Deliverable D8.2
Recommended scientific end-to-end demonstration services

Work Package	WP8
Lead partner	UVSQ
Status	Final
Deliverable type	Report
Dissemination level	Public
Due date	31 December 2019
Submission date	23 April 2020

Deliverable abstract

This ENVRI-FAIR Deliverable 8.2 provides high-level requirements for two scientific end-to-end demonstration services. Beyond the IT FAIR activities of the project these demonstration services are meant to serve scientists. Their implementation will make use of the technical development of the project towards more FAIRness. End to end means from data collection to the actual production of a scientific analysis product. Demonstration is understood to mean both that the service shall need further development before a full large-scale deployment, but also that if successful in demonstrating attractiveness; they can serve as showcases for the project. This deliverable is meant for internal use within ENVRI-FAIR and more specifically in WP8 Atmosphere. The actual implementation is part of the activities of Task 8.5 that started beginning of 2020.

ENVRI-FAIR (www.envri-fair.eu) has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 824068

DELIVERY SLIP

	Name	Partner Organization	Date
Main Author	Rivier L.	UVSQ	
Contributing Authors	C. Lund Myhre; D. Boulanger; M. Fiebig; Lara Ferrighi; J. Tarniewicz; A. Tjulin	NILU, IAGOS, ICOS, EISCAT-3D, SIOS	
Reviewer(s)	Ari Asmi, Markus Stocker	UHEL, TIB	March 2020
Approver	Andreas Petzold	FZJ / IAGOS	23 Apr 2020

DELIVERY LOG

Issue	Date	Comment	Author
V 0.1	Jan/2020	First draft	Cf above
V3.0	Feb/2020	Second draft	

DOCUMENT AMENDMENT PROCEDURE

Amendments, comments and suggestions should be sent to the Project Manager at manager@envri-fair.eu.

GLOSSARY

A relevant project glossary is included in Appendix A. The latest version of the master list of the glossary is available at http://doi.org/10.5281/zenodo.3465753.

PROJECT SUMMARY

ENVRI-FAIR is the connection of the ESFRI Cluster of Environmental Research Infrastructures (ENVRI) to the European Open Science Cloud (EOSC). Participating research infrastructures (RI) of the environmental domain cover the subdomains Atmosphere, Marine, Solid Earth and Biodiversity / Ecosystems and thus the Earth system in its full complexity.

The overarching goal is that at the end of the proposed project, all participating RIs have built a set of FAIR data services which enhances the efficiency and productivity of researchers, supports innovation, enables data- and knowledge-based decisions and connects the ENVRI Cluster to the EOSC.

This goal is reached by: (1) well defined community policies and standards on all steps of the data life cycle, aligned with the wider European policies, as well as with international developments; (2) each participating RI will have sustainable, transparent and auditable data services, for each step of data life cycle, compliant to the FAIR principles. (3) the focus of the proposed work is put on the implementation of prototypes for testing pre-production services at each RI; the catalogue of prepared services is defined for each RI independently, depending on the maturity of the involved RIs; (4) the complete set of thematic data services and tools provided by the ENVRI cluster is exposed under the EOSC catalogue of services.

TABLE OF CONTENTS

1. Motivation

This ENVRI-FAIR Deliverable 8.2 provides high-level requirements for two scientific end-to-end demonstration services. Beyond the IT FAIR activities of the project these demonstration services are meant to serve scientists. Their implementation will make use of the technical development of the project towards more FAIRness. End to end means from data collection to the actual production of a scientific analysis product. Demonstration is understood to mean both that the service shall need further development before a full large-scale deployment, but also that if successful in demonstrating attractiveness, they can serve as a showcases for the project. This deliverable is meant for internal use within ENVRI-FAIR and more specifically in WP8 Atmosphere. The actual implementation is part of the activities of Task 8.5 that started beginning of 2020.

2. WP8 Atmospheric subdomain

This deliverable 8.2 is part of WP8 of ENVRI-FAIR which regroups five European Research Infrastructures (RIs) of the Atmospheric domain. These RIs study the composition of the atmosphere and its physical state, from the ground level to ionosphere, including space weather. The RIs involved in WP8 are ACTRIS, EISCAT, IAGOS, ICOS-Atm, and SIOS[1].
The overall aim of WP8 is to improve the level of FAIRness (Figure 1) of the involved RIs. To this aim, the work is organized in six tasks: coordination as task 8.1, task 8.2 provides a gap analysis in FAIRness, task 8.3 organises detailed implementation plans for each RI based on the analysis of 8.2. The actual implementation work related to increasing FAIRness at the RI level is regrouped in task 8.4 to promote synergies in the solutions that are chosen. Task 8.5 is set up to demonstrate the new interoperability-based atmospheric services, and task 8.6 performs an assessment and provides recommendations for the future strategy.

Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

Figure 1: The FAIR Guiding Principles (Wilkinson et al., 2016) [2].

[1] A brief description of each RI is given in Appendix 1.
[2] Wilkinson, M.D., et al. (2016): The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data 3, 160018, DOI: 10.1038/sdata.2016.18.

The FAIR concept comes from a group of academic and private stakeholders that proposed a set of guiding principles to optimize data usage in a machine actionable way, i.e.; without or only limited human intervention. FAIR is the acronym for **F**indable, **A**ccessible **I**nteroperable, **R**eusable.
In a nutshell, it is possible to relate these four concepts to the following questions: **Findable**: Can I search the data and can I find it? i.e. is it properly identified? Accessible: Once found, can I for example download the data? **Interoperable**: How is the data formatted, what is its syntax? or is the associated metadata provided in a standard format? Is the (meta)data actionable? **Reusable**: Does the data license allow for reuse; how I should cite or give attribution when using the data. The notion of Provenance of the data is addressed under this concept. Each of these four principles is further divided into more technical sub principles.

