EOSC Future Science Project 9
Climate Neutral and Smart Cities

Joachim Wackerow
Integrate climate data from Copernicus ERA5 and air quality data from the European Environmental Agency (EEA) with data from the European Social Survey (ESS)

Integrated Data in ESS Labs
EOSC Future Science Project Climate Neutral and Smart Cities

Climate data
Social scientific survey
European Air Quality
Metadata Requirements for Cross-Domain

• The integration of data from multiple domains has higher demands on metadata than reuse of data in a single domain
  • It increases the requirements on the quality and scope of the metadata
• This applies especially for provenance information

• Following slides show screenshots of the process description application prototype
Overview of Process Activities
Description: The process involves creating a date variable from timestamps based on the time zone of each region, considering that the data is recorded hourly. It also addresses unit differences, converting Kelvin to Celsius and meters to millimeters. The data is then grouped by date, variable, and region, and temperature is averaged while also obtaining maximum and minimum values, accumulating precipitation by date, and identifying the maximum wind gust value. Moving averages are calculated for variables using different time windows (7-day, 30-day, 90-day, 365-day). Baseline values for temperature, precipitation, wind gust, and deviations from the baseline (anomalies) are determined based on the period from 1991 to 2020. Data older than 2015 is removed, and a group-by-operation is performed, collapsing the data by region using population-weighted averages. It is important to note that the ERA5 data may contain imputed and missing values. In memory, each row corresponds to a region, with mesh-blocks aggregated per day to calculate region-level values by taking the average of all variables weighted by the population of each block. The resulting data is stored to disk in CSV, SAV, or other suitable formats, as the data size remains manageable.

Diagram of the Process Activity

- ERA5 Intersim data for review (era5-grids)
- ERA5 Data Processing (current activity)
- Integratable datasource for ERA5 (era5-regions)

Activities (in sequential order):

1. **Create variable date**
   - Compute target variable ‘date’ based on variable ‘time’ and ‘region’ to convert from UTC to local time zones.

2. **Create variable tmpdca**
   - Compute target variable ‘tmpdca’ based on variable ‘tmpdca’, ‘region’, ‘id’, ‘date’ and ‘pop’. Calculate average temperature by date for each grid cell to be included in the data. Temperature measured in Kelvin is converted to
Create Variable 'paccta'
Create Variable ‘pacctcm’
Code for the Creating Variable ‘pacctam’
Created Variable ‘pacctcm’ in the Data Repository
Process Information in Standardized Form

- The description of the process/provenance information is made in DDI-CDI
- DDI Cross Domain Integration (DDI-CDI) is an emerging standard for the domain-independent description of data
- The DDI-CDI Process description is aligned with
  - Prov Ontology of World Wide Web Consortium (W3C)
  - Business Process Model and Notation (BPMN) of the Object Management Group (OMG)
Further Project Information at ESS Labs

https://europeansocialsurvey.org/esslabs/