain and provide access to their data holdings indefinitely.

And while there has been considerable investment in middleware platforms intended to enable interoperability among data spaces, there has been no corresponding investment in the more foundational layer: shared infrastructure for storing and stewarding public data assets over time. Middleware can facilitate data exchange, but it cannot solve the underlying problem of where public data will be stored, who will steward it, and how its long-term accessibility will be guaranteed.

This infrastructure gap is particularly acute for public data held by cultural heritage institutions. Many such institutions participated in digitization programmes funded by Framework Programmes or national initiatives, only to find that ongoing storage and access costs exceed their operational budgets. Some have entered into partnerships with commercial providers that offer storage in exchange for preferential or exclusive access. While such arrangements may address immediate capacity constraints, they risk transferring control over publicly-funded resources to private actors whose interests may not align with public access goals—effectively enabling a form of re-enclosure of publicly-funded digital assets.

The European Open Science Cloud (EOSC) demonstrates that sustainable public data infrastructure is achievable when properly funded and governed.<sup>3</sup> EOSC employs a tripartite governance structure that balances European Commission coordination with Member State participation and bottom-up involvement of the research community—a model that prioritizes inclusivity and ensures the infrastructure remains driven by those who use it. Findings from the NGI Commons project<sup>4</sup> indicate that EOSC represents a significant attempt to build shared infrastructure embedded in the values of open science, though it also illustrates the sustainability challenges that arise when infrastructure development remains tied to project-

---

<sup>3</sup> European Commission (2025). European Open Science Cloud (EOSC). Directorate-General for Research and Innovation. Available at: [https://research-and-innovation.ec.europa.eu/strategy/strategy-research-and-innovation/our-digital-future/open-science/european-open-science-cloud-eosc\\_en](https://research-and-innovation.ec.europa.eu/strategy/strategy-research-and-innovation/our-digital-future/open-science/european-open-science-cloud-eosc_en).

<sup>4</sup> Open Future (2024). Digital Commons as Providers of Public Digital Infrastructures. Available at: [https://openfuture.eu/wp-content/uploads/2024/11/241113\\_Digital-Commons-as-Providers-of-Public-Digital-Infrastructures.pdf](https://openfuture.eu/wp-content/uploads/2024/11/241113_Digital-Commons-as-Providers-of-Public-Digital-Infrastructures.pdf).

based funding. EOSC's mandate, however, is limited to research data and the research community. No equivalent infrastructure exists for the broader categories of public data that the Data Union Strategy identifies as strategic assets: digitized cultural heritage, administrative data, linguistic resources, and public sector information more broadly.

Without such infrastructure, the Data Union Strategy's ambitions risk remaining unrealized. Data spaces will lack the foundational layer they need to function effectively. Valuable public data sets created with public funding will remain siloed within institutions lacking capacity for sustainable provision, or will migrate to commercial platforms where access terms may not serve public interests. This latter outcome amounts to a re-enclosure of publicly-funded digital assets. The substantial public investments in digitization over the past two decades will continue to depreciate rather than compound.

These dynamics are reinforced by the Data Union Strategy's emphasis on transitioning mature data spaces to market-driven models. While this approach may be appropriate for commercial data flows, it is not well-suited to public data whose value lies precisely in serving purposes that markets will not adequately support. Public data requires sustained public infrastructure governed as a commons—not as a subsidy to commercial activity, but as the foundation for public interest uses that generate social rather than private returns.

## **PROPOSED INTERVENTION: A PUBLIC DATA STORAGE LAYER WITH COMMONS-BASED GOVERNANCE**

This building block proposes the establishment of a dedicated public data storage infrastructure, governed according to commons principles, to serve as the foundational layer for common European data spaces and related initiatives.

Commons-based governance represents an alternative to both pure market provision and traditional state administration. Rather than treating data as a commodity to be traded or as an administrative asset to be managed, the commons approach recognizes data as an infrastructural resource—a shared means for many ends that can support the creation of both public and private value. The design of data commons rests on three pillars: stewarding access through clear rules for managing who can use data and under what conditions; collective governance through participation and democratic oversight by a defined community of stakeholders; and public value orientation through mission-driven approaches that ensure benefits flow to society rather than being captured by narrow interests. A data commons developed based on these principles also offers a pathway to developing public AI infrastructures and solutions.<sup>5</sup> These principles provide a framework for preventing appropriation while enabling productive use—balancing openness with the protections needed to maintain the resource over time.

