White Paper

How to use the power of AI to reduce the impact of climate change on Switzerland

Recommendations for the Swiss society and economy to become more resilient against the impact from a radically changing climate

Make key technologies broadly available and overcome challenges through key methodologies in climate- and AI-related fields.

Swiss Academy of Engineering Sciences SATW
St. Annagasse 18 | 8001 Zürich | +41 44 226 50 11 | info@satw.ch | www.satw.ch

6 Vision, Gaps, Opportunities, and Actions to Enable AI to Reduce Climate-Change Impact

Technical disruptions like geospatial data and AI at scale to effectively reduce climate-change impact can only unfold their potential with infrastructure, methodologies, and policies in place. However, if done right, the opportunity for the Swiss society is vast. Thus, in this chapter, we discuss the vision, gaps, opportunities, and actions along six key enabling dimensions: i) data discovery and accessibility, as the foundation of AI applications, ii) robust machine learning at scale, discussing the need to successfully transition AI models from research to production, iii) Swiss communities and activities, supporting the uptake and guidance of AI model exploration, iv) Swiss geo & climate ICT infrastructure, needed to provide such services at scale, v) recommendations for sustainability transition applications, to result profitable business models and finally, vi) responsible and inclusive AI, to proliferate climate recommendations relevant for any stakeholder in Switzerland.

6.1 Data Discovery and Accessibility

Authors: Gregory Giuliani[1,2], Marvin Höge[3], Andreas Scheidegger[3], Claudia Röösli[4], Diana-Denisa Rodila[1,2], Christian Spindler[5], Marcio Dos Reis[6], Frank Emorisier[7], Benjamin Stocker[8], Apolonio Huerta[9], Manuel Kugler[10], Alisa Freyre[11]

[1]University Geneva, Switzerland

[2]UNEP/GRID-Geneva, Switzerland

[3]EAWAG, Switzerland

[4] University of Zurich, Switzerland

[5] Sustainaccount AG, Zurich, Switzerland

[6] Agroscope, Switzerland

[7] Picterra, Switzerland

[8] University Bern, Switzerland

[9]2Blades, Switzerland

[10]SATW, Switzerland

[11] PanData, Switzerland

6.1.1 Vision

Earth Observations (EO) are increasingly produced by many environmental monitoring systems at all scales (from local to global). These systems are generating huge data volumes on a daily and weekly basis making their management and processing a major challenge. EO data can significantly contribute to establishing the baseline for determining trends, defining present conditions, and informing future evolution of the earth and its climate. EO data have the potential to drive progress against key national and international development agendas providing new insights and support better policy making across diverse issues of environmental sustainability. However, the vision of EO data-driven decision making has not been fully addressed, challenges remain, and therefore the full information potential of EO data has not been yet realized. Consequently, efforts have to be made to facilitate the discoverability, accessibility and use of EO data for a more sustainable future. Actions should promote open, coordinated, and sustained data sharing and infrastructure to strengthen research, policy making, decisions and action across different disciplines.

6.1.2 Current state

There is an increasing need for translating the massive amount of climate data and information that already exists into customized tools, products, and services to monitor the range of climate change impacts and their evolution. It is crucial to pre-process the recorded raw data to be readily available for the various analysis tasks of different stakeholders, to unfold the maximal utility.

There is an unprecedented array of new satellite technologies with capabilities for advancing our understanding of environmental processes and the assess changes at scales from local plots to the entire planet. Currently, almost 50 instruments and more than 10 satellites with multiple instruments that are of broad interest to the environmental sciences that either collected data in the 2000s, were recently launched, or are planned for launch in this decade (Figure 6.1). If we look only at Landsat (NASA) and Sentinels (ESA) they generate 30'000 images per day. In 2021, the total Sentinel data volume available for retrieval from the Copernicus Data Access System was 41.86 Petabyte, with a total download volume of 80.5 Petabyte[54].

