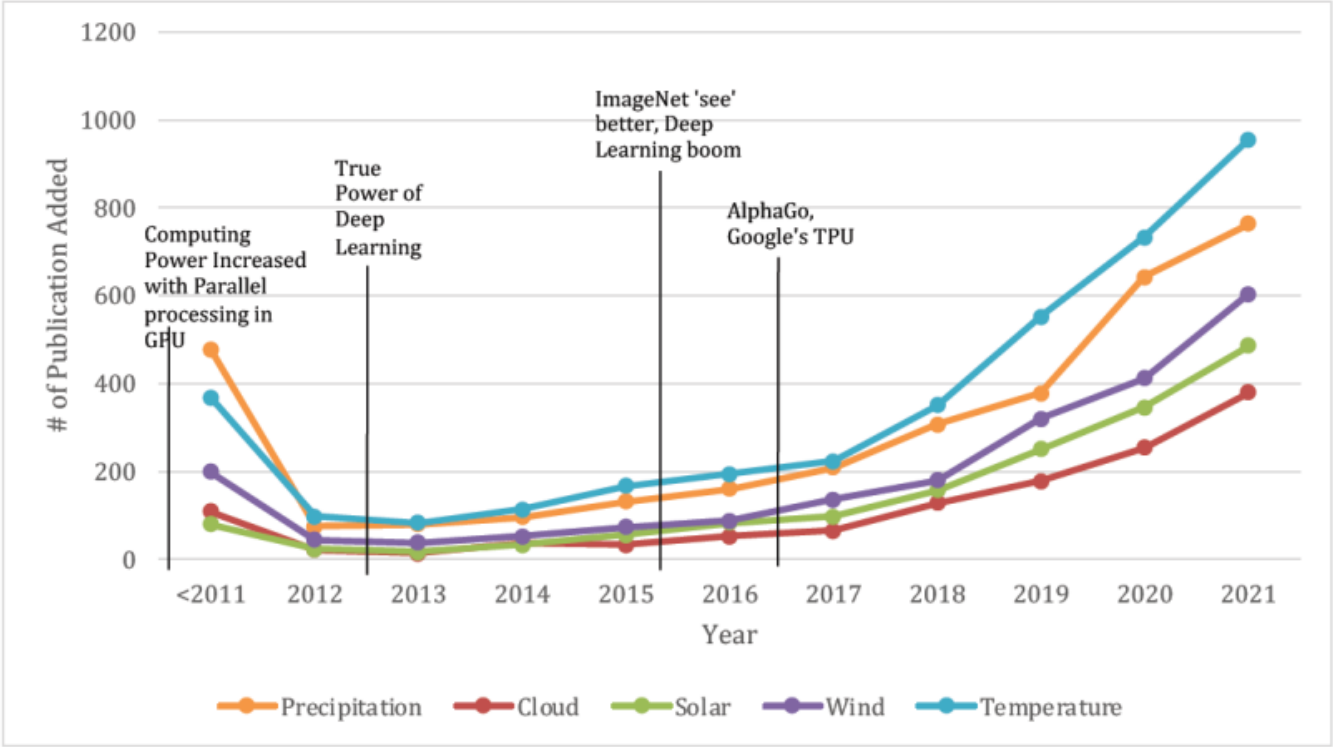
EUMETSAT/NOAA plans for AI/ML exploitation

Rick Miner, Tyler Christensen, Rob Redmon (NOAA)
Paolo Ruti (EUMETSAT)

… the fast revolution ?



noaa.gov/AI

Our Vision

Benefiting NOAA's mission by proliferating the use of Responsible AI at NOAA.