Deliverable 8.2 provides high level requirements for two scientific end-to-end demonstration services:

1) FAIR Atmospheric data provision service
2) Modelled/observed concentration time series comparison service

These two demonstration services respond to end user needs and were thought through with scientific end users. The first one emerged upon the feedback of an initiative looking at the impact of the 2018 European drought on the carbon cycle. The second one emanates from colleagues in the modelling community. It provides comparison between modelled and observed concentration time series.
For each demonstration service, objective and impacts, functional and technical requirements are described.

This deliverable is in line with D8.1 “Atmospheric subdomain FAIRness assessment” and D8.3 “Atmospheric subdomain implementation plan” so that the recommended services described below make use of the planned development in the RIs toward improved FAIRness. The use cases will be implemented in Task 8.5.

3. FAIR Atmospheric data provision service

3.1 Context, objective and impacts

With climate change, extreme events (e.g. drought, fires) are becoming more frequent. There are also human induced extreme events such as viral pandemics that can have extreme impact on the state of the atmosphere. There is also more pressing demand from society to shorten the time required to make available scientific information concerning those events. We propose here to develop an Atmospheric data provision service that will allow discovery, access, download and initial analysis of atmospheric data from an extreme event, using a single entry point. We give here the example of the impact on atmospheric composition of the reduced activity in early 2020 in relation to the COVID-19 pandemic.

This proposition of FAIR atmospheric data provision service is built from the feedback of looking at the impact of the 2018 European drought on the carbon cycle. It should improve upon a certain number of shortcomings identified during the drought exercise and listed hereafter:

1. Manual compilation to get all available data
2. Manual reformatting with the need to code for individual patches for specific measuring sites
3. Manual QC diagnostics
4. Provenance information added manually: which data infrastructure (WCGG, Obspack, ICOS, …)
5. Manual addition of minimum metadata: PI name, email, institute
6. Manual Data selection/analysis

The aim is to shorten time response in providing scientific analysis of an extreme event and harmonized dataset and tools. This FAIR-compliant service will greatly improve the scientific user experience across the WP8 RIs. It will enable more comprehensive analysis of extreme events, giving access at once to all atmospheric RIs data. The assessment can then be done at the same time in terms of air quality and climatic impact.

3.2 Brief descriptive scenario of the service as an example

Objective: Assess the impact on atmospheric composition of a reduced activity due to a pandemic. The FAIR Atmo data provision service will allow to:

1. Go to a web application to search and then determine data availability in terms of the period of interest and the spatial coverage of the measurements from the various WP8 RIs
2. Access to a VRE (most likely a Jupyter notebook) and fetch the selected data in step 1 from the various WP8 RIs, including automatic previews of the datasets.
3. Perform first statistical analysis on the downloaded time series.
4. The data process will be implemented using a tool that enables automated compilation of provenance information
5. Data used in the session will be downloadable as a bundle, and minted with a PID

3.3 Basic requirements in terms of needed utilities

1. Search for concentration data across all WP8 RIs
    - Input: measured parameter
    - Output: map of the sites performing the measurement, data availability in time

2. Data fetch and download tool
    - Output in a common data format
    - Define and implement a minimal set of metadata

Technical proposition: develop a common API to fetch data from each atmospheric RI. Each RI develops interface implementations for their data centre APIs. This solution avoids developing a common data index.

3. Diagnostic tool
    - Simple quality control (e.g. min/max, thresholds on concentration time series)
    - Provide initial statistical analysis associated with the data (e.g. climatologies, anomaly detection)

Technical proposition:
Look at existing libraries that harmonize programmatic access to data (see next section: examples of existing solutions). Note that our communities often use R or Python languages for their data analysis. Make use of virtual research environments and in particular[3] Jupyter notebook to improve collaboration under a common environment. With Jupyter notebooks, it is possible to combine input data, programming, prewritten simulation/calculation programs together with visualization and text.

[3] Note: Jupyter notebooks are only one element of VRE. VREs support much more than data analysis. Often they include file management, social communication, etc. as additional features. See for instance the D4Science VREs.

Example of an initial statistical analysis product

Characterisation of the anomalous atmospheric signal due to the Covid-19 related confinement; analysis in terms of CO/CO_2 concentration ratios.

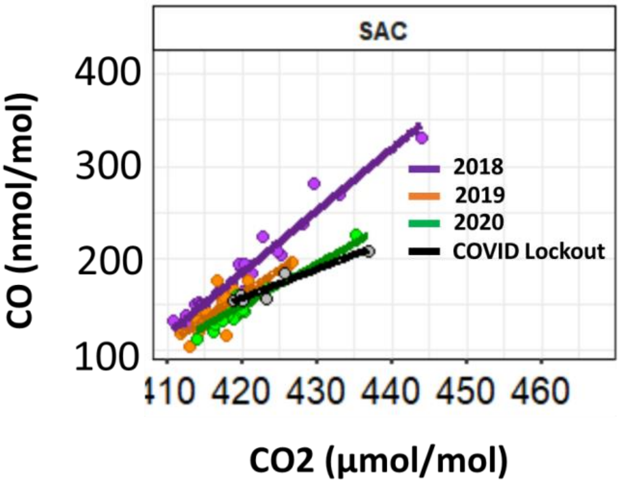

Figure 2: CO, CO2 plot for daily afternoon values shown for the month of march 2020 for different years in different colours (see figure legend). Data is from a measurements station in Saclay (SAC) near Paris, France. Pers. Comm. M. Ramonet, michel.ramonet@lsce.ipsl.fr, pers. comm.

The slope of the lines, corresponding to the CO:CO2 ratios show that it is smaller in March 2020 and even more if the period of the lockout due to the containment of the population is considered (black line COVID lockout). This smaller slope is induced by the reduced transport activity especially from cars due to the confinement in France at that period. Considering ratio/slope of the two species factors out to some extent the influence of the atmospheric transport in the variation of the atmospheric concentrations.

