ICOS Carbon Portal

Progress Report 2014 & 2015

Alex Vermeulen, Margareta Hellström, Harry Lankreijer, Oleg Mirzov,
Ute Karstens, Ingrid van der Laan-Luijkx

Lund University, Sweden
Department of Physical Geography and Ecosystem Science

ICOS Carbon Portal

Progress Report 2014 & 2015

Alex Vermeulen, Margareta Hellström, Harry Lankreijer, Oleg Mirzov,
Ute Karstens, Ingrid van der Laan-Luijkx

Lund, April 2015

Preface

On 1 January 2014 the Integrated Carbon Observation System research infrastructure's Carbon Portal (ICOS CP) entered into its construction & implementation stage.

This document reports on the progress in the years 2014 and 2015. Until the end of 2015 the Swedish part of the ICOS CP was directly funded by the Swedish research Council (VR, Vetenskapsrådet). Starting 2016 the national contribution of Sweden to ICOS CP will flow through ICOS ERIC, the legal representative of ICOS RI that has been established 24 October 2015. From then on ICOS CP will directly report to the ICOS General Assembly.

Table of Contents

Introduction

The Carbon Portal is the "one-stop shop" for all ICOS data products. As such, the CP is envisioned as a virtual data centre, i.e. a place where ICOS data can be discovered and accessed along with ancillary data [M. Köchy, A. Freibauer: Specification of external data products and providers, ICOS-PP deliverable 4.2, 2009] and where users can post elaborated data products that are obtained from ICOS data. The CP will also have the ability to address all the requirements stemming from these aspects, including:

- data security,
- enforcement of the ICOS data policy,
- user-friendly (and machine-friendly) internet-based and other computer-network interfaces.

The CP is the data platform of ICOS. All relevant ICOS data and ancillary data sets from external sources will be published and be accessible through the facilities of the Carbon Portal. The CP system design is based on standard data interfaces and is an integrative access point for all ICOS users, ranging from experts to stakeholders and the general public. The CP shall support standardized data exchange protocols and techniques. The Carbon Portal **is** responsible for the following aspects of handling ICOS data products:

- user registration,
- bibliometrics,
- publishing of the data monitoring data usage and citation,
- developing an integrated metadata system, maintaining metadata consistency inside the RI and proving a device for external metadata systems (e.g. DataCite),
- ensuring long-term archiving
- setting up and operating suitable web-based interfaces for data discovery, visualization and access.

The CP also provide**s** the capability of advanced service composition techniques for web-based distributed processing of ICOS data to generate useful information (e.g. risk maps, alarm maps, and integration and analysis with other types of datasets) for research, public users and decision makers**.**

ICOS data levels

All types of ICOS data need to be easily discoverable, accessed, visualized, and available for further analysis by all interested parties. In addition, provisions shall be made by ICOS to provide standardized and comprehensive synthesis products that summarize the ICOS data, e.g. on annual and seasonal basis. An extended list of ICOS data (that may be covered by the Carbon Portal) is provided in Table 1 below.

ICOS in-situ data by themselves need processing (e.g. filtering) which defines Level-1 and Level-2 data. Further application of models and analysis procedures for producing so called elaborated data-products, such as maps of regional GHG fluxes, multisensor atmospheric concentration datasets, and derived ecosystem properties like water use efficiency, constitute Level-3 data, see Table 1 below. Note that the distribution of specific data types between these different levels of data adapted in this document differs somewhat from that outlined in e.g. the ICOS Stakeholder's Handbook.

In general, the data need to be associated with metadata, compatible with international data formats (following the INSPIRE directive, ISO 19115, etc), such as the ones suggested by the DataCite group [http://www.datacite.org]. In addition, the ICOS data need to be traceable to station and instrument acquisition parameters (Level-0 data), to the algorithms used for data processing at the Thematic Center in change (Level-1 and Level-2 data), to the calibration scale used to report data values in physical units and to the routines used to generate elaborated products (Level-3).

Table 1. ICOS data types and their origins. NRT refers to "near real time" data that have undergone only an initial automatic processing and quality control.

