Digital Public Infrastructure for Environmental Sustainability

UN environment programme

ISBN No: 978-92-807-4157-5
Job No: DTI/2647/GA

Disclaimers

Suggested citation: United Nations Environment Programme (2024). Digital Public Infrastructure for Environmental Sustainability. Geneva.

Production: United Nations Environment Programme
URL: https://wedocs.unep.org/20.500.11822/45181

Acknowledgements

"Digital Public Infrastructure for Environmental Sustainability" is the product of the generous dedication and effort from numerous individuals, whose knowledge, expertise, and insight helped shape this body of work. UNEP acknowledges the contributions made by individuals and institutions to the preparation and publication of this report.

This report was financially supported by the Government of Norway and the UNEP Digital Transformation Subprogramme.

Lead authors:

UNEP: Yaxuan Chen (former UNEP) and Ana Fernandez Vergara
ETH Zurich: Angus Hamilton
Zurich University of Applied Sciences (ZHAW): Kurt Stockinger

Overall supervision (UNEP):

Rowan Palmer, Asad Naqvi, Joy Aeree Kim and Fulai Sheng.

Contributors:

The report benefited from the inputs of experts on the multiple stakeholder consultations with Inna Amesheva (Arabesque), Douglas Arner (University of Hong Kong), Oliver Bealby-Wright (Consumers International), Serge Bignens (University of applied science Bern (BFH)), Sener Celik (Joindata), Sofie Clausen (former UNEP), David Dao (ETH Zurich, GainForest), Crystal Davis (World Resources Institute), Brandon Day (The Yield Lab Institute), Jan Dirk Wegner (ETH Zurich), Simon Evenett (Global Trade Alert, University of St. Gallen), Robert Guralnick (University of Florida), Susanne Katus (Datamaran), Mark Mulligan (King's College London), Michal Nachmany (Climate Policy Radar), Dimitris Papadopoulos (Infili), Bruce Pon (Ocean Protocol), Maria Ricart (Vizzuality), Meike Schröder (Hamburg University of Technology), Vian Sharif (FNZ), Fulai Sheng (former UNEP), Thomas Streinz (New York University), and Chris West (Trase, the Stockholm Environment Institute).

In addition to the experts mentioned before, UNEP is also grateful for the review, comments and additional contributions of the expert group from the data-science community for the development of the technology innovations that includes Martin Braschler (ZHAW), Gaelle Cheruy (AgDATAHUB), Alexis Chun (Singapore Management School, Legalese), Josep Domingo-Ferrer (Universitat Rovira i Virgili), Henry Franks (Climate Policy Radar), Per Ola Kristensson (Cambridge University), Iveta Lohovska (Hewlett Packard Enterprise [HPE]), Megan Ma (Stanford University), Justin McCall (HPE), Fatemeh Nargesian (University of Rochester), Lam Nguyen (The Commonwealth Scientific and Industrial Research Organisation), Jian Pei (Duke University), Sebastien Picardat (AgDATAHUB), Mark de Reveur (The Delft University of Technology), Hartmut Schultze (HPE), Meng Wong (Singapore Management School, Legalese).

Special thanks are also due to David Jensen (UNEP), Sally Radwan (UNEP), Shivam Kishore (UNEP), Adriana Zacarias Farah (UNEP), Helena Rey de Assis (UNEP), Laetitia Montero Catusse (UNEP), Claire Marion Thiebault (UNEP), Joseph Price (UNEP), Sihem Amer-Yahia (Université Grenoble Alpes) and Jonathan Fürst (ZHAW) for their contributions as peer reviewers.

Editor: Anna Brown and Amanda Lawrence-Brown
Design and layout: Reagan Sirengo, Virginia Njoroge and Bev McDonald

The report was prepared as part of the Digital Enabling of Circularity, Innovation, Development, and Environment project. The report advances the objectives outlined in Resolution 9 of the fifth session of the United Nations Environment Assembly (UNEA/E.A.5/Res.9). It calls upon Member States to give due consideration to the pivotal role played by digital infrastructure in promoting sustainable consumption and production patterns, as well as enhancing the sustainability and efficiency of other infrastructure systems. This endeavour is an integral component of comprehensive approaches to sustainable development. Furthermore, the report lends support to the development and prioritization of the Global Environmental Data Strategy (GEDS) as mandated by the UNEA (UNEA/E.A.4/Res.23). This strategy aims to establish a long-term data framework and involves extensive consultation with governments, United Nations agencies, funds and programmes, the secretariats of Multilateral Environmental Agreements, as well as international and regional scientific organizations.

Zurich University
of Applied Sciences

Acronyms

AFI: Accountability Framework Initiative

AI: Artificial intelligence

CDP: Carbon Disclosure Project

DPP: Digital product passports

DPI: Digital public infrastructure

DT: Digital Transformation

EC: European Commission

EU: European Union

GCEP: Green and circular economy policies

GDPR: General Data Protection Regulation

GEDS: Global Environmental Data Strategy

GHG: Greenhouse gas

HCAI: Human-centered artificial intelligence

ILO: International Labour Organization

IoT: Internet of Things

LLM: Large language models

ML: Machine learning

MRV: Monitoring, reporting and verification

OECD: Organisation for Economic Co-operation and Development

PPP: Public-private partnership

SCP: Sustainable consumption and production

UI: User interfaces

UNDP: United Nations Development Programme

UNEP: United Nations Environment Programme

WEF: World Economic Forum

WESR: World Environment Situation Room

Glossary

Artificial Intelligence (AI) system: a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy (Organisation for Economic Co-operation and Development [OECD] 2022).

Big data: collected data sets that are so large and complex that they require new technologies, such as artificial intelligence, to process. The data comes from many different sources. Often they are of the same type, for example, GPS data from millions of mobile phones is used to mitigate traffic jams; but it can also be a combination, such as health records and patients' app use. Technology enables this data to be collected very fast, in near real time, and get analysed to get new insights (European Parliament 2021).

Circular economy: one of the current sustainable economic models, in which products and materials are designed in such a way that they can be reused, remanufactured, recycled or recovered and thus maintained in the economy for as long as possible, along with the resources of which they are made, and the generation of waste, especially hazardous waste, is avoided or minimized, and greenhouse gas emissions are prevented or reduced, can contribute significantly to sustainable consumption and production (UNEP/EA.4/Res.1).

Computational law: the branch of legal informatics concerned with the codification of regulations in precise, computable form (Genesereth 2015).

Cybersecurity: concerns protecting data or information systems against damage, unauthorized use, modification or exploitation (United States, Committee on National Security Systems [CNSS] 2022).

Data lake: is a centralized repository that allows to store all structured and unstructured data at any scale (Amazon Web Services 2023).

Data marketplace: a platform on which anybody (or at least a great number of potentially registered clients) can upload and maintain data sets. Access to and use of the data is regulated through varying licensing models (Schomm *et al.* 2013, p. 15-26).

Data owner: individuals or institutions who make decisions such as who has the right to access and edit data and how it is used (European Union [EU] General Data Protection Regulation [GDPR] 2023).

Data privacy: the rights of individuals, organizations, and businesses over how their data is collected, shared and used, as well as the systems that are implemented to protect the right to privacy (OECD 2013).

Digital ecosystem: is a complex network of people, businesses, and systems that use technology to interact with one another. Such digital ecosystems are significantly different from traditional business ecosystems because they take advantage of physical layers (devices), the information layers (data), and the application layers (apps) (International Institute for Management Development [IMD] 2022).

Digital inclusion: equitable, meaningful, and safe access to use, lead, and design of digital technologies, services, and associated opportunities for everyone, everywhere (United Nations Roundtable on Digital Inclusion 2021).

Digital public infrastructure (DPI): a set of shared digital systems which are secure and interoperable, built on open standards, and specifications to deliver and provide equitable access to public and/or private services at societal scale and are governed by enabling rules to drive development, inclusion, innovation, trust and competition and respect human rights and fundamental freedoms (India's G20 Presidency and United Nations Development Programme [UNDP] 2023).

Greenhouse gases (GHGs): The atmospheric gases responsible for causing global warming and climatic change. The major GHGs are CO_2, methane (CH_4) and nitrous oxide (N_2O). Less prevalent, but very powerful, GHGs include hydrofluorocarbons (HFCs), perfluorocarbons (PFCs) and sulphur hexafluoride (SF_6) (UNEP 2023).

Green economy: one that results in improved human well-being and social equity, while significantly reducing environmental risks and ecological scarcities (UNEP 2011).

Innovation: a system that is designed for customization, encourages competition and promotes the development of new solutions on top of existing platforms and digital systems (UNDP 2023).

Interoperability: the ability to freely exchange and use data between services, regardless of origin, programming language or interface (UNDP 2023).

Language model: assigns probability to a piece of unseen text, based on some training data (Hiemstra 2019).

Machine learning (ML): a subfield of AI that studies "how computer agents can improve their perception, knowledge, thinking, or actions based on experience or data. For this, ML draws from computer science, statistics, psychology, neuroscience, economics and control theory (Stanford 2020).

Metadata: information that is needed to be able to use and interpret statistics. Metadata describe data by giving definitions of populations, objects, variables, the methodology and quality (European Commission 2023).

Non-traditional data sources: types of data that are derived from sources that have historically not been used to generate statistical products. Earth observation data derived from satellites and other remote sensors; administrative data sets compiled by governmental ministries, departments, and agencies, private sector companies, and civic groups; citizen-generated data produced through citizen science initiatives; big data and metadata sets generated and processed primarily by the private sector; and data outputs produced by automated algorithmic processing are some of the main non-traditional data sources which NSOs are experimenting with to complement more traditional survey and census-based data collection methods (Orrell 2021).

Open-domain question answering (OpenQA): the natural language processing problem of finding answers in collections of unstructured documents on diverse topics (Chen *et al.* 2020).

Open data: data that can be freely used, reused and redistributed by anyone – subject only, at most, to the requirement to attribute and share alike (Open Knowledge International 2023).

Privacy enhancing technologies: software and hardware solutions, i.e. systems encompassing technical processes, methods or knowledge to achieve specific privacy or data protection functionality or to protect against risks of privacy of an individual or a group of natural persons (European Union Agency for Cybersecurity [ENISA] 2015).

Sustainable consumption and production: use of services and related products which respond to basic needs and bring a better quality of life while minimizing the use of natural resources and toxic materials as well as the emissions of waste and pollutants over the life cycle of the service or product so as not to jeopardize the needs of further generations (UNEP/PP/INC.1/6).

Traceability: the process by which enterprises track materials and products through the supply chain (OECD 2018).

Triple planetary crisis: the three main interlinked issues that humanity currently faces: climate change, pollution and biodiversity loss (UNFCCC 2022b).

Table of Contents

Tables

Figures

Executive summary

The report examines common information challenges stakeholders face when making decisions related to environmental sustainability and explores the role that Digital Public Infrastructure (DPI) can play as a key part of the solution.

To tackle the interconnected triple environmental planetary crises, it is critical to have accessible, timely, credible, and insightful information that can support environmental sustainability decision-making. Developing interconnected data exchange mechanisms has become a necessity, but sole reliance on private solutions will likely fail to comprehensively address the challenges and may result in further data fragmentation. A blend of private and public solutions is essential. However, there is currently a notable gap in DPI to facilitate the flow of environmental sustainability information to different stakeholders.

This report analyses three cases related to the agrifood sector and identifies six categories of technology innovations (TIs) that could help tackle information challenges:

- **Open data discovery for environmental sustainability**
- **Privacy enhancing technologies to enable flow of environmental sustainability information**
- **Data markets for environmental sustainability-related data**
- **Computational law and data integration of green and circular economy policy measures**
- **Using Large Language Models to 'speak' with green and circular economy policy**
- **Tools and techniques for human-centred artificial intelligence in environmental sustainability decision-making**

Together, the six categories of TIs can help enable DPI for the exchange of sustainability-related data information.

The proposed publicly supported digital ecosystem would enhance transparency and data availability. It would also create incentives and mechanisms for more efficient data generation and collection, ease the discovery of data sources, and reduce barriers to data sharing. The result would be an enhanced flow of information with improved data quality and inter-operability. It would also empower stakeholders by facilitating informed decision-making and promoting greater inclusiveness. Finally, DPI for environmental sustainability would serve as a critical foundation for transformative digital applications, such as digital product passports.

Concluding with a call to action, the report urges stakeholders to actively participate in data exchanges, leveraging DPI to maximize data's potential for environmental sustainability. It lists key recommendations for policy, standardization, financing, innovation, and collaboration.

Policy and regulations: to foster enabling conditions, policy support is required for the development of DPI for environmental sustainability, as well as policy safeguards for public interest. For example, policymakers are encouraged to align with global efforts, establish national priorities to close DPI gaps, incentivize data exchanges, and integrate real-time data into institutional decision-making.

Standardization: guiding the development of DPI for environmental sustainability, through collaborative efforts between various stakeholders, balancing the need for security, privacy, inclusivity, accessibility, interoperability, and adherence to regulatory and legal frameworks. Among others, standardization efforts could focus on a global environmental data strategy (GEDS) and harmonized methodologies to ensure data quality and inter-operability.

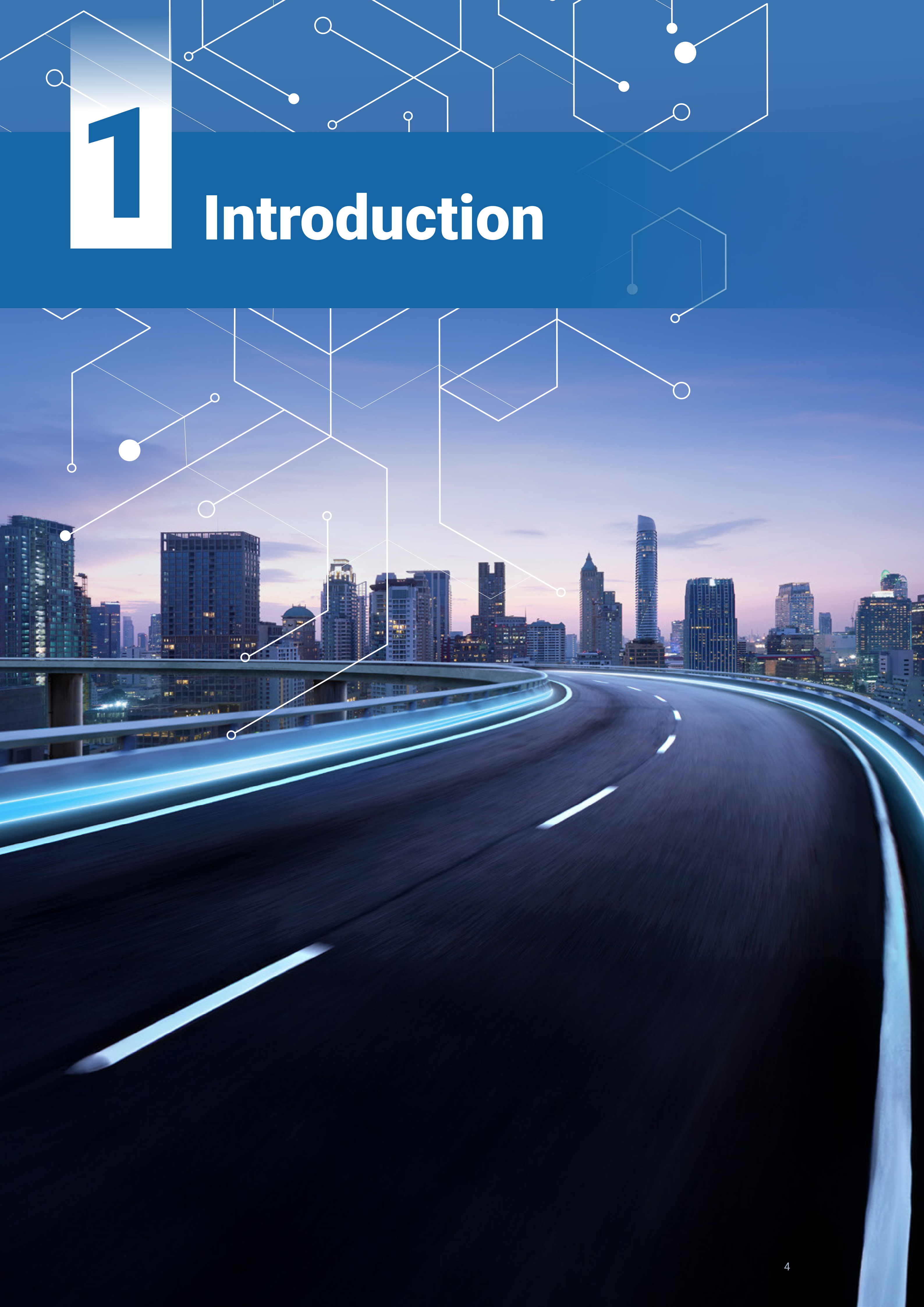
Finance: to scale up efforts on DPI it is essential to leverage public and private investment for paving the infrastructure and bringing life to business models for environmental sustainability via DPI, through reformed global financing mechanisms.

Innovation facilitators: To foster the progress of the development of DPI for environmental sustainability, it is essential to take steps to overcome barriers to innovation, while ensuring integrating inclusivity considerations. Innovation can be spurred by relying on regulatory sandboxes, promoting hackathons, and fostering new market entrants.

Collaboration and partnership: International collaboration and public-private partnerships (PPPs) play a crucial role in ensuring the successful development of DPI for environmental sustainability. These should include direct partnerships with the data science community and technology developers to harness innovation.

UNEP and its international and country partners will continue to promote and support the development of DPI for environmental sustainability by raising awareness amongst key stakeholder groups, conducting research and bridging knowledge gaps between experts from different domains, and supporting pilot projects at the country level.

1 Introduction



To address the interconnected triple planetary crises of climate change, biodiversity loss, and pollution, multi-stakeholder efforts are crucial for transforming how economies interact with the environment. Current socioeconomic systems follow a linear "take-make-waste" model, leading to resource overexploitation, excessive waste, environmental degradation, and heightened greenhouse gas emissions. This model not only depletes ecosystems and biodiversity but also exacerbates social inequalities and climate change, proving environmentally, economically, and socially unsustainable in the long run.

Reform advocates call for a shift towards a green and circular economy, which eradicates waste and pollution through design, drastically reduces the depletion of natural resources, and promotes practices that are low in carbon emissions, resource-efficient, and socially inclusive. To achieve these goals, all actors would need to make fundamental shifts. Governments need to redesign economic incentives and implement transformative policies. The private sector must rethink investment strategies, accelerate innovation, and adopt new circular business models. Consumers must shift their behaviours towards sustainability and make more informed purchasing decisions. Technology developers and the data science community can be pivotal drivers of sustainable innovation, particularly when integrated into transformation processes.

However, there are practical challenges that must be solved. For instance, can consumers access trusted and relevant information about the ecological footprints of products? Can companies optimize value chains to minimize their environmental footprint? Do governments have the necessary data and capacity to assess economic activities to design and enforce effective public policies? These related challenges primarily stem from a lack of accessible, credible, and insightful information that can support environmental sustainability decision-making. Moreover, non-traditional data sources face significant undervaluation and underutilization (Orrell 2021).

Data is the backbone of sustainable economic models like the green economy and circular economy, as well as for SCP. For green economy, a wide range of data is used, from environmental data on biodiversity and ecosystems to data on energy use and emissions. This is essential for sustainable natural resource management, encouraging the use of renewable energy, and reducing climate change impacts. For circular economy, data helps track and manage resources and materials, providing crucial information on how materials flow, are used, and wasted. This information is key to designing sustainable products, maintaining transparent supply chains, and developing new business models like offering products as a service. For SCP, comprehensive data on resource use, energy efficiency, product lifecycles, and consumer behaviour is crucial. It’s used to monitor resource use, shape energy policies, and encourage sustainable consumer habits.

While each area has its specific data needs, the overall requirement is for data that is easy to find, access, share, understand, and use. A robust data infrastructure is crucial for making informed decisions that support environmental sustainability. It ensures that policies and actions in these areas work well together and balance environmental care with societal well-being. Effectively using this data is key to making decisions that support environmental sustainability.

This report therefore focuses on the pivotal role of three types of critical information - environmental monitoring data, insights into economic activities (particularly value chain operations and traceability), and information on policy effectiveness in advancing a green and circular economy. The analysis in these areas aims to gauge the availability, processing, analysis, and practical application of relevant data and information.

The report begins with an analysis focusing on the challenges of identifying environmental hotspots containing significant ecosystem degradation. It then delves into the economic drivers that contribute to these hotspots and examines the scope of policy interventions aimed at governing different aspects of these driving factors. The information challenges identified in this analysis underscore a critical, yet often overlooked, necessity for DPI in supporting decision-making for environmental sustainability.

This is followed by an exploration of six technology innovations that can enable a DPI as a data exchange system for environmental sustainability with the potential of providing transformative solutions for the identified information challenges.

While DPI is still considered an evolving concept, one of the more widely used definitions is that of “a set of shared digital systems which are secure and interoperable, built on open standards, and specifications to deliver and provide equitable access to public and/or private services at societal scale and are governed by enabling rules to drive development, inclusion, innovation, trust and competition and respect human rights and fundamental freedoms” (India’s G20 Presidency and UNDP 2023). Among the key categories of foundational DPI—such as digital identity and digital payments—the one pertaining to data exchange emerges as the most vital for environmental sustainability. In recent years, the first two categories have received more global attention, but DPIs dedicated to data exchange are just now starting to emerge (WEF 2024). As a data exchange system, DPI can facilitate the follow of critical information and help bolster informed decision-making processes.

Building on that, it is noteworthy that a digital ecosystem refers to a complex network of digital technologies, platforms, and environments where different entities (such as businesses, consumers, governmental institutions, systems, and data) interact (see Figure 1). This interaction typically includes the exchange of information, services, and goods, facilitated by digital processes and technologies. DPI acts as a critical intermediary layer within the digital ecosystem, bridging the gap between physical infrastructure elements (i.e. connectivity infrastructure, devices, servers, data centres, and routers) and the application layers (Chakravorti 2023). DPI’s role in environmental sustainability is pivotal; it streamlines the exchange of sustainability data, seamlessly integrating this information into a wide array of applications, including personal carbon tracking, Digital Product Passports (DPPs), and automated environmental compliance systems, among others.

Figure 1: **DPI within a digital ecosystem**

The report concludes with a key message on the need to use DPI as a data exchange ecosystem that supports the environmental sustainability decision-making of governments, consumers, and the private sector.

It should be noted that this research primarily focuses on underscoring the critical, yet often overlooked, necessity for DPI in supporting decision-making for environmental sustainability. However, the research does not delve into the details of financing such infrastructure. Further investigation is imperative to determine the financial requirements needed to unlock transformative applications enabled by DPI. Additionally, the report does not cover the dimension of greening DPI (e.g. reducing carbon intensity or increasing resource efficiency of DPI), a topic that represents a vital area for future research and discussion.