---

<sup>5</sup> Tarkowski, A., et al. (2025). White Paper on Public AI: A Public Alternative to Private AI Dominance. Open Future. Available at: <https://openfuture.eu/publication/white-paper-on-public-ai/>.

The governance of the proposed infrastructure would require an institutional vehicle capable of coordinating Member States, public institutions, and stakeholder communities. Emerging instruments such as the Digital Commons EDIC could provide such a vehicle.

This infrastructure would provide long-term stewardship for public interest data, ensuring that investments in digitization and data generation translate into durable, accessible resources rather than temporary project outputs. Like physical infrastructure, data infrastructure requires ongoing maintenance and care: curation, quality control, format migration, and adaptation to evolving needs. Commons-based governance institutionalizes this stewardship as a core function rather than an afterthought, embedding it within structures that have both the mandate and the capacity to maintain resources over time.

## *Core Functions*

The proposed infrastructure would fulfil three interconnected functions:

- **Long-term storage and preservation.** Providing secure, redundant storage for public data assets, with clear preservation commitments and capacity to maintain accessibility as formats and technologies evolve. This addresses the sustainability gap that has undermined previous digitization investments.
- **Access provision.** Offering standardized interfaces (APIs, bulk download, streaming) that enable data spaces, data labs, researchers, and other legitimate users to access stored data. Access mechanisms should support both individual item retrieval and aggregate dataset provision, recognizing the shift from individual access to bulk access that characterizes AI-era data use.
- **Governance and stewardship.** Implementing commons-based governance that ensures public interest orientation, prevents capture by any single actor, and enables collective decision-making by data contributors about access terms and use conditions. This governance layer is what distinguishes public infrastructure from commercial cloud storage.

The infrastructure should adopt a federated architecture, with storage capacity distributed across participating institutions and Member States rather than centralized in a single facility. This approach aligns with EOSC's model, enables integration with existing national infrastructures, and avoids creating a single point of failure or control.

Importantly, commons-based governance does not require that all access be fully open. As we have argued elsewhere,<sup>6</sup> openness should be understood as lying on a spectrum of possible governance choices, with different access conditions appropriate for different types of use. A commons approach can enable differentiated access—for example, maintaining open access for individual items and research use while applying conditional terms for bulk commercial access, particularly for AI training at scale. This reflects the principle that public content can remain

---

<sup>6</sup> Keller, P. (2025). Impulse Paper: Publishing Cultural Heritage Data in the Age of AI. Open Future. Available at: <https://openfuture.eu/publication/impulse-paper-publishing-cultural-heritage-data-in-the-age-of-ai/>.

freely available while the infrastructure that delivers it at scale may legitimately require contributions back to the commons. The proposed derogation from the Open Data Directive's non-discrimination principle for very large enterprises, introduced in the Digital Omnibus package, signals growing recognition of this approach at the policy level.

## *Relationship to Existing Infrastructure*

The proposed infrastructure should build on and connect with existing initiatives rather than duplicating them:

- **EOSC as a reference point:** The European Open Science Cloud offers relevant experience in developing federated public data infrastructure with distributed governance. Its tripartite governance model and stakeholder-driven approach provide useful lessons, though EOSC's experience also highlights the importance of moving beyond project-based funding. The proposed infrastructure should learn from both EOSC's achievements and its difficulties, while serving a broader mandate encompassing public data beyond research.
- **Data spaces as users:** Common European data spaces would connect to this infrastructure as a storage and access layer. The infrastructure would not replace data space governance but would provide the technical foundation that enables data spaces to function sustainably.
- **Data Labs as processors:** Data labs would draw on stored data for curation, processing, and preparation for AI training. The infrastructure provides the raw material; data labs add value through processing and quality enhancement.
- **Sectoral initiatives as an intermediate layer:** We would expect sector-specific data commons initiatives to build on top of this storage infrastructure, providing domain expertise, curation, and community governance for particular types of public data. A European Books Data Commons,<sup>7</sup> currently being explored by Open Future and a number of cultural heritage institutions, illustrates this model: a sectoral initiative that would operate within the common European data space for cultural heritage while potentially serving users across domains. Such initiatives mediate between the general-purpose storage layer and the specific needs of particular communities and use cases.

