

Multi-source Multi-wavelength Data Challenges in Earth System & Environment

Jean-Pierre Vilotte⁽¹⁾ & Frédéric Huynh⁽²⁾

(CNRS-INSU)

(2) Director IR Data Terra & PIA3 Gaia Data, IRD Montpellier



Séminaire Aristote: Données de la Recherche, Nature, Cultures et Valeurs Sorbonne Université, 18 Octobre 2023



- ⁽¹⁾ Scientific Deputy Intensive HPC and HPDA, AI, & Open science, Institut des Sciences de l'Univers



Earth System, Climate & Environment: From observation to Models

A complex dynamical system

- Coupled geophysical, bio-chemical and environmental processes,
- With different spatial and temporal scales,
- And permanent interactions between the solid Earth, continental surfaces, and fluid envelopes (ocean, atmosphere),
- As with the anthroposphere.

Evolution of the needs

- > Integrated approaches to complexity and uncertainty quantification
- > Multi-physics & multi-scale stochastic simulations
- > Multi-domain, multi-sensor data enabling multi-wavelength analysis (*spatial, spectral, temporal & textual*)
- > In-situ & space data coupling: e.g. data assimilation, inference
- \succ Long-term data series and services
- > Integrated Artificial intelligence and big data services
- > Inter- and trans-disciplinary science/societal challenges

Geomagnetic eismometers & GPS measurement Volcani Generation of omagnetic field

Scientists and decision makers need to have an integrated interdisciplinary approach for multi-domain data access and services together with ondemand AI-enabled data analysis and modelling of increasing resolution, uncertainty quantification, and extreme natural events characterisation



Explosion of Data fluxes and diversity

Data Ubiquity



InSight mission



Copernicus/Sentinel







New generation of discoveries relies upon:

- **Data acquisition**: streaming data processing/reduction/compression, primary data delivery
- **Observation continuity:** long time / on-demand short time duality
- **Data Archiving and management**: long-term archiving, curation (interoperable data reference models, provenance, distribution)
- **Data FAIR services**: multi-source virtual observatory services, DOA









Volcano observatories





Data explosion (rate, volume, diversité):

- Edge environnements: observation, monitoring systems
- **Centralised environments (**Cloud, HPC): large ensemble of simulations, data assimilation, High-end Data Analytics

- High-performance statistical data analysis: Al-based distributed multisource data analysis
- **ELarge ensemble simulations**: multi-physics and multi-scale systems, uncertainty quantification, extreme events
- **Probalistic inversion/assimilation:** high-dimensional probabilistic inference methods
- Wide-area workflows (HPC/HDA): data logistics in a multi-provider Edgeto-Centralised infrastructure continuum

A multi-dimensional challenge



Data -> Extraction/cleaning -> Integration/aggregation -> Learning models -> trigger / question -> predict

Data Terra: French Research E-Infrastructure for Earth System & Environment

Х

GOALS

Provide an integrated system to access data and data products with processing and analysis services for observing, understanding and forecasting in an integrated way the history, functioning and evolution of the Earth system & Environment subject to global changes

Integrated platform : Earth system sciences data, services and products.

- → 26 Research Organisations and Universities
- → 30 Observing Data and Services Centres (CDOS)
- → 32 Consortium of Scientific Expertise (CES)
- → 450 scientists, engineers and technicians







Through the CNRS-INSU:

→ Leverage, within the Centres of observational data and services (CDOS), the SNOs articulated by the territorial network of the Observatoires des Sciences de l'Univers (OSUs)





Data Terra: A unique access to the Earth System Data and Services





- project added values



actices g and each	Distributed data access, interoperability and processing	National reference centre for Earth System & environment		Earth sys & environ commun stewards	
Software repository	Identity & Access management	Secu- rity	Hypervision metrics		Ma Actio D
	Data Lakes Multi-sources and distributed data repositories	Federated Dat a & Processing services		Long-terr data arch & steward	
Services	Distributed processing				
RID / CLOUD DAT/FranceGrille) (EuroHPC, Prace/GENCI)					

Data Terra: Services operated through the GAIA Data platform

Integrated multi-domain data and computing services across the infrastructures continuum

Ongoing Digital Twins projects

Extremes, Climate, digital Ocean, Geo-hazards, Territory

• Thematic applications:

- combining Copernicus, EO data, socio-economic information, scenarios ...
- From fundamental research to decision making and economic activities (risk mitigation, sustainable urban areas, energy transition, water resources)

Related National Priority Programs: e.g.