4. Provide minimal provenance information
    - define the appropriate level of provenance (for raw data, derived data, …)
    - Provide DOI, Attribution, Licence

Technical proposition: Follow ENVRIplus recommendation on provenance; make use of PROV-O standard (http://www.envriplus.eu/wp-content/uploads/2015/08/D8.5-Data-provenance-and-tracing-for-environmental-sciences-system-design.pdf).

Note: Jupyter notebook don’t give provenance info automatically. For automated provenance tracking, one needs a workflow system such as Taverna or Kepler. In such systems, the workflow is fully automated; inputs and outputs of all processes are specified as well as the process pipeline. The execution can thus be completely automated and provenance can be tracked accordingly.

3.4 Examples of existing solutions

Many elements already exist that could serve as examples towards this service. Here are a few examples close to our communities.

New Particle Formation Event Analysis

https://marketplace.eosc-portal.eu/services/new-particle-formation-event-analysis/information

Strength: VRE environment, semantic structuration, automated provenance added
Limitation: very specific test case

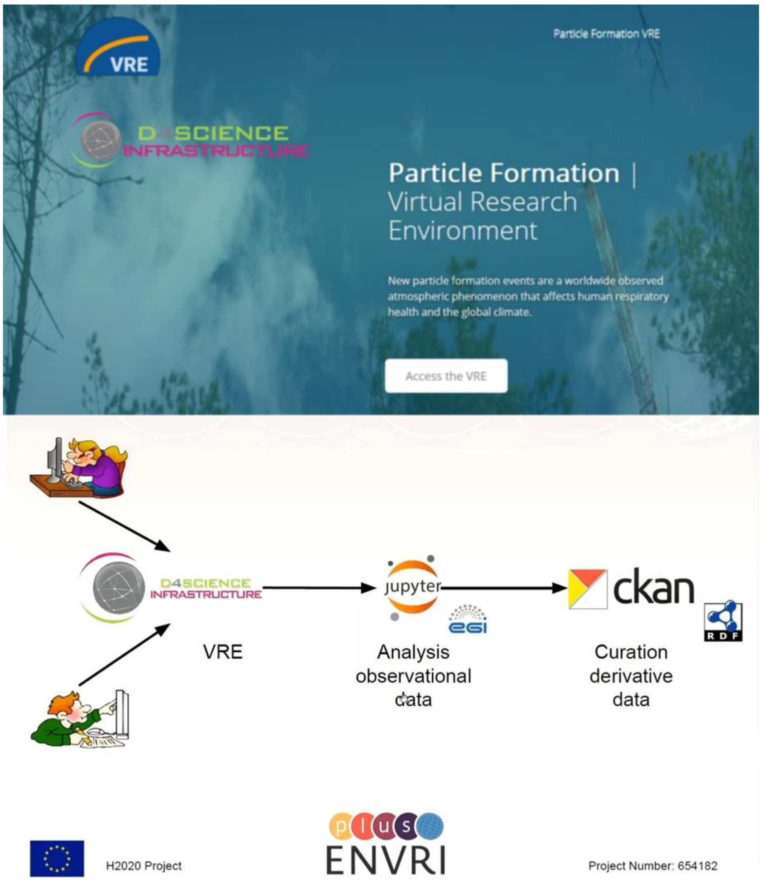

Figure 3: Home page of the "New Particle Formation Event Analysis". https://marketplace.eosc-portal.eu/services/new-particle-formation-event-analysis/information

Resources:

- Markus Stocker, Pauli Paasonen, Markus Fiebig, Martha A Zaidan, and Alex Hardisty (2018). Curating Scientific Information in Knowledge Infrastructures. Data Science Journal, 17:21. https://doi.org/10.5334/dsj-2018-021
- Video about the demonstrator : https://youtu.be/ra9W7b5DbgI
- The Virtual Research Environment https://marketplace.eosc-portal.eu/services/new-particle-formation-event-analysis

PANGAEA

Some data repositories have developed python libraries that support the automated download and analysis of metadata and data. PANGAEA is an example data repository and provides the pangaeapy library, which is available online at https://pypi.org/project/pangaeapy/. Such libraries can inspire the development of similar libraries for other data repositories.

NEON

The National Science Foundation's National Ecological Observatory Network (NEON) is a continental-scale observation facility designed to collect long-term open access ecological data to better understand how U.S. ecosystems are changing. NEON collects environmental data and archival samples that characterize plant, animals, soil, nutrients, freshwater and atmosphere from 81 field sites strategically located in terrestrial and freshwater ecosystems across the U.S.

NEON developed open access utilities in form of R packages to access and analyse data as well as data tutorials that can serve as guiding examples for the FAIR atmo data access service. More can been read at https://www.neonscience.org/get-started-neon-series.