## **COMMONS-BASED GOVERNANCE MODEL**

As outlined above, commons-based governance provides a framework for managing shared resources that prevents capture while enabling productive use. For data infrastructure specifically, this approach addresses a fundamental challenge: public data represents a collectively produced resource whose governance must balance the interests of both those who contribute data and those who benefit from its use. Neither pure market provision nor traditional bureaucratic administration adequately serves this dual constituency—the former

---

<sup>7</sup> Keller, P. (2025). Outline for a European Books Data Commons. Open Future. Available at: <https://openfuture.eu/publication/outline-for-a-european-books-data-commons/>.

risks commercial capture, the latter may fail to represent the diverse communities whose data is being managed.

This governance model would have three dimensions:

- **Contributor governance:** Institutions that contribute data to the infrastructure should have governance rights proportional to their participation. This ensures that stewardship decisions reflect the interests of those whose data is being managed.
- **Access governance:** Access terms should be determined collectively, with the capacity to distinguish between different user categories and use types. Non-commercial, public interest uses (research, education, journalism, civic purposes) should receive preferential access. Commercial uses may be subject to conditions or contributions that support infrastructure sustainability.
- **Public interest orientation:** Governance structures should include mechanisms to ensure that the infrastructure serves broad public interests, not only the interests of contributing institutions. This might include civil society representation, transparency requirements, and periodic public interest assessments.

The Digital Commons EDIC could provide an institutional vehicle for governing this infrastructure, connecting it to other commons-based digital initiatives and ensuring durable governance backed by sustained public investment. As a legal instrument designed to enable Member State cooperation on strategic digital infrastructure, the EDIC framework is particularly well-suited to stewarding sovereign data infrastructure that must operate across jurisdictions while remaining anchored in public-interest objectives.

## **FIT WITH THE MULTIANNUAL FINANCIAL FRAMEWORK**

The proposed storage layer should be understood as enabling infrastructure—foundational capacity that supports a wide range of downstream activities and initiatives. Without shared public infrastructure for storing and stewarding public data, the EU's ambitions to leverage data as a driver for AI development will remain constrained. Data spaces cannot function sustainably, AI factories lack access to the curated public datasets they require, and valuable data resources created through digitization and research projects degrade or become inaccessible because no institution has the mandate or capacity to maintain them. This infrastructural character has implications for how such investment should be positioned within the MFF: not as a time-limited project but as a durable public asset requiring sustained commitment.

The absence of such infrastructure also represents a pattern of wasteful public investment: digitization projects, data space pilots, and AI initiatives repeatedly create valuable data resources that subsequently cannot be sustained. A dedicated storage layer would ensure that outcomes from publicly funded research, innovation, and digitization projects realise their full potential by remaining accessible and usable over time.

The **Digital Leadership** window of the **European Competitiveness Fund**, which consolidates the Digital Europe Programme and related instruments, provides the natural funding channel for data infrastructure. The proposed storage layer would complement ECF investments in data spaces and AI factories by providing the foundational capacity these initiatives require. The proposed EU Preference mechanism could ensure that infrastructure development strengthens European providers and avoids creating new dependencies.

Research and innovation funding through **Horizon Europe** would be strengthened by the availability of an underlying storage infrastructure. Such infrastructure can ensure the sustainability of outcomes from Horizon Europe-funded research and development projects—data resources, tools, and methods that currently risk becoming inaccessible once project funding ends. The connection to EOSC creates natural synergies with Horizon Europe's Open Science priorities.

Member States' **Partnership Plans** could include commitments to connect national public data holdings to European infrastructure, ensuring that the federated architecture achieves broad coverage across the Union.

The critical requirement is that funding be structured for long-term sustainability rather than time-limited projects. Infrastructure that exists only for the duration of a funding cycle cannot provide the durability that distinguishes public infrastructure from temporary initiatives. This requires dedicated budget lines that recognize data storage as an ongoing infrastructure investment rather than a discrete project expenditure. The alternative, continued reliance on project-based approaches, perpetuates the cycle in which public investment repeatedly creates data resources without the means to sustain them.