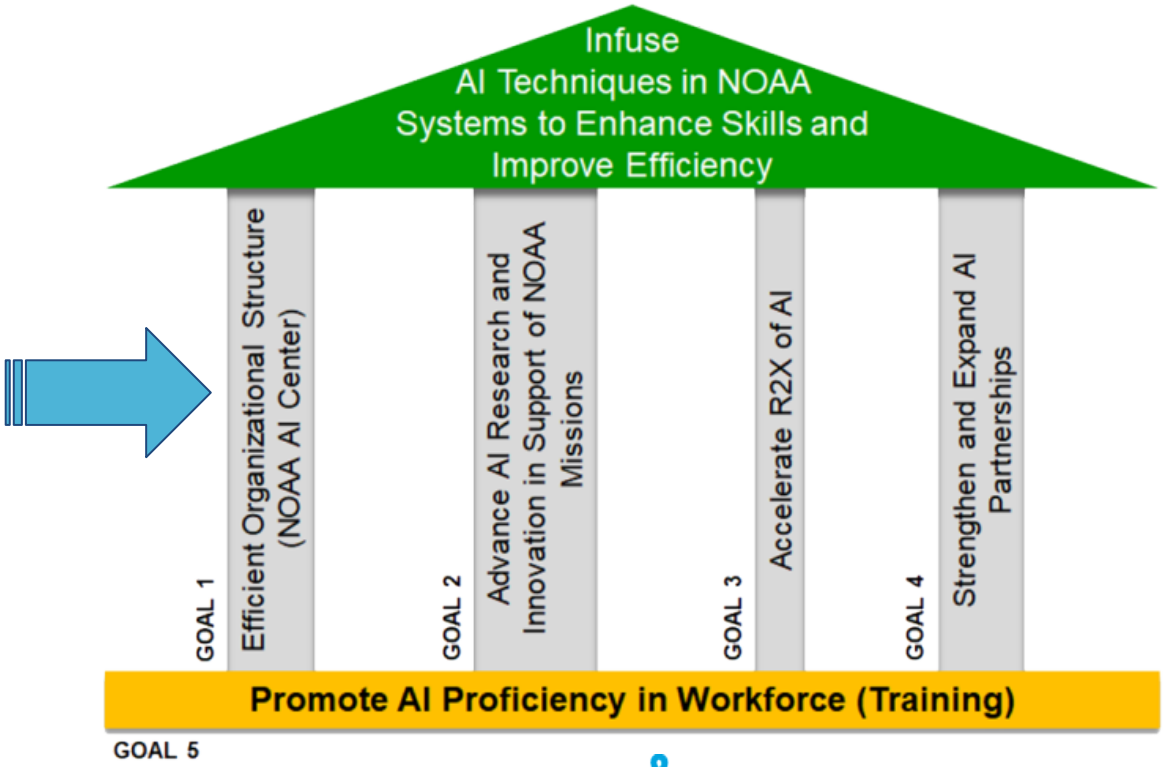
How?

We'll do this by lowering the cost of engaging curiosity for our Community of Practice

Coordination Group for Meteorological Satellites

CGMS Add CGMS agency logo here (in the slide master)

Coordination Group for Meteorological Satellites - CGMS



Coordination Group for Meteorological Satellites

CGMS Add CGMS agency logo here (in the slide master)

Coordination Group for Meteorological Satellites - CGMS

A major question for EUMETSAT is about the applicability and utility of Artificial Intelligence (AI) and Machine Learning (ML) to the upstream part of the weather/climate value chain (i.e., thinning of satellite data, gap analysis, retrieval of L2 data) to strengthen the products.

At the same time, it will be strategic to pursue collaborations and studies on key downstream applications where EUMETSAT can play a strong role in working with member's countries and international agencies and multinational initiatives to further exploit the EUMETSAT satellite products and synergies with Copernicus.

Links to Strategy 2030

As for the delivery of operational services, the roadmap supports EUMETSAT in exploiting the latest advances in remote sensing and algorithmic sciences

The AI and ML roadmap will contribute to the integration of the Satellite Application Facilities (SAF) into EUMETSAT product's development

The AI and ML roadmap will guide EUMETSAT in planning IT infrastructure and software evolution to guarantee agility and flexibility towards a more AI/ML oriented data access.

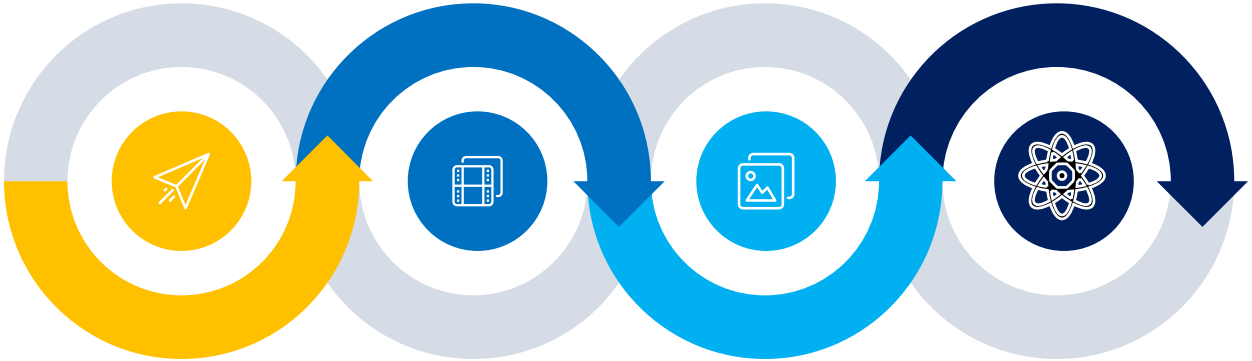
As for the cooperation with other satellite operators and the contribution to global partnership, the AI and ML roadmap will contribute to further enhance the EUMETSAT international role and strengthen the bilateral portfolio

Coordination Group for Meteorological Satellites

CGMS Add CGMS agency logo here (in the slide master)

THE ROADMAP

The EUMETSAT AI/ML roadmap approved by Council in 2022, will guide EUMETSAT in supporting member states. The roadmap provides an initial stimulus to further explore application areas. As an example, ESOC and EUMETSAT organized a joint workshop to explore operational applications.