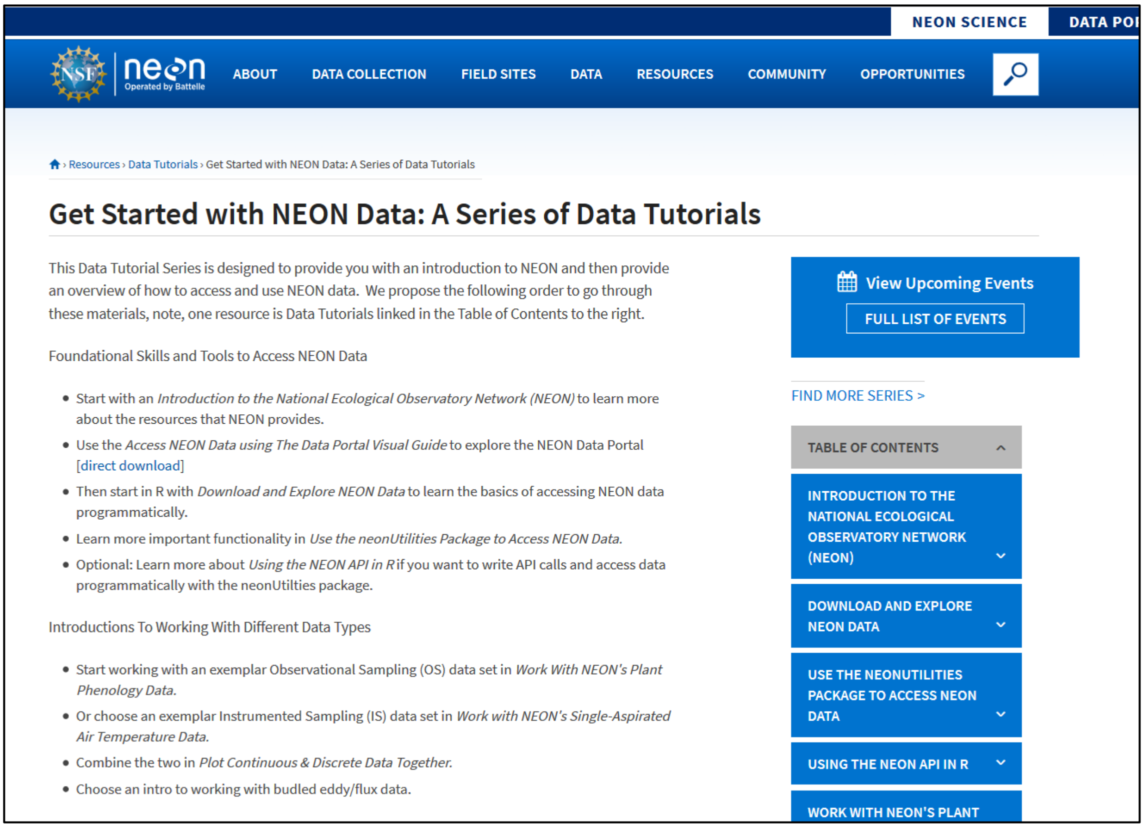

Figure 4: Webpage of NEON access utilities tutorials. https://www.neonscience.org/get-started-neon-series.

ICOS data portal access

The Integrated Carbon Observation System (ICOS) is a European Research Infrastructure and the European measurement system for high quality and high precision greenhouse gas observations. The ICOS Carbon Portal provides free and open access to all ICOS data.

The Carbon Portal serves data from over 130 stations, processed by the Thematic Centers as raw, near real time and final quality-controlled data, supplemented with elaborated (model) data and analyses. By the end of 2020, practically all stations will be in full operation.

ICOS develops and offers:

- discovery, preview and download of quality-controlled observational data
- advanced visualizations such as animated flux maps
- popular scientific products for policy makers, authorities, teachers and students.

Strength: Discovery and access, previsualisation, shopping chart, license, landing pages, metadata, DOIs, machine access.

Limitation: ICOS data only.

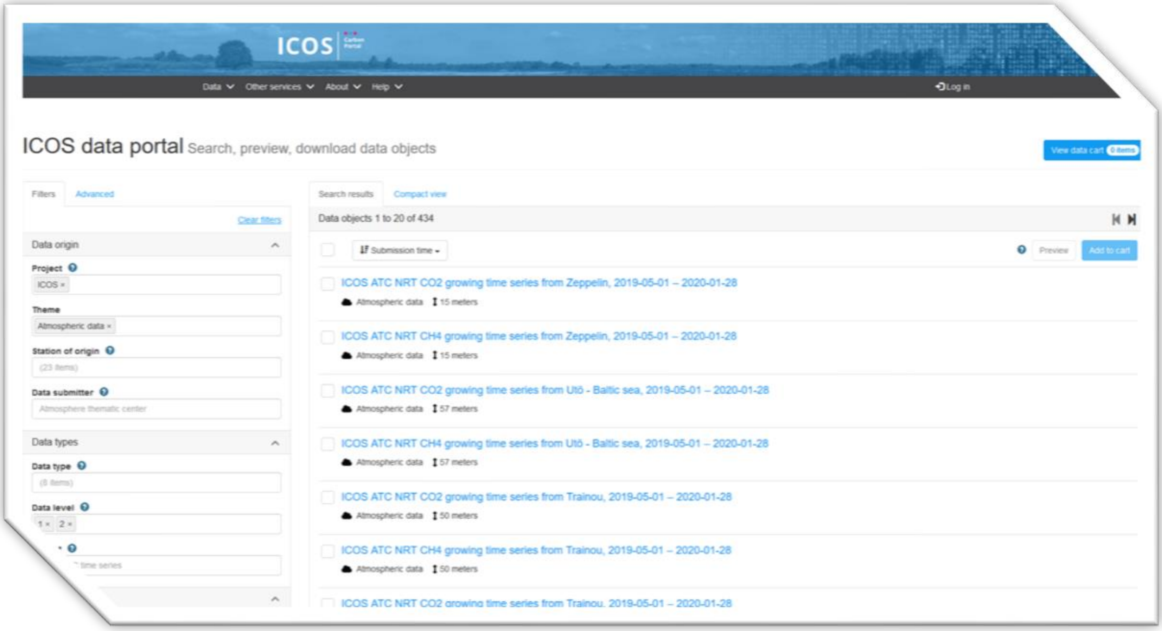

Figure 5: Webpage ICOS Carbon Portal data download. https://data.icos-cp.eu/portal/#{%22filterCategories%22:{%22level%22:[1,2],%22theme%22:[%22atmosphere%22],%22project%22:[%22icos%22]}}

4 Modelled/observed concentration times series comparison service

4.1 Context, objectives and impact

An exercise common to many atmospheric RIs is to compare modelled concentration time series with actual concentration observations.

In terms of modelling, atmospheric concentration time series are related to surface fluxes by the following transport equation:

$$\frac{D\phi}{Dt} + \phi\vec{\nabla}\cdot\vec{v} = S$$

It formalizes a conservation law where the atmospheric concentration φ is related to some source term *S (e.g. surface fluxes)* via the wind field $\vec{v}$. In a modelling framework, considering S an entry, one can compute φ(t) using an atmospheric transport model.
We propose to develop a generic service in which maps of surface fluxes (gridded, netcdf) will serve as entry to a transport model to compute time series. The modelled time series are then compared to actual observed time series at observation sites of ENVRI RIs (see general flow diagram below).