- One Water: "Water, Common Good",
- FairCarbonN: Carbon in continental ecosystems, trajectories for carbon neutrality
- Sous-Sol: géothermie, H2 and CO2 storage, water
- TRACCS: Transforming Climate modelling for climatic services
- Digital Ocean: toward a digital twin of the ocean system
- NumPEx: Numerics for exascale











Data Terra: National Thematic Reference Centre for Earth System & Environment

An ecosystem for sharing open research data: Federated, Accessible, Interoperable & Reusable (FAIR)









Data Terra: European and International Positioning





Data Terra within EOSC

- → A French EOSC thematic node for Earth System, **Environment and Biodiversity**
- ➡ A set of services integrated into the EOSC portal

Contributes to a number of EOSC projects through the CNRS-INSU

FAIR EASE (01/09/2022, 5M€, 26 partenars, coord: Data Terra/CNRS A. Rizzo, IRD). **Objectif: develop et leverage within EOSC the distributed and integrated observation** and modelling for the Earth System, Environment and Biodiversity, with the user communities and the European Science Research Infrastructures. 143 PM/1,2M€

FAIR IMPACT (01/06/2022, 10M€, coord: DANS). Objectif: develop a FAIR EOSC « EOSC de données et de services FAIR », by supporting the practical implementation of the FAIR principles in the scientific communities and for research results at the national, European and international levels. 14PM

GEO INQUIRE (01/10/2022, 15M€, coord: GFZ). Objectif: Access to observations, data products and services for monitoring and simulating the geosphere dynamic processes, at unprecedented levels of detail and accuracy. 8PM

DT GEO (01/09/2022, 10M€, good: BSC). Objectif: Deploy a digital twin prototype for geophysical and human-made environmental extreme events: Coupled Digital Twin Components (DTC) for risk associated to earthquakes and induced seismicity, volcanic eruptions, and triggered tsunamis (earthquakes, landslides). 4PM

others: BioDT, InterTwin





EOSC infrastructure node at the European level





Data Terra: characterise, analyse and model the Earth atmosphere

Dynamics, physics and chemistry of the Earth atmosphere Climate changes and air quality services

Data are produced by instrumental platforms, observation systems The pole of data and services **AERIS** federates at the national level the data (satellites, ballons, aircrafts, in-situ distributed sensors), scientific field management activities and the scientific expertise for the Earth atmosphere. campaigns, laboratory experiments, earth system model simulations.

Mutualise multisource air quality Pic-du Mid-Observati monitoring. ATTAL - ATmospheric Observations in CESAM, Paris (NRS-1,15) HELIOS, Drivers CNRS-CARE Browne Mitchie (MA), Toulous Site EMEP-1815 Tandana

Estimate and quantify the carbon emissions and particules





What thematics for the territoires? Air quality example





Create a European monitoring network and services for the air quality

2021 07 01

LATMOS - ULB IASI CO total column (x 10¹⁸ molec./cm²)

4 4.5 5 5.5 6 3.5

Monitor and track the effects of forest fires for air quality







AERIS: Tracking smoke plumes

Smoke plumes can be observed by different means during forest fires



Space observation

Forest fires can be observed by geostationary satellites. Data are collected, processed and made available to users by the ICARE data centre.

Animation of the forest fire in Canada, July 2022, realised by AERIS with the application GEORING https://www.icare.univ-lille.fr/ geostationary-satellites-data/









Ground networks



The European research infrastructure ACTRIS for the observation and exploration of aerosols, clouds and réactif gazes delivers lidars data that can be used to observed and monitor forest fires from ground. AERIS is the data centre of ACTRIS-FR.

Data Terra: Earth structure, composition and telluric risks

Understand the Solid Earth structure, composition and dynamical processes Telluric hazards and risk mitigation, geo-resources

The pole of data and services **FORM@TER** federates at the national level the data management activities and the scientific expertise for the solid Earth.



GDM-OPT GDM-OPT signifie Ground Deformation Monitoring using OPTical image time series

Merci de vous connecter avant de commencer. (Lors de la première connexion, suivre les instructions)



DSM-OPT signifie «Digital Surface Models OPTICAL stereoscopic



GDM-SAR signifie Ground Deformation Monitoring using inSAR image time series



Spatio-temporal monitoring of critical zones



Data are multi-soupe and multi-wavelengths (space observation, satellite missions, in-situ instrumentation, lab experiments, and model simulations results. Workflows and computing/visualisation services are made available.

Which thematics for the territories ?



Ground deformation calculation from spatial imagery



FLATSIM

ion avec des équipes des laboratoires ISTerre, LGL-

Surface Déformation at continental scales Spatio-temporal monitoring of critical zones (local, regional)

- active faults and magmatic systems
- Iandslides
- Regional lithospheric deformations





ForM@Ter: computing and visualisation services

Le pôle de données et de services pour l'océan Odatis fédère au niveau national des activités de gestion de données et d'expertise scientifique pour les océans.

GDM-OPT (SLIDE, ETQ, ICE)

Optical Correlation Sentinel 2

Visualisation of temporal series (<u>TIO</u>, Volat et al., 2017)

Results made available and public

DSM-OPT : **MNS** from Pléiades images

Predefined parameters depending on the landscape type of the studied zone

Stereo images requested directly via DINAMIS

MNS catalogue

Automatisation in the CIEST framework

Economical model

FLATSIM : inSAR with complet Sentinel-1 archive Second AI 07/2022 superficie > 10 Millions km2 Start to be in production au CNES Use case in the PEPR NumPEx







www.poleterresolide.fr







Data Terra: Ocean dynamics and biodiversity

The pole of data and services Odatis federates at the national level the data management activities and the scientific expertise on the Oceans

to Topor (Triscopy Coastline tracking DYNALIT







www.odatis-ocean.fr

The data source are from satellite missions, in-situ instruments (fixed platforms, float arrays , gliders, radars sea campaigns, laboratory measurements, ocean models simulations)



ODATIS: Colors, temperatures, turbidity

Studies issued from the recommandations of the CES Ocean colours of the ODATIS pole, funded by the CNES, development ACRI.