## LEGAL AND POLICY CONTEXT

The proposed intervention aligns with and reinforces several ongoing policy initiatives, offering a practical mechanism to operationalize commitments that currently lack implementation pathways. The intervention builds on the logic of the Data Union Strategy's emphasis on scaling strategic data assets, extends the **Open Data Directive's**<sup>8</sup> ambitions for public sector data availability, and complements the (forthcoming) **Cloud and AI Development Act's** focus on sovereign infrastructure. For policymakers working across these files, the proposed storage layer represents a missing piece that would enhance the effectiveness of existing instruments and increase the return on ongoing investments in data spaces, AI development, and digitization.

More broadly, the intervention aligns with emerging discussions around public AI infrastructure:<sup>9</sup> the idea that AI development should be supported by publicly governed data, compute, and models that serve public purposes alongside commercial innovation. Embedding public data

---

<sup>8</sup> Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast). Official Journal of the European Union, L 172, 26.6.2019, pp. 56–83. Available at: <https://eur-lex.europa.eu/eli/dir/2019/1024/oj/eng>.

<sup>9</sup> Tarkowski, A., et al. (2025). White Paper on Public AI: A Public Alternative to Private AI Dominance. Open Future. Available at: <https://openfuture.eu/publication/white-paper-on-public-ai/>.

storage within this broader public AI framework would strengthen the coherence and effectiveness of current policy initiatives.

The **Data Union Strategy**<sup>10</sup> explicitly identifies the need to scale up strategic data assets and create trusted environments for data pooling and AI training. It acknowledges that data spaces require federated infrastructure and emphasizes the transition from pilot initiatives to sustainable, scaled operations. The proposed storage layer would operationalize these commitments by providing durable capacity for the public data assets the Strategy identifies as strategic priorities.

The forthcoming **Cloud and AI Development Act** (CAIDA) represents the most significant near-term opportunity to embed public data infrastructure within the EU's digital policy framework. The Act's focus on sovereign cloud capacity and AI development creates a natural opening for provisions that ensure public data assets are stored on European infrastructure under European governance. Positioning the proposed storage layer as a component of CAIDA would anchor it within the broader push for digital sovereignty while ensuring that public data infrastructure develops in coordination with—rather than separately from—commercial cloud and AI capacity.

## **RESOURCE REQUIREMENTS AND FUNDING INFRASTRUCTURE**

Establishing public data storage infrastructure requires sustained investment, but the costs are commensurate and modest relative to the scale of the policy ambitions it would support and the investments already being made in related initiatives.

- **Storage costs:** Cloud storage costs have declined substantially and continue to fall. For a federated architecture using European providers, storage costs for petabyte-scale collections are manageable –likely in the range of tens of millions of euros annually for comprehensive coverage of strategic public datasets. This is modest relative to the investments in creating these data sets.
- **Personnel and governance:** Sustainable infrastructure requires ongoing technical management, user support, and governance coordination. A core team of around 20 FTE could manage European-scale infrastructure, with additional capacity distributed across participating institutions. This estimate draws on the staffing models of comparable initiatives such as EOSC and Europeana, though the precise requirements would depend on the scope and architecture of the implementation.
- **Development costs:** Initial setup of storage architecture, access interfaces, and governance tools would require a one-time investment. Since the core technical components are largely solved problems with mature open source implementations, development costs would be

---

<sup>10</sup> European Commission (2025). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions – A Data Union Strategy: Unlocking Data for AI and Innovation. COM(2025) 835 final. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52025DC0835>.

modest—the primary investment is in integration, federation protocols, and governance design rather than fundamental technical development.

- **Sustainability model:** Long-term sustainability could combine several funding sources: base funding from the ECF for public interest functions; contributions from data spaces using the infrastructure; and fees for commercial access to data sets where appropriate. Crucially, infrastructure that serves AI development objectives could be cross-subsidised through AI-related investments. AI Factories and Data Labs require access to high-quality, curated data sets; rather than funding data preparation separately within each initiative, a portion of AI infrastructure investment could flow to the underlying storage layer that makes such data available. This would recognize public data infrastructure as an integral component of Europe's AI capacity rather than a separate expenditure. The goal should be a mixed model where public funding covers core public interest functions while AI and commercial uses contribute to operational costs without driving governance decisions.

