Data Considerations for AI Applications
in Environmental Context

Hamed Alemohammad
Clark University

The advancements and integration of artificial intelligence (AI) present a transformative opportunity to address pressing global environmental challenges, including climate change, biodiversity loss, and pollution. However, realizing AI's full potential in environmental services is contingent upon a robust and ethically sound data ecosystem. This article synthesizes existing literature to explore critical data considerations across four key dimensions: access, governance, quality, and ethics. Analysis reveals that while AI offers unprecedented capabilities for environmental monitoring, prediction, and management, significant challenges persist in data fragmentation, proprietary barriers, quality inconsistencies, and the inherent ethical complexities of AI deployment. Overarching opportunities lie in fostering open data initiatives, developing adaptive governance frameworks, implementing inclusive data curation and preprocessing, and prioritizing human-centered, sustainable AI development. The article concludes with strategic recommendations aimed at cultivating a collaborative, transparent, and equitable data infrastructure essential for harnessing AI effectively for a healthier planet.

Introduction

AI is rapidly emerging as a transformative technology with immense potential to address some of the most pressing global environmental challenges of the 21st century, including climate change, biodiversity loss, and disaster response and mitigation among others. The inherent capacity of AI to process vast amounts of complex data, identify intricate patterns, and optimize multifaceted systems makes it particularly well-suited for these challenges. AI applications are already demonstrating significant value across diverse environmental domains, from enhancing real-time environmental monitoring and optimizing energy systems to advancing sustainable agriculture and informing disaster response (Figure 1).[1]

For instance, AI has shown remarkable effectiveness in identifying potential violators of environmental regulations, with one study indicating more than 600% improvement in detecting water pollution violators through machine learning algorithms compared to random selection.[2]

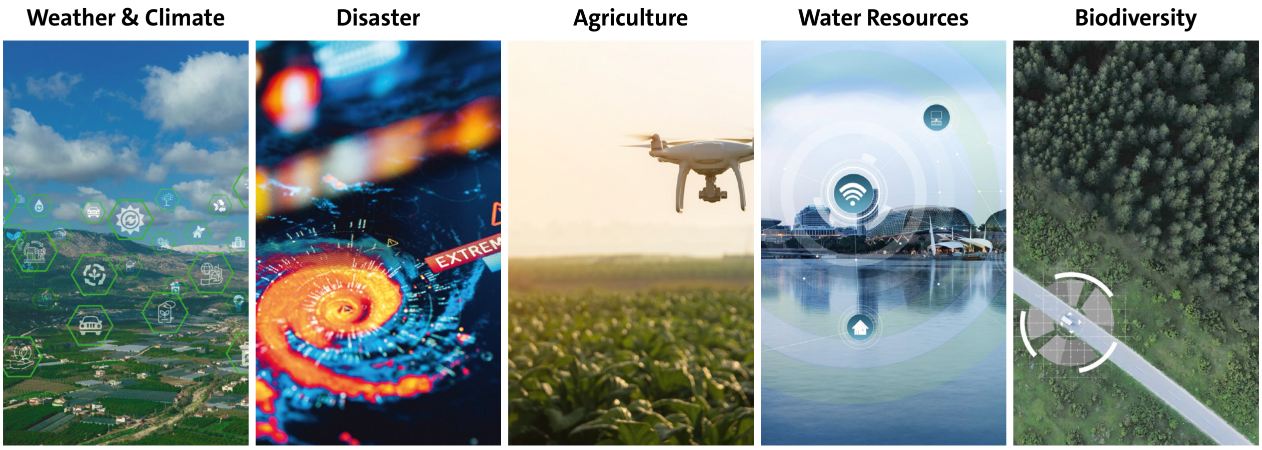
In biodiversity conservation, AI has been effective in improving species tracking and identification.[3] It also aids pollution control and resource management through applications in precision farming, waste management, and air and water quality monitoring, while enhancing climate resilience.[4]

Recent advances demonstrate AI's impact on disaster response and mitigation. For example, Google's AI-driven flood forecasting system now protects over 700 million people across 100+ countries, predicting flood extent up to seven days in advance and saving millions in economic damages annually.[5] Similarly, IoT-enabled wildfire detection networks in Europe have reduced response times from hours to minutes, preventing catastrophic losses.[6] These cases illustrate how robust data ecosystems and AI integration can deliver measurable benefits beyond theoretical promise.