2022/23

- M1 - Establishment of a dedicated AI/ML call for joint project (supported through fellowships)
- M2 - Building EUMETSAT AI/ML coordination team
- M3 - Explorative workshops for new action areas

2023/24

- M4 - Running initial projects as Pathfinders for AI/ML
- M5 - Tailoring EWC software for AI/ML
- M6 - Developing training data set for Pathfinders

2024/25

- M7 - Workshop on AI/ML EUMETSAT applications
- M8 - Large AI/ML context (i.e., MOOC)
- M9 - Evaluation from STG and SWG

2025/26

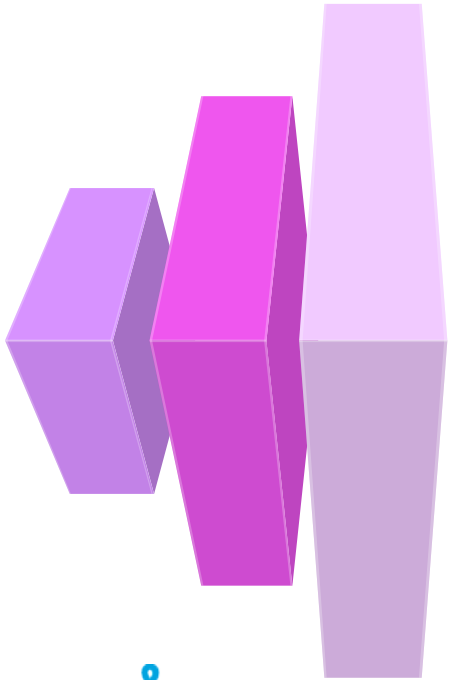
- M10 - Evaluation of AI/ML roadmap by PAC
- M11 - Approving and establishing Phase 2

What does AI-ready mean for NOAA and EUMETSAT?

Community driven AI ready Data

Cloud AI ready Data

Operationally AI ready Data

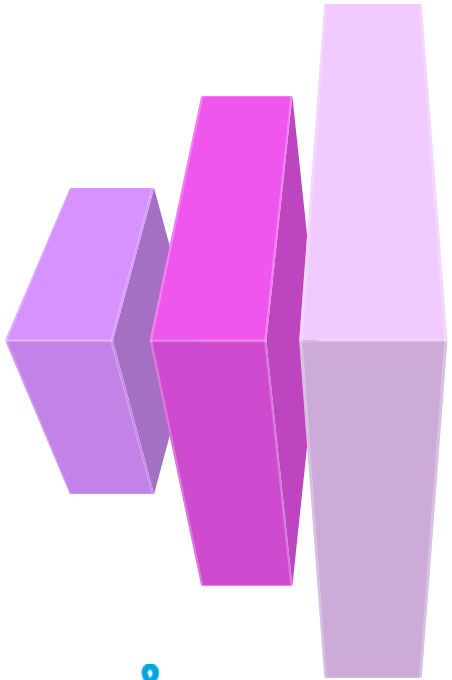


What does AI-ready mean for NOAA and EUMETSAT?

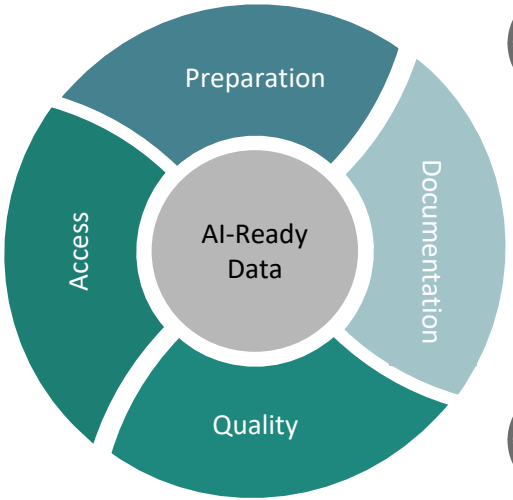
Community driven AI ready Data

Cloud AI ready Data

Operationally AI ready Data



Community-driven AI-Ready Data



Documentation enables users to understand the contents of a data set and should provide information and tools to ***increase data usability***.

Quality* information** determines the dataset's "fit-for-purpose***" for AI & affects the trustworthiness of AI applications.

Access affects the ***efficiency & reproducibility*** for AI R&D process.

Preparation help identify common ***data services and tools*** to reduce preprocessing burden for users including AI practitioners.