The service can be used as validation tool when for example two surface fluxes are transported and compared with observed times series. One would score better than the other in comparison to the actual observations.

This modelled/observed concentration times series comparison service can also serve as diagnostic for model development and improvement of atmospheric transport models. When a concentration tracer with a surface emission is considered well known, discrepancies between modelled and observed time series can suggest closer examination of specific parts of the model. Tracers as used in this context are substances (trace gases, aerosols) that are affected by atmospheric motions.

4.2 Brief descriptive scenario of the service as an example:

Objective: Compare two CO surface emission maps [4] in terms of their ability to reproduce actual atmospheric concentration observations. The model/observed concentration time series comparison service will allow to:

1. Upload two CO surface emission maps via a user friendly interface that will then re-grid files so that they can be used as input to the transport model.
2. Choose the set of observation sites that will be used to perform the comparison between observed and modelled time series.
3. Generate figures comparing modelled and observed time series including scoring figures (e.g. taylor plos) that tell which CO surface emission map generated atmospheric concentration time series that best compare to actual observations. A transport model is used to compute modelled concentrations from the two CO surface emission maps.
4. Download the plots and generate a catalogue of comparison exercises

4.3 Basic requirements in terms of needed utilities

Figure 6 below describes the flow of the service starting on the left with two branches: one for the model, and the other for observation, which meet in a comparison module.

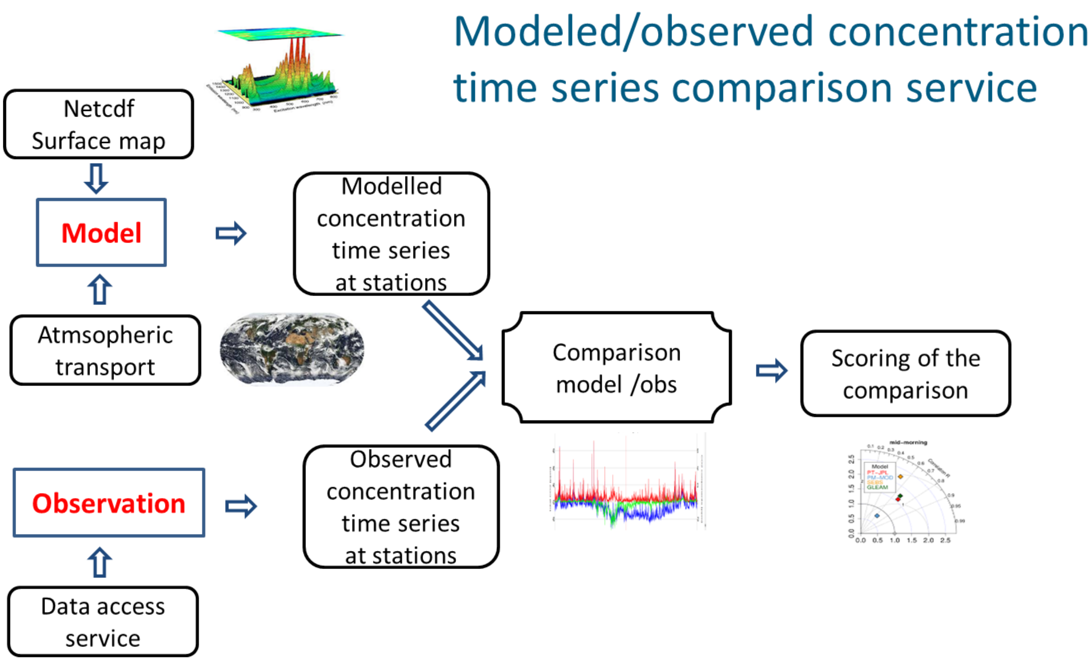

Figure 6: Modelled/observed concentration time series comparison service: diagram showing the different modules of the service.

1. User friendly web interface to upload input surface fluxes to be transported by the model
   - Input: should support a few different formats, netcdf being one of them
   - Output: the surface flux is re-gridded to the computational grid used by the atmospheric transport model

2. User friendly data provision tool
   - Make use of tools developed in the "FAIR data atmo provision service" (see section 3 above)

[4] CO is given as an example here but any atmospheric species could be considered.

3. Develop module of atmospheric transport
    - Output: modelled time series at observation sites
    - Provide model uncertainty evaluation

Technical Proposition: make use of cloud computing to allow high-resolution simulations. Rationalize the way wind forcing is used/shared (most likely ECMWF winds).

4. Develop module that compare modelled time series to actual observations
    - Define and implement efficient scoring tools

Proposition: One such scoring tool, often used, is the so-called Taylor diagram. It compares different models in a synthetic way. It is used to quantify the degree of correspondence between the modelled and observed behaviour in terms of two statistics: the Pearson correlation coefficient and the standard deviation (Figure 7).

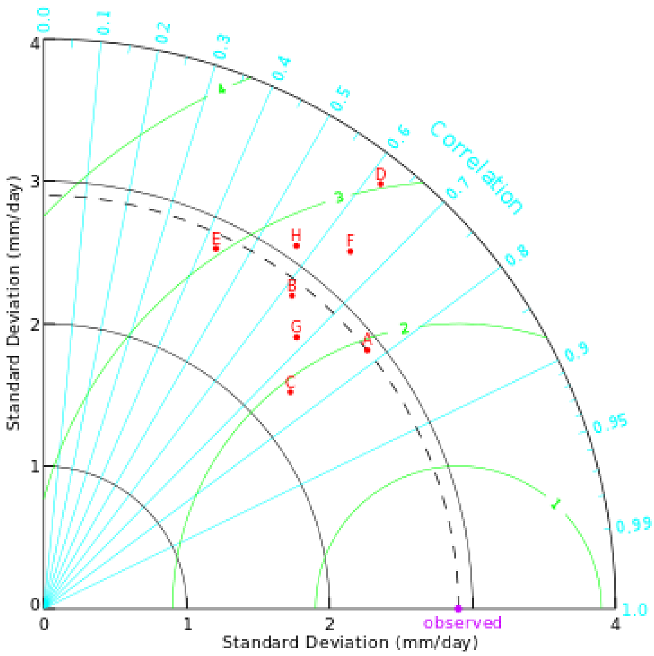

Figure 7: Example of Taylor diagram displaying a statistical comparison with observations of eight model estimates (points in red) of the global pattern of annual mean precipitation. Model A shows the best comparison with observation in this scoring. It has the same standard deviation of the observation (radii on the plots) and the highest correlation with observation (angle on the plot).