Unlike traditional statistical models – which typically rely on predefined assumptions and smaller, structured datasets – modern AI approaches are dynamic and data-driven, requiring large-scale, diverse, and high-quality inputs to function effectively. This shift amplifies the need for robust data governance because, while statistical models could often operate with limited or curated data, AI systems depend on comprehensive and interoperable datasets. Therefore, ethical frameworks become essential to ensure fairness, reliability, trust, and impact.

This article aims to provide a synthesis of existing research and practical considerations regarding the critical data aspects for applying AI in environmental contexts. It specifically focuses on four foundational areas: data access, governance, quality, and ethics and offers a critical review of the literature to inform policymakers, researchers, and practitioners on how to responsibly and effectively leverage data for AI-driven environmental solutions.

***Figure 1:** AI applications in environmental domains*



Data Access in Environmental AI

AI's performance relies heavily on the breadth and quality of data it learns from. High-quality and accessible environmental data is essential for powering AI tools and models to address global environmental challenges.

Environmental data comes in many forms. *Structured data* is information organized in a clear, tabular format – similar to a spreadsheet – where each row represents an observation (such as a weather station reading) and each column represents a specific attribute (like temperature, date, or location). This predictable structure makes it easy for AI systems to process and analyze. In contrast, *unstructured data* includes items like photos, videos, or free-form text, which do not follow a fixed format and require additional steps for AI to interpret. Increasingly, environmental analysis relies on *multimodal data*, which combines different types – such as images, sensor readings, and text – to provide a richer and more complete picture for AI models.

Creating a collaborative, interoperable data ecosystem is therefore critical to enable AI to support system-level environmental responses, beyond localized efforts.

There is also the notion of AI-ready data which does not have a fully defined and adopted definition across the community. The widely accepted best practice is that AI-ready data is high-quality, well-structured, and governed information that is accurate, accessible, representative of the population, and enriched with context, making it immediately usable for training and deploying AI models at scale.

As environmental problems become more complex and interconnected, comprehensive AI analysis requires integrating diverse datasets, and data silos are no longer sufficient. For AI to be fully effective, inputs from Internet of Things (IoT) sensors, satellites, meteorological stations, and citizen science must be harmonized. In this regard, UNEP's Global Environmental Data Strategy (GEDS) is designed to make environmental data easier to access, share, and use by improving its quality, governance, and interoperability. GEDS helps countries and organizations build the skills and systems needed to collect and manage data effectively – especially through capacity-building programs that support technical training, institutional development, and policy alignment.[7]

So far, GEDS has engaged over 500 stakeholders globally, led to the redesign of UNEP's World Environment Situation Room, and launched 11 prototypes and 9 digital solutions that test its principles in real-world settings.[8] These efforts have helped make environmental data more AI-ready, supported evidence-based policymaking, and fostered international collaboration to tackle climate change, pollution, and biodiversity loss.

Multi-stakeholder collaborative models demonstrate how open, machine-readable data pipelines – combined with secure access controls – can accelerate discovery while protecting intellectual property. Large biopharma programs exemplify this approach: for instance, the multi-cloud data platform Nerve Live, built by Novartis, reorganized petabyte-scale research and operational data to break silos and support AI at scale – an approach directly portable to environmental observatories seeking to unify satellite, in-situ, and administrative datasets.[9]

When data remains isolated, AI's capacity to generate robust insights and predictive models is limited, resulting in fragmented solutions. Creating a collaborative, interoperable data ecosystem is therefore critical to enable AI to support system-level environmental responses, beyond localized efforts.[10]

Challenges to Data Access

Recent literature highlights that the integration of AI in environmental monitoring and decision-making is fundamentally challenged by data silos, heterogeneous standards, and limited interoperability:[1, 11]

- **Fragmentation and Silos:** Environmental data are frequently stored in isolated systems using varying standards and formats, which restricts discoverability, interoperability, and seamless integration. This fragmentation and lack of interoperable data formats and metadata hinder unified AI workflows, complicate the aggregation and comparison of datasets across domains, and limit the ability of AI to generate comprehensive environmental insights and support coordinated action. Empirical and perspective pieces on environmental data identify interoperability and siloing as primary barriers to integrated analysis in the age of AI.[12]

- **Discoverability and Usability:** Many datasets that are presumably available publicly lack proper metadata, license or are not machine-readable, making them difficult to be discovered and integrated into AI systems.