Data levels	General definition	Atmosphere	Ecosystem	Ocean
0	Raw data	(Electrical) signals directly from instruments		
1	Geophysical data	Data in physical units, either NRT or flagged	Fluxes and other data in physical units, NRT and validated	Fluxes and other data in physical units
2	Enhanced data	Quality controlled and validated data	Gap filled data using several processing algorithms	Gap filled data using several processing algorithms
3	Elaborated products	Flux datasets and maps with results from ecosystem and inverse modelling, data assimilation results, atmospheric concentration datasets, and derived ecosystem properties		

The ICOS Carbon Portal, a little history

The ICOS Carbon Portal was created on the basis of a white paper produced by the ICOS community, describing the needs and requirements for the ICOS data portal. Two different consortia were formed that independently both wrote a proposal to implement the CP. The proposal from the Swedish-Netherlands consortium (Lund University, ICOS-NL) was selected by the members of the ICOS PP project and approved by the ICOS Infrastructure Stakeholders Interim Committee in November 2012.

The national Swedish application was then approved by the Swedish Research Council VR in January 2013. The work at the CP headquarter at host institute Lund University, Department of Physical Geography and Ecosystem Science, started in September 2013 with a 4 month planning phase, followed by a one year development phase in 2014. The planning phase has been reported in a separate document (Hellström et al, 2014).

ICOS was established only November 2015 as an ERIC, a legal representation for a research infrastructure in Europe. The Carbon Portal, together with the Head Office, is part of ICOS ERIC.

This document reports on the development and following operational/implementation phase in 2015. Because the ERIC was established only late 2015, in fact most of the 2015 period can still be considered as an extended development phase.

Human resources at CP

In the planning phase only 1.5 fte divided over 9 persons worked on the planning of the activities for the development phase. The installation of the Carbon Portal Director, originally foreseen for January 2014 was delayed and took place only June 2014. Only after that the further fulfilling of the positions at the Carbon Portal could took place. A full-time system architect and a programmer were hired starting February 2015 and a full-time scientist for the elaborated products started September 2015. An additional 1.8 fte is hired from existing personnel at Lund University and since January 2015 one full-time scientist is working, also on elaborated products, at the CP branch in the Netherlands at Wageningen University.

Products and services

General

For developing the services, CP works with light-weight and independent separated services that mostly run in self-contained environments, so called “containers”, of which we use a special form called Docker. This also will make it easy to integrate ICOS services with Open Cloud systems in the future and allows for tremendous scalibility.

A Linux container is a set of processes that are isolated from the rest of the machine. A container can encapsulate any application dependency. For example, if a website relies on a particular version of the PHP scripting language, the container can encapsulate that version. As a result, multiple versions of the same scripting language can co-exist in the same environment - without the administrative overhead of a complete software stack, including the OS kernel. Containerized applications perform about as well as applications installed traditionally. Containers are transforming application development and delivery with lightweight and massively scalable resources. Containers are a great tool to create a virtual infrastructure that provides resources to build, deploy and instantiate applications. Proven to accelerate application delivery, containers simplify the packaging of applications along with their dependencies. A virtual network infrastructure that is scalable, secure and highly resilient is crucial to delivering on the potential of containers.

Docker is an open platform for developers and sysadmins to build, deliver and run distributed applications. The Docker Engine container comprises just the application and its dependencies. It runs as an isolated process in user space on the host operating system, sharing the kernel with other containers. Docker is currently used for simple application deployment, continuous integration, application lifecycle management, Platform-as-a-Service (PaaS) and distributed applications composition.

All developments of CP are shared through Github, including all code and documentation. The repository can be accessed at: https://github.com/ICOS-Carbon-Portal.

All software used at CP that is not home-developed is open source and thus far also free software in order to remain vendor independent and neutral.

Services infrastructure

Most of the CP services will be implemented through single Docker containers. A proxy server (Nginx) distribute the user requests to the applicable services.

Currently the CP runs through a single server, located at the Lund University supercomputing centre LUNARC. The server is a dual XEON Dell Plexserver with 64 GByte RAM and 30 TByte diskspace (RAID-50), operating on CentOS 7.0. The server is connected to the university’s internet connection with a 1 Gb/s network connection. The servers RAID harddisk can be read and written to at transfer rates around 1 GB/s.

Both Nginx and the services are completely scalable, that is, if more capacity is needed more Nginx processes can be distributed over more servers, and also the Dockers providing the other services can also be distributed over one or more servers each, for example in one of the EGI clusters.

Figure 1 shows an overview of the currently developed services. All services run in Dockers, except for the authorization module and the data central services that run in the Java Virtual Machine (JVM).