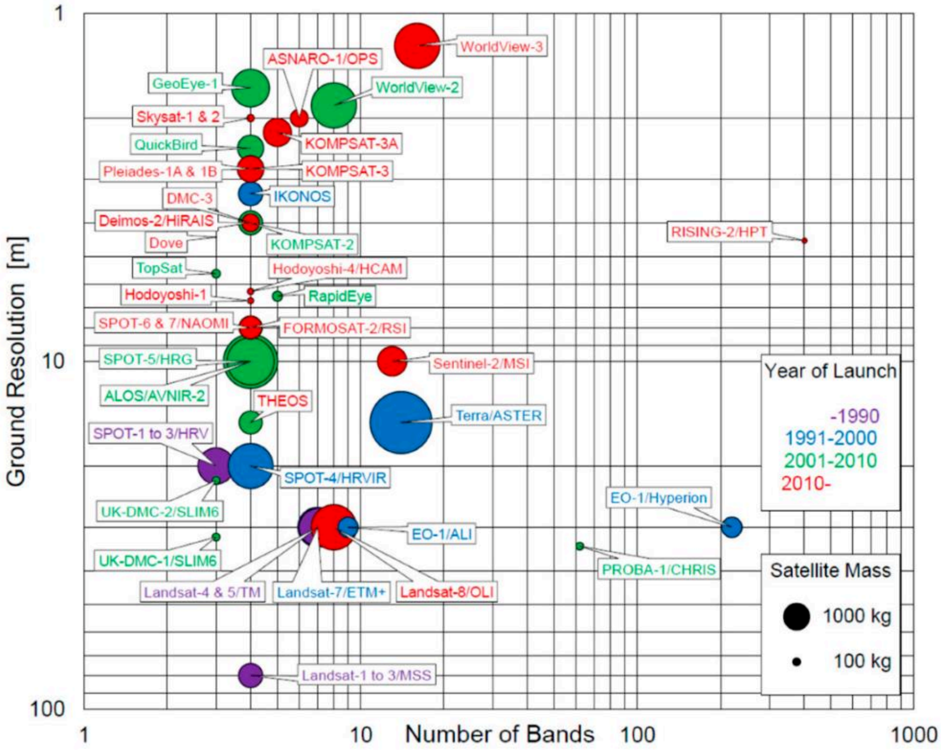

Figure 6.1: Overview chart depicting spatial and spectra (number of bands) resolution of Earth observation satellites (Kurihara et al., 2018).

However, most of these systems are designed for a specific purpose and therefore are operating in silos contributing to the fragmentation of the EO landscape and ultimately making hard to discover and access these invaluable resources. This in particular is true for commercial satellites fleets with high spatio-temporal resolution.

Further, EO data is clearly posing challenges related to Big Data handling, especially Volume (amount of data); Variety (diversity of data types & source); Velocity (speed of data generation); Value (extraction of meaningful information) and Veracity (accuracy of data). Addressing Big Data challenges

---

[54] https://scihub.copernicus.eu/twiki/pub/SciHubWebPortal/AnnualReport2021/COPE-SERCO-RP-22-1312_-_Sentinel_Data_Access_Annual_Report_Y2021_merged_v1.0.pdf

requires a paradigm change away from traditional local data processing approaches towards data-centric processing, to lower the barriers caused by data size.

To tackle the mentioned issues emerging global trends of (1) free and open data access policies for Landsat and Sentinel data; (2) the increasing provision of Analysis Ready Data (ARD) from EO satellites and (3) the distribution of open source software for managing and exploiting EO data, enables monitoring environmental changes at various spatial and temporal scales while complementing traditional data sources such as national statistics, administrative data or census information. Significant work has recently been done to lower barriers and facilitate the access of end-users to harness the full potential of EO data, and to address mandates, national processes, or reporting obligations. Earth Observation Data Cubes (EODCs such as Digital Earth Australia, Digital Earth Africa, or the Swiss Data Cube) and cloud-based processing platforms (such as the Copernicus DIASes, Earth on Amazon, Google Earth Engine, MS Planetary Computer) have emerged as technology enablers to manage, access, and analyse Big EO Data, thereby strengthening connections between data providers, applications, and end-users**.**

In addition, the advancement of data management and sharing principles such as FAIR (Findable-Accessible-Interoperable-Reusable) together with new cloud-optimized data formats (Cloud-Optimized Geotiff, Zarr) and Application Programming Interfaces like the Spatio-Temporal Asser Catalog (STAC) are paving the way to efficient and effective discovery and access to large volumes of EO data while facilitating spatio-temporal analysis over a given region anywhere in the world.