- Create a catalogue of comparison exercises:

The aim of the catalogue is to record comparison exercises to produce a bank of knowledge with metadata on the comparison exercises, allowing for basic information on the comparison and contact for further information.

4.4 Examples of existing solutions

Many elements that could serve towards this service already exist. Here are a few examples close to our communities.

1. **IAGOS CO contribution tool: Carbon Monoxide contribution: quantification of IAGOS CO origin calculated with SOFT-IO v1.0**

http://www.iagos-data.fr/#CMSConsultPlace:ANCILLARY_DATA

Source attribution using FLEXPART and carbon monoxide emission inventories (SOFT-IO version 1.0 ; Sauvage et al., 2017) is a tool based on the FLEXPART particle dispersion model (Stohl et al., 2005) coupled with emission inventories provided in the scientific community (e.g. GVAS v1.2, GFED4, MACCity, EDGAR). It has been developed for the atmospheric community to quantify source/receptor links for atmospheric trace gases considered as passive, such as for in-situ measurements of carbon monoxide (CO). SOFT-IO has been first applied and evaluated for the IAGOS data sets.
SOFT-IO simulates the global contributions of anthropogenic and biomass-burning emissions from the ECCAD emission inventory database (http://eccad.aeris-data.fr/) for all measured CO mixing ratios. This tool helps the users in the interpretation of the substantial IAGOS data sets.
In particular, it helps quantifying the CO geographical origin and emission sources that drive the observed CO distributions in the troposphere and in the lower stratosphere. For each CO measurement (every 0.5° in latitude or longitude at cruising altitude, every 10hPa during ascent or descent of the flight), CO origin is automatically calculated and separated by type of source (biomass burning and anthropogenic emissions) and by geographical region (14 over the globe).

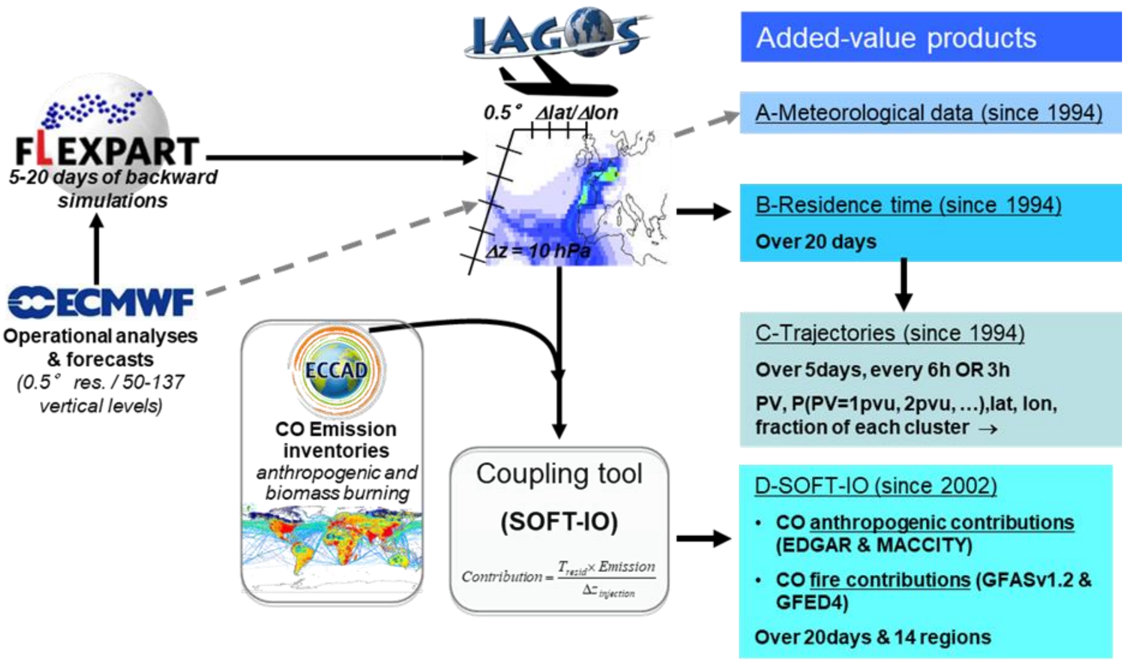

Figure 8: Elements of the IAGOS CO contribution tool: quantification of IAGOS CO origin calculated with SOFT-IO v1.0

2. ICOS CP Footprint tool

https://stilt.icos-cp.eu/viewer/
This is an online tool to analyse potential contributions of natural fluxes and anthropogenic emissions to the atmospheric CO2 concentrations at a selection of ICOS atmospheric stations. The tool has two active parts: The STILT on demand calculator and STILT results visualisation. The 'on-demand calculator' allows one to run the simulation tool for any point within Europe and add the results to the data-list available in the STILT results visualisation for visualization. The STILT results visualisation visualizes the foot print calculations. There are already a number of simulations available in the viewer for several ICOS measurement towers. You can go direct to the viewer to see examples of the footprint tool.