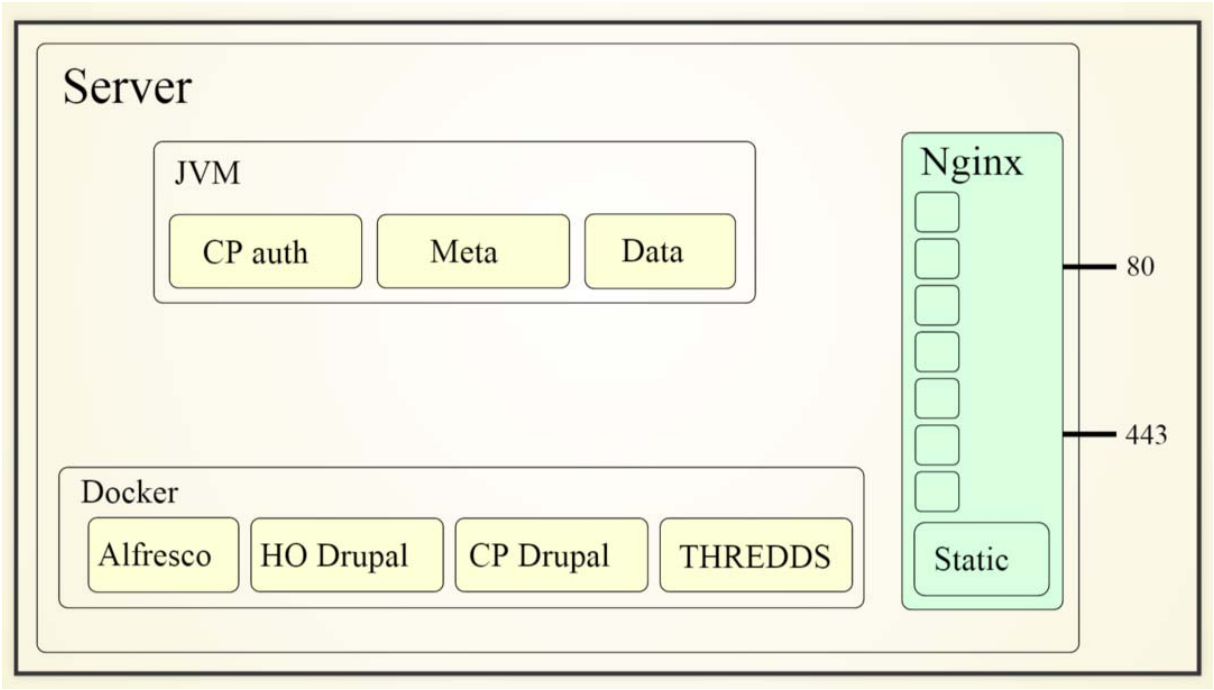


Figure 1

All services, including the web pages, are transmitted securely through SSL, where needed controlled through the authorization module CPauth. The individual services in Fig. 1 are further described in the following sections. The only ports open on the system are ports 80 and 443. All access through these ports are handle by Nginx.

Single-Sign-On

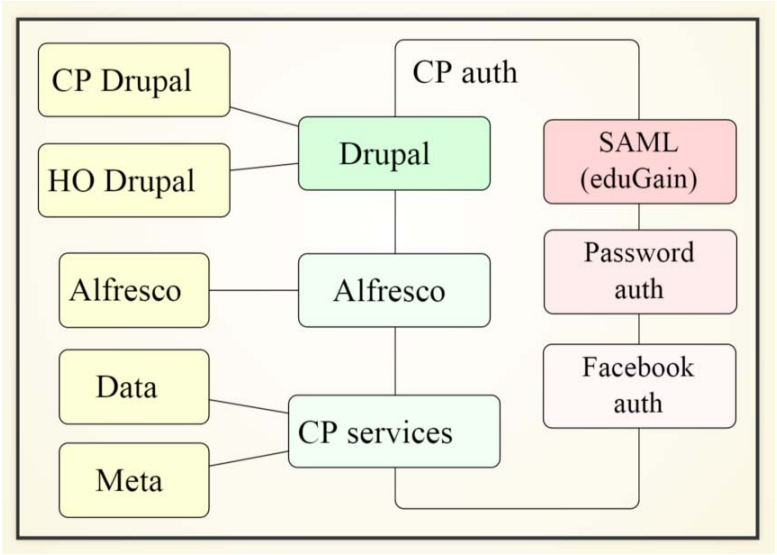


Figure 2

Fig. 2 displays the relation between the CPauth service and the other services. Through the SSO service in SSO user can be authenticated through SAML at eduGAIN as trusted

identity at their own institute or university. Each service then can apply its own authorization scheme to the user's identity. This has been implemented at this moment for Drupal and the CP services Meta and Data. The Alfresco service still relies on its own authentication scheme. It is also possible to use CP provided identities to authenticate, for example to allow administrator login to certain services.