Ultimately, having access to open data will provide many opportunities/benefits[55], such as (1) supporting broad bconomic benefits and growth, (2) enhancing social welfare, (3) growing research and innovation opportunities, (4) facilitating the education of new generations and (5) benefits for effective governance and policy making.

6.1.3 Gaps, limitations, and concerns

Data interoperability ensures a seamless exchange of data between systems, applications, and services. It enables data discovery and transfer across diverse sources, facilitating analytics, data integration, and sharing. There are two key types of interoperability, which are:

1. ***Data-level interoperability:*** Enables data sharing across applications and platforms.
2. ***Semantic-level interoperability:*** Ensures data is correctly interpreted by various systems.

These interoperability levels can be achieved following a three steps procedure:

1. Data-level interoperability is achieved through infrastructures or platforms. They collect data from diverse sources, store it in a dedicated repository, and provide access in various formats. This ensures common data formats (e.g., COG, Zarr) and protocols (OGC APIs & webservices) for analytics and machine learning.
2. Semantic-level interoperability is attained by adding metadata and linking data to a standardized vocabulary. Data standards ensure data are correctly interpreted by different AI systems, leading to uniform, consistent datasets.

---

[55] https://www.earthobservations.org/documents/open_eo_data/GEO-XII_09_The%20Value%20of%20Open%20Data%20Sharing.pdf

3. Establishing a data vocabulary linked to an ontology can be done through two methods: data mapping (unifying data elements) and data federation (sharing data as if from a single source). These standards facilitate data sharing across businesses without relying on other information systems.

Other data exchange concerns and solutions are:

- ***Common APIs and Protocols***: Standardized APIs and protocols like REST are vital for data discovery and accessibility and supporting machine learning tool interoperability. REST, using HTTP and JSON, is widely adopted for exposing services as web APIs, allowing seamless communication for model training, inference, and data sharing. This is the case for openEO[56] or PANGEO[57], as well as the emerging OGC APIs[58].
  By embracing such interoperability standards, data providers and ML practitioners create interoperable services, promoting ease-of-consumption by other systems. These APIs establish a common communication language and facilitate the integration of multiple frameworks and tools into a unified system.
- ***Metadata and Standards:*** Interoperability goes beyond models and APIs, encompassing data exchange and metadata. Standardized metadata and data formats are crucial for sharing, understanding, and using data across various machine learning frameworks. Initiatives like Data Catalog Vocabulary (DCAT), Data Package, ISO 19115/19139 series and Open Geospatial Consortium (OGC) establish common metadata standards, enhancing data-level interoperability. In particular the Spatio-Temporal Asset Catalog (STAC) is becoming the de facto standard for meta data of spatio-temporal data[59].
  Metadata standards like DCAT[60] offer a common vocabulary for describing datasets, simplifying discovery, and understanding. Data Package focuses on packaging data and metadata for easier sharing. OGC CSW and OGC Records API allows implementing catalogs and search interfaces for metadata while ISO standards provide the necessary elements to produce standardized data descriptions also known as metadata. Adopting these standards ensures seamless integration of datasets into different frameworks and tools, improving data discovery, interoperability, and collaboration. Adopting these standards ensures seamless dataset integration across frameworks, improves data discovery, enhances interoperability, and fosters efficient data-driven collaborations.