Background information:
The current model framework consists of the Lagrangian transport model STILT (Stochastic Time Inverted Lagrangian Transport; Lin et al., 2003) together with emission-sector and fuel-type specific emissions from a pre-release of the EDGARv4.3 inventory (EC-JRC/PBL, 2015) and biospheric fluxes from the diagnostic biosphere model VPRM (Vegetation Photosynthesis and Respiration Model; Mahadevan et al., 2008 [5]). The model framework has been developed at the Max Planck Institute for Biogeochemistry and is implemented as a web-based service at the Carbon Portal.

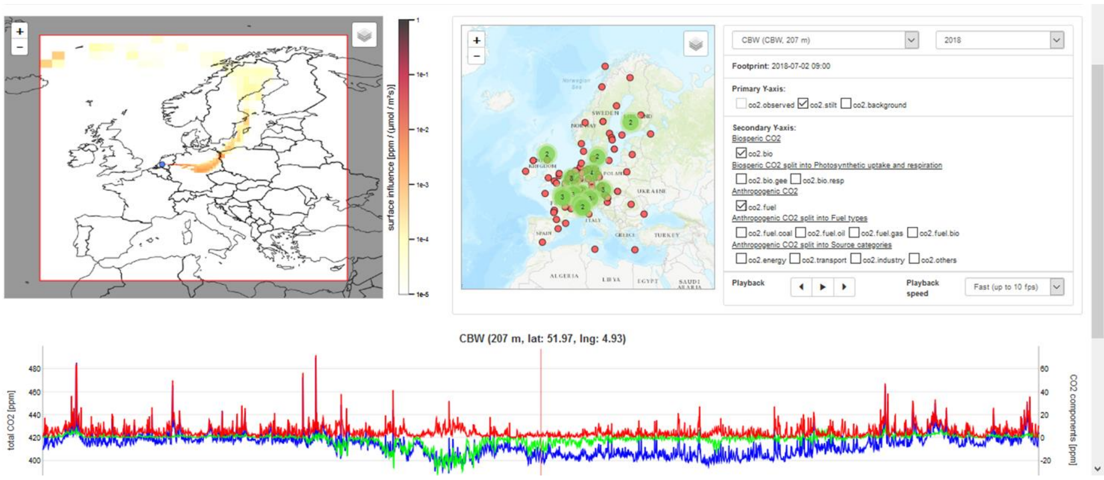

Figure 9: Webpage from the ICOS Carbon Footprint tool

[5] Mahadevan, P., et al. (2008): A satellite-based biosphere parameterization for net ecosystem CO_2 exchange: Vegetation Photosynthesis and Respiration Model (VPRM), Global Biogeochemical Cycles 22(2), 17, DOI: 10.1029/2006gb002735.

Appendix 1: RIs in WP8 Atmospheric subdomain

ACTRIS – Aerosol Cloud and Trace gases Research Infrastructure

The ACTRIS Research Infrastructure operates 5 topic databases and one integrating unit that are all linked via a metadata portal. In the upcoming years, more data centres will be added. All data are available through a common portal. First dataset dates back to 2001. They are from FP5 pre-project such as CREATE, EUSAAR, EARLINET, EARLINET-ASOS, CLOUDNET, ACTRIS-FP7 and ACTRIS-2. The number of active data users (unique IPs) per month is around 3-400, and around 40 000 data sets are currently downloaded per year where 1 data set is 1 full year of data
ACTRIS includes currently ca sites (50 operational on a continuous basis) and around 110 different atmospheric variables are measured, among the most important:

- 8 cloud profile variables, 12 aerosol profile variables, 12 (50) aerosol in situ variables, about 75 different trace gases measured at the surface,
- 30 different methodologies, time resolution ranging from seconds to 1 week(s),
- NRT data available is available from 25 sites and for around 10 variables,
- Many in situ sites are collocated with ICOS.

EISCAT_3D European Incoherent Scatter Scientific Association

EISCAT_3D is an ESFRI Landmark, it is in its implementation phase and the first set of data is expected in 2021. However, the already operating EISCAT Scientific Association contains data from the present systems (in operation since 1981). The treated variables and types of data are incoherent scatter radar data, bulk of stored data: Auto-correlations (Time lags, in a number of view directions, ranges, and sites). Data derived from the auto-correlations are physical ionospheric parameters (n_e, T_e, T_i, v_d, …) in voxels. Specialized data include Meteors, space debris, and more.
EISCAT_3D counts three sensor sites, where a common volume is observed, with centralized data storage and access.
The EISCAT_3D system is under construction and the first data is expected to be obtained at the end of 2021, which means that there are no EISCAT_3D data available at the present. However, EISCAT Scientific Association has been operating incoherent radar systems and collected data since 1981. The FAIRness assessment here is based on the status for the present data sets.
There are two different fundamental types of archived EISCAT data: Low-level data which consists of receiver voltage levels and spectral data, and high-level data consisting of the ionospheric physical parameters. These data types are handled in slightly different manners

IAGOS - In-service Aircraft for a Global Observing System

IAGOS performs in situ observations on mobile platforms (airliners). Measuring about 10 variables (CO, O_3, H_2O, NO_x, NO_y, clouds particles, aerosols, CH_4, CO_2), as well as aircraft measurements (air temperature, wind, etc.). The data are available in high resolution, meaning 4 seconds. Two packages are available: Package 1: one instrument per variable (O_3, CO, H_20, clouds) or Packages 2 optional: one instrument per variable (NO_y, GHG, aerosols). The same level of quality is assured, whereas uncertainties are specific to each instrument. Available are Near-Real-time (NRT) data (3 days) for Copernicus: data assimilation, model validation; and Real-Real-time (RRT) are in progress (3 hours) but only for vertical profiles.
The IAGOS Data Centre is operational since July 2011. The IAGOS RI is based on former projects, namely MOZAIC (1994-2014) and CARIBIC (since 1997). The number of users per month evolves around 15. The type of users being mainly operational services (CAMS) and academic scientists (trends, model validation and assimilation, process studies, satellite validation). The infrastructure currently counts a global coverage of 8 airliners.