Audience

This report is intended to serve as a bridge between two broad communities. The first is the wide spectrum of policymakers, who, while perhaps not deeply entrenched in the technical nuances of digitalization, ICT, and infrastructure, play pivotal roles in shaping the landscape across these and other critical sectors such as environmental and economic planning. For example, policymakers in digitalization, ICT, and infrastructure are key to this dialogue due to their influence over the technological frameworks and policies that enable or hinder DPI's deployment. But equally critical are the policymakers focused on environmental and economic planning. All their roles are crucial for integrating digital solutions with environmental objectives.

The second community is technology developers and the data science professionals who are outside the traditional policy sphere. The report aims to provide them with information about the types of environmental sustainability challenges that their expertise and innovation can help to address.

The interconnectedness of these domains underscores the report's interdisciplinary nature. It reflects the understanding that successful DPI development for environmental sustainability requires a harmonized approach, integrating technological expertise and innovation with environmental and economic foresight. This realization points to the need for a unified strategy that involves collaboration across all these sectors, and policies that support cross-disciplinary initiatives. The target audience is thus envisioned as a coalition of forward-thinking policymakers and innovators from across these critical domains, united in their commitment to driving DPI for environmental sustainability.

Methodology

The research for this study was exploratory and descriptive, employing a mixed-methods approach, combining qualitative data from expert consultations with desktop research. The preparation of the report had two distinct phases, each contributing to a comprehensive understanding of the subject matter.

The first phase focused on the development of use cases within three specific domains: environmental monitoring, value chain traceability and transparency of green and circular economy policies. This phase aimed to identify key challenges in the specified domains through a combination of desktop research and expert consultation (Green Policy Platform [GPP] 2022).

A comprehensive literature review encompassed academic papers, books, reports, and online resources. Complementing the desktop research, expert consultations were conducted to gain valuable insights from subject matter experts in the three domains (GPP 2022). The project team identified a panel of thematic experts with diverse backgrounds and relevant expertise around the topics of environmental data, value chain intelligence and economic policymaking. In the form of multi-stakeholder exchanges, the project team conducted five in-depth, semi-structured roundtable discussions to gather expert opinions, insights, and recommendations.

All meeting discussions were recorded and systematized for precise data capture, followed by qualitative content analysis to extract key themes, patterns and recommendations. These are referred to as thematic expert consultations throughout the report.

Before expert consultations, a draft of each technology innovation was prepared, which involved an extensive review of existing technologies and innovations with potential applications within the domains under investigation.

The draft was subsequently presented to experts on data science for their valuable insights and feedback on the technology solutions proposed and their maturity in the application, through one-to-one semi-structured interviews. Their comments and recommendations were integrated into the draft to refine and enhance the proposed technology innovations. In the report, these are called data science expert interviews.

To ensure the reliability and validity of our findings and recommendations, a triangulation approach was employed, cross-referencing insights from both phases of research. Any discrepancies or inconsistencies were meticulously addressed to refine the findings.

During the consultations, participants' privacy and consent were given priority. Experts were fully informed about the study's purpose, and their anonymity was preserved when requested.

2 Identifying challenges through three use-cases

Critical information for environmental sustainability decision-making encompasses environmental data, economic activities insights (i.e. value chain operations) and information on policy effectiveness in advancing a green and circular economy.

To illustrate information challenges for the effectiveness of environmental sustainability initiatives, this section presents three use cases from the agrifood sector covering: 1) environmental monitoring; 2) value chain traceability; 3) the transparency of policies related to green and circular economy.

This sector was selected as an area of focus as it traverses all aspects of the triple planetary crisis. The expansion of agrifood systems, based on land use conversion such as deforestation and practices of overexploitation and intensified agriculture, has been a leading cause of biodiversity loss (Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services [IPBES] 2020). Agricultural commodities such as beef, soy, palm oil, coffee and cocoa, are a pre-eminent contributor to current deforestation rates (Food and Agriculture Organization of the United Nations [FAO] 2019; 2020).

In addition, it is estimated that globally, deforestation and forest degradation account for around 11 per cent of GHG emissions (FAO 2022b). Emissions generated within the farm gate by crop and livestock production and related land use change contribute about one-fifth to one-quarter of total emissions from all human activities, when measured in CO_2 equivalents (Intergovernmental Panel on Climate Change [IPCC] 2014; IPCC 2019). Once pre- and post-production activities along agrifood systems' value chains are included, food and agriculture activities generate up to one-third of all anthropogenic emissions globally (Tubiello *et al.* 2022). The food value chain in many countries is on course to overtake farming and land use as the largest contributor to GHGs from the agrifood system (United Nations 2021).

Adding to the sustainability complexity of the sector is the governance challenge associated with the value chain structure. Only one per cent of farms in the world are larger than 50 hectares, but they use over 70 per cent of the world's farmland (FAO 2021a). The challenge of chemical pollution from fertilizers and pesticides in agriculture from the sector is also globally significant with 64 per cent of global agricultural land at risk (Tang *et al.* 2021).

All these reasons emphasize the suitability of using the agrifood sector to develop use cases for DPI for environmental sustainability. The research uses beef as a reference point for data gathering and the model application, as it is resource-intensive and an important driver of pollution, soil degradation and GHG emissions (World Wildlife Fund [WWF] 2023).

It is important to emphasize that the analytical framework used in this report is not limited to these examples; it can be equally applied to other environmental hotspots, their associated value chains and relevant public policies. To this end, each subsection concludes with general remarks that translate information-related challenges into data science queries.

2.1 Environmental monitoring

This section explores whether and to what extent environmental monitoring can accurately assess the extent of different environmental problems and their drivers in order to inform response strategies. Environmental monitoring actors, ranging from formal institutions to professionals and individual citizens that work on environmental data, contribute to the observation of environmental change and the identification of environmental hotspots and risks. While their motivations for monitoring may vary, from political directives to personal interest, the efficacy and accuracy of their methods heavily influence policy decisions and public perception.

In light of this, the following section offers a deeper dive into the key sources from which they derive their environmental data, the tools they employ for analysis and the reception of their findings among the intended target users.

2.1.1 Identifying environmental problems and trends

International organizations collect and manage environmental data on climate, nature and pollution[1]. Since 1946, FAO has monitored global forests using a process of national data collection, international review, and validation (2004b). Data sources include national forest inventories, remote sensing assessments, full-cover forest maps and questionnaires (FAO 2018b; 2022a). This monitoring provides insight into forest-related land-use changes. Between 1990 and 2020, global forests as a proportion of total land areas decreased from 32.5 per cent to 30.8 per cent, representing a net loss of 178 million hectares of forest (FAO and UNEP 2020).

The interpretation of imagery from Earth-observing satellites at national and international levels has changed the landscape for global forest monitoring, by providing spatially detailed and timely information on forest dynamics. For instance, the Global Land Analysis & Discovery (GLAD) lab at the University of Maryland and Google (Global Land Analysis & Discovery 2022) provides time-series analysis of Landsat[2] Earth-observing satellite images to characterize forest extent and change (United States Geological Survey 2022). It features data sets of global forest cover at approximately 30x30m resolution with continuous data updates from 2000. User-friendly tools such as Global Forest Watch, Earth Map and SEPAL, can enable complex land cover monitoring. Using statistical analysis to identify the intensity and the temporal trend, areas with trends of deforestation that are new, sporadic, intensifying or persistent can be located to enable timely intervention by decision-makers (Hansen *et al.* 2013; Harris *et al.* 2017).

Not all environmental monitoring can benefit from the application of digital tools to the same level as described above. It depends on the feature being monitored, how it can be accurately detected in a satellite image and on the frequency needed to inform decision-making. There is also a large capacity gap among stakeholders in making use of the tools and variants associated with a specific sectoral or regional context.

2.1.2 Identifying the drivers of forest loss

The analysis of direct drivers of deforestation is conducted through an assessment of the transition from one class of land use to another. Identifying specific activities driving deforestation at the subnational level is challenging due to time-lags before the first harvest on cleared land and possible land-use changes. Researchers use economic models comparing crop category expansion with deforestation to determine deforestation embodied in production for chosen areas (Pendrill *et al.* 2019).

Additionally, satellite images (Figure 2) help differentiate deforestation drivers based on forest and land-use dynamics (Curtis *et al.* 2018).[3] Machine-learning decision-tree models recognize visual patterns in 10km x 10km grid cells, predicting the most likely cause of forest disturbance (Curtis *et al.* 2018).

1. The Food and Agriculture Organization of the United Nations gathers environmental data on agriculture, forestry and fisheries; the World Meteorological Organization (WMO) on the atmosphere; the World Health Organization (WHO) on pollutants as they effect human health; the International Labor Organization (ILO) and WHO on the working environment; the International Atomic Energy Agency (IAEA) on safe use of nuclear energy; the Inter Governmental Maritime Consultative Organization (IMCO) on marine pollution; and the United Nations Economic and Social Council (UNESCO) has a continuing, long-term programme of scientific and technical information exchange. For a detailed account of information collection and interpretation by international organizations, see United Nations General Assembly, Consolidated Document on the United Nations System and the Human Environment Submitted by the Administrative Committee on Co-ordination, at 36-39, U.N. Doc. A/CONF.48/12 (1971).
2. The Landsat Program is a series of Earth-observing satellite missions jointly managed by NASA and the United States Geological Survey, starting from 1972.
3. This includes (i) commodity-driven deforestation, defined by the long-term, permanent conversion of forest and shrubland to a non-forest land use such as agriculture (including oil palm), mining or energy infrastructure; (ii) shifting agriculture, defined as small- to medium-scale forest and shrubland conversion for agriculture that is later abandoned and followed by subsequent forest regrowth; (iii) forestry, defined as large-scale forestry operations occurring within managed forests and tree plantations with evidence of forest regrowth in subsequent years; (iv) wildfire, defined as large-scale forest loss resulting from the burning of forest vegetation with no visible human conversion or agricultural activity afterward; and (v) urbanization, defined as forest and shrubland conversion for the expansion and intensification of existing urban centres.

Figure 2: **Satellite images show deforestation drivers** (Source: Curtis *et al.* 2018, p. 1108–1111)[4]

Being widely applied in the context of precision agriculture, spectral vegetation indices applications combined with open-source satellite data can generate insights to recognize the crop type and to determine pasture quality on the land. Pixel-based or object-based image analysis together with convolutional neural networks (CNNs) are enabling models to automate detecting livestock from imagery (Mücher *et al.* 2022). The multibillion satellite imaging services market is expanding, driven by both private providers like Google, DigitalGlobe and Planet, and public initiatives like EU's Copernicus Program and those of NASA and USGS, offering varying resolutions with commercial satellites often providing higher resolution images around 30-50cm per pixel. This development provides opportunities to promptly recognize the specific category of commodity or agricultural product or automatically detect cattle herds associated with deforestation.[5]

The MapBiomas is an application combining some of these data science innovations for the purpose of environmental monitoring. For instance, by tracing the spectral behaviour of land-cover changes (MapBiomas 2020) between 2001 and 2019 in one area of South America, the MapBiomas provides details on the vegetation on the land-cover (Figure 3). It shows that in the past two decades, over 90 per cent of the deforestation in the traced area was caused by forest conversion to pasture associated with cattle ranching (MapBiomas 2022).

4. The authors have provided the following explanation on the images. "Categories were assigned according to dominant disturbance type (Figure 1), with each representing a different forest and land-use dynamic."

5. Note that detection of vegetation cover and farming activities are different from concluding them as the direct driver of deforestation. For instance, the land can be cleared initially for pastures and cattle ranching but later converted into soy production. It is also possible that crops such as soy and corn are grown in a rotary manner (Trase 2020b, p. 4).

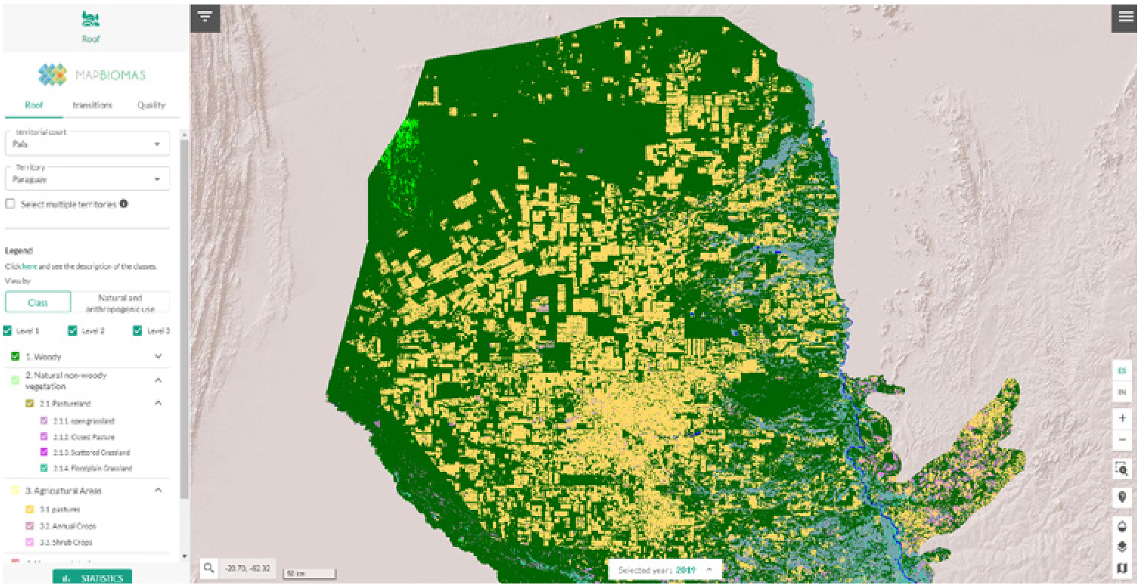


Figure 3: **Vegetation cover of one area in South America in 2019** (Source: MapBiomas 2022)

2.1.3 Connecting the drivers of deforestation with value chains

Understanding "hotspots" of environmental impacts and the drivers of deforestation involves identifying the stakeholders behind certain activities and connecting them to a specific part of the value chain. This helps determine how businesses and economies can shift incentives towards a green and circular economy.

For example, in the beef value chain, it is essential to trace cattle and cattle products' origins, ownership and movement through space and time (Veit and Sarsfield 2017).[6] Satellite images and land registry data can help identify landowners responsible for deforestation, while data on Indigenous communities and environmental licences can provide insights into the legality of deforestation and associated cattle ranching.

Connecting pastureland with cattle ranching requires data sets on ranch locations. Traceability systems for agricultural products can help in this regard. Paraguay, for instance, has implemented SITRAP (FAO 2004a; FAO 2018c) and SIGOR II (Eisele 2021; Paraguay, Servicio Nacional de Calidad y Salud Animal [SENACSA] 2017) to trace livestock from origin to destination, while private producers within the Rural Association of Paraguay (ARP) use software to register farms and livestock (United Nations Framework on Climate Change [UNFCCC] 2018).

Public monitors, such as trase.earth, use millions of data points to trace connections between production and international trade at subnational level. This mapping indicates influential consumer groups, trade players and governments that can help change deforestation status (Trase 2020a).

2.1.4 Observations on the environmental data ecosystem

This use case demonstrates how digital technology supports environmental monitoring through data generation methods in addition to traditional ground-measurement, census and statistics. Remote sensing, in particular, generates rich data sources, providing a broad and consistent coverage of environmental features. Table 1 provides an illustrative list of remote sensing data products that can be used for different environmental monitoring applications (Sun *et al.* 2022). International organizations, governments and think tanks actively use these tools to produce digital public goods and access pooled environmental intelligence.

6. Note the differences between legal ownership and economic ownership. "The legal owner is the person who is recognized in law to own the asset or good in question. At the same time the economic owner could be another person who exercises control over the asset and ultimately benefits from its use." (United Nations Economic Commission for Europe [UNECE] 2013).

Category	Data sets	Data products
Land	Land cover	FROM-GLC, ESA WorldCover, Esri LandCover, GlobelLand30
	Soil moisture	SMAP, MODIS, Landsat, Sentinel
	Land surface temperature	FY-3, FY-4, Landsat, Sentinel-3, AMSR2, MODIS
	Digital elevation	SRTM, ASTER GDEM, ALOS DSM, Tan DEM
	Surface reflectivity	GF, MODIS, Landsat, Sentinel-2
	Soil material content	HWSD
Vegetation	Forest biomass	GF, MOD17A2H, GLAS, Landsat, Sentinel-2, PALSAR, LiDAR
	Vegetation coverage	GF, Landsat, Sentinel-2, GLASS FVC, GEOV3 FVC
	Forest height	ICESat, Landsat, ALOS PALSAR
	Vegetation evapotranspiration	MOD16A2
	Normalized vegetation index	GF, MOD13A1, Landsat, Sentinel-2
Climate	PM2.5, PM10, O3	Himawari, MODIS, CALIPSO, MISR, POLDER
	Evapotranspiration	SSEBop, MOD16, GLEAM

Table 1: **An illustrative list of remote sensing data products** (Sun *et al.* 2022)

Despite these benefits, environmental monitoring stakeholders express concerns about data access, management and coverage to the desired granularity (GPP 2022). In some cases, high-resolution imagery in the public domain ranges from 1m to 3m. While a higher resolution could increase the accuracy of monitoring especially in small-scale projects (noting the diminished commercial value of archival data for satellite companies), the technical challenges concern storage space for exponential growth of volume of data and technical capacity for data analysis. Moreover, to access higher resolution data, e.g. those at 30cm, the data procurement cost could be commercially prohibitive.

While lower resolution satellite data can provide a basic layer for environmental intelligence, the quality of data globally is unequal and must be combined with localized data resources to reflect the realities of local ecosystems and communities. For instance, some remote sensing applications can measure the species and height of a tree. This needs to be combined with additional information from other sources about the tree's role within a localized ecosystem, or the value of a tree to local communities or its role in the rituals of Indigenous people. However, this can have limited impact if data owners do not realize the value of their data, nor do they have the platforms to share it. Incorporating these multi-faceted elements of environmental information into decision-making can improve ground-truth understandings of links between value chains, policy measures and the environment, while improving justice and representation.

To balance the need for granular intelligence at a reasonable cost, there are technological innovations that first use low resolution data to identify targets or hotspots and then access data of high resolution for more granular and targeted monitoring. Experts also emphasize the importance of working with partners on the ground to ask the right questions and avoid bias in the data analysis, which could lead to flawed solutions.

While various quantitative and qualitative sources of environmental data offer a range of technical options for more comprehensive and accurate Earth observation, there is a lack of a unified global system that would provide a stable and long-term infrastructure for data collection and guarantee consistency and coherence in the data gathered. This is a systemic problem that could be addressed through digital public infrastructure.

There are also major capacity challenges for stakeholders to process big data while addressing the problem of data gaps in country or project applications. Moreover, public resources, especially national investment in environmental data, are shrinking (Hoekstra 2022, p.6). The donation-driven funding model for environmental monitoring forces monitoring institutions to compete for resources rather than collaborating. This has the effect of fragmenting the knowledge base and putting it in the hands of short-term projects that come and go, causing loss of information.

Environmental data and intelligence have facilitated the establishment of a credible scientific baseline for global environmental governance in past decades. However, challenges exist for stakeholders to extract knowledge from environmental data and convert it into decisions and behaviour changes. Without connecting environmental intelligence with information of human society and economic activities, even the most perfect monitoring scheme would not be sufficient to address the root causes of environmental problems. The absence of a data-knowledge-action pipeline is associated with the reality of the data ecosystems of value chains and economic policies and the lack of effective connection between the three domains.

2.2 Value chain traceability

Modern value chains are increasingly complex and intricate networks of value chain operators, individuals, and businesses with operations in different countries (Akin *et al.* 2022). The beef value chain will be used to illustrate different stakeholders, including those in charge of inputs, breeding, rearing, fattening farms, abattoirs or meatpackers, exporters/traders, transporters, retailers, packaging, etc. With this level of complexity, achieving reliable and timely data of entire value chains, to inform environmental sustainability decision-making is a challenging goal.

The demand for measurable environmental impact data on value chains is escalating, notably to have trustworthy traceability systems. This surge is underscored by a proliferation of ongoing initiatives, which are indicative of progress but also contribute to a dense and sometimes confusing landscape of information (OECD 2023). The growing number of due diligence legislations suggests that regulatory bodies are progressively adopting a supply chain perspective to address sustainability concerns (World Economic Forum 2022). In climate reporting, there is a growing trend for companies to not only report emissions from their own operations (referred to as "Scope 1" in the terminology of the GHG Protocol) as well as emissions from purchased energy (Scope 2) but also emissions occurring elsewhere in their supply chain, both upstream and downstream (Scope 3) (GHG Protocol 2014). Furthermore, it is anticipated that there will be cooperative interactions between the requirements for disclosure and the practices of due diligence, as projected by Norton Rose Fulbright (2022).

Having explored the intricacies of environmental monitoring, this use case on value chain traceability analyses how companies within the beef value chain identify, measure, and report the environmental impact of their operations. At its core, this task requires connecting environmental data with value chain information. Starting from the premise that value chain monitoring can serve a variety of goals, this use case explores the data requirements for different types of value chain monitoring efforts to identify the major information challenges.

2.2.1 The variety of value chain monitoring goals

Companies require monitoring systems to strengthen their decision-making with reliable information on their operations, their impact and that of the wider value chain. Corporate needs may vary for developing, or hiring a provider of, monitoring systems. For many companies, the main goal in improving their data ecosystems is to increase their efficiency, optimize their benefit-cost strategies and adjust their business strategies.

Food value chains are especially vulnerable to environmental change, so companies seek the information to prevent supply chain disturbances and build resilience. A study showed that in the past 50 years, half of all shocks to crop production systems were a result of extreme weather events (Cottrell *et al.* 2019).

Other companies gather monitoring information across their supply chains for due diligence processes, to avoid reputational risks, corporate social responsibility reporting and regulatory compliance purposes (UNEP 2013). The increased public pressure to be sustainable and socially responsible has led many companies to publicly commit to sustainable practices (UNFCCC 2022a). In particular, anti-deforestation commitments have gained popularity in recent years (FAO 2018a). To report on the fulfilment of those pledges, companies require reliable and complete information on the environmental performance of value chains, including their relationship with forest loss. Companies need to track a commodity's origin to ascertain that it did not contribute to deforestation (The New York Times 2021).

Alongside those commitments, there is a growing trend to establish Monitoring, Reporting and Verification (MRV) systems for corporate sustainability performance (Chever *et al.* 2022). Despite challenges created by the significant differences between the existing frameworks that lead to sustainability disclosures (Pucker 2021), companies continue to deal with multiple voluntary sustainability standards and legally binding disclosure requirements. In the EU, a survey found 198 certification schemes for sustainable agriculture at the farm-level (Chever *et al.* 2022). More recently, the OECD also mapped the growing number of reporting initiatives to describe a "fast and furious" trend on the demand and supply of quantified environmental impact information in food systems, taking a supply chains perspective (2023). At the global level, noteworthy frameworks include the Carbon Disclosure Project (CDP), the Climate Disclosure Standard Board, the Global Reporting Initiative, the Value Reporting Foundation, the Sustainability Accounting Standards Board, the Task Force on Climate-related Financial Disclosure and the International Integrated Reporting Council.