6.1.4 Opportunities for Swiss stakeholders

Improving the discoverability, interoperability, and accessibility to quality-controlled data using web-based technologies could optimize the regular update of inventories of current and planned climate data records. This could reduce the time invested in finding datasets and could allow linking data records with relevant documentation and user feedback. Furthermore, this could facilitate the browsing, filtering, retrieval, and ingestion of these records into automated data processing pipelines, in order to help the generation of community-driven tools, libraries (e.g., Python client libraries) and added value applications across catalogues.

---

[56] https://openeo.org/
[57] https://pangeo.io/
[58] https://ogcapi.ogc.org/
[59] https://stacspec.org/
[60] https://www.w3.org/TR/vocab-dcat-3/

6.1.5 Recommendations and actions

Best Practice: To tackle the challenges of data and machine learning interoperability, consider these approaches:

- ***Community Collaboration:*** Promote cooperation among developers, researchers, and practitioners through industry-wide consortiums, open-source initiatives, and forums. This fosters consensus, standardization, and knowledge sharing, providing a platform for collective problem-solving.
- ***Standards and Best Practices:*** Encourage adopting established standards for interoperable data discovery and access (e.g. OGC) as well as for model interchange. Develop and document best practices for data storage, management, discovery, access and model integration and deployment, emphasizing performance optimization and compatibility. Comprehensive guides and documentation can aid practitioners.
- **Tooling and Integration Libraries:** Create tools and libraries to simplify model conversion and framework communication. These should automate common interoperability tasks, easing the workload for developers. Integration libraries can abstract framework complexities, allowing developers to focus on their applications.
- **Research and Innovation:** Invest in ongoing research for improved interoperability. Explore techniques for model adaptation, cross-framework optimization, and efficient serialization to minimize performance trade-offs. Collaboration between academia and industry can drive innovative approaches and address emerging challenges.
- **Interconnection of infrastructures:** different models such as data federation[61], system of systems (GEOSS[62]) or data spaces[63] can help enabling interoperability between systems, infrastructures, and platforms.

6.1.5.1 Recommendations for decision makers

- Ensure that open government data principles are implemented by agencies.
- Facilitate the discovery and access to data made available by governmental agencies (e.g. MeteoSwiss, swisstopo, BAFU, FSO).
- Bring interoperability along the entire data value chain. It will facilitate storing, visualizing, accessing, processing, analyzing, and integrating climate data and information and enables users to add create value-added products and services.
- Delivering climate services using interoperable web services can lower the barriers for both data providers and data users. In particular, it can enhance the reusability of data and components in various applications, and get increased return on investment.
- Interoperable climate services together with corresponding technical and scientific capacities can play a crucial role in ensuring quality, integrity and availability of datasets and consequently promoting, contributing and supporting research activities and a trusted open science.
- Governance, policies, and institution: Open Data policies and Data Sharing and Management Principles are spreading in various communities and this will strongly influence climate community as well as institutions that are providing data and delivering climate services.

---

[61] https://earthserver.xyz/
[62] https://www.earthobservations.org/geoss.php
[63] https://dataspaces.info/

- Development of capacities at human, institutional, and infrastructure levels: Building capacities will help to reach large adoption, acceptance and commitment on data sharing principles. It will also strengthen the capacity of scientists to provide usable and understandable information to decision makers and convince data holders to make available their data to a wide audience facilitating data discovery, access, and processing.

6.1.5.2 Recommendations for stakeholders

- Make as much data available under FAIR principles and keep as little as possible proprietary.
- Publish data on trusted public digital repositories (like Zenodo), ideally that provide interoperability arrangements.
- Develop capacities in Data Science at the human, institutional, and infrastructure levels.
- Adhere to Open Data policies and Data Sharing and Management Principles.
- Contribute to relevant governance bodies.

References

Kurihara, J.; Takahashi, Y.; Sakamoto, Y.; Kuwahara, T.; Yoshida, K. HPT: A High Spatial Resolution Multispectral Sensor for Microsatellite Remote Sensing. Sensors 2018, 18, 619.
https://doi.org/10.3390/s18020619

Ustin, Susan L., and Elizabeth M. Middleton. "Current and near-term advances in Earth observation for ecological applications." Ecological Processes 10 (2021): 1–57.