Spatial data



Couleur de l'eau - Concentration en Chlorophylle-a, capteur MODIS



- Surface temperatures and turbidity
- Chlorophylle concentration

Diffusion end September with graphical interface: export of data file fichier, sélection of in-situ point and comparaison with satellites. data ...







Température de surface, capteur MODIS



Turbidité, capteur MODIS







Data Terra: Continental surface dynamics and natural resources

The pole of data and services Theia federates at the national level the data management and the scientific expertise on ecosystems and environmental resources of continental surfaces and interfaces.

Which thematics for the territories?

Urban Land use spot









The multi-sources data come from satellites missions and in-situ instruments. They support the development of data products, models and software with added value on the there agriculture, biodiversity, climate, water, forest, littoral, snow and ice, natural risks, health, urban added values.



Sentinel-2 FAPAR over the Canvargue, Rance (2018-09-2)







Snow cover





Biomass production

Theia: Hydroweb.next service

Centralised access to free hydrological data: satellites, in-situ, model simulations.



Satellite data products, model simulations results from the the CES water

The Vosges exam ple: Search, online access and visualisation













End date of snowfall in the Vosges



tMap contributors, CARTO



Theia: Digital cartography of soils

Visualisation of soil data in Occitanie

Examples of available cartography data products

- Soil properties properties (clay content, limon, PH and soil organic carbon)
- Soil reservoir map



Lagacherie, P., Arrouays, D., Walter, C., 2013. Cartographie numérique des sols : principe, mise en œuvre et potentialités. Etude Gest. des Sols 20, 83-98







Grille de 90mx90m



Data Terra: Enable and ease access to HR spatial images

DINAMIS: a national institutional device for mutualised supply of high resolution satellite imagery

• French authorised institutional users

 Scientific researchers, Ministry of education and territorial actors

• Authorised foreign institutional researchers

For cooperation with South countries, European union and international collaborations

• **Private users (France)**

For R&D projects leaders and public providers











dinamis.data-terra.org

Couverture métropolitaine 2020. Copyright ADS - IRD, INRAE, IGN

Littoral Sénégal. Pléiades Copyright CNES



Montpellier. Pléiades © CNES 2020, distribution Airbus DS

An Institutional offer

Specific Pléiades covers

Sandy littorals sableux (metropolitan, Guyane, DOM-TOM,..)

Millenium SPOT 6-7 France metropolitain covers

Acquisitions ad-hoc Pléiades, and SPOT 6-7 Worldwide - Expressed by the UIAs, whatever is the affectation.

Complementary HR imageries

Relay of the program CNES Spot World Heritage (Spot 1-5), Plateform CNES PEPS (Sentinel 2), Geosud, programme CNES Kalideos : RapidEye, CosmoSkyMed, TerraSar,-X, Aster...





VAGUE DE CHALEUR EN EUROPE

Eté 2022 Les scientifiques travaillent à comprendre les vagues de chaleur en Europe, leur fréquence ainsi que leur intensité afin de permettre aux sociétés civiles de mieux se préserver et de réussir à inverser la tendance.

La chaleur extrême persistante amène également à mieux appréhender la notion de climat au sein de la société civile.



Feux de forêt et vaque <u>de chaleur – un été</u> particulièrement <u>intense</u>

Estimation du volume de glace détaché lors de leffondrement du glacier de la Marmolada

Estimation du volume de glace détaché lors de l'effondrement du glacier au Kurdistan



<u>Vague de chaleur</u> juin 2022 en mer <u>Méditerranée</u>

L'ampleur de la sécheresse rendue visible par les satellites

Data Terra: Some new challenges

Transdisciplinary context of the Earth System and environment societal

- Multi-type of data (quantitative and qualitative, including textual data): new challenges for data veracity, cross AI-enabled data analysis and data interoperability
- **Augmented Data ecosystem:** not only research data but data produced by a wide diversity of territorial and private actors New challenges in terms of data access policy (privacy, confidentiality), data veracity, provenance, and data licensing
- - New territorial landscapes: zones ateliers, observatories, ...

New pressing technical and societal issues in terms of

- * **Ethics and epistemology:** uncertainty quantification, explainability and social adhesion to be developed with the SHS
- * Al-enabled data analysis in operational context: robustness, extreme events and statistical estimators
- * New observation systems and digital objects: distributed AI-enabled data analysis across the Edge-to-Cloud/HPC continuum, data logistics
- * *Efficiency versus frugality:* optimisation at the system level of the energy and environmental impact of large distributed observation E-infrastructures: not only technical but related to energy transition, new territories, and society (need to include the SHS)
- * **Resilience to the energy fluctuation and volatility:** a new paradigm