With different approaches, audiences and materiality, these frameworks guide companies to disclose key information to investors, executives, consumers and the public about their environmental performance, financial information, social impact and governance. Companies frequently face legal obligations to adhere to sustainability disclosure requirements at both national and international levels (UNEP 2015). While there are many other frameworks that are voluntary, many businesses are using environmental labels, entailing some level of data disclosure, to support and increase the credibility of their impact claims. The sources of these obligations can vary from trade or financial policies, industrial and sanitary regulations or environmental laws.

In this context, governments are also driving the development of monitoring systems through regulatory tools by establishing mandatory requirements to electronically register, process and share key information among supply chain businesses, authorities and consumers (OECD 2023). Under the EU Green Deal legislation, specifically the EU Strategy for Sustainable Textiles and Ecodesign within the European Commission's Ecodesign for Sustainable Products Regulation, the DPP will be fully implemented by 2030 to provide clear, structured and accessible information on the environmental sustainability characteristics of products. Another compliance requirement soon applicable to European markets relates to the ban on deforestation, which imposes due diligence obligations on companies who place, make available or export beef to disclose key data, such as the geolocation of all supply chain operations, that can be later verified (EU 2022). The EU also recently passed regulations on European Sustainability Reporting Standards and is assesing a Green Claims Directive proposal (European Commission 2023; European Union, Directorate-General for Environment 2023). While drafts of the reporting standards for this Directive are still under discussion, the overall framework creates stringent sustainability reporting requirements on large firms as well as on publicly traded SMEs.

The result is an increasingly complex landscape of governance made up of often overlapping measures without a shared policy language. A poorly aligned governance landscape leaves companies with compliance burdens that require complex data collection, inefficiency for having to disclose information for multiple reporting systems and risk non-adherence that leads to legal consequences, financial penalties, and reputational damage.

For smaller businesses, comprehending relevant policies can also pose a significant challenge (UNEP 2013, p.17). To adhere to MRV systems, companies need a thorough understanding of several aspects: which policies apply to them based on their industry and location, the nature of these policies (such as taxes, subsidies, or license bans), the specific actions or inactions required, deadlines for compliance or mandatory periodic verifications, monitoring methods for implementation, penalties for non-compliance or repeat offenses, and the necessary documentation for proving compliance.

companies need to make **significant investments** in their data ecosystems

To comply with voluntary and legally binding requirements, it is increasingly important for companies to have strong monitoring systems with accurate and verifiable information. Deforestation, for example, poses a massive financial and compliance risk to companies, as lack of traceability of the origins of products risks companies being associated with negative environmental impact, harming their reputation, risking investment withdrawal, or triggering legal consequences (Erling *et al.* 2022). For the year 2022, 60% of companies disclosing through CDP reported some level of forest-related risks, with the potential financial impact of these risks averaging US$330 million per disclosing company. In comparison, the average projected costs of mitigating these risks were just US$17 million per company (2023b).

To respond to all these monitoring needs, companies need to make significant investments in their data ecosystems, to strengthen data collection, analysis, integration, and report preparation, not to mention the costs of implementing new processes, including staff training (UNEP 2013).

2.2.2 Data requirements and challenges for monitoring deforestation from agrifood sector

One of the main challenges to achieve these monitoring goals is to identify all companies involved in the value chain, including their size, level of intervention and location. As recently demonstrated by the COVID-19 pandemic with many businesses struggling to respond to global supply chain disruptions due to limited visibility into their value chains. Few companies have sufficient information about their business relationships and upstream suppliers (Schrage 2020). While this has started to change, a survey revealed that little more than half of the respondents have value chain visibility systems in place, with 45 per cent still reporting limited visibility (Alicke *et al.* 2022). In fact, only 16 per cent reported having a good view of third-tier suppliers.

The challenge of achieving end-to-end traceability in agrifood, among other reasons, is due to the complexity of the value chains and the need to access information from a large number of individuals and companies that are not in direct business relationships. Smaller companies face resource constraints to develop monitoring tools but may benefit from direct relationships with suppliers since the identification of first-tier suppliers is simpler. In contrast, larger companies can invest in monitoring but struggle with complex structures and higher scrutiny. However, regardless of size, agrifood value chains in general face a big challenge to identify indirect suppliers up to farm location. In the best-case scenario, this can be verified with public registries of rural land tenure or a national inventory of property.

To tackle this, larger companies are developing in-house software for value chain management (Alicke *et al.* 2022) while others outsource this needed to specialized companies that monitor global value chains, using artificial intelligence (AI) and machine learning to analyse big data. Conversely, individual players and Small and Medium-sized Enterprises (SMEs) have limited capacity and accessibility to the type of digital infrastructure needed for supporting decision-making as well as for providing the necessary data to other players within their value chains (Winter *et al.* 2023). In fact, they can be regarded as bottlenecks - influencing the environmental performance of an entire value chain due to their lack of information, infrastructure, and expertise (Ramakrishnan *et al.* 2017). For example, a farmer might not have the resources or incentives to invest in IoT devices to start collecting key data on its operations. In many cases, they do not know the value of this data and do not have platforms to share it. Additionally, a lack of capacity around data means that small businesses risk marginalization in markets due to the data-processing challenge of contributing to value chain sustainability schemes (Lambin *et al.* 2018).

It should be noted that there is an increasing availability of calculation tools and platforms for environmental impact information sharing (OECD 2023), like the Cool-Farm tool and the Partnership for Carbon Transparency, which significantly enhances the capacity of agrifood companies to estimate their carbon footprints and engage in Scope 3 reporting, thereby creating infrastructure solutions that could help address some of the key challenges in monitoring deforestation. However, these would need to be widely utilized by the relevant stakeholders, furthering the challenge of capacity among the relevant stakeholders.

Apart from this overarching issue of visibility and unbalanced capacities between key players, data accessibility and ensuring the quality of information are other major hurdles. Regardless of their monitoring goal, companies would be required to have information on their operations and correlate it with environmental data.

As explained in the previous use case, environmental data is becoming increasingly available despite the persistent challenges. In that context, it is possible for players of different sizes to rely on those data sets to obtain climate data, biodiversity metrics, water supply information and other environmental metrics. However, to integrate them into their information systems and build digital infrastructural solutions, questions of discoverability, interoperability and data integration would need to be addressed. Companies already face significant technical difficulties to integrate any monitoring data or software into existing information and communication systems (Schroeder and Lodemann 2021; May 2019). Integrating environmental data ecosystems would only exacerbate those difficulties.

AFI and CDP reported that from a total of **675** reporting companies, **38%** accept not having origin information for at least half of their commodity volumes.

Lacking a robust DPI for origin traceability and the resources to consistently access high-resolution private satellite imagery makes gathering necessary information quite challenging. Tracking forest loss, for example, would require sourcing geospatial data from public platforms like Earth Map, Global Forest Watch, the Global Agricultural Drought Monitoring and Forecasting System, or NASA's Earth Observing System Data and Information System (EOSDIS). This data must then be manually compared to the estimated locations of value chain operations. Such manual comparison and estimation are necessary due to the absence of publicly available data on property ownership throughout the value chain. AFI and CDP reported that from a total of 675 reporting companies, 38 per cent accept not having origin information for at least half of their commodity volumes (2022).

Moreover, those stakeholders accessing open satellite imagery, without the automatization that specialized services offer, would require a constant manual update of images to perform some degree of monitoring, which would make it virtually impossible for those without the financial resources and institutional capacity.

Furthermore, it is necessary for value chain players to rely on complete, accurate and trustworthy information. From an internal perspective, it is important to have good quality data to make informed decisions; while from an external perspective, it can be critical to avoid reputational, legal and financial risks. There have been cases in which financial authorities initiate investigations against companies on suspicion of investment fraud over allegations of "greenwashing" in sales statements concerning environmental protection and sustainability aspects of investments (Erling *et al.* 2022). Despite these risks, a study by the European Commission revealed that over half (53.3 per cent) of the environmental claims reviewed in the EU were determined to be unclear, misleading, or unfounded, with 40 per cent lacking substantiation (2020).

Finally, it must be emphasized that few mid-tier companies and SMEs have the institutional capital to build a well-designed digital system to monitor their value chains. Without vast resources to invest in the technology, access private services or engage with the rest of the value chain, smallholders face significant competitive disadvantages in this information asymmetry (Montgomery 2022).

Even without deliberate efforts to restrict SMEs' access to essential infrastructure, this situation leads to market power imbalances. A few dominant players end up controlling the digital infrastructure for value chain information flows. For businesses where control over data infrastructure is core to their business model, data privacy is also a competitive strategy. In many value chains, businesses with a high market share over a specific stage of a value chain control the key infrastructures for data generation, transfer, and use (Fisher and Streinz 2021). While this data would be highly valuable for decision-making in environmental sustainability, control over data is a core part of business models and so openly revealing data is perceived as a risk to competitive advantage.

> A few dominant players end up **controlling the digital infrastructure** for value chain information flows

For some experts, when control over the means of data production is unevenly distributed it exacerbates data inequality because those in control also decide which other actors can access and use it; at the same time, it creates significant risks for society as a whole, namely when the interests of those in control do not necessarily align with societal interests (Fisher and Streinz 2022, p. 844-847). This imbalance can lead to a range of repercussions, from granting undue influence to select economic entities, altering competition and innovation dynamics in the market, to potentially facilitating censorship of crucial public information.

2.2.3 Observations on value chain data ecosystem

Companies of every size still struggle to exercise complete visibility over their value chains, given the complexity of corporate structures today. Some companies have developed their own tools and practices to monitor their value chains, but companies without the resources to invest or hire private monitoring systems are left out due to a lack of open-source software or open data sharing in value chains. In turn, this deepens power imbalances and inequality, and harms effective competition between the different players in a market because few players end up in control over data infrastructure for value chain information flows.

Opacity in this domain is so prominent that even technology developers have developed business models in which features for protecting strict confidentiality and privacy to incentivize the use of their tools (Visipec 2021). In this context, apart from reporting frameworks and due diligence obligations, little to no information on value chains is shared openly.

Furthermore, public tracking of value chains relies on limited information disclosed by companies due to legal requirements, policies or voluntary initiatives, often resulting in estimations and educated guesses (Trase 2018). Unfortunately, this often generates distrust in the results of the monitoring by different stakeholders as well as risks of greenwashing. In turn, this also hinders public monitoring since it is challenging to make comparisons, which creates risks of misinterpretation of disclosed information. Fragmentation of data ecosystems also arises as companies use different methodologies, standards and definitions, leading to distrust and confusion among stakeholders. To address this diversity, tailored private solutions like add-on tools for internal monitoring systems have been introduced, catering to the company's methodology and standards (Visipec 2021).

Specific data requirements will increase over time as reporting and disclosure frameworks require corporates or financial institutions to report and disclose their full set of material impacts and dependencies (TNFD 2023). But, establishing monitoring systems involves hidden costs, often unevenly distributed, including those associated with limited institutional capacity for data collection, analysis and uptake. Without obligations nor support mechanisms to generate those digital systems and infrastructure, there are no incentives for value chain data sharing and transparency promotion; each company has complete freedom to decide what mechanisms to establish to monitor its value chain. A traditional business perspective favours the restriction of data sharing for competition reasons (GPP 2022). But the lack of appropriate monitoring tools exposes all companies to financial, operational, legal, reputational and environmental risks. This also hinders the connecting of this data ecosystem to environmental data and policies on green and circular economy.

2.3 Transparency of green and circular economy policies

Policymakers are confronted with a significant information challenge when formulating and implementing public policies, particularly in the realm of green and circular economy policy (GCEP). The effectiveness of many of these policies hinges on the availability and analysis of accurate information, such as on the environmental impact of different economic activities across their life cycles and value chains. This is on top of the need for monitoring the implementation of public policies and their impacts to help ensure government accountability and facilitate continual improvement. "Transparency" in this regard is not only about publishing public policies, but about all information that allows for assessing the effectiveness of policy decisions and the accountability of governments.

In the context of GCEP, the aim is for policies to shift the incentives at each stage of the life cycle of a particular economic activity, from resource extraction to end-of-life, to minimize environmental impact and maximize resource efficiency. This needs to be viewed from the perspective of global value chains - from conception to end-of-life, the interconnected, international network of activities, entities, and processes, which contribute to producing, distributing and managing these goods and services across different markets and regions. National policy impact over the global value chains is intricate, contingent on how the targeted actors are embedded within production networks (OECD 2019).

Building on this intricacy, the information challenge persists for decision-makers, particularly as the GCEP cycle necessitates a comprehensive understanding of the environmental sustainability effects of value chains to be targeted by the sets of policies. The multifaceted dynamics within global value chains, which involve diverse actors from different regions and span products' life cycle, demand robust data collection and analysis frameworks, especially when aiming for accurate sustainability assessments. The cross-border nature of these chains further complicates the information landscape, requiring nuanced policy evaluation and stakeholder engagement to ensure that economic activities of all stages of the life cycle align with environmental sustainability goals.

Having explored challenges for stakeholders through use cases of identifying hotspots of deforestation, and the connection between deforestation and the agrifood value chain, the following use case will cover information problems around GCEP to tackle the issue of deforestation.

2.3.1 Information requirements for green and circular economy policies

At the UNFCCC COP26 World Leaders Summit "Action on Forests and Land Use", over 130 leaders, representing more than 90 per cent of the world's forests, committed to work together to halt and reverse forest loss and land degradation by 2030 (UNFCCC 2022a). Governments of production, trading and consuming markets agreed to collaborate to tackle deforestation hotspots through public policies that influence the agriculture value chain.

On the production side, it is important to strengthen environmental law enforcement and to redesign agricultural policies to deal with the challenge of sustainability transition. On the importing side, countries must facilitate trade policies that do not drive deforestation and land degradation (UNFCCC 2022a). In practice, access to information in the agriculture value chain and the capacity to analyse that information is highly relevant to the design and implementation of effective public policies.

Taking the importing market as an example, the policy formulation can be built on the distinction of the product being deforestation-associated, deforestation-free, or even sustainably managed. However, this necessitates the collection and verification of product information far beyond the physical characteristics relating to its immediate function.

There are two information approaches. One is delegating the work of excluding the deforestation-associated product from the value chain or recognizing sustainability efforts through a third party, such as a voluntary standard body or certification schemes. For example, Switzerland relies on four globally recognized and third party-verified certification schemes as a condition of preferential market access to palm oil from Indonesia (Fedlex 2021). The other approach is directly requesting data from operators or other relevant value chain players based on a defining matrix, along with means of verification by the authority. The EU law to fight global deforestation uses this second approach (EUR-Lex 2023) (See Figure 4).

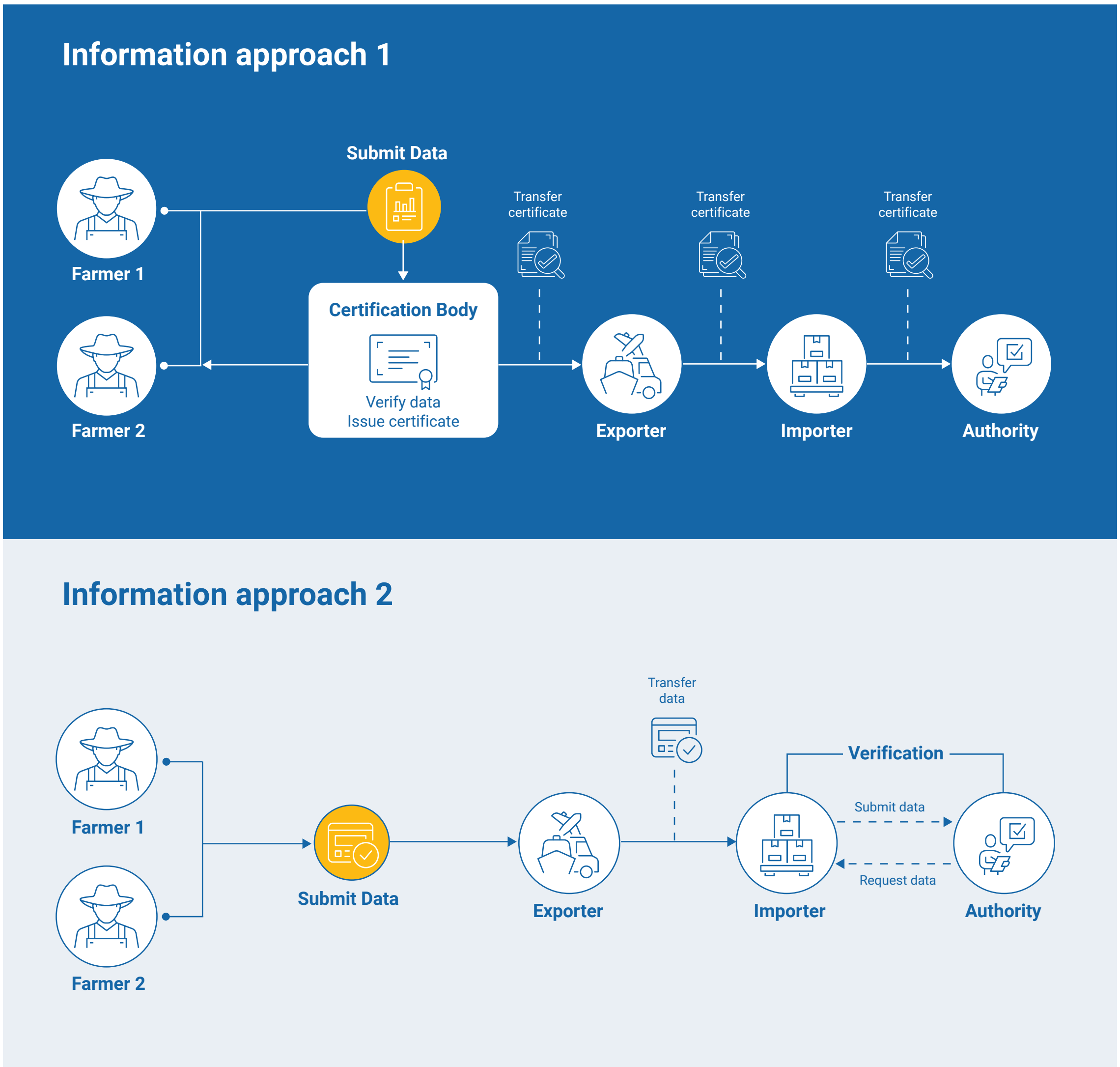


Figure 4: **Information approaches under GCEPs**

Among other things, the EU law includes a ban on imports and domestic use of deforestation-related products, requirements of due diligence and compliance of law and regulations of production countries. The bill not only covers beef products but also downstream products including leather, chocolate, and furniture. And it is not only concerned with cattle raised on deforested land but also cattle that have consumed feed from deforested areas.

The law requires more information than is typically required in a conventional trade setting. To fulfil the due diligence statement, operators are required to conduct at least an annual risk assessment, which includes an understanding of the deforestation status of the commodity, reliability of suppliers' information, including whether indirect ones are verifiable, risk associated with country governance, and to set up policies and an internal structure to mitigate the risk.

Operators are required to be able to identify the geolocation-based origins of specific amounts of cattle and cattle products, and the ownership of that product, over space and time. Upon request, operators need to submit information to allow the authority to have the same visibility and verify claims.

The authority can, for example, use satellite monitoring tools, spot checks including field audits or DNA analysis to check where products come from (European Commission 2023, art. 18 (2)). EU authorities would have access to relevant information, such as geographic coordinates. Anonymised data would be available to the public. The authorities would conduct annual checks on the operators and traders for products at a frequency of 9 per cent from a high-risk jurisdiction (European Commission 2023, art. 16). The risk levels are to be further defined but would be based on the deforestation status, forest-agriculture land conversion status and production trend of relevant commodities products, as well as the deforestation-related UNFCCC NDC commitment, international cooperation agreements on deforestation and the strengthening of domestic environmental law.

The European Commission estimated that overall costs of due diligence for companies could total between **€158 Mn** and **€2.4 Bn** a year.

The legislation has triggered various debates, at the international level and across different industries (World Trade Organization 2022). One core debate concerns the costs - and who should bear them. The European Commission estimated that overall costs of due diligence for companies could total between €158 million and €2.4 billion a year (European Parliament 2022). Depending on the complexity and risk associated with deforestation in the operator's value chain, setting up the due diligence system would involve one-off payments of between €5,000 and €90,000. Large operators are expected to bear the costs. While there are mitigation provisions for SMEs and small farmers, they would still need to comply with certain information requirements and thus the necessary technical capacity would be needed for compliance.

However, there are opportunities to address the information cost, when viewing the EU law in conjunction with other green and circular economy legislative efforts. The deforestation law and the Ecodesign for Sustainable Products Regulation exhibit a systemic interplay in their information approaches that could facilitate compliance if the digital infrastructure and related incentives are well integrated. The latter regulation, introducing the DPP, aims to enhance product circularity, energy efficiency and other environmental sustainability facets by setting ecodesign requirements for various product groups. The DPP, accessible via a data scan, provides crucial product information like durability, reparability, recycled content, and spare parts availability, aiding informed purchasing decisions, facilitating repairs and recycling, and enhancing product life cycle environmental impact transparency.

Although this report primarily focuses on agriculture and food use cases, both laws help illustrate the centrality of information requirements for GCEP as they intersect significantly on downstream products like leather and furniture. The deforestation law's enforcement exemplifies raw material sourcing information requirements, aligning with the DPP's data capture across the value chain. This synergy underscores cross-agency legislative collaboration and enforcement, offering options of cost-sharing with downstream business players along the product life cycle and efficiency benefits for businesses, especially if the DPP is integrated into the deforestation law enforcement. Access to the DPP also incentivizes business compliance with the deforestation bill, opening broader market opportunities.

2.3.2 Observations on the data ecosystems of green and circular economy policies

Digital technology has impacted most aspects of the public policy process. Social media has influenced agenda-setting, allowing political actors to increase the salience of certain issues and shape their framing. In policy formulation and adoption stages, digital technology has simplified crowdsourcing ideas and engaged citizens through various "civic tech" tools (Gilardi 2022). During implementation and evaluation stages, online publication, open government data and information exchange through global governance frameworks have enriched public data on national policies. This has enabled the growth of policy trackers, databases, and compliance technologies to support business players. It should be noted that other groups, such as civil society organizations, also need to understand policies and have access to this data to engage in civil processes for policy change.