- **Proprietary Concerns and Data Ownership:** Private sector actors may treat data as proprietary, restricting broader use. AI models, however, require large, diverse, and high-quality datasets to function effectively. Restrictions on data access, can stifle innovation, reduce the diversity of training data, and lead to less robust or biased AI models. Data ownership on the hand is deeply tied to questions of trust, equity, and justice. For example, when local communities generate environmental data, but external actors (companies, researchers) use it without ongoing consent or benefit-sharing, it can erode trust and reduce local agency. This dynamic can discourage future participation and undermine the effectiveness of community-driven environmental and climate responses.[13]

The tension between private control over data and its public value is a fundamental challenge. AI's power lies in its access to vast, diverse datasets. But when key data is proprietary or ethically mismanaged, it limits equitable AI applications. This situation demands more than technical fixes – policy changes, data commons, and shared governance models are needed to reframe environmental data as a collective resource that advances both innovation and equity.

Opportunities for Enhanced Data Access

Despite the obstacles, there are clear paths to improve data access for environmental AI:

- **Open Data Initiatives and Platforms:** Open access data is critical for the success of AI in addressing environmental challenges. A successful example of such a solution is the Open Targets Platform in life sciences.[14, 15] Open Targets is a consortium of research organizations, nonprofit foundations, and for-profit pharmaceutical companies. They pool genomics, proteomics, and bioinformatics data in a shared data common, enabling both innovation and public benefit. This model allows private actors to contribute proprietary data while benefiting from shared infrastructure and governance, balancing commercial interests with public good. The governance structure is participatory, with clear rules for data access, use, and benefit-sharing.

© by Municipal Council of Maputo

- **Machine-Readability and AI-Readiness:** Ensuring environmental data and metadata are machine-readable is vital for seamless AI integration. This includes using standardized formats and clear, accessible metadata.[16] Darwin Core is an example of a community effort to provide a standardized, interoperable framework for biodiversity data, allowing machine learning algorithms to process and integrate diverse datasets more effectively.[17]

- **Leveraging Citizen Science:** Citizen science is a growing source of valuable data, enriching understanding of ecological systems. For example, programs like MOPA in Mozambique allow citizens to report waste issues via mobile phones, improving response times from five days to fourteen hours.[18] However, such data must be ethically managed, ensuring consent, representation, and accuracy before being used in AI workflows. This topic will be explored in greater detail in the following sections.

- **Incentivizing Data Sharing:** Policies can encourage data providers – from academia to business – to generate open, AI-ready data.[19] Strategies may include linking funding to digitization efforts or open data sharing. For example, the Digital Europe Programme (DIGITAL) funds public administrations and research organizations to digitize infrastructure and services. Funding is contingent on implementing digital transformation projects, including data sharing and interoperability initiatives.[20]

- **Interoperability Standards:** To measure progress and promote accountability, widely adopted data standards are needed. The FAIR (Findable, Accessible, Interoperable, and Reusable) principles offer a framework to improve data discoverability and reusability, while respecting ethical constraints.[21] The geospatial community's adoption of FAIR principles illustrates the growing demand for well-documented, accessible environmental data.[22]

Data Governance in Environmental AI

Robust data governance is key to ensuring that AI is applied responsibly and effectively in environmental contexts. It includes policies and standards to manage data throughout its lifecycle. Governance models must prioritize accountability, fairness, provenance, transparency, and privacy.

Given the complexity of AI systems and their broad implications, governance must be adaptive – capable of keeping pace with rapid technological change while ensuring responsible innovation. This means balancing progress with equity and risk mitigation. One-size-fits-all policies are ineffective; instead, flexible, evolving structures are needed to support ethical use and maintain public trust. Governance must allow for continuous updates to keep pace with AI's shifting capabilities across environmental sectors.

Challenges to Data Governance

Many challenges hinder effective data governance for environmental AI:

- **Regulatory Fragmentation and Gaps:** Although some global recommendations exist, the policy landscape is still fragmented and lacks harmonized international standards, making consistent governance difficult.[23]

- **Opacity and the "Black Box" Problem:** As AI becomes more complex, it becomes harder to understand or audit its decision-making, raising risks related to compliance, ethics, and accountability. The opacity of complex AI models makes explainability and accountability difficult – yet essential.