The costs outlined above are modest when set against three comparisons. First, the investments already made in creating public data assets that now lack sustainable infrastructure—digitization programmes, research data, administrative data sets—represent sunk costs whose value depreciates without preservation capacity. Second, the commercial value currently being extracted from public data by actors better positioned to store and provide access to it represents a transfer of public investment to private benefit. Third, the EU is committing substantial resources to AI infrastructure. Investing in public data storage is not an additional cost but a necessary complement—one that ensures AI infrastructure has access to the European data resources it needs.

## **POSSIBLE RISKS AND MITIGATIONS MEASURES**

The primary risk is that the proposed intervention will go unattempted. The absence of shared public data infrastructure carries substantial costs that are largely invisible because they manifest as foregone opportunities rather than direct losses: digitized collections that degrade without stewardship, research outputs that become inaccessible after project funding ends, data spaces that cannot function sustainably, and AI development that proceeds without access to high-quality European public data. These costs compound over time as the gap between Europe's data ambitions and its infrastructural capacity widens. The current pattern—of repeated investment in data creation without corresponding investment in data stewardship—leaves Europe dependent on commercial providers for access to resources created with public funding.

Mitigating this risk entails explicitly recognizing public data storage and stewardship as essential public infrastructure within EU funding frameworks. Long-term investment should be anchored within the Multiannual Financial Framework, and a public data storage layer should be positioned as durable infrastructure rather than a time-limited project outcome.

Implementation does present challenges—securing sustained funding commitments, designing governance that balances efficiency with inclusivity, achieving technical interoperability across a federated architecture, and preventing capture by commercial interests. These challenges are

real but may be mitigated through careful institutional design, building on experience from initiatives like EOSC. The governance and technical challenges of commons-based infrastructure are well-documented and addressable.

## **IMPLEMENTATION PATHWAY**

Building public data infrastructure is inherently a long-term, multi-stage process. Unlike project-based initiatives that can be designed, funded, and completed within a single funding cycle, infrastructure requires sustained commitment across planning, development, and operational phases. The pathway outlined below recognizes this reality while identifying concrete steps that can be taken within current and forthcoming policy windows. It also acknowledges that infrastructure development must proceed incrementally—beginning with priority sectors and use cases before expanding to broader coverage.

In the short term, this means advocating for explicit recognition of public data infrastructure as a funding priority within the ECF Digital Leadership window during the MFF negotiations. It also means working to embed infrastructure objectives and architectural requirements within the Cloud and AI Development Act proposal so that public data storage is recognized as a component of Europe's sovereign AI capacity. Alongside these advocacy efforts, the focus should be on supporting the development of sectoral pilots that demonstrate the approach in practice and generate operational learning, while engaging with EOSC governance to explore whether an expanded mandate or parallel structures could accommodate broader infrastructure ambitions.

Once the new MFF comes into force in 2028, the conditions will exist to move from advocacy and piloting to active development. An infrastructure development programme can be launched, with governance structures established—potentially through the Digital Commons EDIC—and priority data spaces progressively connected to shared storage infrastructure. This phase will also be the moment to develop a sustainability model grounded in real operating experience rather than theoretical projections.

By 2031 and beyond, the ambition is to reach operational scale, with comprehensive coverage of strategic public datasets across priority sectors. At that point, the focus shifts to transitioning towards a sustainable funding model that combines base public funding with contributions from AI development initiatives and cost recovery from commercial access. The longer-term trajectory can then be shaped by evidence—evaluating whether and how to expand coverage to additional data categories based on demonstrated demand and available capacity.

The Data Union Strategy sets ambitious goals for European public data, but ambition without infrastructure is insufficient. This brief proposes a practical intervention—dedicated public storage infrastructure governed as a commons—that would provide the missing foundational layer for data spaces, AI development, and the long-term stewardship of Europe's public data assets. The investment required is modest relative to both the value of the assets at stake and the cost of continued inaction. What is needed now is political commitment to treat public data infrastructure as what it is: essential public infrastructure deserving sustained public investment.