However, the data ecosystem of this domain suffers capacity challenges in data processing and analysis. Scholars and analysts differ in the collection, categorization and evaluation of public policy, and the level of detailed policy information that should form part of the data sets. The collection and categorization of data related to circular economy policies, such as recycling rates, product lifespans and waste generation, can be varied and complex due to the multifaceted nature of circular activities. This creates difficulties in assessing policy effectiveness systemically across cases due to the cross-region and cross-sector nature of those policies, which in turn can limit implementation and compliance. There is often insufficient understanding of the temporal dynamics of policy change, how and why specific policies work (or not) and how policy choices interact in an increasingly complex mix of policies (UNEP 2019, p. 284). Without fully mapping and understanding those interactions, policymakers cannot benefit from aligning policy measures to improve the design of effective new policies that fit within existing regimes and reduce the administrative burden of implementation.[7]

Attributing environmental outcomes to specific policies is also challenging, as most environmental policies do not have one-to-one correlations with their outcomes (UNEP 2019, p. 288). Measuring the policy outputs alone (e.g. adoption of policy instruments) would not adequately capture the preferences of different countries for one or another instrument. For example, one country may impose recycling rate requirements, while another country might implement product-as-service models, and a third country uses product design related standards. In each of these cases, the expected impact will be reducing negative environmental impact and improving resource efficiency. The indicator is influenced, however, by the industrial structure, natural conditions, level of income and other factors that are not, or not directly, impacted by (environmental) policies (UNEP 2019, p. 279). Therefore, the evaluation of policy effectiveness often comes down to expert judgment, as there is no commonly agreed approach to assessing effectiveness (UNEP 2019, p. 284). As a result, it is challenging to establish a feedback loop to support the policy cycle.

7. This could be instrumental for several different policy-making processes. In Preferential Trade Agreements, where parties spend months, if not years, negotiating the language and technicalities of environmental regulations and standards (Bellman and van der Ven 2020). Additionally, integrating environmental considerations into national policy would ideally involve the development of complex regimes where environmental policy is built across government bodies operating in different domains (Nunan, Campbell and Foster 2012, p. 263-265).

The introduction of automation tools such as computer-based models have been useful for policy analysis and effectiveness. However, noting that "models can also be instrumentalized to justify already decided policies and targets" (Süsser *et al.* 2021), this technology development would not be able to solve methodology flaws in policy evaluation.

The EU deforestation-free bill, which combines command-and-control and economic incentives by setting a trade ban on deforestation-related products, offers an example that may bypass the effectiveness challenges of attributing environmental outcomes to specific policies. It also addresses the problems of diversified measurement frameworks on business operations without mandating a single standard. Value chain operators are given the flexibility to decide how to organize the information to achieve compliance. This policy innovation would not be possible without enhanced public monitoring capacity on the environment and technology readiness in value chain traceability.

If viewed in a siloed manner, this approach could trigger information costs in addition to the direct cost of transforming value chain operation practices. Further concerns include how this cost will be distributed among economic players and if it will further enhance the concentration of power for big players in relevant value chains. Systemic approaches such as the Ecodesign Sustainable Product Regulation include the requirements of DPPs to provide opportunities to address some of the concerns by allowing for information cost saving and generating value from product sustainability information, which is a precondition for accessing green and circular economy markets.

2.4 Addressing challenges

Although the challenges in environmental sustainability decision-making are recognized (see Table 2), domain experts find it difficult to devise a strategic plan to address these issues, often resulting in only broad, general recommendations for solutions (GPP 2022). In the domain of environmental monitoring, experts emphasize the necessity for digital public infrastructure for support and technical coordination. This will foster a more certain and cost-effective environment for innovation and pilot testing.

Use case challenge	Data problem
Measuring deforestation driven by the beef value chain	Inaccessibility, varying quality of high-resolution data, and undervaluation of data
	Existing data can be hard to find (discoverability)
Monitor beef value chain's economic activities despite company data opacity	Disconnected data ecosystems
Facilitate beef value chain companies compliance with GCEPs, addressing small players' policy comprehension gaps	Difficulties in querying interrelated policy information in natural language (data uptake).
Overlapping measures and MRV frameworks	Process natural language information complexities
Informed policy design for GCEPs	Insufficient access to relevant data and challenges for data collection.
	Existing data can be hard to find (discoverability)
Data's lack of inclusiveness	Insufficient data collection and generation Development of tools without a user focus

Table 2: **Summary table of identified data challenges**

When it comes to value chain traceability, experts underline the importance of having effective incentives, strong safeguards and finding common ground through standardization; all to overcome the hurdles of data sharing.

In the GCEP domain, experts stress the need to establish information mechanisms to aid the evaluation of the effectiveness impact of public sustainability policies. However, there has been limited consensus on the design of such mechanisms.

A recurring theme across these recommendations use cases is the call for systemic support and incentives to enhance data collection, sharing and uptake within and across these three use cases.

These persistent information challenges unequivocally result in limited information exchanges, leading to a landscape of disconnected and siloed information systems. These issues are also associated with lack of open data systems and interoperability.

Underlying this complicated context of information challenges, one issue stands out. There is an absence of adequate DPI to facilitate the flow of information through exchange mechanisms. To overcome these systemic information challenges for environmental sustainability, enabling DPI as a data exchange system is crucial. The subsequent section of the report delves into six key technology innovations (TI) that can enable DPI for environmental sustainability decision-making.

3 Technology innovations on the rise

In this section, the report presents a series of promising technology innovations with the potential to transform decision-making by enabling DPI for environmental sustainability. The six Technology Innovations match emerging solutions with the data challenges as presented in Section 2.

Specifically, this section focuses on technology innovations that aim to address the information challenges highlighted, with some targeting specific challenges more directly than others. However, it is important to recognize that these innovations possess a broader transformative potential when they operate collaboratively, functioning as an integrated infrastructure system. Together, they enable DPI for environmental sustainability as a data exchange system dedicated to sustainability-related information.

As abovementioned, DPI acts as a critical intermediary layer within a digital ecosystem, bridging the gap between physical infrastructure elements (like connectivity, devices, servers, data centres, and routers) and the application layers (See Figure 1). DPI's role in environmental sustainability is pivotal; it streamlines the exchange of sustainability data, seamlessly integrating this information into a wide array of applications. Such integration paves the way for innovative uses, including personal carbon tracking, DPPs, and automated environmental compliance systems, among others. The technology architecture of DPI should facilitate inclusive, flexible digital ecosystems, with individual solutions built according to principles of interoperability, scalability, modularity, and security (India's G20 Presidency and UNDP 2023).

In this context, this report advocates for harnessing DPIs for environmental sustainability decision-making, as a way to enhance transparency and accountability through enabling data exchanges, thereby improving data availability. The proposed digital system would create incentives and mechanisms for more efficient data generation and collection, ease the discovery of data sources, and reduce barriers to data sharing. The result would be an enhanced flow of information, which balances the demand and supply in the data value chain, leading to a self-reinforcing process that enhances data quality.

Achieving this goal requires more than just utilizing digital technologies for sustainability; it is crucial to actively influence the development processes of these technologies and systems. Involving a diverse range of stakeholders in the development process is essential, including close collaborations with the data science community and technology developers to ensure that digital systems and technology solutions are explicitly designed for environmental sustainability purposes.

> A DPI for environmental sustainability will **empower stakeholders** through enabling informed decision-making and promoting greater inclusivity.

A DPI for environmental sustainability will empower stakeholders through enabling informed decision-making and promoting greater inclusivity. Furthermore, this digital public infrastructure serves as a critical foundation for transformative applications, such as Digital Public Passports. Ultimately, the vast data requirements of SCP, green, and circular economies can be met more effectively through such a comprehensive system.

Additionally, this would set a foundational standard for cross-sectoral applications, upon which sector-specific solutions can be developed to meet particular needs and ensure a baseline for inclusive solutions. While the public nature of this infrastructure guarantees a basic level of inclusivity and functionality, private solutions can further enhance and build upon this foundation. Policymakers are advised to consider the governance implications of integrating private solutions into DPI frameworks.

In turn, sole reliance on private solutions will likely fail to comprehensively address these challenges and may result in further data fragmentation. Therefore, a blend of private and public solutions is essential. In the ideal scenario, private solutions could grow on top of DPI and in turn, private solutions could be building blocks for DPI.

What is each technology innovation needed to enable DPI for environmental sustainability?

Each of the six TIs identified in this section (See Figure 5) responds to "specific challenges" drawn from the use cases in Section 2, which is then connected to an identified emerging solution. The TI identifies the opportunity presented by a technology group, includes best practice examples, and highlight current limitations. Each technology innovation aims to highlight a potential use case and jump-start innovation on DPI for environmental sustainability.

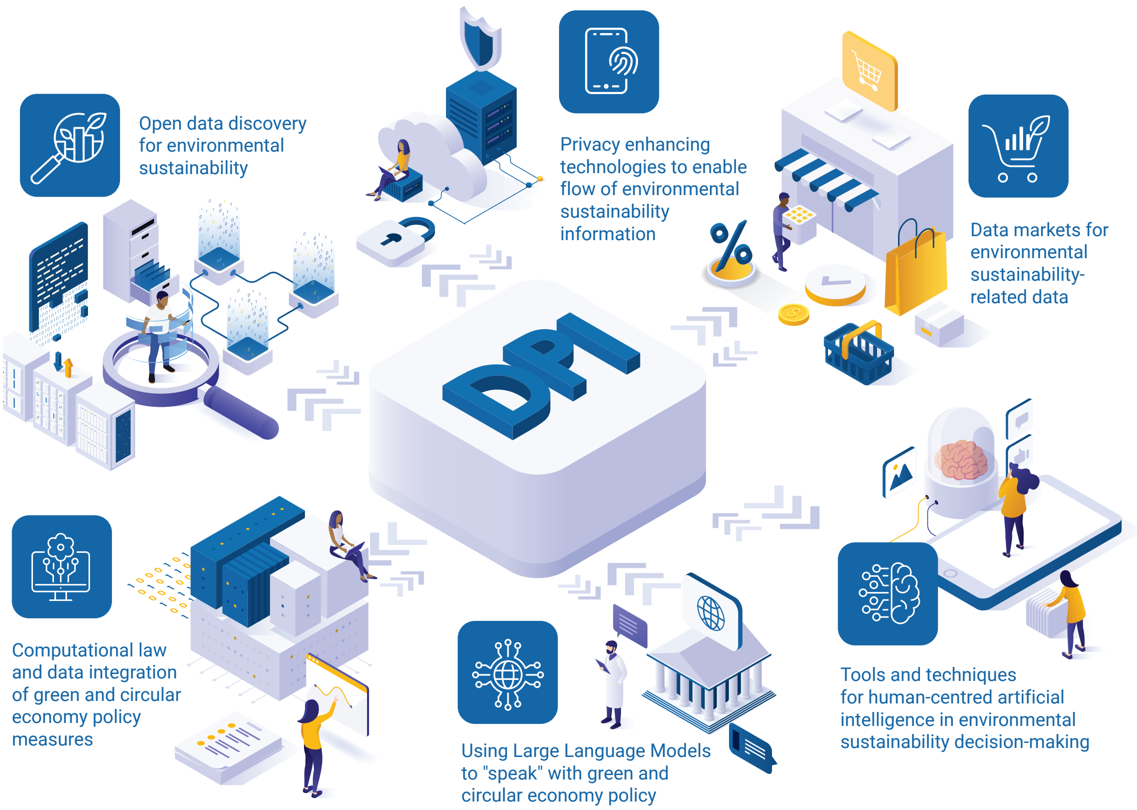


Figure 5: **Technology innovations for a DPI for environmental sustainability**

3.1 TI1: Open data discovery for environmental sustainability

Open data is vital for driving innovation and transparency in environmental sustainability. In DPI, open data ecosystems and (closed) data marketplaces fill distinct and complementary roles. On the one hand, open data takes the approach of being a digital public good: providing an essential base level of information for environmental sustainability analysis. On the other hand, data marketplaces can meet demand for more specific and timely data or larger volumes of data often combined with analysis. Open data approaches tend to favour the FAIR principles: findable, accessible, interoperable, reusable (Wilkinson *et al.* 2016). The aim is that open data "can be freely used, modified, and shared by anyone for any purpose".

However, many individuals and organizations report challenges in finding and unlocking the value of open data for sustainability. As part of DPI, solutions for data set search can bridge the gap between open data and data users. A primary challenge is data set discovery and retrieval: both finding the correct online data repository, then efficiently navigating the repository to locate the correct data sets. While discovery is also challenging in closed data markets, the vast number of repositories and volume of open data makes discovery an acute challenge. In the computer science field, research and development has focused on improving search and indexing techniques, and these advances must be adapted to fit environmental sustainability-related data set search.

Specific challenge

Data is a vital input for the three use cases covered in section 2 including for environmental monitoring, establishing multi-tier transparency in value chains, or evaluating the outcomes of public policy. Despite this, much of the value of open data for sustainability analysis goes unrealized due to discoverability challenges. The data needed for analysis crosses jurisdictions and data types, meaning that data is fragmented across many repositories, which often exist in the "long tail" of the Internet.[8,9] Data discovery is both time-consuming and requires analysts that understand the landscape of repositories in their domain. For scalability, better techniques are needed for data set retrieval.

A core feature of this challenge are the ways data is described and the domain knowledge of users when searching for data. The descriptions for data sets, known as metadata, include critical information on the data set, such as: temporal, spatial and subject coverage; keywords describing the data; the format of data; and the organization that collected the data. Many repositories make data accessible only through keyword search, which leads to possibilities of metadata being missing or inaccurate, or situations where users do not know the right keywords to express their data needs.

Technology trend

DPI is needed to improve the availability and discovery of open data sets, therefore enabling uptake. In thematic expert consultations, some recommended that international organizations or governments take a central role in storing and managing data sets for environmental sustainability. However, while centralized data lake deployment could improve accessibility and discoverability by providing a single, clear location for search, in practice this is a mammoth task, especially when considering the broad range of data sets that could be useful for environmental sustainability decision-making. In the near term, infrastructural solutions that promote a federated but decentralised data ecosystem combined with improved data set search can offer robust solutions to this challenge.

Data set search techniques fall into two main categories: metadata search and data-driven data set search. Currently, metadata search is most common, with the main advantage being that searching keywords in metadata is far simpler and more efficient than searching across different data types, including tabular data, images, or geospatial data (Chapman *et al.* 2020). The release of Google "Dataset Search" beta version in 2018 improved the reach of data set search solutions through an Internet crawler that collects and indexes Schema.org metadata on data sets, and allows a userfriendly search function (Brickley, Burgess and Noy 2019). This in turn incentivized data publishers to describe data using schema.org markup, causing the number of data sets described to increase from 500,000 to 30 million between 2016 and 2020 (Brickley, Burgess and Noy 2020).

8. The "long tail" of the Internet is used to describe resources that are only available following a specific query on a repository, meaning that data will not be picked up by an Internet crawler, for example.
9. TRASE for instance uses a Spatially Explicit Information on Production to Consumption Systems (SEI-PCS) approach that systematically links individual value chain actors with subnational production regions and associated environmental risks. Data sources include national and subnational export or production data where available, and the use of independent data sets including the logistics of producing companies, production and taxation (Trase 2020a).

However, metadata may not encompass the information a user needs about a data set. Moreover, a user may lack the specific language needed to describe the specific data set they require (Chapman *et al.* 2020). In most cases, as with Schema.org, the schema used for metadata are domain-agnostic, and do not adequately meet the needs of data users in specific domains. In biodiversity data, key interests for biodiversity researchers - including ENVIRONMENT, MATERIAL, ORGANISM, PROCESS and LOCATION - are not typically included in metadata and thus only appear in metadata searches where they are included as keywords (Löeffler *et al.* 2021). In environmental sustainability, data set search depends on the specific subjects and qualities of a value chain or policy field, and thus similar work is needed to establish domain-specific interests and how they can be included in metadata. From a technological perspective, researchers have proposed the automated generation or enrichment of metadata from the contents of data sets using AI (Suhara *et al.* 2022).

Data-driven data set discovery is useful when the user needs to discover data sets that can be joined together to augment the value of a data set and the insights that can be generated. Consider a use case where the owner of a value chain data set - including data on locations, product verticals and operations - wishes to integrate data on land-use change or pollution in the same geolocations. This approach is called constructive data set search, and it is often needed where answers to questions lie across two or more databases. However, it is technically challenging due to differences in metadata and data schema across the data sets (Chapman *et al.* 2020). Statistical and AI-based approaches can assess the "unionability" of data sets by measuring the likelihood that two attributes contain values from the same domain (Nargesian *et al.* 2018). Unionability techniques have laid a foundation for "Goal-Oriented Data Discovery", which could lead to automated discovery and joining of tables based on a user query.

Key remarks

The open data ecosystem is providing a growing wealth of open data that is valuable for environmental sustainability decision-making. However, for better uptake of data, DPI is needed to help users discover and access data sets based on their application. Metadata is the core ingredient; a vital next step is to build standardized metadata vocabularies and semantic taxonomies for the environmental sustainability domain which can act as enablers for powerful data discovery tools. While assistive tools such as automated metadata generation can be deployed to support metadata coverage, human oversight and validation will still be needed, especially for high-quality application of taxonomies in the form of data annotation. To address scenarios where keyword search is ineffective, a range of approaches is also needed to improve data discovery. This includes mapping of data connections to enable data-driven data set search, and the restructuring of data repositories to allow users to explore and find relevant data sets when the right keywords are challenging to express (Oullette *et al.* 2021).

3.2 TI2: Privacy enhancing technologies to enable flow of environmental sustainability information

Data-driven decision-making for environmental sustainability depends on the flow of environmental data, value chain data and policy information from a wide range of stakeholders. Infrastructural solutions are needed to allow the use of data for sustainability-related purposes while maintaining the privacy of data subjects and data sharers. Opportunities are provided by data architectures that improve security in data sharing, such as homomorphic encryption or differential privacy, or federated learning techniques that train models without needing data transfer.

Specific challenge

In today's economy where "data is the new oil", it is challenging to facilitate flows of data between institutions while upholding data sovereignty and protecting the right to privacy for both data-sharing institutions and the subjects of data.

Accessing value chain data presents a particular challenge in understanding how resource dependencies and value chain activities connect with environmental impacts and responsibilities. For example, an analysis of the environmental impact of beef must look at the role of meat-processing companies and retailers as well as cattle ranches, which requires value chain information. Companies with capacity may collect the data needed for internal supply chain visibility but are reluctant to share data openly due to the privacy concerns of suppliers that are the subjects of data, reputational risks and the lack of trust in data systems.

While ensuring data flow from value chain actors with high control over data infrastructure is crucial, granular understandings of value chains also depend on data contributions from smaller businesses, which have their own privacy concerns. Privacy concerns may include questions of "who" will be able to access data once shared, and "how" will it be used.

The data privacy and sovereignty concerns of all stakeholders, including Indigenous groups, local communities or small businesses, must also be addressed. For example, attention should be paid to frameworks such as the CARE principles, which uphold that Indigenous groups should be able to control the content of data that is collected and shared by or about Indigenous groups (Carroll *et al.* 2021, p. 108).

Technology trend

The field of computer science offers a wide range of solutions for privacy across different use cases. Privacy enhancing technologies (PETs) are data architectures that manage or modify data to enhance privacy while retaining as much utility as possible (Garrido *et al.* 2022). This is a broad group of technologies, and the choice of a PET solution will differ according to the use case, the level of privacy concerns around data sets and the capacity of stakeholders. PETs must work in tandem with governance and policy: data privacy is a subject of many evolving regulations designed to protect individuals, while mandatory data reporting is a component of many public policies or certification schemes.

A key group of PETs aim to improve access to potentially sensitive data sets by masking confidential information while sharing data. Anonymized data sets, with key identifying information such as names or addresses removed, are still at risk of privacy breach where data is insufficiently anonymized. Differential privacy is an emerging and widely-discussed technology designed to deal with this challenge. These techniques add "random noise" to statistics, which obscure the details of individual data points while retaining patterns in data for accurate queries across data sets (Dyda *et al.* 2021). When looking at the output data sets, differential privacy means it is impossible to tell if one individual's data was in the original data set or not (Harvard 2023). This has shown value in some use cases, but is dependent on the data: for example, where longitudinal data is needed to observe changes in value chains or businesses over time, the level of "noise" must be increased to prevent the identification of patterns, which reduces the usefulness of data. Well-known applications of differential privacy, such as in the 2020 US Census, have attracted criticism for reducing data value (Domingo-Ferrer, Sánchez and Blanco-Justicia 2021).

An alternative PET is homomorphic encryption, which enables data sharing without depending on trust. Under conventional data encryption, data is vulnerable following decryption by the end user, who could then share data openly (Acar *et al.* 2018). The innovation of homomorphic encryption is that end users can conduct analyses on data without needing to decrypt the data, meaning data can remain confidential while it is processed (Hellwig and Huchzermeier 2022). Moreover, data can be encrypted by stakeholders using commercially available software packages, then can be shared to a cloud server afterwards, meaning the data owner never has to expose unencrypted data (Acar *et al.* 2018). Applications analysing encrypted medical data have shown the potential opportunity of this technique, but the scalability is limited, as despite recent advances, analysis of homomorphically encrypted data remains slow, requires large amounts of computational resources, and does not allow ad hoc data queries.

Similar to homomorphic encryption, federated learning techniques are important in mitigating the privacy risks of data transfer for training machine-learning models. Federated learning architectures work by "bringing the code to the data, instead of the data to the code", meaning models can be trained on decentralized data without data ever being shared or transferred by data owners (Bonawitz *et al.* 2019, p.1). Big tech companies have used these techniques to train models using data from mobile devices, where model training is brought to the device to construct a global model, which is then redistributed to devices (Google 2017).[10] Recent advances include swarm learning, which removes the need for a centralized server in federated learning by sharing parameters through the "swarm network" and building models independently on the sites of the data. A Blockchain "smart contract" is used to orchestrate model training among nodes in the swarm network and guard against malign actors. However, incorporating Blockchain decreases speed and efficiency (Warnat-Herrestal *et al.* 2021).

Key remarks

Improving the analysis of privately-held data for sustainable decision-making requires caution, but can unlock valuable insights. Digital technologies can provide safeguards for the use of data that can work in tandem with regulatory measures that maintain data privacy where necessary, and the release of data were useful. The choice of technology depends on the task, the nature of data and the computational resources available. Differential privacy is promising as it could mean sensitive data can be released with small relative errors and strong privacy guarantees. This means that queries can be conducted on large sensitive data sets without breaching individual privacy, though the software is difficult for non-expert users to implement, and the level of noise needed for privacy guarantees may make it impractical for time-series data (Dyda *et al.* 2021). Homomorphic encryption is also promising, but innovation is needed to increase the efficiency of methods while providing robust security guarantees. For training machine-learning models, federated learning techniques will prove valuable for accessing big data resources without requiring data transfer and consolidation, which improves trust. Use cases must be developed to understand how best protect the privacy of the most valuable private data sets needed for environmental sustainability decision-making.