- **Bias and Privacy:** AI trained on biased historical data may replicate discriminatory patterns, affecting areas like environmental justice. Simultaneously, the need for large-scale and diverse datasets to train AI models raise concerns about privacy and security.[24]

In environmental contexts where AI might influence regulation, disaster response, or resource use, it is crucial to understand how decisions are made. Lack of transparency can perpetuate bias and worsen inequities, particularly when historical data reflects systemic injustice. A biased AI might unfairly target communities or businesses, reinforcing disparities. Governance must enforce explainable AI (XAI) and establish oversight to trace decisions and audit systems. These measures are vital to build trust and ensure AI works for the public good, not against it.

Data quality, therefore, is not a secondary concern but a foundational pillar for trustworthy environmental AI.

Opportunities for Improved Data Governance

Despite these challenges, several opportunities exist:

- **Multi-Stakeholder Governance:** Engaging stakeholders across sectors – industry, academia, civil society, and affected communities – can foster inclusive and ethical governance.

- **Tailored Regulation:** Countries should adapt AI regulation to their context, considering sector-specific needs rather than applying uniform rules. The proposed EU AI Act for example uses a risk-based approach, setting different requirements for AI systems depending on their application and potential impact.[25] For instance, AI used in healthcare or critical infrastructure faces stricter requirements compared to those used in entertainment, reflecting sector-specific risks and needs.

- **International Collaboration:** Harmonizing global standards is critical. The Coalition for Sustainable Artificial Intelligence, led by France, UNEP, and ITU (International Telecommunication Union), is a prime example. This coalition brings together governments, international organizations, industry, and civil society to standardize how AI's environmental impact is measured and to create adaptive, evolving governance frameworks that respond to rapid technological change.[26]

- **Data Stewardship:** Promoting responsible, participatory, and rights-based data stewardship can shift power toward more equitable and public-interest-driven data ecosystems.[27]

Data Quality in Environmental AI

The effectiveness of AI in environmental monitoring and decision-making is heavily dependent on the quality of its input data. This is even more important in supervised AI frameworks where models are trained using pairs of input and target (aka labels) data, and the model should predict the target using the input. In these scenarios, accurate and consistent label data is essential, otherwise AI models risk producing flawed predictions and unreliable guidance for environmental action.

Challenges to Data Quality

Despite its importance, environmental data often suffers from serious quality issues:

- **Missing Data:** Common in satellite, IoT, and sensor data due to equipment failure, transmission issues, or sampling gaps. Missing data can happen in time or space and hinder model training and the detection of long-term trends, extreme events or localized phenomena.

- **Uncertainty and Inconsistency:** Sensor drift, human error (in measurement or labeling), and variability in the surrounding environment introduce noise in data. While AI models need to be trained to tolerate inherent noise in real-world data, high-quality, accurate label data is essential to ensure that models can learn meaningful patterns and deliver reliable predictions, rather than amplifying errors or noise present in the input.[28]

- **Bias:** Historical or imbalanced datasets can reinforce existing inequities, particularly in environmental justice contexts.

Environmental data is inherently irregular – sourced from satellites, ground sensors, citizen science, and archival records, each with unique protocols and uncertainties. Raw datasets are rarely usable without significant processing. Problems such as missing values, environmental noise, inconsistent standards, and spatial/temporal gaps directly reduce the accuracy and trustworthiness of AI models.[28] When trained on biased or low-quality data, AI can produce misleading outputs, undermining the very environmental goals it aims to support.

Opportunities for Improvement in Data Quality

Several strategies can significantly improve the quality of environmental data:

- **Automated and Reproducible Preprocessing Pipelines:** Flexible, adaptive pipelines can automatically detect and correct common data quality issues across diverse sources. Building reproducible pipelines ensures that others can assess any quality issues in the data, and revise the pipeline accordingly.

- **Investment in Data Curation:** Proactive funding for data collection, cleaning, and maintenance is vital for long-term AI effectiveness and an essential public good investment.[29]

- **Metadata and Naming Standards:** Datasets should be searchable, indexed, and accompanied by rich, persistent metadata to ensure usability, transparency, and traceability. The establishment of SpatioTemporal Asset Catalog (STAC) for geospatial data assets is an example of use case-driven metadata standard that has fueled a range of geospatial AI models and applications.[30]

- **Real-Time AI-Enhanced Fusion:** Development of real-time, AI-driven fusion tools allows seamless integration of heterogeneous data streams to improve responsiveness and decision-making.