Earth Science Information Partners (ESIP) Data Readiness Cluster is a forum for community members from agencies (NOAA, NASA, USGS, DOE, USAF, etc.), private sectors, and academia to:

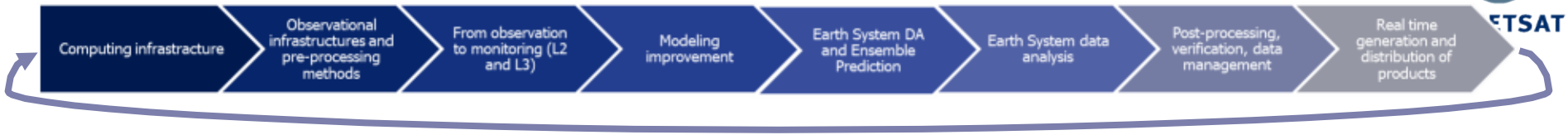
- Understand users' data need for AI/ML R&D with environmental data
- Develop community standards, leading practices & tools for AI-ready data

AI-ready data checklist
https://doi.org/10.6084/m9.figshare.19983722.v1

Coordination Group for Meteorological Satellites

CGMS Add CGMS agency logo here (in the slide master)

Community-driven AI-Ready Data



Cloud preparedness

Accelerating the satellite algorithm production through AI/ML

Objects and Processes Detection

Working on Training and Training Data Sets

Developing Nowcasting and very short range forecasting methods

Climate reconstruction and Environmental parameters

Improving the Earth System modelling system

Scouting user requirements for future missions and new products

EUMETSAT members' priority areas

Coordination Group for Meteorological Satellites

CGMS Add CGMS agency logo here (in the slide master)

Community-driven AI-Ready Data

Create Tropical Cyclone Model Training Dataset

Challenge

AI-ready and accessible benchmark satellite datasets are needed to drive the future of tropical cyclone trajectory, intensity and coastal impact prediction (e.g. coastal flooding and other infrastructure damage).

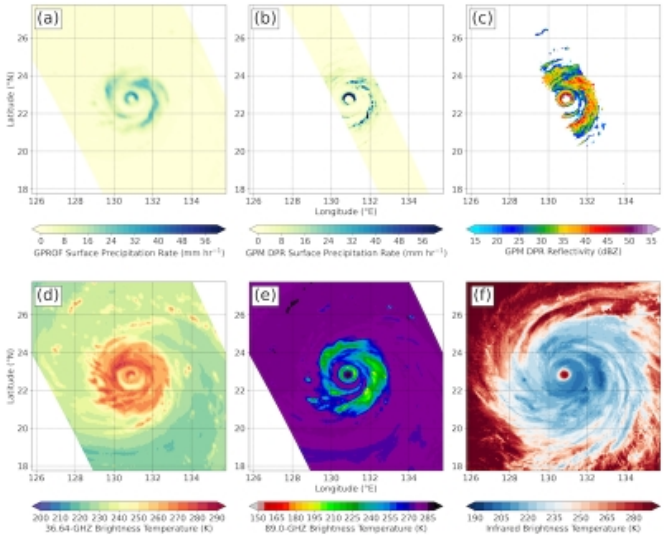
Description and Expected Outcomes

Evaluate a new dataset's AI-readiness against NCAI draft standards and make necessary changes to brand it as AI-ready. The new dataset, Tropical Cyclone PRecipitation, Infrared, Microwave, and Environmental Dataset (TC PRIMED), collocates and subsets LEO/GEO satellite imagery with ancillary model information to create a 22-yr dataset of TC-centric scenes. This dataset will supersede NCEI's HURSAT.

NCAI Benefits

AI-ready standard maturation; Lesson Learned via interactive Python notebook ("Learning Journey"); increased collaboration with NOAA's NODD (previously BDP).

POC: Chris Slocum (NOAA/NESDIS/STAR), christopher.slocum@noaa.gov



A sampling of TC PRIMED products from Typhoon Maria (2018) at 10:13 UTC on 9 July 2018 in the western Pacific, where a) is GPROF, b) is GPM DPR precipitation rate, c) GPM DPR reflectivity, d) 36.6 GHz, e) 89 GHZ, and IR from Himawari-8.

Coordination Group for Meteorological Satellites

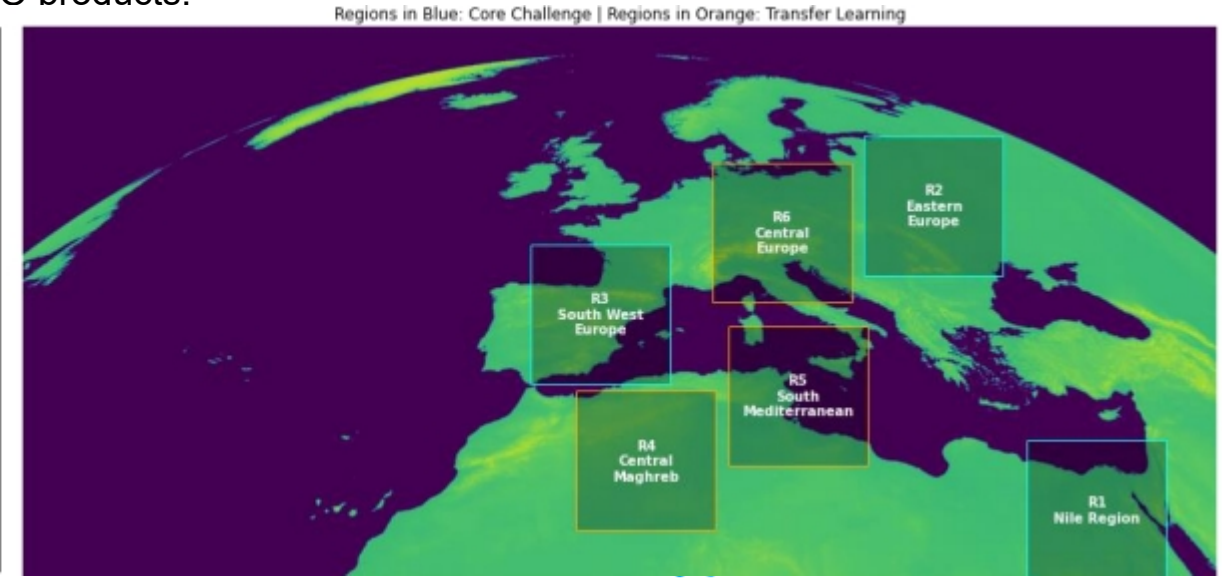
CGMS Add CGMS agency logo here (in the slide master)