10. In those cases, the process begun when a device retrieved the latest model and enhanced it by learning from the data present on the phone itself. It then condensed the improvements into a compact, focused adjustment. This adjustment is the only thing transmitted back to the cloud, secured through encrypted communication, where it was combined with adjustments from other users to refine the collective model. The original training data stayed exclusively on the user's device, ensuring no individual modifications were retained in the cloud.

3.3 TI3: Data markets for environmental sustainability-related data

Well-functioning data markets allow institutions with heavy data needs to access the data they require, while suppliers of data are incentivized to collect new data or share key data sets they control. In decision-making for environmental sustainability, data marketplaces can play a crucial role by enhancing rewards and incentives for collecting, curating, and exchanging data that drives sustainable actions. They can be connected to DPIs for environmental sustainability, facilitating the supply and demand dynamics of sustainability-related data.

Specific challenge

High demand for environmental sustainability data does not currently translate into sufficient incentives and rewards for data collection and sharing in the data value chain. Data collection is fragmented and inconsistent. Many non-profit organizations depend on short-term grant financing from universities or governments, which leads to problems of continuity and longevity in their data collection efforts. This reliance also results in limited interoperability beyond the specific objectives of a given data set (Biber 2013). A mechanism is needed that encourages and rewards stable, long-term data collection, providing accurate baselines and regular updates for the use of businesses, governments and citizens.

Another challenge is faced by some stakeholders, such as individuals or organizations, who manage their own data over a very limited domain. Examples could include a local community with potential to collect data about a local natural ecosystem, or a small business that could sell elements of business information as a secondary revenue stream. Additionally, the growing abundance of Internet of Things (IoT) devices in value chains and in everyday life means that valuable data is being generated at every moment (Ramachandran, Radhakrishnan and Krishnamachari 2018). However, in most scenarios, small-scale data owners[11] do not realize the potential value of their data, nor do they have the opportunity to share it. Improving the supply of crowdsourced or citizen science data has the potential to improve understanding of real-world conditions of ecosystems, value chains and policy impacts.

For data users, the challenge is to access the needed data, ensure that data is relevant, credible, of sufficient-quality, and that it is provided at a reasonable cost considering the financial means of the data user and the use case.

To overcome these challenges, it is essential to clearly establish the value of data in environmental sustainability and to develop DPI that facilitates the exchange of data value among stakeholders. Ensuring the availability of data is crucial for driving more effective actions in environmental sustainability.

Before starting with the analysis of the technology trends for this TI, it should be noted that data marketplaces have been historically commercial solutions, facilitating transactions but not directly enabling DPI. However, there are a few emerging market signals and government actions that hint at the transformation of data marketplaces into pivotal DPI enablers. Governments, notably China's, are actively working to encourage and regulate the marketization of data, signalling a move towards recognizing data as a strategic asset (Global Times 2022). These examples could indicate an appetite from governments to fund, subsidize, control, or develop data marketplaces. Different operational models would need be explored, however, data markets under public authority stewardship could leverage these platforms to incentivize data collection, enhance data quality, and advance open data as a digital public good.

11. Under the EU GDPR the term "data owner" refers to either individuals or teams who make decisions such as who has the right to access and edit data and how it is used.

This government intervention suggests a future where data marketplaces transcend their commercial origins to become foundational elements of DPI. For now, the incentives for data exchanges are very limited, relying mostly on regulations and policies, meaning that a market for selling and buying data could unlock a whole new model for incentivizing data exchanges. Examining the specific use case in Section 2 sheds light on the potential to dismantle siloed data ecosystems and encourage the sharing of private data within a data marketplace that is publicly financed, subsidized, or government controlled. Such marketplaces would enable the enhanced flow of high-resolution satellite imagery, data from local communities and Indigenous groups, and private supply chain information. This TI takes this premise to advocate for relying on data marketplaces to enable DPI for environmental sustainability.

Technology trends

Data marketplaces can enhance data flow by offering infrastructures for sharing, discovering, and monetizing data. This, in turn, can encourage the generation of more data together with value-added analytics and services. Such marketplaces are therefore a promising technology for bolstering data flow into businesses and applications, thereby facilitating environmental sustainability decision-making.

A crucial question of technology design in data marketplaces is the choice of centralized or decentralized platform architecture. In a centralized approach, data products are offered via a predefined centralized data storage location, such as a cloud repository (van der Ven *et al.* 2021). The advantage of this approach is that data contributions can be combined and packaged into new data products, which adds value to data, though the downside is that a trusted data broker is needed. Decentralized data marketplaces, on the other hand, involve only metadata being shared with a centralized repository, while data is exchanged directly between partners following an agreed transaction using APIs (van der Ven *et al.* 2021). A trend in data marketplace solutions has been the adoption of Blockchain to ensure fair transactions through smart contracts and create an immutable record of historical transactions, removing the need for a trusted source party (Banerjee and Ruj 2018).

Innovation in data marketplaces has also focused on improving security and control over data sets. For data owners, there is a constant threat that a user may misuse data, for example by publishing data openly and thus diminishing the scarcity value of the data set. Technological approaches have provided some protection. A good example is the Snowflake Marketplace, which uses a role-based access control (RBAC) layer to let data owners give users access to data sets directly where they are stored on the cloud rather than through file exchange, which prevents the illegal copy of data after purchase (Snowflake 2022). However, it is practically impossible to fully prevent the misuse of data, as limited data access or anonymized data reduces the value for legitimate users. This makes it challenging to build trust among data owners in open data marketplaces, resulting in a trend towards closed data marketplaces with access limited to trusted partners operating in specific regions or industries.

Finally, the valuation of data and distribution of rewards and incentives for data sharing is an open challenge, despite being an important focus in the data markets literature. There is a problem of information asymmetry: buyers want to be sure that data is useful, while sellers do not want to reveal data sets prior to purchase. This prevents effective valuation of data, as the value of data is not inherent but dependent on how it is used. In machine learning, researchers suggest distributing rewards for data based on a Shapley data value, which quantifies the value of each training datum to the performance of a predictor or model (Ghorbani and Zou 2019). The use of Shapley value has grown in popularity, and in the future, it could allow for experimental architectures such as "data-blind interfaces" where data buyers describe a task they need to accomplish, then data is automatically assigned by the data marketplace solution (Kennedy *et al.* 2022).

These techniques hold promise in tandem with governance approaches, such as data standards, protection for Intellectual Property (IP) and taxonomies which provide frameworks to ensure consistency, interoperability, and quality in data markets as well as ownership protection. A critical concern, however, is that the use of data and the frequency of its transaction is not necessarily equal to the robustness of the data content (such as greenwashing) or the legitimacy of the data sharing (such as data privacy breach)[12].

As a result, the question for data market design is how to effectively incentivize the sharing of sufficient quality data, while not triggering fake, poor quality or illegal data provision that risks undermining the data value market in the long run. In addition, data value distribution needs to fairly reward the stakeholders that conduct the data work so that the power of data is not monopolized by a few big players. Validation mechanisms that prevent and remove falsified or poor-quality data are also essential to protect the integrity of data markets.

Key remarks

Data marketplaces have the opportunity to improve incentives and rewards around valuable data for improving environmental sustainability decision-making. In principle, the architectures could improve the capacity of diverse stakeholders to share and access valuable data through secure exchange mechanisms. Data valuation mechanisms, which are the core of the data marketplace, remain an open challenge as the value of data is dependent on the use case. As a next step, data market builders must consider how to value data for environmental sustainability. This could include identifying and surveying potential user groups; economic analyses of costs of compliance versus non-compliance; and the development of pricing models that potentially differentiate between commercial and non-commercial use cases. A key enabler for data marketplaces are data standards and validation mechanisms including high-quality data annotation, which will improve consistency, quality and trust in data exchange.

3.4 TI4: Computational law and data integration of green and circular economy policy measures

Policies that govern the sustainability of value chains are frequently updating and growing in number as regulators realize the need to tackle climate change.[13] This creates a challenge for policymakers, who need to strengthen and align policy measures across borders and sectors for effective environmental governance. This also means most policies are not readable or understandable for machines, which makes it hard for governments or startups to build the platforms or apps that can take policy implementation and compliance into the digital era.

An innovative solution is Computational Law (or CompLaw), which involves writing law in coding languages, to make policy unambiguous, machine-readable and programmable. The adoption of CompLaw promises significant benefits: policymakers can craft and assess policy measures more efficiently, businesses can automate various compliance processes, and the door is opened for the application of AI to assist in monitoring policy outcomes.

Recognizing the governance complexities associated with CompLaw, an alternative strategy involves harnessing modern data extraction and integration techniques to enhance policy comprehension and alignment.

12. Note that the sharing of certain sources of data can be found illegal due to political decisions or privacy concerns that materialize after the decision of data sharing.

13. "Policies" is used here to cover a range of different types of action by government or non-government actors. Examples include: government reporting requirements or policy measures such as green subsidies; industry standards; voluntary certification schemes.

Specific challenge

As the number of green and circular economy policies and standards grows, the result is an increasingly complex governance landscape made up of often overlapping measures and rules without a shared policy language.

To avoid policy silos and enable effective policymaking and implementation, adopting a value chain approach, it is essential to be able to align and compare policy measures as well as understand interactions and trade-offs. Key components for alignment include the semantics of policy measures, machine readability, information on how a policy functions and schema for data collection.[14] For policymakers, the benefit is to identify potential overlaps, contradictions and conflicts, improve policy synergies and their impact and reduce the administrative burden of implementation. It also reduces the burden of compliance for businesses by simplifying data collection, avoiding inefficiencies of multiple reporting systems and streamlining the policy implementation process.

Technology trend

CompLaw is a technology that can help improve alignment between relevant policy measures. As discussed above, CompLaw is an emerging field that advocates for the representation of legal rules and processes in programmable languages. Where legal language can be ambiguous, keywords in programming languages have a determined meaning and function when used together, which would improve the interpretation of laws and policies by machines. Some practitioners have proposed the use of a "domain-specific language" (DSL) for law, which would be formal programming languages specifically designed to capture the semantics and syntax of law (Chun 2022). For others, CompLaw can use any programming language, with formal logic programming languages like Prologue, or widely used programming languages such as Java or Python, being adequate for encoding the logic of law.

CompLaw has benefits in policymaking and for building legal applications that are enabled by DPI. For policymakers, CompLaw could improve the ability to design, assess and revise policies based on their effectiveness and fit within broader policy frameworks. CompLaw could also improve the responsiveness of legal software systems. Currently, any website or application for compliance needs to be reprogrammed whenever there is a change in relevant law or applications, whereas with law represented in code any changes can be implemented automatically (Genesereth 2015). In the fast-moving world of GCEP, this has an obvious benefit as a back-end enabler for software solutions targeting compliance challenges. Future applications could include compliance monitoring using AI, though this depends on the type of compliance and the availability of high-quality reporting data.

The uptake of CompLaw is slow due to monolithic volumes of law and legacy legal processes, but innovations are improving its scalability. In particular, the rapid improvement of Large Language Models (LLMs) capable of natural language processing and generation, such as GPT-4, offer potential for translating large bodies of natural language into code or translating law from one programming language to another. The performance of LLMs on tasks such as legal examinations suggest this type of application could be realized in the near future (OpenAI 2023). However, a clear problem with natural language generation - particularly in a policy situation where reliability is essential - is hallucination, where models generate text that is nonsensical or unfaithful to the original source material (Ji *et al.* 2022, p.3); thus generative AI must be used with careful human oversight and validation.

14. In the green economic policy context, governmental policies and standard requirements are featured with their connection to environmental performance. There could be transparency requirements that companies are required to disclose including their energy, water and land consumption, the emission as a result of their production and the relationship between the value chain and endangered species zones, among other things. It could also be about substantive performance requirements that set technical requirements on the pollution and waste, no deforestation association in the value chain or no use of pesticide in order to receive rewards or to avoid being panelized. Typically for policymaking and compliance, it necessitates the flow of data along the flow of materials.

While CompLaw can improve future policymaking, another set of technologies deal with aligning and integrating policies as they exist now. This is not a new approach in the policy context - for instance, trade delegates spend years negotiating certification and verification schemes to recognize each other's policy requirements for coherence. Technically, this means identifying relevant policy measures, identifying shared items in policy measures, noting the language used to describe shared items and providing a reference guide for the end user.

A key technology is information extraction (Smith *et al.* 2022). This automatically extracts relevant information from different text documents and integrates it into a joint, homogenized database. The main challenges are how to find relevant information and how to ensure the extraction algorithms work with high precision. Humans could do the same task with high precision, but it is tedious and time-consuming to carefully read and fully understand hundreds or even thousands of documents. Algorithms, on the other hand, are able to read documents extremely fast but might make mistakes when extracting information or summarizing the main content. A particular challenge is represented by modifiers, such as the word "not", which is not content-bearing independently and so could be ignored by an information extraction algorithm, but has a vital effect on a sentence's meaning.

To produce resources that improve the comparison of policies, another key technology is data integration (Brunner and Stockinger 2020). Unlike information extraction, which deals with unstructured data such as text, data integration typically deals with structured data that is stored in relational databases, Excel-sheets or CSV-files. The main challenges are how to integrate data sets with different formats or different granularity. For example, one country might measure the CO_2 footprint of companies on a yearly basis while others report on a monthly basis but only for selected regions. Again, machine-learning algorithms leveraging LLMs show promising results (Dong and Rekatsinas 2018). However, typically, any automatic data integration problem requires detailed feedback from humans and labelled data (which is often hard to obtain) to help the integration algorithms learn the right patterns. In some cases, policy language is also intentionally ambiguous, often for political purposes, allowing a wide scope for interpretation by different constituencies. The meaning of words in policy or law is the product of political processes and compromises, making it difficult to accurately extract the underlying intent and meaning. While information extraction and data integration are increasingly capable, additional research is required to make the approaches practical in real policy-making processes or legal applications.

Key remarks

There is no "magic bullet" that can make complex policies simple to work with, but technology solutions can improve the available tools. CompLaw seems promising to help streamline legal semantics and provide better building blocks for policy design, analysis, implementation and comparison. Meanwhile, advances in information extraction mean that tools are increasingly available to make sense of huge volumes of policies in natural language. A crucial limit to these techniques occurs when policy language has different meanings in different contexts and cultures. As a result, natural language-based analysis needs to be carefully developed and validated by experts. A technology solution might instead return sections from a policy or legal text to let a user compare for themselves, rather than trusting the automatic analysis. In reality, extraction and integration is a viable short-term solution, but depends on effective "human-in-the-loop" design, while CompLaw should be viewed as a long-term goal that deals with the more fundamental issues of policy language and inter-operable semantics. Each scenario depends on collaboration between wide groups of policymakers and the subjects of policy, both in technology design but also more fundamentally in governance.

3.5 TI5: Using Large Language Models to "speak" with green and circular economy policy

Tools and techniques related to natural language processing can change the way stakeholders interpret and respond to policy measures. New techniques from data science and AI provide the opportunity to improve policy comprehension by allowing users to query policy data in natural language. Policy comprehension is a possible use case for LLM tools, popularized by ChatGPT, which allow users to ask questions and receive a response in conversational language. To be specific, this is a challenging task of Open-domain Question-Answering (OpenQA or ODQA), which involves answering factual questions from a large knowledge corpus of unstructured text.

Specific challenge

Comprehension of policy is essential for stakeholders who are affected by policy or engage in policy-making processes. GCEPs can take a variety of forms, differ in terms of scope and objectives, target different sectors of the economy, impose obligations, and offer benefits to a wide range of actors and aim to manage different environmental risks and ecological scarcities. Moreover, some of the regulations can be highly technical, lengthy and convoluted for subjects of the laws, regulations and policies to comply; especially in the context of the increasing reporting and compliance schemes on sustainability. Often, they rely on multiple prescribed definitions with defined applications; reference multiple instruments not integrated into the same policy, rather build upon each other; the regulatory regime can vary depending on predefined conditions or ranges contingent on the subjects; concrete applications require interpretation; and use non-fixed periodic economic benchmarks as the basis for calculating payments, obligations or penalties owed.[15]

This heterogeneity and complexity are barriers to policy querying. Predefined categories would need to be able to capture a multiplicity of criteria and models would need to be trained to consider different application scenarios. If a subject of these policies is seeking to identify those cases that are applicable to its conditions, external inputs would be required to determine direct applicability, which would mean that users must know how to connect policy with information from their own specific circumstance.

Technology trend

In recent years, and in particular since the release of GPT-3.5 and ChatGPT in late 2022, the availability and usefulness of natural language processing (NLP) tools has boomed. An important technology leading to recent breakthroughs in NLP - and all machine learning - has been the discovery of transformer architectures, which allow language models to consider the words in sentences simultaneously, which makes it easier for training models to understand the relationships between words than the word-by-word approach of previous neural networks (Vaswani *et al.* 2017). Transformer technologies paved the way for the LLMs of today, such as the GPT-3.5 (Generative Pre-trained Transformer 3), which was pre-trained efficiently using transformer architectures on huge volumes of online text.

15. For instance, a country might provide a reduction of the Property Tax under their tax code for owners of real-state property intended for residential use, such as a 25 per cent reduction, to those who have on their property adult and living trees, as long as the trees receive the necessary maintenance in accordance with environmental regulations. In this example, the tax code would be a lengthy document with mostly irrelevant information for green economy purposes, the application would be dependent on definitions of concepts such as "residential use" and "necessary maintenance" from other instruments, the amount of property tax might not even be calculated based on the same tax code, and the applicability of the incentive would be contingent on subjective conditions such as ownership of real-state property. Not to mention that its applicability would be subject to interpretation in many instances, like the case of a farmer that uses part of their property for residential purposes and the other for trade.

Advances in LLMs have improved the feasibility of OpenQA tasks, with models capable of interpreting complex policy questions and generating the policy guidance, policy synthesis or other needs of a user (Chen and Yih 2020). OpenQA systems use a range of NLP techniques to understand the semantics of questions, such as information retrieval, passage ranking and text summarization. Results are impressive - in early 2023, GPT-4 passed the Bar Exam in the United States, placing in the 90th percentile (Katz *et al.* 2023).

However, experts advise extreme caution when using LLM tools for OpenQA in a consequential decision-making context. A well-documented problem with LLMs in their current form is hallucination, which models generate incorrect answers to questions. Researchers have experimented with techniques to improve the truthfulness and reliability of generated answers (Vaghefi *et al.* 2023), but ongoing hallucinations mean that OpenQA tasks for policy measures are risky, as misleading responses could have real negative effects. As of 2023, the meticulous yet ambiguous nature of policy language, along with a relatively small global body of policy documents for training data, means the accuracy of OpenQA in the policy domain is currently relatively low and requires more active research[16].

An alternative solution for improving the accuracy and reliability of OpenQA policy tasks could involve storing policy information in structured databases. The challenge then becomes a question of how to query large databases with specific information in natural language. Unlike in the OpenQA scenario where data is stored as unstructured text, in this scenario data is stored in relational databases, i.e. information is structured in tabular form and linked with other database tables. Querying these databases typically requires high proficiency in domain-specific database query languages - a skill that policymakers and subjects often do not have.

However, recent trends in generative AI-based research tackle this problem by automatically translating the natural language of non-tech savvy end-users into the correct database query language (Brunner and Stockinger 2021; Amer-Yahia *et al.* 2022). The approach is similar to translating English to French, for example. However, the major difference is that each database has a different structure, different semantics, etc. (see also TI4). To tackle these challenges, there must be enough and relevant feedback for training complex machine-learning algorithms to ensure accurate responses.

Key remarks

Advances in NLP technologies and improvements in Question-Answering models can transform the way that stakeholders interact with policies. In practice, this could strengthen the implementation of GCEP by improving policy comprehension. Any solutions, though, must be built with the intricacy and ambiguity of policy in mind, as well as a recognition of the risks of inaccuracy in consequential scenarios. A safer and more reliable solution could first improve the identification of relevant sections of policy text and return relevant sections in their original form. Although it is not an OpenQA solution, this approach can be found in Climate Policy Radar, which provides a natural language search function on top of a database containing the full text of more than 3,500 climate policies, returning the relevant sections of policies to users (Climate Policy Radar 2023). When this can be done accurately, including for some of the more complex policy-related questions discussed in this TI, generated QA answers could be linked with direct excerpts of policy text, providing a dual explanation to users.

16. Climate Policy Radar has aggregated 3,500 climate policy documents globally. While this is a huge body of text for human policy analysts, this is comparatively small for training LLMs.

3.6 TI6: Tools and techniques for human-centred artificial intelligence in environmental sustainability decision-making

As applications and platforms for data-driven sustainability emerge across all sectors and geographies, it is essential to keep humans "in-the-loop". While the prospects of AI are enticing, they come with substantial risks, including the potential for inaccurate or biased outcomes due to inadequate training data and low transparency in the development processes underlying these models.

This underscores the necessity for oversight and accountability, which depends upon humans' ability to interact with, understand and troubleshoot the technology. Additionally, the design of new platforms and processes must be inclusive, ensuring the incorporation of knowledge and needs of a wide range of users, including marginalized groups.

Strong principles for human-centred artificial intelligence (HCAI) in sustainability are essential and DPI must incorporate technologies or services that support implementation. Current examples include human experiment-based benchmarking that aligns technologies with human insights, or interactive machine-learning interfaces that allow the iterative building and refining of models by expert groups without deep programming capabilities.

Specific challenge

A clear problem for AI, including AI applications for sustainability goals, has been inclusiveness in the development, deployment and use of digital technologies.

Digital solutions risk reinforcing pre-existing biases and inequalities in decision-making in environmental sustainability, including gender occupational stereotypes. Women often have fewer educational opportunities for reskilling than men and less access to resources for adapting to climate change, including technologies (ILO 2022, p. 15-16). Training initiatives should be consciously designed to tackle this sort of gender-specific challenges. For truly inclusive digital solutions, it is imperative not only to focus on its creation but to ensure that individuals of all genders can effectively harness these tools. Integrating women's perspectives into the design process of digital solutions and trainings on the use of those solutions is essential. Equally, solutions must also meet the needs of local or Indigenous groups.

Moreover, achieving inclusiveness in decision-making technologies for environmental sustainability requires close attention to how stakeholders receive and interpret information.

Technology trend

As AI systems have developed that appear capable, at face value, of making automated decisions typically reserved for humans, organizations have considered how technologies should be designed to keep humans "in-the-loop". New principles have spurred a fresh agenda for the design of inclusive AI systems, although technological implementation of many of these principles remains challenging. In the Human-Computer Interaction field, this has been called "human-centered artificial intelligence" (HCAI), where AI combines high levels of automation with high levels of human control (Shneiderman 2020).