Ultimately, the success of AI in addressing environmental challenges is inseparable from the integrity and usability of the data it relies on. Investments in data infrastructure, quality standards, and automated processing are essential to unlock AI's full potential in environmental science. Rigorous preprocessing – imputation, harmonization, denoising – and validation are essential to make data usable and AI outputs reliable. Data quality, therefore, is not a secondary concern but a foundational pillar for trustworthy environmental AI.

Data Ethics in Environmental AI

The deployment of AI in environmental contexts raises critical ethical concerns that must be proactively addressed:

- **Algorithmic Bias:** AI models trained on biased or incomplete data that does not provide a fair representation of the population risk reproducing and amplifying societal inequalities, particularly in sensitive areas such as environmental justice.[31] Without proper oversight, AI could unintentionally direct enforcement or resources in a manner that disproportionately harms vulnerable communities.

- **Data Privacy:** The use of large datasets in environmental monitoring brings data privacy concerns, especially where personal or geolocated data is involved. Privacy-preserving techniques and strong data protection standards are necessary to uphold the rights of those represented in the data (individuals, communities, organizations, etc) and build trust.

- **Transparency and Accountability:** True transparency and accountability in AI systems begin with the data itself. Establishing clear data provenance – tracking where data originates, how it is collected, and what transformations it undergoes – is essential for understanding and auditing AI-driven decisions. Comprehensive metadata and rigorous documentation practices ensure that every dataset used in model development is accompanied by information about its source, quality, and limitations.

- **Social Equity and Inclusion:** Ensuring equitable distribution of AI's benefits requires not only engaging underserved regions, such as Low- and Middle-Income Countries (LMICs), but also addressing disparities in data representation. Many LMICs face significant barriers to contributing and accessing high-quality environmental data, which can reinforce digital and environmental divides. To promote true inclusion, it is essential to invest in building local data infrastructure, support the collection of contextually relevant datasets, and ensure that data from diverse regions and communities is integrated into AI systems.

AI trained on biased data risks perpetuating discrimination, especially in environmental regulation. For instance, if environmental enforcement data has historically targeted certain communities unfairly, AI could continue this trend. Additionally, transparent data governance and participatory data stewardship can empower local communities, enabling them to shape how their data is used and to benefit from AI-driven insights.

To prevent this, ethical AI development must take a collaborative, inclusive, and careful approach to data curation, implement fairness-enhancing techniques, and condust regular ethics reviews. The goal should be to design systems that advance equity and justice, offering equal opportunity and benefit to all. Addressing these data-related challenges is as critical as reskilling efforts to mitigate job displacement, ensuring that AI advances social equity at every stage of the data lifecycle.

Geospatial datasets that are accessible, searchable, and indexed support critical environmental management work.

Conclusion and Recommendations

The integration of AI into environmental services holds significant promise, but its success depends on how data is accessed, governed, managed, and ethically used. Data fragmentation, proprietary constraints, inconsistent quality, and ethical concerns limit AI's effectiveness. The paradigm shift towards data-intensive AI solutions for environmental problems underscores that data is not merely an input but a critical infrastructure component whose characteristics directly determine the efficacy and fairness of AI outcomes.

Lessons from mature sectors such as life sciences and finance highlight the value of rigorous data governance and interoperability. In drug discovery, AI has accelerated development timelines by 40%, reducing costs and improving precision medicine outcomes.[32] Similarly, financial institutions leverage AI for fraud detection, saving billions annually. These examples demonstrate that structured frameworks and ethical data stewardship are critical for scaling AI responsibly – principles that environmental science can adopt to achieve similar transformative gains.

Nevertheless, AI offers numerous opportunities for advancing environmental services. Interdisciplinary collaboration among environmental researchers, AI developers, social scientists, local communities, and policymakers is critical to ensuring solutions are scientifically rigorous, ethically responsible, and usable in practice.

To unlock the full transformative potential of AI for environmental services, a concerted, multi-faceted approach is required. The following recommendations are crucial for policymakers, researchers, and practitioners:

1. **Prioritize Open and Interoperable Data Ecosystems:** Invest in open platforms and enforce universal interoperability standards following FAIR principles to overcome proprietary barriers and data silos. Holistic environmental solutions require accessible, integrated data to empower AI with comprehensive insights.