Community-driven AI-Ready Data

EUMETSAT

Wheater4cast competition “The goal is a short-term prediction (8 hours in 15 minute intervals) of the selected weath products The weather movies consist of multi-channel images encoding the cloud properties, temperature, turbulence, and rainfall, based on meteorological satellite data obtained in collaboration with AEMET/NWC SAF” W have provided 1 year of NWC SAF GEO products.

- Regions R1, R2 and R3 for developing the ML models
- Regions R4, R5 and R6 for testing the models



Magnetic Navigation: MagNet Challenge

X-Prize to forecast Space Weather effects on Magnetic Navigation

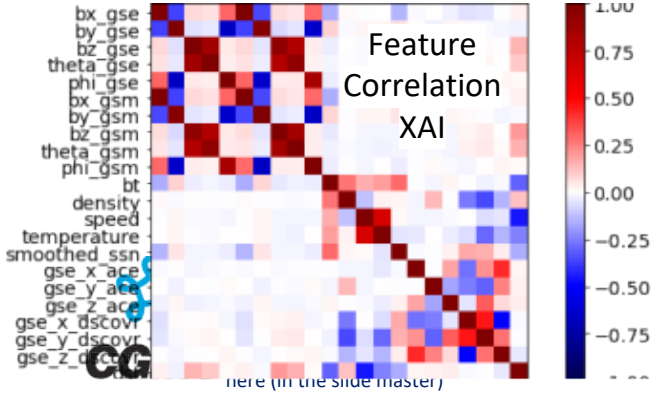
- ***Challenge*** *- to improve precision magnetic navigation during times of heightened space weather, e.g. for efficient critical mineral exploration.*
- ***NCEI Innovates*** - funded a competition to improve an ML model to predict a key Space Weather index from DSCOVR satellite solar-wind data, mitigating impacts on magnetic navigation; *600 participants and ~1200 model submissions*

www.ncei.noaa.gov/products/geomagnetic-data

DRIVENDATA

	User or team		RMSE
	Ammarali32	1	11.1311
	belinda_trotta	2	11.2532
	LosExtraterrestres	3	11.2944
	k_squared	4	11.5293

Dr. Belinda Trotta
Melbourne, Australia
Senior Software Eng.
2nd Place



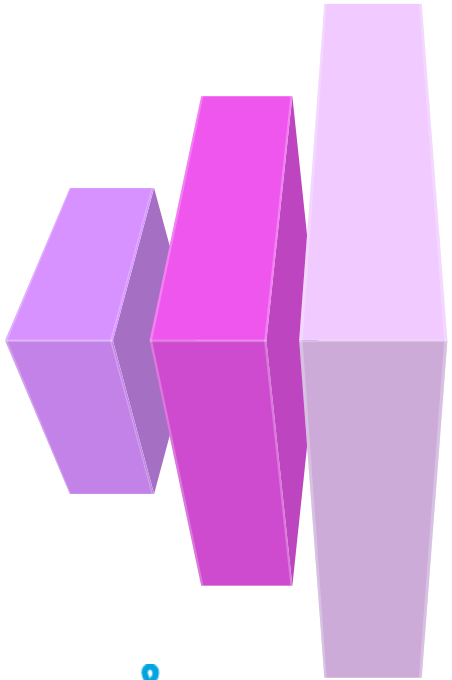
here (in the slide master)

What does AI-ready mean for NOAA and EUMETSAT?