Governments, international organizations, businesses and civil society organizations alike have published "principles for AI" with implications for digital inclusion in AI tools. In a white paper by the Berkman Klein Center for Internet & Society at Harvard, a number of shared categories of principles were identified that illustrate the trends in principles (Fjeld *et al.* 2020). Categories with particular relevance for environmental sustainability include: "Accountability", including principles to ensure that there is accountability for the impact of AI systems; "Transparency and explainability", including principles advocating for oversight of AI; "Fairness and Non-discrimination", which calls for AI to promote inclusivity; and "Human Control of Technology". These principles recognize how AI systems are not only measured against performance, but also other criteria that prioritize the concerns of the human users and subjects of AI systems.

In response to principles relating to accountability and transparency, research has focused on how AI can be made interpretable - meaning that AI systems have “the ability to explain or present in understandable terms” (Doshi-Velez and Kim 2017, p.2). Interpretability would allow stakeholders to have oversight of the reasoning for decisions that may affect ecosystems or value chain outcomes or improve fairness by allowing stakeholders to detect bias in the data that informs environmental sustainability decisions. Unfortunately, the training of algorithms on large volumes of data makes it very difficult to evaluate why models deliver certain outcomes.

Experiment-based benchmarks can be used to evaluate explanations provided by models. A suggested method for testing AI is through human experiment methods that evaluate whether explanations improve human performance on measurable tasks or match the predictions of human experts given an input and explanation (Doshi-Velez and Kim 2017, p.4-5). To aid the training of explainable models, an open repository could be created containing real-world problems, different methods used to address the problem and the performance of methods on the task. Repositories with real-world testing data could train models to identify the factors that are most important for explanations and assist domain experts who have been trained to evaluate AI systems used in their domain of expertise. These efforts improve transparency, help identify bias and align systems with human control.

Advancements in technology are also introducing novel methods that improve the capabilities of domain experts to train and understand AI systems, aligning with principles of human control over technology. For model training, techniques for Interactive Machine Learning (IML) allow the building and refinement of models through iterative cycles of input and review by user groups, for example by providing indicative samples, describing features, or selecting high-level model parameters (Dudley and Kristensson 2018). Domain knowledge can be applied through graphic interfaces, as seen in the medical or biological space where domain experts can annotate diagrams with labels to train the analytical processes and classification results of algorithms (Berg *et al.* 2019).

Advances in technologies have also provided opportunities for personalized information delivery, which can improve understanding and information uptake. Generative AI has been a breakthrough, as users can interact with information tools in natural language and express needs more clearly. Researchers have addressed the challenges stakeholders face in navigating applications or information services through adaptive user interfaces (Adaptive UI), which automatically adapt the organization or presentation of the user interface in response to some characteristic of the user or context (Gajos and Chauncey 2017). Results have shown that Adaptive UI can improve inclusiveness by providing human-computer interaction that is more responsive to an individual’s cognition.

Key remarks

AI is changing the ways that humans analyse situations and make reasoned responses. While the architects of AI advise that AI systems should be used cautiously as "co-pilots", rather than trusted decision-makers, AI tools are entering the toolkits of stakeholders, from companies, to governments, NGOs and citizens. This means the effect of tools will interact with existing inequalities in decision-making in environmental sustainability, including across genders and social groups.

In this context, HCAI approaches may be essential to ensure inclusiveness within DPI. They are also a vital element for DPI to drive widespread impact. The principles guiding HCAI encompass crucial aspects like explainability, transparency and accountability. These principles are not just theoretical; they can be put into practice through a growing array of practical tools. Examples include human-experiment data sets tailored for evaluating AI performance and tools facilitating interactive machine learning in collaboration with domain experts. While HCAI represents a relatively new field, nurturing its development and application is essential, especially when considering its applications within the field of environmental sustainability. This is significant due to the broad reach and distinct capabilities of stakeholder groups involved and the imperative for DPI to have a far-reaching impact across the broader economy. Furthermore, beyond establishing inclusive DPI, it is pivotal to ensure that individuals, irrespective of gender, are adequately equipped to leverage these tools effectively. This underscores the need for trainings on digital tools to prioritize the HCAI approach.

3.7 Technology innovations for environmentally sustainable outcomes

This report has explored six categories of technological innovations that hold the promise of overcoming information challenges for stakeholders as highlighted in Section 2, aiming to achieve desired outcomes that facilitate informed environmental sustainability decision-making (see Table 2).

The establishment of data markets for environmental sustainability-related data can revolutionize real-time environmental monitoring, especially for beef-driven deforestation. These markets are crafted to create strong incentives for the collection and sharing of data, thereby providing stakeholders with consistent and reliable information which is essential for informed decision-making and strategic interventions in environmental management.

The second significant innovation is the development of open data discovery for environmental sustainability, designed to overcome the perennial challenge of data discoverability. This system is expected to significantly facilitate the identification of relevant data, empowering stakeholders to efficiently locate and utilize the information needed to address sustainability challenges effectively.

Privacy-enhancing technologies are being proposed to navigate the delicate balance between the free flow of information and the protection of privacy. By introducing such technologies, this technology innovation is addressing the barriers to data sharing, ensuring that stakeholders can collaborate without the risk of compromising sensitive information.

To further support compliance and understanding of environmental standards, Large Language Models can "speak" with green and circular economy policies. This innovation aims to foster improved compliance with environmental regulations across companies of all sizes by providing a better understanding and easier policy interpretation, thus simplifying complex regulatory landscapes.

In the intricate realm of legal and policy frameworks, computational law and data integration for green and circular economy policy measures can streamline Monitoring, Reporting, and Verification (MRV) frameworks. This approach is expected to eliminate redundant processes and reduce confusion among stakeholders, thereby facilitating a more efficient and clear-cut approach to environmental policy compliance and monitoring.

Finally, to address the challenge of inclusivity in data, it is proposed to harness the power of techniques focused on human-centered artificial intelligence. This integrative approach is intended to generate a more inclusive data environment that not only represents diverse groups but also enhances the user experience, enabling stakeholders to interact with data collection and generation tools in a more user-friendly and productive manner.

DPI as a data exchange system for sustainability-related data	
Technology innovation	**Desired outcome**
Data markets for environmental sustainability-related data	Real-time monitoring of deforestation with data readily available for stakeholders by creating incentives for collecting and sharing. This process is expected to bring about consistent and reliable data across the board
Open data discovery for environmental sustainability	Facilitate identification of relevant data
Privacy enhancing technologies to enable flow of environmental sustainability information	Addressing the barriers for data sharing without compromising privacy
Using Large Language Models to "speak" with green and circular economy policy	Improved compliance with environmental standards across all company sizes due to better understanding and easier policy interpretation
Computational law and data integration of green and circular economy policy measures	Streamlined MRV frameworks that reduce redundancy and confusion among stakeholders
Data markets for environmental sustainability-related data	Data-driven policy making by creating incentives for data collecting and generation
Open data discovery for environmental sustainability	Facilitate identification of relevant data
Data markets for environmental sustainability-related data and tools and techniques for human-centred artificial intelligence in environmental sustainability decision-making	A more inclusive data environment that represents diverse groups and allows for user-friendly interaction with data-generating and collection tools

Table 3: **DPI as a data exchange system for sustainability-related data**

The next step concerns the role of stakeholders in making those technologies work for environmental sustainability as part of ongoing efforts to build DPI and related applications.

4 Conclusions and recommendations

4.1 Key message: using the data exchange system for environmental sustainability

As shown in Section 2, the need for an improved data exchange ecosystem follows a growing demand for data in environmental governance. A series of pressures have put environmental sustainability at the top of the agendas of governments, businesses, and the global public at large. Data is the key ingredient for accountability, for empowerment, and proof of sustainability: for governments to meet international obligations, for businesses to access markets and secure investment, for consumers to make conscious choices between products.

Each of the six data science technologies discussed in this report not only facilitates the flow of data for decision-making by generating value for users but also enables DPI as a data exchange system. This report recommends these six TIs to address underlying information challenges within the data ecosystem, thereby closing the gap in DPI that hinders the flow of environmental sustainability information to different stakeholders. This system empowers actors to proactively address environmental challenges, public policies, and the realities of value chain operations. As highlighted in Section 3, this DPI for environmental sustainability enables more efficient data generation, collection, and sharing, thereby improving the availability and quality of data crucial for decision-making. What does this mean for decision-makers?

decision-makers must be **active participants** in data exchange

In short, decision-makers must be active participants in data exchange. This requires a change of mindset, from simply responding to data demands and treating data sciences as merely tools to promoting data as an asset and working with non-traditional collaborators on ground data. Every type of stakeholder, no matter how big or small, can take part by considering the data they create and its value to others, the information they need to act sustainably, and then how data can be put at the center of business models utilizing data from other parties. The roles can go from data subject, data broker, data owner, to data platform organizers, data consumers, or others. The give and take of data exchange for environmental sustainability must feature in mindsets in the same way as the use of natural resources or the flow of finance.

In most cases, stakeholders could benefit from the synergies between different technologies but in practice, each use case will depend on the availability of technologies, and stakeholders taking an active mindset to incorporate technologies and tools into their business models.

Returning to the case of the agrifood sector, consider the following examples. Farmers in a collective could utilize ground truth data about their cropland's environmental conditions, sharing regular sensor data through a data marketplace for a fee, while employing privacy techniques for sensitive information. This data could serve upstream entities in food processing or logistics for nature-related financial disclosures or in developing consumer apps that evaluate purchase sustainability using AI with human-centered design. Similarly, a regulator focused on the industry's environmental sustainability could leverage large language models in a managed application, enabling businesses to navigate complex regulations by asking specific policy questions. By coding key regulations, compliance apps could automatically adapt to legal changes, enhancing transparency and clarifying data set value. This approach, along with open data discovery, equips supply chain operators with proprietary and open data for crafting sustainable production strategies and data-driven reporting capabilities.

For the advancement of DPI as a data exchange system for environmental sustainability, it is also crucial to put the challenge of inclusiveness in focus, including gender issues. Overall, gender disaggregated data is still scarce, and insufficient consideration is given to the consequences of lacking gender- responsive data analysis, as well as its uptake among environmental sustainability stakeholders (United Nations

Women 2018). Gender-related data could be boosted to accelerate the realization of value of this data, driving gender-related data pricing and incentivizing data supply on the ground. Additionally, big data could be used to discover gender trends and related intelligence (Data2x 2019). The human-centred approach could reflect gender-based differences in user demand and user adoption since user qualities like gender can differentiate how they consume information (Fallows 2005). For example, to achieve a truly inclusive digital landscape in AI, the design of training programmes and tools must address gender-specific challenges, prioritize women's perspectives, and emphasize the HCAI approach, ensuring all can utilize these tools.

4.2 Action points: enabling a digital public infrastructure for environmental sustainability

To facilitate stakeholder participation in data exchange systems for environmental sustainability, it is crucial to develop DPI with environmental sustainability as a core outcome. Global policy initiatives, such as those under the G20 and UN SDG Summit, are already laying the groundwork for DPI development, which can be extended to include environmental sustainability goals. There is already a strong foundation to build on these existing initiatives and institutional support as backed by political will. Going beyond the current focus of those initiatives on digital economy and financial inclusion, this report has explored what is required for a DPI for environmental sustainability.

> DPI will help stakeholders meet policy requirements and **further environmental sustainability**, which in turn will increase usage and help evolve DPI for other applications

This report showed how DPI as a data exchange system for sustainability-related data can be harnessed for decision-making, but it also emphasizes the importance of integrating sustainability into DPI development. The relationship is multi-directional: DPI will help stakeholders meet policy requirements and further environmental sustainability, which in turn will increase usage and help evolve DPI for other applications. In this evolving context, efforts and resources must be devoted in the short and medium term to ensure an early inclusion of sustainability considerations into ongoing and future development of DPI.

In parallel, political and institutional alignment must be capitalized for DPI global agendas to be introduced into wider sustainability goals. To address the gap in DPI for environmental sustainability policymakers can foster an enabling environment that supports its development. Policymakers may rely on international organizations and other key partners to help raise awareness on the relevance of DPI, promote its adoption in national contexts and bridge knowledge gaps between experts.

Similarly, NGOs, data scientists, external investors and other experts can collectively scale up efforts to establish a DPI for environmental sustainability. Collaboration between various stakeholders is key to harness DPI for environmental sustainability, including direct partnerships with the data science community and technology developers to harness innovation.

Intrinsically connected, the following action points illustrate some of the keyways forward to enable a DPI for environmental sustainability:

Policy and regulations: to foster enabling conditions, policy support is required for the development of DPI for environmental sustainability, as well as policy safeguards for public interest and risk mitigation. Attaining this policy objective necessitates the incorporation of a synergistic technology-domain perspective in all stages of the policy cycle, from agenda-setting and formulation through to implementation and evaluation.

- Join and be part of the evolving global policy and collaboration efforts around DPI, such as those spearheaded by the G20, the UN Secretary General and others originating from the Global South.
- Integrate national priorities of harnessing DPI for environmental sustainability within national digital transformation plans, infrastructure national plans, strategies, roadmaps, public investment programmes, nationally determined contributions, or other instruments for achieving the Sustainable Development Goals by 2030.
- Incorporate real-time sustainability data into the formulation process of circular and green economy policies with the aim to contribute to its effectiveness and to better tailor responsive strategies for addressing environmental challenges (e.g. leverage the use of existing data resources like the World Environment Situation Room).
- Test metrics and benchmarks to measure the performance of DPI for environmental sustainability.
- Leverage advancements in NLP and Question-Answering models to enhance stakeholder engagement with policies, particularly to improve understanding and implementation of GCEP.
- Actively support research and development (R&D) initiatives and foster collaborative partnerships with technology developers and the data science community.
- Prioritise environmental sustainability use cases for DPI that also mitigate gender disparities and other biases, benefiting vulnerable groups such as women, the elderly, youth, micro and small enterprises, and rural communities.

Standardization: guiding the development of DPI for environmental sustainability, through collaborative efforts between various stakeholders, balancing the need for security, privacy, inclusivity, accessibility, interoperability and adherence to regulatory and legal frameworks.

- Accelerate efforts for the adoption and implementation of a Global Environmental Data Strategy by 2025 for managing, utilizing and sharing environmental data, which will enable a DPI for environmental sustainability.
- Expedite global consensus on taxonomies around key concepts on emerging technologies and amplify these efforts for categorization systems that enable interoperability on DPI for

environmental sustainability (e.g. convening international dialogues with a dedicated mandate on this space).

- Initiate sectoral consensus-building on the methodologies employed for data collection, analysis, and data sharing for sustainability-related data disclosure purposes, to ensure consistency in these processes to enhance the accuracy, comparability and reliability of disclosed information.
- Standardize data verification protocols to enhance data quality, consistency, relevance, and accuracy of sustainability-related information disclosures.
- Advocate for the establishment of sustainability-specific metadata for emerging data sets and enhancement of this metadata in existing relevant data resources to create large open data repositories that enable experimentation, facilitate data discovery, and augment its uptake (e.g. champion the introduction of keywords "environment", "material", "organism", and "location" to existing metadata of biodiversity relevant data sets).
- Develop a robust risk management framework and sound risk management practices for DPI for environmental sustainability, a pivotal element to maintain trust in DPI ecosystems.
- Standardize pricing mechanisms of sustainability-related data to stimulate the emergence of new business models within the data exchange ecosystem.

Finance: to scale up efforts on DPI it is essential to leverage public and private investment for paving the infrastructure and bringing life to business models for environmental sustainability via DPI, through reformed global financing mechanisms.

For Infrastructure Financing (Public Finance):

- Allocate public funds to develop and maintain DPI that supports the integration of Human-Centred Artificial Intelligence (HCAI) in sustainability applications.
- Financially back the development of data standards and validation mechanisms to ensure high-quality data annotation and trust in data exchange.
- Mobilize sustainable investment for R&D on DPI for environmental sustainability.
- Assign public resources to raise awareness and build capacities among key stakeholders around potential applications of DPI for environmental sustainability and enhance public monitoring capacities.

To promote blended finance for DPI:

- Develop applications that utilize DPI to enhance decision-making in environmental sustainability, ensuring they are tailored to specific user groups and promote the incorporation of diverse knowledge sources.
- Construct business models that facilitate the secure exchange, valuation, and pricing of data in data marketplaces, considering different use cases and user groups.

To reform Major Global Financing Mechanisms:

- Re-evaluate and reform the risk portfolios of major global financing mechanisms to accommodate the unique challenges and opportunities presented by DPI investments.
- Align the financing mechanisms for DPIs with environmental sustainability goals to ensure that investments are directed towards initiatives that genuinely contribute to sustainability and are measurable against defined metrics and benchmarks.

Innovation facilitators: to foster the progress of the development of DPI for environmental sustainability, it is essential to take steps to overcome barriers to innovation, while ensuring integrating inclusivity considerations in innovations.

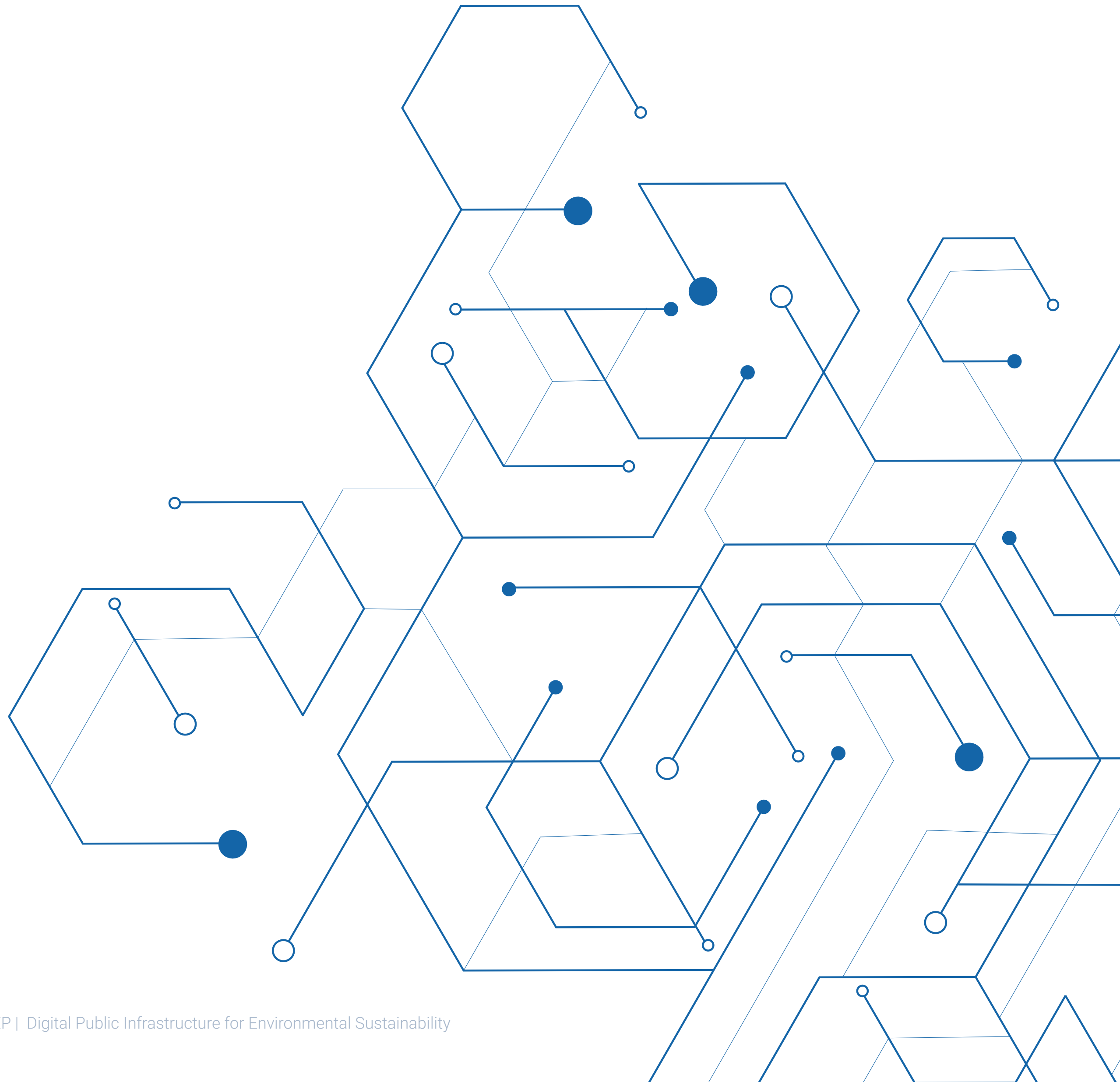
- Champion the establishment of regulatory and technology sandboxes in the environmental sustainability domain that grant exemptions or provide regulatory flexibility for piloting emerging technological solutions.
- Organize sustainability-related hackathons to accelerate innovation and problem solving of real-world challenges and test novel approaches for DPI.
- Invest in building the capacities of government officials, technologists and other stakeholders involved in the developing and implementation of DPI.

Collaboration and partnership: International collaboration and public-private partnerships (PPPs) play a crucial role in ensuring the successful development of DPI for environmental sustainability.

- Foster platforms to facilitate intergovernmental dialogues that scale up ongoing efforts for global and regional frameworks for the ethical development, design, and implementation of emerging technologies; as well as on other key areas like safeguarding individual privacy rights, enhancing cybersecurity measures and ensuring the harnessing of AI's capabilities for the benefit of humanity on a global scale.
- Capitalize on existing multi-stakeholder alliances, like the Coalition for Digital Environmental Sustainability (CODES) and the One Planet Network, to steer and prioritize the use of emerging digital technologies for DPI.
- Coordinate the work between international agencies on the advancement of DPI for environmental sustainability to complement their expertise and contribute to better solution development and enable cross-pollination of ideas.

- Facilitate collaborative platforms and teams where domain experts, data scientists and industry players work cohesively from the ideation to the implementation phase, ensuring robust and relevant DPI for environmental sustainability.
- Implement training and capacity-building programs to boost data literacy, analytical skills, and monitoring abilities across public and private sectors, with a specific focus on addressing gender-specific challenges and incorporating women's perspectives to ensure equitable DPI utilization.

UNEP and its international and country partners will continue to promote and support the development of DPI for environmental sustainability by raising awareness amongst key stakeholder groups, conducting research and bridging knowledge gaps between experts from different domains, and supporting pilot projects at the country level.



References

Acar, A., Aksu, H., Uluagac, A.S. and Conti, M. (2018). A survey on homomorphic encryption schemes: Theory and implementation. ACM Computing Surveys 51(4), 79:1-79:35. https://doi.org/10.1145/3214303.

Akın Ateş, M., Suurmond, R., Luzzini, D. and Krause, D. (2022). Order from chaos: A meta-analysis of supply chain complexity and firm performance. Journal of Supply Chain Management 58(1), 3–30. https://doi.org/10.1111/jscm.12264.

Alicke, K., Barriball, E., Foster, T., Mauhourat, J. and Trautwein, V. (2022). Taking the pulse of shifting supply chains. https://www.mckinsey.com/capabilities/operations/our-insights/taking-the-pulse-of-shifting-supply-chains.

