EUMETNET E-AI WG Data Curation

Kick-off, 14 Oct 2024

AGENDA

13:00 - 13:15 Greetings and background

13:15 - 13:30 Round-table / Get to know each other poll

13:30 - 14:30 Proposed ToR / tasks for the WG

- What do we want to do as a group? *"I would need this"/ "I would like to participate with this N hours per month"*

- First deliverables

14:30 - 14:50 Working practices, tools, meetings

14:50 - 15:00 Next steps

Live minutes at:
WG1-minutes_14102024 - Google Docs

- Everyone are welcome to contribute to the minutes!
- Minutes will be without names and gender neutral

Minutes will be left for edit for the rest of the week and then curated to GitHub

EUMETNET

Round-table questionnaire

Where are you coming from?

What are you doing in the field of AI/ML?

What's your relationship with the data (producer / user)?

What are you looking from this working group? How much time you estimate to have to contribute?

Poll: https://www.menti.com/altmce3sswaa

Please put organisation, github credential and email to the chat or put them to Email and Github handle collection | DFN-Terminplaner

EUMETNET

Proposed Terms of Reference (ToR)

WG1s primary intention is to lower barriers in E-AI activities data usage by analysing available and missing data holdings, obstacles in using them, and creating examples of using and converting the data into a suitable form for Machine Learning. The considered applications cover weather forecasting (from nowcasting to seasonal), climate applications, and downstream products and services

Participation

- Active Group Members:

  Individuals with a keen interest in contributing to deliverables

- Passive Group Members:

  Individuals with an interest in the topic of data curation to gain general insights. No hands-on implementation is necessary

EUMETNET

Proposed Terms of Reference (ToR) – First deliverable

Gap analysis of used and missing data in AI/ML applications relevant to E-AI content

1. Recognising the datasets and considering why they are used or not used

If you are interested to contribute, please write to the chat or via email to
roope.tervo@eumetsat.int;
stephan.siemen@ecmwf.int;
arianna.valmassoi@dwd.de

Later to be extended to cover aspects such as

2. Data access, including data-proximate processing capabilities
3. License/access policies
4. Available format(s)
5. Access and usage examples
6. Tools to re-format and available converters
7. Need to prepare other formats (for 'hot' datasets)
8. Need to develop converters
9. Suitability (for which purpose it can be used)
10. Processing chain, how the data has been produced (e.g. quality control)

EUMETNET

Proposed Terms of Reference (ToR) – Other activities

We, as a Working Group, will decide on other activities and deliverables

If you are interested to contribute, please write to the chat or via email to
roope.tervo@eumetsat.int;
stephan.siemen@ecmwf.int;
arianna.valmassoi@dwd.de

1. More comprehensive tutorials on data access, loading, and transformations

2. Zarr data structure standardisation and/or profiles for different use cases

3. Create a mechanism for sharing data inside the community; registry of relevant data sets focused on E-AI activities with aspects mentioned in analyses above

4. Your topic?

EUMETNET

Practicalities

- Live minutes at Google Docs → to be moved to GitHub after the meeting

- All resources (including presentations, deliverables, minutes, examples, code…) in GitHub: https://github.com/eumetnet-e-ai/wg1_data_curation

- Email lists to be set up

- Does the group have interest for a chat channel on https://chat.europeanweather.cloud ?

- Meeting frequency proposal

  - Working meetings every 2 weeks

  - Interest group meetings every 3 months

EUMETNET

Facilitators

Roope Tervo, EUMETSAT, roope.tervo@eumetsat.int

Arianna Valmassoi, DWD, arianna.valmassoi@dwd.de

Stephan Siemen, ECMWF, stephan.siemen@ecmwf.int

Interested?

If you are interested to
1. contribute to the first deliverables
2. future work
3. Join interest group

Please write to the chat or via email to
roope.tervo@eumetsat.int;
stephan.siemen@ecmwf.int;
arianna.valmassoi@dwd.de

EUMETNET