ICOS-Atm Integrated Carbon Observation System ICOS Atmosphere

ICOS Research Infrastructure is coordinated and integrated by the ICOS European Research Infrastructure Consortium (ERIC). ICOS ERIC was established in 2015 and is operational since 2016. The first year of observations being 2016. Variables for the Atmospheric part of ICOS are in situ time series of greenhouse gases at 21 existing atmospheric sites today, around 30 sites will be available by the end of 2020. Several sites are collocated with ACTRIS. Several sampling heights are often available along mast often exciding 100m height. Continuous data is available for the following variables: CO_2, CH_4, CO, meteorological data. On a weekly basis, flasks are sampled for the following variables: CO_2, CH_4, N_2O, SF_6, CO, H_2, 13C and 180 of CO_2.
The instrumentation used includes infrared spectroscopy for the GHG, with stringent QAQC multi-level procedures. NRT data are available on a daily basis.

SIOS - Svalbard Integrated Arctic Earth Observing System

The SIOS Research Infrastructure entered its operational phase in September 2018. It includes distributed data management with contributing data centers hosting the data. The central node harvests discovery metadata and builds services on top of data. Currently harvesting from 4 data centers, 3 further data centers are in process of being integrated. However, the RI is also relying on existing datasets and sites, for which some very long time series are available. The treated variables and types of data are In-situ and remote sensing observations. Interdisciplinary, core data are long time series. Real time data are available. The number of core observation facilities is not yet decided. Data may be connected to other RIs such as ACTRIS, ICOS or operational programs such as WMO, EMEP. Currently the number of registered users of the portal evolves around 170.

SIOS is a regional observing system for long-term measurements in and around Svalbard, with the scope of integrating existing data centers through a distributed data management system (SDMS) which harvests, indexes and makes available data from different contributing centers. Each data centre has its own procedures and technical solutions tailored to the needs and the use of that data centre. SIOS is multidisciplinary and promotes integration of data through a dedicated working group involving all partners.

Appendix 2: Glossary

ACDD	Attribute Convention for Data Discovery (for NetCDF)
API	Application Programming Interface
B2HANDLE	EUDAT minting, storing, managing and accessing persistent identifiers
CAS	Central Authentication Service
CC-BY	Creative Commons Attribution License
CDI	Common Data Index (metadata format and data access system by SeaDataNet)
CF	Climate and Forecast (semantics for NetCDF)
CMEMS	Copernicus Marine Environment Monitoring Service
COPERNICUS	A major earth observation programme run by European Commission and European Space Agency
CSR	Cruise Summary Report
CSW	Catalogue Service for the Web
DMP	1) Data Management Plan 2) Data Management Platform (WP9)
DOI	Digital Object Identifier
DSA	Data Seal of Approval
ECV	Essentia Climate Variable
EDIOS	European Directory of ocean Observing Systems
EDMED	European Directory of Marine Environmental Datasets (SeaDataNet)
EDMO	European Directory of Marine Organisations
EDMERP	European Directory of Marine Environmental Research Projects
EML	Election Markup Language
EMODNET	European Marine Observation and Data Network
EMSO	European Multidisciplinary Seafloor and water column Observatory
ENVRI	1) A environmental RI cluster FP7 project 2) Environment research infrastructures (in ESFRI level or upcoming) as a community
ENVRIplus	A environmental RI cluster H2020 project
EOSC	European Open Science Cloud
EOV	Essential Ocean Variable(s)
ERDDAP	NOAA developed science data server technology
ERIC	European Research Infrastructure Consortium (legal entity type)
EUMETNET	Grouping of 31 European National Meteorological Services
ESFRI	European Strategy Forum on Research Infrastructures
FAIR	Findable Accessible Interoperable Reusable
FAQ	Frequently Asked Questions
FORCE11	a community to help facilitate the change toward improved knowledge creation and sharing
GBIF	Global Biodiversity Information Facility
GCMD	Global Change Master Directory
GDAC	Global Data Assembly Center
GEMET	GEneral Multilingual Environmental Thesaurus
GEO	Group on Earth Observation (System of Systems)
GEOSS	Global Earth Observation System of Systems
GOFAIR	An international programme on FAIR implementation
GOOS BGC	Global Ocean Observing System Biogeochemistry Panel
GUI	Graphical User Interface
ICOS	Integrated Carbon Observation System
ICT	Information and Communications Technology
IMIS	Integrated Marine Information System
INSPIRE	Infrastructure for Spatial Information in the European Community
iRODS	Open Source Data Management Software
JCOMM	Joint Technical Commission for Oceanography and Marine Meteorology
LW	LifeWatch
Marine-ID	Registration and authentication services for marine data services
MDA	Marine Data Archive
NetAPP	Hybrid cloud service
NetCDF	Network Common Data Format
NVS	NERC Vocabulary Services
NOAA	US National Oceanic and Atmospheric Administration

OAUTH	Open Authorization (standard)
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OBIS	Ocean Biogeographic Information System
ODIP	Ocean Data Interoperability Platform
OGC	Open Geospatial Consortium
OpenDAP	Open-source Project for a Network Data Access Protocol
ORCID	Open Researcher and Contributor ID
OWL	Web Ontology Language
PID	Persistent Identifiers
PROV-O	Web Ontology Language encoding of the PROV Data Mode
QA/QC	Quality Assurance/Quality Control
RDF	Resource Description Framework
RI	Research Infrastructure
RSS	Really Simple Syndication
SAML	Security Assertion Markup Language
SEADATANET	SeaDataNet pan-European infrastructure for marine data management
SME	Small or medium Enterprise
SparQL	SparQL Protocol And RDF Query Language
SWOT	Analysis on Strengths, Weaknesses, Opportunities and Threats
VRE	Virtual Research Environment
W3C	World Wide Web Consortium
WMO	World Meteorological Organisation
WoRMS	World Registry of Marine Species
WPS	Web Processing Services
YAML	Yet Another Mockup Language