Amer-Yahia, S., Koutrika, G., Braschler, M., Calvanese, D., Lanti, D., Lücke-Tieke, H. *et al.* (2022). INODE: Building an end-to-end data exploration system in practice. ACM SIGMOD Record 50(4), 23–29. https://doi.org/10.1145/3516431.3516436.

Banerjee, P. and Ruj, S. (2019). Blockchain Enabled Data Marketplace -- Design and Challenges. arXiv. https://doi.org/10.48550/arXiv.1811.11462.

BDO RCS Auditores Independientes SS (2022). Report on Agreed-Upon Procedures regarding the verification of Minerva Foods' Cattle Purchase Monitoring System, included in the Third-party Audit Reference Term denominated 'BEEF3_AUDIT_PY_JUL19_FINAL'. Sao Paolo. https://www.minervafoods.com/wp-content/uploads/2022/10/Relatorio-Auditoria-Independente-PY-2021_EN.pdf.

Bellmann, C. and van der Ven, C. (2020). Greening regional trade agreements on non-tariff measures through technical barriers to trade and regulatory co-operation. 2020/04. Paris: OECD. https://doi.org/10.1787/dfc41618-en.

Benjelloun, O., Chen, S. and Noy, N. (2020). Google dataset search by the numbers. In) The Semantic Web – ISWC 2020. J.Z. Pan, V. Tamma, C. d'Amato, K. Janowicz, B. Fu, A. Polleres, O. Seneviratne, and L. Kagal (eds). Cham: Springer International Publishing (Lecture Notes in Computer Science). 667–682. https://doi.org/10.1007/978-3-030-62466-8_41.

Berg, S., Kutra, D., Kroeger, T., Straehle, C.N., Kausler, B.X., Haubold, C. *et al.* (2019). ilastik: Interactive machine learning for (bio)image analysis. Nature Methods 16(12), 1226–1232. https://doi.org/10.1038/s41592-019-0582-9.

Biber, E. (2013). The challenge of collecting and using environmental monitoring data. Ecology and Society 18(4). https://www.jstor.org/stable/26269441. Accessed 5 July 2023.

Bolfe, E.L., Jorge, L.A.C., Sanches, I., Costa, C.C. da, Luchiari Jr., A., Victória, D., Inamasu, R., Grego, C., Ferreira, V., & Ramirez, A. (2020). Agricultura digital no Brasil: tendências, desafios e oportunidades: resultados de pesquisa online. Campinas: Embrapa. Available at: https://www.embrapa.br/agropensa/produtos-agropensa. Accessed 14 Oct. 2020.

Bonawitz, K., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., Ivanov, V. *et al.* (2019). Towards Federated Learning at Scale: System Design. arXiv. https://doi.org/10.48550/arXiv.1902.01046.

Brandão, R. and Arbix, G. (2022) 'Artificial Intelligence, Ethics and Public Policy— The Use of Facial Recognition Systems in Public Transport in the Largest Brazilian Cities', Journal of Service Science and Management, 15(05), pp. 551–575. Available at: https://doi.org/10.4236/jssm.2022.155032.

Brickley, D., Burgess, M. and Noy, N. (2019). Google dataset search: Building a search engine for datasets in an open Web ecosystem. In The World Wide Web Conference. New York, NY, USA: Association for Computing Machinery (WWW '19). 1365–1375. https://doi.org/10.1145/3308558.3313685.

Brunner, U. and Stockinger, K. (2020). Entity matching with transformer architectures - a step forward in data integration. In 23rd International Conference on Extending Database Technology, Copenhagen, 30 March - 2 April 2020. OpenProceedings 463–473. https://doi.org/10.5441/002/edbt.2020.58.

Brunner, U. and Stockinger, K. (2021). ValueNet: A natural language-to-SQL system that learns from database information. In 2021 IEEE 37th International Conference on Data Engineering (ICDE). 2177–2182. https://doi.org/10.1109/ICDE51399.2021.00220.

Buřivalová, Z., Yoh, N., Butler, R.A., Chandra Sagar, H.S.S. and Game, E.T. (2023). Broadening the focus of forest conservation beyond carbon. Current Biology 33(11), R621–R635. https://doi.org/10.1016/j.cub.2023.04.019.

Carbon Disclosure Project and Accountability Framework Initiative (2022). From commitments to action at scale: Critical steps to achieve deforestation-free supply chains. https://cdn.cdp.net/cdp-production/cms/reports/documents/000/006/368/original/CDP_AFI_Forest_Report_2022_%2814%29.pdf?1654614758.

Carbon Disclosure Project (2023a). Carbon Disclosure Project responses. https://www.cdp.net/en/responses?page=2&per_page=5&queries%5Bname%5D=minerva&sort_by=project_year&sort_dir=desc&utf8=%E2%9C%93.

Carbon Disclosure Project (2023b). Global Forests Report. https://cdn.cdp.net/cdp-production/cms/reports/documents/000/007/182/original/CDP_Global_Forest_Report_2023.pdf?1688396252.

Carroll, S.R., Herczog, E., Hudson, M., Russell, K. and Stall, S. (2021). Operationalizing the CARE and FAIR Principles for Indigenous data futures. Scientific Data 8(1), 108. https://doi.org/10.1038/s41597-021-00892-0.

Chakravorti, B. (2023) 'The Case for Investing in Digital Public Infrastructure', Harvard Business Review, 22 May. Available at: https://hbr.org/2023/05/the-case-for-investing-in-digital-public-infrastructure. Accessed: 7 February 2024.

Chapman, A., Simperl, E., Koesten, L., Konstantinidis, G., Ibáñez, L.-D., Kacprzak, E *et al.* (2020). 'Dataset search: a survey. The VLDB Journal 29(1), 251–272. https://doi.org/10.1007/s00778-019-00564-x.

Charalabidis, Y., Zuiderwijk, A., Alexopoulos, C., Janssen, M., Lampoltshammer, T. and Ferro, E. (2018). Open data evaluation models: Theory and practice. In The World of Open Data: Concepts, Methods, Tools and Experiences. Y. Charalabidis, A. Zuiderwijk, C. Alexopoulos, M. Janssen, Lampoltshammer,T. and Ferro, E (eds). Cham: Springer International Publishing (Public Administration and Information Technology). 137–172. https://doi.org/10.1007/978-3-319-90850-2_8.

Chen, D. and Yih, W. (2020). Open-domain Question Answering. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts. Online: Association for Computational Linguistics. 34–37. https://doi.org/10.18653/v1/2020.acl-tutorials.8.

Chever, T., Gonçalves, A. and Lepeule, C. (2022). Research for AGRI Committee - Farm certification schemes for sustainable agriculture: State of play and overview in the EU and key global producing countries, concepts and methods. Brussels: European Union, Policy Department for Structural and Cohesion Policies. https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU(2022)699633. Accessed 5 July 2023.

Chun, A.N. (2022). CLAWs and Effect, The Singapore Law Gazette, 15 December. https://lawgazette.com.sg/practice/tech-talk/claws-and-effect/. Accessed 5 July 2023.

Columbia Engineering (2022). What Is a Sandbox in Tech and Cybersecurity? Columbia Engineering Boot Camps, 3 August. https://bootcamp.cvn.columbia.edu/blog/what-is-a-sandbox/. Accessed 6 July 2023.

Committee on National Security Systems (CNSS) (2022). Committee on National Security Systems (CNSS) Glossary. https://www.cnss.gov/CNSS/openDoc.cfm?a=oIVxEol2VpOzkQEB%2FbYUFw%3D%3D&b=A42156D0D35F40932C8C22E1AD056FFE36A8AA68821B797D-3256BF742DD4A78BD5BF4B77E9122BBBCBB9B281AE55C691. Accessed 5 July 2022.

Council of the European Union (2022). Draft Regulation of the European Parliament and of the Council on the making available on the Union market as well as export from the Union of certain commodities and products associated with deforestation and forest degradation and repealing Regulation. https://data.consilium.europa.eu/doc/document/ST-10284-2022-INIT/en/pdf.

Cottrell, R.S., Nash, K.L., Halpern, B.S., Remenyi, T.A., Corney, S.P., Fleming, A. *et al.* (2019). Food production shocks across land and sea. Nature Sustainability 2(2), 130–137. https://doi.org/10.1038/s41893-018-0210-1.

Curtis, P.G., Slay, C.M., Harris, N.L., Tyukavina, A. and Hansen, M.C. (2018). Classifying drivers of global forest loss. Science 361(6407), 1108–1111. https://doi.org/10.1126/science.aau3445.

Data2x (2019). Big Data, Big Impact? Towards Gender-Sensitive Data Systems. https://data2x.org/resource-center/big-data-report/. Accessed 6 July 2023.

Democractic People's Republic of Korea, Ministry of Trade, Industry and Energy (MOTIE) (2021). 'K-ESG Guidelines'. http://www.motie.go.kr/motie/gov3.0/gov_openinfo/sajun/bbs/bbsView.do?bbs_seq_n=631&bbs_cd_n=30.

Diaz, S. and The Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) (eds) (2020). The Global Assessment Report on Biodiversity and Ecosystem Services: Summary for Policymakers. Bonn (Germany). https://wfp.sharepoint.com/sites/LRCDissemination/Catalogue/Docs/ENGLISH/ELR%20792.pdf. Accessed 1 August 2023.

Domingo-Ferrer, J., Sánchez, D. and Blanco-Justicia, A. (2021). The limits of differential privacy (and its misuse in data release and machine learning). Communications of the ACM 64(7), 33–35. https://doi.org/10.1145/3433638.

Dong, X.L. and Rekatsinas, T. (2018). Data integration and machine learning: A natural synergy. In Proceedings of the 2018 International Conference on Management of Data. New York, NY, USA: Association for Computing Machinery (SIGMOD '18). 1645–1650. https://doi.org/10.1145/3183713.3197387.

Doshi-Velez, F. and Kim, B. (2017). Towards A Rigorous Science of Interpretable Machine Learning. arXiv. https://doi.org/10.48550/arXiv.1702.08608.

Dudley, J.J. and Kristensson, P.O. (2018). A review of user interface design for interactive machine learning. ACM Transactions on Interactive Intelligent Systems 8(2), 8:1-8:37. https://doi.org/10.1145/3185517.

Dyda, A., Purcell, M., Curtis, S., Field, E., Pillai, P., Ricardo, K. *et al.* (2021). Differential privacy for public health data: An innovative tool to optimize information sharing while protecting data confidentiality. Patterns 2(12), 100366. https://doi.org/10.1016/j.patter.2021.100366.

Eisele, J. (2021). Traceability at the Farm Level in Argentina, Brazil, Paraguay & Uruguay. Traceability Webinar, May. https://textileexchange.org/app/uploads/2021/05/Traceability-Webinar-1-presentation-May-4-2021.pdf. Accessed 1 May 2022.

EOS (2022). EOS SAT Satellite Constellation To Serve Agricultural Needs. https://eos.com/eossat/. Accessed 1 August 2023.

Erling, U.M., Willms, N. and Stark, C. (2022). ESG Reporting and Greenwashing as Compliance Risk. News from Pohlmann & Company. https://www.pohlmann-company.com/en/esg-reporting-and-greenwashing-as-compliance-risk/.

EUR-Lex (2023). Regulation (EU) 2023/1115 of the European Parliament and of the Council of 31 May 2023 on the making available on the Union market and the export from the Union of certain commodities and products associated with deforestation and forest degradation and repealing Regulation (EU) No 995/2010 (Text with EEA relevance), OJ L. http://data.europa.eu/eli/reg/2023/1115/oj/eng. Accessed 1 August 2023.

European Commission (2020). Final report on Environmental claims in the EU: Inventory and reliability assessment. https://ec.europa.eu/environment/eussd/smgp/pdf/2020_Greenclaims_inventory.zip. Accessed: 20 January 2024.

European Commission (2023). The Commission adopts the European Sustainability Reporting Standards. News article. https://finance.ec.europa.eu/news/commission-adopts-european-sustainability-reporting-standards-2023-07-31_en. Accessed 4 February 2024.

European Parliament (2022). Towards deforestation-free commodities and products in the EU. Briefing: EU Legislation in Progress. https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/698925/EPRS_BRI(2022)698925_EN.pdf. Accessed 9 January 2022.

European Parliament (2023). Texts adopted - Deforestation Regulation - Wednesday, 19 April 2023. https://www.europarl.europa.eu/doceo/document/TA-9-2023-0109_EN.html. Accessed 6 October 2023.

European Space Agency (2019). New Financial Resources for Copernicus Space Component. https://www.esa.int/Applications/Observing_the_Earth/Copernicus/New_financial_resources_for_Copernicus_space_component. Accessed 6 July 2023.

European Union, Directorate-General for Environment (2023). Green Claims. New criteria to stop companies from making misleading claims about environmental merits of their products and services. https://environment.ec.europa.eu/topics/circular-economy/green-claims_en. Accessed 20 April 2024.

European Union Agency for Cybersecurity (2015). Readiness Analysis for the Adoption and Evolution of Privacy Enhancing Technologies. https://www.enisa.europa.eu/publications/pets.

European Union High-level Expert Group on Artificial Intelligence (2019). A definition of Artificial Intelligence: Main capabilities and scientific disciplines. Report/Study. https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines. Accessed 6 July 2023.

Fallows, D. (2005). How Women and Men Use the Internet. https://www.pewresearch.org/internet/2005/12/28/how-women-and-men-use-the-internet/. Accessed 6 July 2023.

Fedlex (2021). RS 632.324.27 - Ordonnance du 18 août 2021 sur l'importation au taux préférentiel d'huile de palme de production durable en provenance d'Indonésie. https://www.fedlex.admin.ch/eli/cc/2021/618/fr. Accessed 1 August 2022.

Fisher, A. and Streinz, T. (2021). Confronting Data Inequality. Columbia Journal of Transnational Law 60(3), 829–956. https://heinonline.org/HOL/P?h=hein.journals/cjtl60&i=841. Accessed 5 July 2023.

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. and Srikumar, M. (2020). Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. Rochester, NY. https://doi.org/10.2139/ssrn.3518482.

Food and Agriculture Organization of the United Nations (2004a). Decreto No 2.504/04 - Instituye el sistema de trazabilidad del Paraguay (SITRAP) para animales de la especie bovina. | FAOLEX. https://www.fao.org/faolex/results/details/en/c/LEX-FAOC052635. Accessed 1 August 2023.

Food and Agriculture Organization of the United Nations (2004b). National Forest Inventory - Field Manual. Working Paper 94/E. https://www.fao.org/3/ae578e/ae578e.pdf.

Food and Agriculture Organization of the United Nations (2018a). Zero-deforestation commitments: A new avenue towards enhanced forest governance?: Forestry Working Paper 3. https://books.google.com.eg/books?id=LWy1DwAAQBAJ.

Food and Agriculture Organization of the United Nations (2018b). Global Forest Resources Assessment 2020: Guidelines and specifications. https://www.fao.org/3/I8699EN/i8699en.pdf. Accessed 5 January 2022.

Food and Agriculture Organization of the United Nations (2018c). Resolución No 2.197 – Por la cual se aprueba el Reglamento del Sistema de Trazabilidad del Paraguay (SITRAP) en su cuarta versión. FAOLEX. https://www.fao.org/faolex/results/details/en/c/LEX-FAOC197460/. Accessed 1 August 2023.

Food and Agriculture Organization of the United Nations (2019). FAO - News Article: The biodiversity that is crucial for our food and agriculture is disappearing by the day. https://www.fao.org/news/story/en/item/1180463/icode/. Accessed 1 August 2022.

Food and Agriculture Organization of the United Nations and United Nations Environment Programme (2020a). The State of the World's Forests 2020. https://doi.org/10.4060/ca8642en.

Food and Agriculture Organization of the United Nations (2020b). Transforming agriculture and food systems: halting deforestation and promoting sustainable production and consumption of forest products. In Committee on Forestry, Twenty-fifth session, FAO (COFO/2020/7.1). https://www.fao.org/3/nd565en/nd565en.pdf. Accessed 1 August 2023.

Food and Agriculture Organization of the United Nations (2021a). FAO - News Article: Small family farmers produce a third of the world's food. https://www.fao.org/news/story/en/item/1395127/icode/. Accessed 1 August 2023.

Food and Agriculture Organization of the United Nations (2021b). The share of agri-food systems in total greenhouse gas emissions. 31(FAOSTAT Analytical Brief).

Food and Agriculture Organization of the United Nations (2021c). World Food and Agriculture - Statistical Yearbook 2021. Rome: FAO. https://www.fao.org/3/cb4477en/cb4477en.pdf.

Food and Agriculture Organization of the United Nations (2022a). FRA 2020 Remote Sensing Survey. http://www.fao.org/documents/card/en/c/cb9970en. Accessed 1 August 2023.

Food and Agriculture Organization of the United Nations (2022b). REDD+ Reducing Emissions from Deforestation and Forest Degradation. https://www.fao.org/redd/en/. Accessed 1 August 2023.

Gajos, K.Z. and Chauncey, K. (2017). The influence of personality traits and cognitive load on the use of adaptive user interfaces. In Proceedings of the 22nd International Conference on Intelligent User Interfaces. New York, NY, USA: Association for Computing Machinery (IUI '17). 301–306. https://doi.org/10.1145/3025171.3025192.

Garrido, G.M., Sedlmeir, J., Uludağ, Ö., Alaoui, I.S., Luckow, A. and Matthes, F. (2022). Revealing the landscape of privacy-enhancing technologies in the context of data markets for the IoT: A systematic literature review. Journal of Network and Computer Applications 207, 103465. https://doi.org/10.1016/j.jnca.2022.103465.

Genesereth, M. (2015). Computational Law: The Cop in the Backseat. White Paper. Codex - The Standard Center for Legal Informatics. http://complaw.stanford.edu/readings/complaw.pdf. Accessed 5 July 2023.

Germany Federal Ministry for Economic Affairs and Energy (2019). Making space for innovation: The handbook for regulatory sandboxes. https://www.bmwk.de/Redaktion/EN/Publikationen/Digitale-Welt/handbook-regulatory-sandboxes.html. Accessed 6 July 2023.

GHG Protocol (2014). GHG Protocol Agricultural Guidance. https://ghgprotocol.org/agriculture-guidance. Accessed 12 July 2023.

Ghorbani, A. and Zou, J. (2019). Data Shapley: Equitable valuation of data for machine learning. In Proceedings of the 36th International Conference on Machine Learning. International Conference on Machine Learning, PMLR. 2242–2251. https://proceedings.mlr.press/v97/ghorbani19c.html. Accessed 5 July 2023.

Gilardi, F. (2022). Digital Technology, Politics, and Policy-Making. 1st edn. Cambridge University Press. https://doi.org/10.1017/9781108887304.

Global Land Analysis & Discovery (2022). Global Forest Change. https://glad.earthengine.app/view/global-forest-change#dl=1;old=off;bl=off;lon=20;lat=10;zoom=3. Accessed 1 August 2023.

Global Times (2022). China issues 20 measures to build basic systems for data to boost devt, security - Global Times. Available at: https://www.globaltimes.cn/page/202212/1282215.shtml. Accessed: 7 February 2024.

Godar, J., Gardner, T.A., Tizado, E.J. and Pacheco, P. (2014). Actor-specific contributions to the deforestation slowdown in the Brazilian Amazon. Proceedings of the National Academy of Sciences 111(43), 15591–15596. https://doi.org/10.1073/pnas.1322825111.

Google AI Blog. (2017). Federated Learning: Collaborative Machine Learning without Centralized Training Data. [online] Available at: https://blog.research.google/2017/04/federated-learning-collaborative.html. Accessed 5 June 2023.

Green Policy Platform (2022). DECIDE Events. Digital Enabling of Circularity, Innovation, Development and Environment. https://www.greenpolicyplatform.org/initiatives/digital-enabling-circularity-innovation-development-and-environment. Accessed 1 August 2022.

Guyra Paraguay (2017). Informe extendido. Monitoreo Mensual del Cambio de Uso y Cobertura de la Tierra, Incendios y Variación de la Cubierta de Aguas en el Gran Chaco Americano.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A. *et al.* (2013). High-resolution global maps of 21st-century forest cover change. Science 342(6160), 850–853. https://doi.org/10.1126/science.1244693.

Harris, N.L., Goldman, E., Gabris, C., Nordling, J., Minnemeyer, S., Ansari, S. *et al.* (2017). Using spatial statistics to identify emerging hot spots of forest loss. Environmental Research Letters 12(2), 024012. https://doi.org/10.1088/1748-9326/aa5a2f.

Harvard University (2023). Differential Privacy. Harvard University. https://privacytools.seas.harvard.edu/differential-privacy.

Hellwig, D.P. and Huchzermeier, A. (2022). Distributed ledger technology and fully homomorphic encryption: Next-generation information-sharing for supply chain efficiency. In Innovative Technology at the Interface of Finance and Operations: Volume II. V. Babich, J.R. Birge, and G. Hilary (eds). Cham: Springer International Publishing (Springer Series in Supply Chain Management). 31–49. https://doi.org/10.1007/978-3-030-81945-3_2.

Hiemstra, D. (2009). Language models. In Encyclopedia of Database Systems. L. Liu and M.T. Özsu (eds). Boston, MA: Springer US. 1591–1594. https://doi.org/10.1007/978-0-387-39940-9_923.

Hoekstra, R. (2022). How natural capital accounting can help accelerate finance for nature. https://seea.un.org/sites/seea.un.org/files/how-nca-can-help-accelerate-finance-for-nature.pdf. Accessed 5 January 2023.

India, Ministry of Health and Family Welfare. (2022). Update on Ayushman Bharat Digital Mission. Press Information Bureau, Government of India. Available at: https://www.pib.gov.in/PressReleasePage.aspx?PRID=1813660. Accessed 24 Jan. 2024.

International Labour Organization (2022). Just Transition: An Essential Pathway to Achieving Gender Equality and Social Justice. Technical Paper. https://www4.unfccc.int/sites/SubmissionsStaging/Documents/202204141910---ILO%20submission%20-%20Just%20transition%20-%20An%20essential%20pathway%20to%20achieving%20gender%20equality%20and%20social%20justice.pdf. Accessed 6 July 2023.

International Monetary Fund. (2023). Pix: Brazil's Successful Instant Payment System. In IMF Staff Country Reports, (Issue 289). Available at: https://www.elibrary.imf.org/view/journals/002/2023/289/article-A004-en.xml. Accessed 24 Jan. 2024.

Intergovernmental Panel on Climate Change (ed.) (2014). Climate change 2014: mitigation of climate change: Working Group III contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. New York, NY: Cambridge University Press.

Intergovernmental Panel on Climate Change (2019). Special Report on Climate Change and Land — IPCC site. https://www.ipcc.ch/srccl/. Accessed 1 August 2022.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y. *et al.* (2023). Survey of hallucination in natural language generation. ACM Computing Surveys 55(12), 1–38. https://doi.org/10.1145/3571730.