Community driven AI ready Data

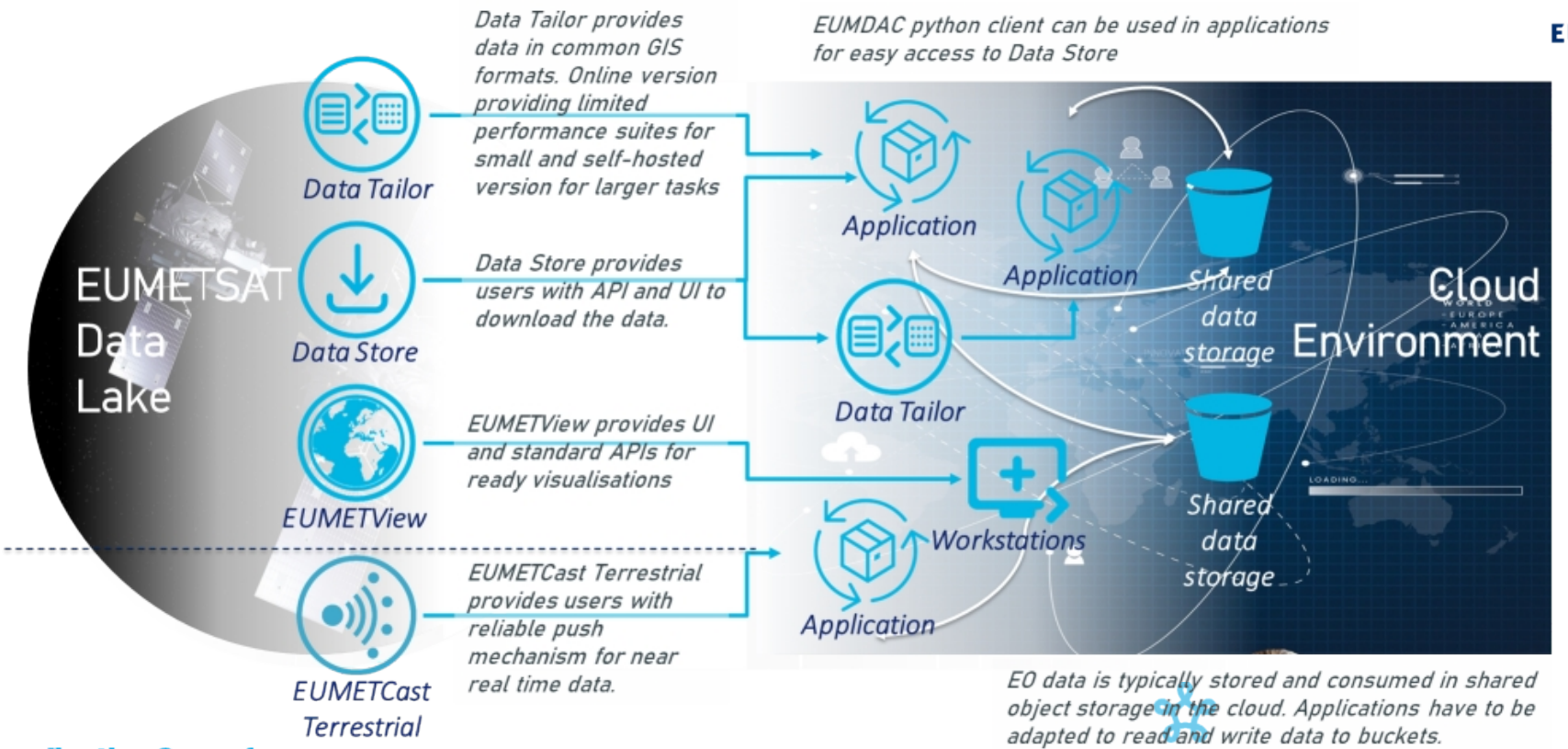
Cloud AI ready Data

Operationally AI ready Data



Cloud AI-Ready Data

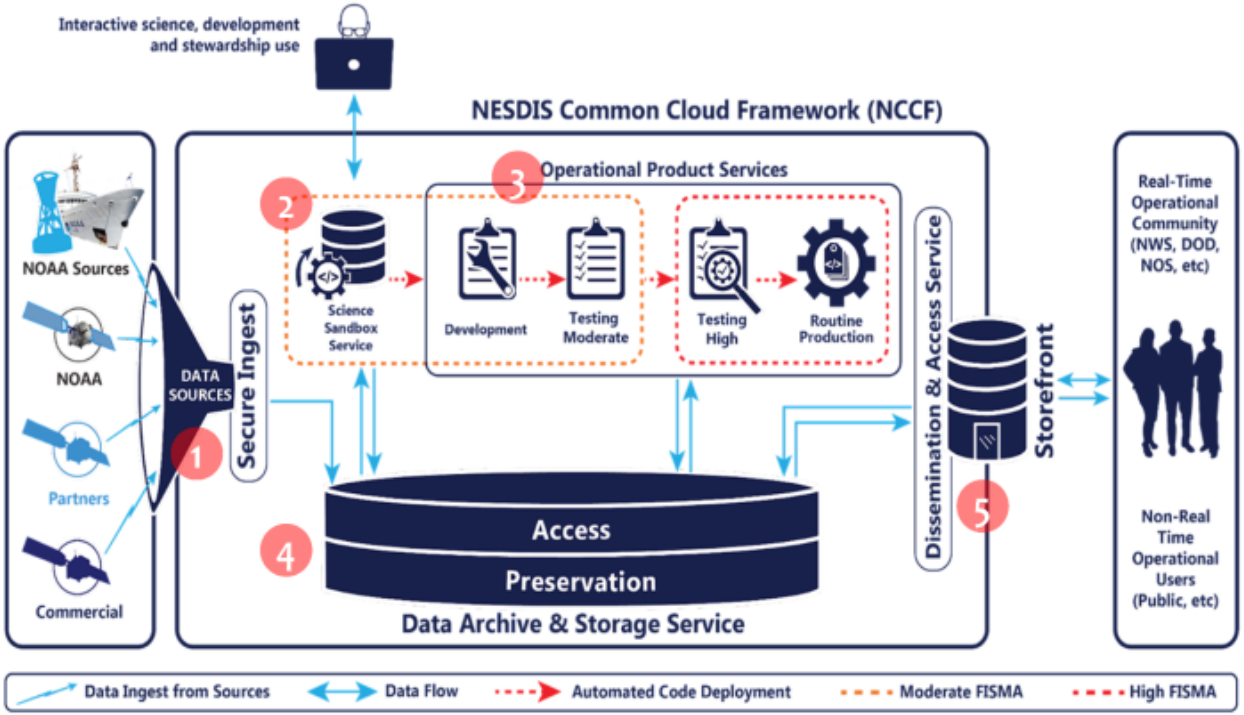
EUMETSAT



Coordination Group for Meteorological Satellites

CGMS Add CGMS agency logo here (in the slide master)

Cloud AI-Ready Data



NESDIS is building an enterprise cloud-based ground system, including cloud data dissemination

Coordination Group for Meteorological Satellites

CGMS Add CGMS agency logo here (in the slide master)

Cloud AI-Ready Data

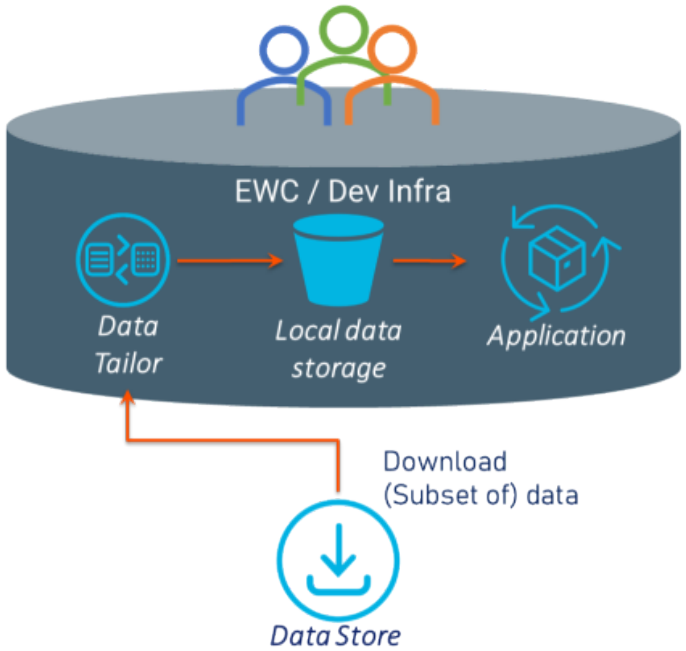
Typical Machine-learning use case

EUMETSAT

Ensure that people get necessary skills:

- Using object storage and other distributed storages
- Building stateless applications
- Containers & Kubernetes
- CI/CD & ansible / cloud formation / terraform
- Monitoring, gathering logs, serverless...

Creating teams that include a person with cloud skills is probably the most practical way right now



1) Developers copy tiny subset of data to their development infrastructure
2) Development of the ML model in local infrastructure
3) Training with larger dataset, probably on larger infrastructure
4) Running the inference

Cloud AI-Ready Data

Cloud Optimized Data Lake of Archived Water Column Sonar

Challenge

AI-ready Water Column Sonar (WCS) data are needed to add additional value in support of fisheries acoustic science and fisheries management.

Description and Expected Outcomes

Develop a data lake of archived WCS data translated into cloud-friendly formats to improve interoperability and scaled processing.