2. **Establish Adaptive, Collaborative and Transparent Data Governance:** Create flexible governance models that emphasize accountability, fairness, and transparency across AI's lifecycle. Promote explainable AI and ensure that decisions are auditable to build trust and avoid bias.

3. **Invest in Robust Data Quality Management:** Fund data cleaning, standardization, and R&D into automated preprocessing and real-time fusion methods. High-quality inputs are essential for reliable and effective AI performance in environmental contexts.

4. **Champion Ethical and Sustainable AI Development:** Embed ethics into development cycles, including mitigation of data bias and equitable access. Encourage energy-efficient models, responsible data use, and sustainable operations to ensure AI's benefits outweigh its environmental footprint.

By strategically addressing these data considerations, the global community can significantly enhance the capacity of AI to deliver impactful solutions for environmental services, fostering a more resilient, equitable, and sustainable future.

About the author

Hamed Alemohammad is an Associate Professor in the Graduate School of Geography and Director of the Center for Geospatial Analytics at Clark University. He is a technical leader and interdisciplinary scholar with extensive expertise in remote sensing, Earth science, and AI. In recent years, his research has been focused on development and application of geospatial foundation models. He also serves as a member of the Technical Advisory Committee of Digital Earth Africa. Prior to Clark University, Hamed was the Chief Data Scientist and Executive Director at Radiant Earth. Hamed received his Ph.D. in Civil and Environmental Engineering from MIT.

1 Olawade, D. B., Wada, O. Z., Ige, A. O., Egbewole, B. I., Olojo, A., & Oladapo, B. I. (2024). Artificial intelligence in environmental monitoring: Advancements, challenges, and future directions. *Hygiene and Environmental Health Advances, 12,* 100114. https://doi.org/10.1016/j.heha.2024.100114

2 Hino, M., Benami, E., & Brooks, N. (2018). Machine learning for environmental monitoring. *Nature Sustainability, 1*, 583–588. https://doi.org/10.1038/s41893-018-0142-9

3 Pollock, L. J., Kitzes, J., Beery, S., et al. (2025). Harnessing artificial intelligence to fill global shortfalls in biodiversity knowledge. *Nature Reviews Biodiversity, 1*, 166–182. https://doi.org/10.1038/s44358-025-00022-3

4 Miller, T., Durlik, I., Kostecka, E., Kozlovska, P., Łobodzińska, A., Sokołowska, S., & Nowy, A. (2025). Integrating artificial intelligence agents with the Internet of Things for enhanced environmental monitoring: Applications in water quality and climate data. *Electronics, 14*(4), 696. https://doi.org/10.3390/electronics14040696

5 Nearing, G., Cohen, D., Dube, V., *et al.* (2024). Global prediction of extreme floods in ungauged watersheds. *Nature, 627,* 559–563. https://doi.org/10.1038/s41586-024-07145-1

6 Dryad. (n.d.). *Revolutionizing wildfire prevention.* Accessed on October 15, 2025. https://www.dryad.net/post/revolutionizing-wildfire-prevention

7 United Nations Environment Programme. (n.d.). *Global environmental data strategy (GEDS).* Accessed on October 15, 2025. https://www.unep.org/topics/digital-transformations/global-environmental-data-strategy-geds

8 Radwan, G. S. (2025). *Briefing on the global environmental data strategy*. United Nations Environment Programme. https://wedocs.unep.org/bitstream/handle/20.500.11822/45360/GEDS-CPR-April-18.pdf

9 Novartis. (n.d.). *The power of transparency*. Accessed on October 30, 2025. https://live.novartis.com/en/article/the-power-of--transparency

10 PRISM Sustainability Directory. (n.d.). *Environmental data interoperability standards.* Accessed on July 30, 2025. https://prism.sustainability-directory.com/area/environmental-data-interoperability-standards/

11 Fraisl, D., et al. (2025). Leveraging the collaborative power of AI and citizen science for sustainable development. *Nature Sustainability, 8*, 125–132. https://doi.org/10.1038/s41893-024-01489-2

12 Walker, D. M., et al. (2023). Perspectives on challenges and opportunities for interoperability: Findings from key informant interviews with stakeholders in Ohio. *JMIR Medical Informatics*. https://doi.org/10.2196/43848