Katz, D.M., Bommarito, M.J., Gao, S. and Arredondo, P. (2023). GPT-4 Passes the Bar Exam. Rochester, NY. https://doi.org/10.2139/ssrn.4389233.

Kennedy, J., Subramaniam, P., Galhotra, S. and Castro Fernandez, R. (2022). Revisiting online data markets in 2022: A seller and buyer perspective. ACM SIGMOD Record 51(3), 30–37. https://doi.org/10.1145/3572751.3572757.

Lambin, E.F., Gibbs, H.K., Heilmayr, R., Carlson, K.M., Fleck, L.C., Garrett, R.D. *et al.* (2018). The role of supply-chain initiatives in reducing deforestation. Nature Climate Change 8(2), 109–116. https://doi.org/10.1038/s41558-017-0061-1.

Löffler, F., Wesp, V., König-Ries, B. and Klan, F. (2021). Dataset search in biodiversity research: Do metadata in data repositories reflect scholarly information needs? PLOS ONE 16(3), e0246099. https://doi.org/10.1371/journal.pone.0246099.

MapBiomas (2020). MapBiomas Chaco General "Handbook" Algorithm Theoretical Basis Document (ATBD). https://mapbiomas-br-site.s3.amazonaws.com/MapBiomas%20CHACO/Colecao_2/ATBD/ATBD_-_MapBiomas_Chaco_Collection2.pdf. Accessed 30 April 2022.

MapBiomas (2022). MapBiomas Chaco. https://plataforma.chaco.mapbiomas.org/?activeBaseMap=9&layersOpacity=70&activeModule=coverage&activeModuleContent=coverage%3Acoverage_main&activeYear=2022&mapPosition=-25.987333%2C-61.741207%2C6&timelineLimitsRange=1985%2C2022&baseParams[territoryType]=1&baseParams[territories]=1%3BChaco%3B1%3BBioma%3B-33.86864986811106%3B-67.7200854863786%3B-17.54059819626697%3B-55.76232946817149&baseParams[activeClassesLevelsListItems]=1%2C7%2C8%2C9%2C10%2C2%2C11%2C16%2C17%2C18%2C19%2C3%2C12%2C13%2C20%2C21%2C22%2C14%2C15%2C4%2C5%2C6. Accessed 1 May 2022.

May, P. H. (2019), Valuing externalities of cattle and soy-maize systems in the Brazilian Amazon; Application of the TEEBAgriFood Evaluation Framework. TEEB for Agriculture and Food, UNEP. https://teebweb.org/wp-content/uploads/2020/12/TEEBAgriFood-Brazil-soy-cattle-FINAL.pdf.

Montgomery, O. (2022). Small Businesses Are Being Left Behind in the Retail Supply Chain Crisis. Software Advice. https://www.softwareadvice.com/resources/small-businesses-left-behind-in-retail-supply-chain-crisis/.

Mücher, C.A., Los, S., Franke, G.J. and Kamphuis, C. (2022). Detection, identification and posture recognition of cattle with satellites, aerial photography and UAVs using deep learning techniques. International Journal of Remote Sensing 43(7), 2377–2392. https://doi.org/10.1080/01431161.2022.2051634.

Narayanan, A. and Shmatikov, V. (2008). Robust de-anonymization of large sparse datasets. In 2008 IEEE Symposium on Security and Privacy (sp 2008). 111–125. https://doi.org/10.1109/SP.2008.33.

Nargesian, F., Zhu, E., Pu, K.Q. and Miller, R.J. (2018). Table union search on open data. Proceedings of the VLDB Endowment 11(7), 813–825. https://doi.org/10.14778/3192965.3192973.

Norton Rose Fulbright (2022). 10 things to know about the EU Corporate Sustainability Due Diligence proposal. https://www.regulationtomorrow.com/eu/10-things-to-know- about-the-eu-corporate-sustainability-due-diligence-proposal/. Accessed 24 February 2024.

Nunan, F., Campbell, A. and Foster, E. (2012). Environmental mainstreaming: The organisational challenges of policy integration. Public Administration and Development 32(3), 262–277. https://doi.org/10.1002/pad.1624.

Open Knowledge International (2023). The Open Definition. https://opendefinition.org/.

OpenAI (2023). GPT-4 Technical Report. arXiv. https://doi.org/10.48550/arXiv.2303.08774.

Organisation for Economic Co-operation and Development (2013). The OECD Data Privacy Framework. https://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf.

Organisation for Economic Co-operation and Development (2018), OECD Due Diligence Guidance for Responsible Business Conduct. Paris: OECD. https://mneguidelines.oecd.org/OECD-Due-Diligence-Guidance-for-Responsible-Business-Conduct.pdf.

Organisation for Economic Co-operation and Development (2019). Measuring distortions in international markets: The aluminium value chain. Paris: OECD. https://doi.org/10.1787/c82911ab-en.

Organisation for Economic Co-operation and Development (2022a). Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449. OECD.

Organisation for Economic Co-operation and Development (2022b). Environmental Impacts Along Food Supply Chains: Methods, Findings, and Evidence Gaps. Paris: OECD. https://doi.org/10.1787/48232173-en.

Organisation for Economic Co-operation and Development (2023). Fast and furious: the rise of environmental impact reporting in food systems. Paris: OECD. https://doi.org/10.1093/erae/jbad018.

Orrell, T. (2021). Towards a Framework for Governing Data Innovation. Fostering Trust in the Use of Non-Traditional Data Sources in Statistical Production. Brief on behalf of SDSN TReNDS, Open Data Watch, and DataReady. Available at: https://static1.squarespace.com/static/5b4f63e14eddec374f416232/t/6061c9994f9237430a22291d/1617021339239/Trust+in+Statistics-L3.pdf.

Ouellette, P., Sciortino, A., Nargesian, F., Bashardoost, B.G., Zhu, E., Pu, K.Q. *et al.* (2021). RONIN: Data lake exploration. Proceedings of the VLDB Endowment 14(12). https://doi.org/10.14778/3476311.3476364.

Pendrill, F., Persson, U.M., Godar, J. and Kastner, T. (2019). Deforestation displaced: Trade in forest-risk commodities and the prospects for a global forest transition. Environmental Research Letters 14(5), 055003. https://doi.org/10.1088/1748-9326/ab0d41.

Phiri, D., Simwanda, M., Salekin, S., Nyirenda, V.R., Murayama, Y. and Ranagalage, M. (2020). 'Sentinel-2 data for land cover/use mapping: A review. Remote Sensing 12(14), 2291. https://doi.org/10.3390/rs12142291.

PlanetScope (2022). PlanetScope. https://developers.planet.com/docs/data/planetscope/. Accessed 1 June 2022.

Pucker, K.P. (2021). Overselling Sustainability Reporting. Harvard Business Review, 1 May. https://hbr.org/2021/05/overselling-sustainability-reporting. Accessed 5 July 2023.

Rajotte, J.-F., Bergen, R., Buckeridge, D.L., El Emam, K., Ng, R. and Strome, E. (2022). Synthetic data as an enabler for machine learning applications in medicine. iScience 25(11), 105331. https://doi.org/10.1016/j.isci.2022.105331.

Ramachandran, G.S., Radhakrishnan, R. and Krishnamachari, B. (2018). Towards a decentralized data marketplace for smart cities. In 2018 IEEE International Smart Cities Conference (ISC2). 1–8. https://doi.org/10.1109/ISC2.2018.8656952.

Ramakrishnan, P., Haron, H. and Goh, Y.-N. (2017). Factors influencing green purchasing adoption for small and medium enterprises (SMEs) in Malaysia. International Journal of Business and Society 16(1). https://doi.org/10.33736/ijbs.552.2015.

Ritchie, H., Roser, M. and Ortiz-Ospina, E. (no date). Internet, OurWorldInData.org. https://ourworldindata.org/internet.

Satellite Imaging Corporation (2022). Land Cover Mapping, Change Detection. https://www.satimagingcorp.com/applications/environmental-impact-studies/land-cover-and-change-detection/. Accessed 1 August 2022.

Schomm, F., Stahl, F. and Vossen, G. (2013). Marketplaces for data: An initial survey. ACM SIGMOD Record 42(1), 15–26. https://doi.org/10.1145/2481528.2481532.

Schrage, M. (2020). Data, Not Digitalization, Transforms the Post-Pandemic Supply Chain. Digital-first enterprise success demands clarity-first supply chain design. The MIT Sloan Management Review [Preprint]. https://sloanreview.mit.edu/article/data-not-digitalization-transforms-the-post-pandemic-supply-chain/.

Schroeder, M. and Lodemann, S. (2021). A systematic investigation of the integration of machine learning into supply chain risk management. Logistics 5(3), 62. https://doi.org/10.3390/logistics5030062.

SENACSA (2022). SIGOR :: SENACSA. https://www.senacsa.gov.py/index.php/Temas-pecuarios/sigor. Accessed 1 August 2022.

Serdarasan, S. and Tanyas, M. (2012). Dealing with complexity in the supply chain: The effect of supply chain management initiatives. SSRN Electronic Journal [Preprint]. https://doi.org/10.2139/ssrn.2056331.

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. International Journal of Human–Computer Interaction 36(6), 495–504. https://doi.org/10.1080/10447318.2020.1741118.

Smith, E., Papadopoulos, D., Braschler, M. and Stockinger, K. (2022). LILLIE: Information extraction and database integration using linguistics and learning-based algorithms. Information Systems 105, 101938. https://doi.org/10.1016/j.is.2021.101938.

Snowflake (no date). Overview of Access Control. https://docs.snowflake.com/en/user-guide/security-access-control-overview. Accessed 5 July 2023.

Stanford University (2020). Artificial Intelligence definitions. Stanford University. https://hai.stanford.edu/sites/default/files/2020-09/AI-Definitions-HAI.pdf.

Suhara, Y., Li, J., Li, Y., Zhang, D., Demiralp, Ç., Chen, C. *et al.* (2022). Annotating columns with pre-trained language models. In Proceedings of the 2022 International Conference on Management of Data. New York, NY, USA: Association for Computing Machinery (SIGMOD '22). 1493–1503. https://doi.org/10.1145/3514221.3517906.

Sun Xiuping and Yang Tianpeng XU Chongbin, H.W.C.Q. (2022). Application of GEP accounting based on remote sensing satellite data. Spacecraft Recovery & Remote Sensing 43(4), 138–147. https://doi.org/10.3969/j.issn.1009-8518.2022.04.013.

Sun, Z. and Liu, B. (2023). Blockchain-based supply chain financial ciphertext retrieval system. In 2023 IEEE 2nd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA). Changchun, China: IEEE. 1431–1436. https://doi.org/10.1109/EEBDA56825.2023.10090690.

Süsser, D., Ceglarz, A., Gaschnig, H., Stavrakas, V., Flamos, A., Giannakidis, G. *et al.* (2021). Model-based policymaking or policy-based modelling? How energy models and energy policy interact. Energy Research & Social Science 75, 101984. https://doi.org/10.1016/j.erss.2021.101984.

Tang, F.H.M., Lenzen, M., McBratney, A. and Maggi, F., 2021. Risk of pesticide pollution at the global scale. Nature Geoscience. [Online] Available at: https://dx.doi.org/10.1038/s41561-021-00712-5. Accessed 4 February 2024.

Taskforce on Nature-related Financial Disclosures (2023). The TNFD Nature-related Risk and Opportunity Management and Disclosure Framework. London.

Tengö, M., Hill, R., Malmer, P., Raymond, C.M., Spierenburg, M., Danielsen, F. *et al.* (2017). Weaving knowledge systems in IPBES, CBD and beyond—lessons learned for sustainability. Current Opinion in Environmental Sustainability 26–27, 17–25. https://doi.org/10.1016/j.cosust.2016.12.005.

Tompkins, L. (2021). Hundreds of Companies Promised to Help Save Forests. Did They? The New York Times, 2 December. https://www.nytimes.com/2021/12/02/climate/companies-net-zero-deforestation.html. Accessed 5 July 2023.

Trase (2018). Supply chain mapping in Trase: Summary of data and methods. http://resources.trase.earth/documents/Trase_supply_chain_mapping_manual.pdf.

Trase (2020a). A vision for Trase: 2016-2020. http://resources.trase.earth/documents/Trase-vision-for-2020.pdf. Accessed 4 March 2022.

Trase (2020b). How Trase assesses commodity deforestation and commodity deforestation risk. http://resources.trase.earth/documents/data_methods/Trase_deforestation_risk_method_final%20Sept%202020.pdf. Accessed 1 June 2022.

Trase (2023). What is Trase? Methods and data. https://supplychains.trase.earth/about.

Tsai, M.-L., Ong, C.W. and Chen, C.-L. (2023). Exploring the use of large language models (LLMs) in chemical engineering education: Building core course problem models with Chat-GPT. Education for Chemical Engineers 44, 71–95. https://doi.org/10.1016/j.ece.2023.05.001.

Tubiello, F.N., Karl, K., Flammini, A., Gütschow, J., Obli-Laryea, G., Conchedda, G. *et al.* (2022). Pre- and post-production processes increasingly dominate greenhouse gas emissions from agri-food systems. Earth System Science Data, 14(4), 1795–1809. https://doi.org/10.5194/essd-14-1795-2022.

United Nations (1971). Consolidated document on the United Nations System and the human environment: submitted by the Administrative Committee on Co-ordination. 36–39. https://digitallibrary.un.org/record/3798019. Accessed 1 August 2023.

United Nations (2021). New FAO analysis reveals carbon footprint of agri-food supply chain | UN News. https://news.un.org/en/story/2021/11/1105172. Accessed 1 August 2023.

United Nations Roundtable on Digital Inclusion (2021). Digital Inclusion. https://www.un.org/techenvoy/sites/www.un.org.techenvoy/files/general/Definition_Digital-Inclusion.pdf. Accessed 5 July 2023.

United Nations Development Programme and India's G20 Presidency (2023). The DPI Approach: A Playbook. https://www.undp.org/publications/dpi-approach-playbook.

United Nations Development Programme (2023). The Human and Economic Impact of Digital Public Infrastructure. https://www.undp.org/publications/human-and-economic-impact-digital-public-infrastructure.

United Nations Economic Commission for Europe (2013). Meeting of Group of Experts on National Accounts – Interim meeting on Global Production, 12th Session, Geneva, 3-4 April 2013. https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.20/2013/Working_Paper_2.pdf. Accessed 31 May 2022.

United Nations Environment Assembly (2019). UNEP/EA.4/Res.1. https://documents-dds-ny.un.org/doc/UNDOC/GEN/K19/010/42/PDF/K1901042.pdf?OpenElement.

United Nations Environment Programme (2011). Towards a Green Economy: Pathways to Sustainable Development and Poverty Eradication - A Synthesis for Policy Makers. Available at: https://wedocs.unep.org/20.500.11822/32245.

United Nations Environment Programme (2013). Frequently Asked Questions on Corporate Sustainability Reporting Tackling the big questions around the global corporate sustainability reporting agenda. https://wedocs.unep.org/bitstream/handle/20.500.11822/26171/FAQ_Corporate_Sustainability.pdf?sequence=1&isAllowed=y.

United Nations Environment Programme (2015). Raising the bar- Advancing Environmental Disclosure in Sustainability Reporting. https://wedocs.unep.org/bitstream/handle/20.500.11822/9807/-Raising_the_Bar_-_Advancing_Environmental_Disclosure_in_Sustainability_Reporting-2015UNEP_Raising_the_Bar_2015.pdf.pdf?sequence=3&isAllowed=y.

United Nations Environment Programme (ed.) (2019). Global Environment Outlook – GEO-6: Healthy Planet, Healthy People. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108627146.

United Nations Environment Programme (2022). Supply chain monitoring: Expert exchange on connecting three monitoring domains. DISCAP initiative page. https://www.greenpolicyplatform.org/event/supply-chain-monitoring-expert-exchange-connecting-three-monitoring-domains.

United Nations Environment Programme (2022). Glossary of key terms. Preparatory works. Intergovernmental negotiating committee to develop an international legally binding instrument on plastic pollution, including in the marine environment. UNEP/PP/INC.1/6. https://wedocs.unep.org/bitstream/handle/20.500.11822/40683/K2221529%20-%20UNEP-PP-INC.1-6%20-%20Advance.pdf.

United Nations Environment Programme (2023). Emissions Gap Report 2023: Broken Record – Temperatures hit new highs, yet world fails to cut emissions (again). https://doi.org/10.59117/20.500.11822/43922.

United Nations Framework Convention on Climate Change (no date). Action Pledges. https://www4.unfccc.int/sites/NWPStaging/Pages/ActionPledges.aspx. Accessed 3 July 2023.

United Nations Framework Convention on Climate Change (2018). Latin America South-South Exchange - Sustainable livestock and its relationship with forests. https://redd.unfccc.int/uploads/2234_33_infobrief_latin_america_en_-_v4_lores.pdf.

United Nations Framework Convention on Climate Change (2022a). [ARCHIVED CONTENT] Glasgow Leaders' Declaration on Forests and Land Use - UN Climate Change Conference (COP26) at the SEC – Glasgow 2021. https://webarchive.nationalarchives.gov.uk/ukgwa/20230418175226/https://ukcop26.org/glasgow-leaders-declaration-on-forests-and-land-use/. Accessed 1 October 2022.

United Nations Framework Convention on Climate Change (2022b). What is the triple planetary crisis? https://unfccc.int/news/what-is-the-triple-planetary-crisis. Accessed 23 January 2024.

United Nations Women (2018). Making women and girls visible: Data for gender equality. In Turning Promises into Action: Gender Equality and the 2030 Agenda for Sustainable Development. 46–69. https://doi.org/10.18356/2956c59a-en.

United States Geological Survey (2022). What is the Landsat satellite program and why is it important? https://www.usgs.gov/faqs/what-landsat-satellite-program-and-why-it-important. Accessed 1 August 2023.

Vaghefi, S., Wang, Q., Muccione, V., Ni, J., Kraus, M., Bingler, J. *et al.* (2023). ChatClimate: Grounding Conversational AI in Climate Science. Rochester, NY. https://doi.org/10.2139/ssrn.4414628.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N. *et al.* (2017). Attention is all you need. In Advances in Neural Information Processing Systems. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html. Accessed 5 July 2023.

Veit, P. and Sarsfield, R. (2017). Land rights, Beef Commodity Chains, and Deforestation Dynamics in the Paraguayan Chaco. Washington, DC: USAID Tenure and Global Climate Change Program. [Preprint]. https://www.land-links.org/wp-content/uploads/2017/06/USAID_Land_Tenure_TGCC_Paraguay_Risk_Assessment_June-2017.pdf.

van de Ven, M., Abbas, A., Kwee, Z. and Reuver, M. (2021). Creating a Taxonomy of Business Models for Data Marketplaces. in. https://doi.org/10.18690/978-961-286-485-9.23.

Visipec (2020). Visipec Frequently asked questions. https://www.visipec.com/wp-content/uploads/2020/02/Frequently_Asked_Questions.pdf.

Warnat-Herresthal, S., Schultze, H., Shastry, K.L., Manamohan, S., Mukherjee, S., Garg, V., *et al.* (2021). Swarm Learning for decentralized and confidential clinical machine learning. Nature 594(7862), 265–270. https://doi.org/10.1038/s41586-021-03583-3.

Wilkinson, M.D. *et al.* (2016) 'The FAIR Guiding Principles for scientific data management and stewardship', Scientific Data, 3(1), p. 160018. Available at: https://doi.org/10.1038/sdata.2016.18.

Winter, M., Dopler, S., Müller, J.M. and Zeisler, A. (2023). Information sharing and multi-tier supply chain management of SMEs in the context of Industry 4.0. Procedia Computer Science 217, 1378–1385. https://doi.org/10.1016/j.procs.2022.12.336.

World Bank (2018). Breaking the 'Grass Ceiling': Empowering Women Farmers. https://www.worldbank.org/en/news/feature/2018/03/06/breaking-the-grass-ceiling-empowering-women-farmers. Accessed 6 July 2023.

World Bank. (2023a). Improved Population and Civil Registration and Digital Identification for Better and More Inclusive Service Delivery in Indonesia. [online] Available at: https://www.worldbank.org/en/news/press-release/2023/05/15/improved-population-and-civil-registration-and-digital-identification-for-better-and-more-inclusive-service-delivery-in-indonesia. Accessed 24 Jan. 2024.

World Bank (2023b). Open Data in 60 seconds. https://opendatatoolkit.worldbank.org/en/data/opendatatoolkit/open-data-in-60-seconds.

World Economic Forum (2022). Supply Chain Sustainability Policies: State of Play. White Paper. https://www3.weforum.org/docs/WEF_Supply_Chain_Sustainability_Policies_ 2022.pdf. Accessed 04 March 2024.

World Economic Forum (2024). Advancing Digital Public Infrastructure for the Agriculture Sector. Briefing paper. https://www3.weforum.org/docs/WEF_Advancing_Digital_Public_Infrastructure_Agriculture_Sector_2024.pdf. Accessed 28 February 2024.

World Resources Institute (2023). Traceability and Transparency in Supply Chains for Agricultural and Forest Commodities. A review of success factors and enabling conditions to improve resource use and reduce forest loss. Report. https://doi.org/10.46830/wrirpt.22.00156. Accessed 10 January 2024.

World Trade Organization (2022). Members note work in three environment initiatives and discuss anti-deforestation efforts. https://www.wto.org/english/news_e/news22_e/envir_02feb22_e.htm. Accessed 1 August 2023.

World Wildlife Fund (no date). Sustainable Agriculture Beef. https://www.worldwildlife.org/industries/beef#:~:text=Beef%20production%20has%20a%20considerable,emissions%20from%20human%2Drelated%20activities. Accessed 1 August 2022.

Xu, X., Sanford, T., Turkbey, B., Xu, S., Wood, B.J. and Yan, P. (2022). Polar transform network for prostate ultrasound segmentation with uncertainty estimation. Medical Image Analysis 78, 102418. https://doi.org/10.1016/j.media.2022.102418.

Yüksel, A.S. and Karabıyık, M.A. (2023). TraViQuA: Natural language driven traffic video querying using deep learning. Traitement du Signal 40(2), 543–553. https://doi.org/10.18280/ts.400213.

This publication is supported by the Environment Fund - UNEP's core financial fund. The Fund is used to provide scientific evidence on the state of the global environment, identify emerging environmental issues and innovative solutions, raise awareness and advocacy, bring together stakeholders to agree on action, and for building capacity of partners. Core funding gives UNEP the strength and flexibility to implement the programme of work (in support of the 2030 Agenda) as approved by its Member States, and to strategically respond to emerging challenges. UNEP is grateful to all the Member States that contribute to the Environment Fund.

For more information: unep.org/environment-fund

UN environment programme

For more information:
unep-communication-director@un.org
United Nations Avenue, Gigiri
P O Box 30552, 00100
Nairobi, Kenya

unep.org