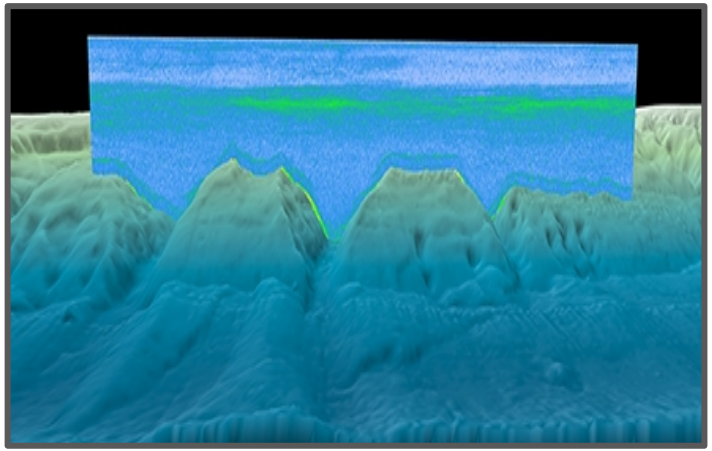
- The complex, binary, and proprietary formatted data are hosted on AWS currently through the NODD and are not AI-ready.
- Data will be transformed and made AI-ready and an interactive notebook will be developed for others to learn from.

Benefits

This effort will lower the learning curve to analyze the data more easily, especially in a cloud environment.

NCAI Benefits

AI-ready data, lessons learned as interactive notebook for NCAI's Learning Journey library.



Water column sonar data collected on the NOAA Okeanos Explorer in the North Atlantic Ocean. Sonar data are overlaid onto coastal relief model bathymetry. (Source)

POC: Carrie Wall Bell (NOAA/NESDIS/NCEI and CIRES), carrie.wall@noaa.gov

Coordination Group for Meteorological Satellites

CGMS Add CGMS agency logo here (in the slide master)

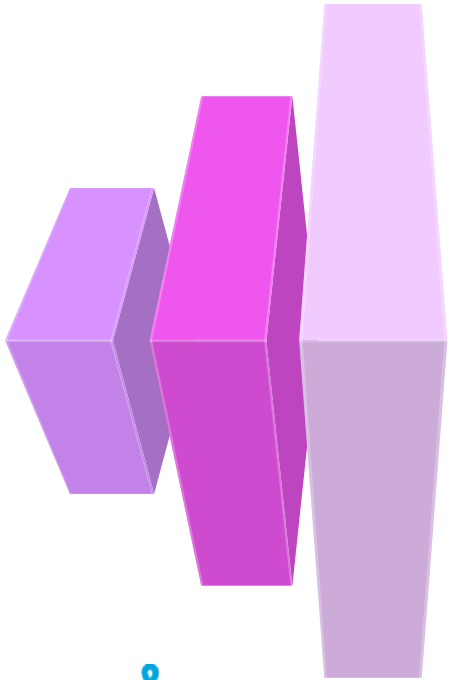
Updated: September 15, 2022

What does AI-ready mean for NOAA and EUMETSAT?

Community driven AI ready Data

Cloud AI ready Data

Operationally AI ready Data



EUMETSAT


Supporting the members
Pathfinder 1 – nowcasting
Pathfinder 2 – future extremes
Pathfinder 3 – processors
Pathfinder 4 – new applications
AI Ready Cloud System
Building the new operations
• Use of AI/ML in satellite operations
• Big data services
• Computational Science and processors
• Quality assessment and software coding analysis
• Data and images exploration – enhancing the user's experience
Being more agile and smart
Will ChatBot help us in finding/analysing internal documents?
Will ChatBot magnify the users experience and facilitate the uptake of our data?

Proposed AI-Ready Data Community Roadmap

ESIP Data Readiness Cluster

01	Developing and maintaining AI-ready data standard	• ***Short term***: Review of common domain data standard • ***Mid term***: Develop proposed AI-ready data standard for publishing process • ***Long term***: Publish and maintain AI-ready data standard
02	Developing automatic tools for AI-readiness assessment	• ***Short term***: Update assessment form based on user feedback • ***Mid term***: Develop metrics to display & synthesize AI-readiness assessment results • ***Long term***: Develop automatic AI-readiness assessment tool
03	Developing and improving AI-ready open environmental data	• ***Short term***: Uplift a pilot set of thematic AI-ready data • ***Mid term***: Develop tools and leading practices to improve data readiness at scale for community adoption • ***Long term***: Provide AI-ready data discovery tools and services
04	Sustaining the engagement with user and capacity building	• ***Short term***: Increase the engagement with private sectors • ***Mid term***: Develop primers for AI-ready data checklist / standards for different user personas • ***Long term***: Develop and maintain training materials on AI-ready data and tools for different user personas

- AI-ready data standards development: Current focus is on open environmental data; will it be different for other types of data?
- Integration between environmental data and socioeconomic data to advance the understanding of environmental quality.
- Welcome test users to apply the checklist to non-environmental data.

Final Joint Remarks

- Data access API – make any step to make sure that the access to the respective cloud system is facilitated – interoperability at usage level
- Key training data sets for improving nowcasting considering also to global-north-south

Thanks


DESPITE OUR GREAT RESEARCH RESULTS, SOME HAVE QUESTIONED OUR AI-BASED METHODOLOGY.
BUT WE TRAINED A CLASSIFIER ON A COLLECTION OF GOOD AND BAD METHODOLOGY SECTIONS, AND IT SAYS OURS IS FINE.