13 PRISM Sustainability Directory. (n.d.). *Community data ownership in climate interventions*. Accessed on July 30, 2025. https://prism.sustainability-directory.com/scenario/community-data-ownership-in-climate-interventions/

14 Pujol Priego, L., & Wareham, J. (2024). Data commoning in the life sciences. *MIS Quarterly, 48*(2), 491–520. https://doi.org/10.25300/MISQ/2023/17439

15 Open Targets. (n.d.). *Open Targets platform*. Accessed on November 4, 2025. https://www.opentargets.org/

16 Data Science Center ETH Zurich. (n.d.). *Enhancing environmental data interoperability across the ETH Domain*. Accessed on July 31, 2025. https://www.datascience.ch/articles/enhancing-environmental-data-interoperability-across-the-eth-domain

17 Biodiversity Information Standards (TDWG). (n.d.). *Darwin Core*. Accessed on October 25, 2025. https://dwc.tdwg.org/

18 United Nations Development Programme. (n.d.). *Use case 3: Real-time monitoring of the environment*. Accessed on June 4, 2025. https://www.smartertogether.earth/how-collective-intelligence-is-used-for-the-sdgs/use-case-3

19 Innovation, Science and Economic Development Canada. (n.d.). *Towards a national digital platform for biodiversity conservation: Data repository and artificial intelligence-enabled analytics*. Accessed on July 29, 2025. https://ised-isde.canada.ca/site/science/en/office-chief-science-advisor/science-advice-support-canadas-national-biodiversity-strategy/data-ai-biodiversity-report

20 European Commission. (n.d.). *The Digital Europe Programme.* Accessed on November 4, 2025. https://digital-strategy.ec.europa.eu/en/activities/digital-programme

21 Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data, 3*, 160018. https://doi.org/10.1038/sdata.2016.18

22 Crystal-Ornelas, R., Varadharajan, C., O'Ryan, D., et al. (2022). Enabling FAIR data in Earth and environmental science with community-centric (meta)data reporting formats. *Scientific Data, 9*, 700. https://doi.org/10.1038/s41597-022-01606-w

23 Bank. (n.d.). *Global trends in AI governance*. Accessed on June 4, 2025. https://documents1.worldbank.org/curated/en/099120224205026271/pdf/P1786161ad76ca0ae1ba3b1558ca4ff88ba.pdf

24 Patel, K. (2024). Ethical reflections on data-centric AI: Balancing benefits and risks. *International Journal of Artificial Intelligence Research and Development, 2*(1), 1–17. https://doi.org/10.2139/ssrn.4993089

25 European Union. (n.d.). *The EU Artificial Intelligence Act.* Accessed on November 4, 2025. https://artificialintelligenceact.eu/

26 Coalition for Sustainable Artificial Intelligence. (n.d.). *Coalition for Sustainable Artificial Intelligence*. Accessed on October 30, 2025. https://www.sustainableaicoalition.org/

27 Modhvadia, R., & Reid, O. F. (2024). *Participatory and inclusive data stewardship: A landscape review.* Ada Lovelace Institute. https://www.adalovelaceinstitute.org/report/participatory-inclusive-data-stewardship/

28 Elmes, A., Alemohammad, H., et al. (2020). Accounting for training data error in machine learning applied to Earth observations. *Remote Sensing, 12*(6), 1034. https://doi.org/10.3390/rs12061034

29 Alation. (n.d.). *Creating a culture of data curation for AI success*. Accessed on November 1, 2025. https://www.alation.com/blog/culture-data-curation-ai-success/

30 SpatioTemporal Asset Catalog. (n.d.). *STAC specification.* Accessed on October 10, 2020. https://stacspec.org/

31 Mergen, A., Çetin-Kılıç, N., & Özbilgin, M. F. (2025). Artificial intelligence and bias towards marginalised groups: Theoretical roots and challenges. In *AI and diversity in a datafied world of work: Will the future of work be inclusive?* (Vol. 12, pp. 17–38). Emerald Publishing.

32 Visan, A. I., & Negut, I. (2024). Integrating artificial intelligence for drug discovery in the context of revolutionizing drug delivery. *Life (Basel), 14*(2), 233. https://doi.org/10.3390/life14020233