Extension of the authentication allowing using OAuth 2.0 (two-factor) and ORCid's is foreseen.

Ontology/Open Linked data

An important aspect of Carbon Portal's technical activities is designing, handling and archiving detailed metadata about ICOS data products, measurement stations, instruments, collaborators and their roles, etc. The technology chosen by CP for managing the metadata is the Semantic Web/Linked Data (https://www.w3.org/standards/semanticweb/), which is a set of standards endorsed by the W3C (World Wide Web Consortium). The most relevant standards include:

- RDF (Resource Description Framework) – a Web-friendly standard for encoding arbitrary metadata. It is built on a few simple core ideas (the following description is intentionally simplified):
    - Every metadata document is a sequence of atomic statements called triples. The triples describe resources – the things of interest that need metadata or are the metadata themselves.
    - Every resource is identified by a URL.
    - Every triple consists of three elements: subject, predicate, and object.
    - Predicate can be thought of as a property (e.g. file name, number of rows, etc.) linking the subject to a property value (the object).
    - The subject and the predicate are always resources. The object can be either a simple literal (e.g. a string, a number, a timestamp) or a resource as well.
    - Finally, the same resource can occur at different positions in different triples. For example, an object of one triple can be a subject of another. This enables chaining of properties and nesting of values, making the data model of RDF a directed labelled graph.
- SPARQL (SPARQL Protocol and RDF Query Language) – the protocol for exposing RDF datasets as Web services, and the query language to query them. These Web services are called SPARQL endpoints, and the protocol describes which HTTP API they must follow. In the simplest case, a SPARQL endpoint replies to HTTP GET requests with the query encoded in the URL as a parameter. The core principle of SPARQL as a query language is formulating graph patterns of relationships between resources, with some of the graph nodes declared as variables. The task of the database is then to find all the matches between the pattern and the actual RDF datasets, bind the variables to the actual values, and report the bindings.

- OWL (Web Ontology Language) – the standard for designing formal, computer-readable knowledge models describing arbitrary knowledge domains. Such data models are referred to as ontologies in computer science. OWL's features include (among others) the following:
    - classes and class hierarchies (subclasses)
    - datatype properties and object properties; subproperties
    - property ranges and domains
    - class restrictions (universal, existential, and cardinality)
    - inference semantics (rules that allow deducing extra statements from the existing ones)

The above standards are widely recognized by both industrial and research institutions. Consequently, they have the benefit of being supported by a large selection of software products, including many open source solutions.

The central element in metadata infrastructure is a database that supports RDF as the data model and is capable of answering SPARQL queries. Such databases are called triple-stores and are available from multiple vendors, both commercially and as open source.

A comprehensive metadata storage system has been developed at Carbon Portal, based on the Sesame triple-store (http://rdf4j.org). The system serves as the back end part of the Meta service (https://github.com/ICOS-Carbon-Portal/meta) and has the following features:

- storing and exposing ICOS metadata
- exposing ICOS metadata profiles (as OWL ontologies)
- SPARQL endpoint
- logical grouping of RDF graphs into so-called instance servers with fine-grained control over read/write operations and persistence
- support for RDF logging – chronological listing of atomic RDF statement assertions and retractions; this feature allows complete traceability of all the changes and possibility to reset the metadata state to any point in the past
- support for OWL inference; this feature is used to support the generic metadata entry application (see Metadata services below)

Data lifecycle/Data flow

In extensive discussion with Head-Office, Central Facilities and Carbon Portal work has been performed to systematically define the data lifecycle of all ICOS data products. This has led to the ICOS data lifecylce document (currently in final draft). Also extensive discussions have been performed in the ENVRI and ENVRIplus projects (see chapter: related project work) and these EU projects have delivered many useful insights for the development of the Carbon Portal backbone data services.

Essential for all actions, like registration of data, tracking provenance and curation, attribution and licensing is the identification of the data using unique and persistent identifiers. The standard developed for this is to use unique Persistent Digital Identifiers (PID's). The global system that issues these handles is called the Global Handle Regsitry (the Handle System), its main administrator is CNRI in the USA. They allot prefixes to the users. ICOS is registered since autumn 2015 and has handle prefix 11676. ePIC is the European provider that registers, stores and resolves persistent identifiers. When PID's are minted the user registers a so called landing page that the handle system will resolve to when called with the PID. The user is responsible to maintain this landing page, even if the underlying data object has been removed or changed, then the landing page is called a tombstone page or should refer to the PID of the new version of the data object.

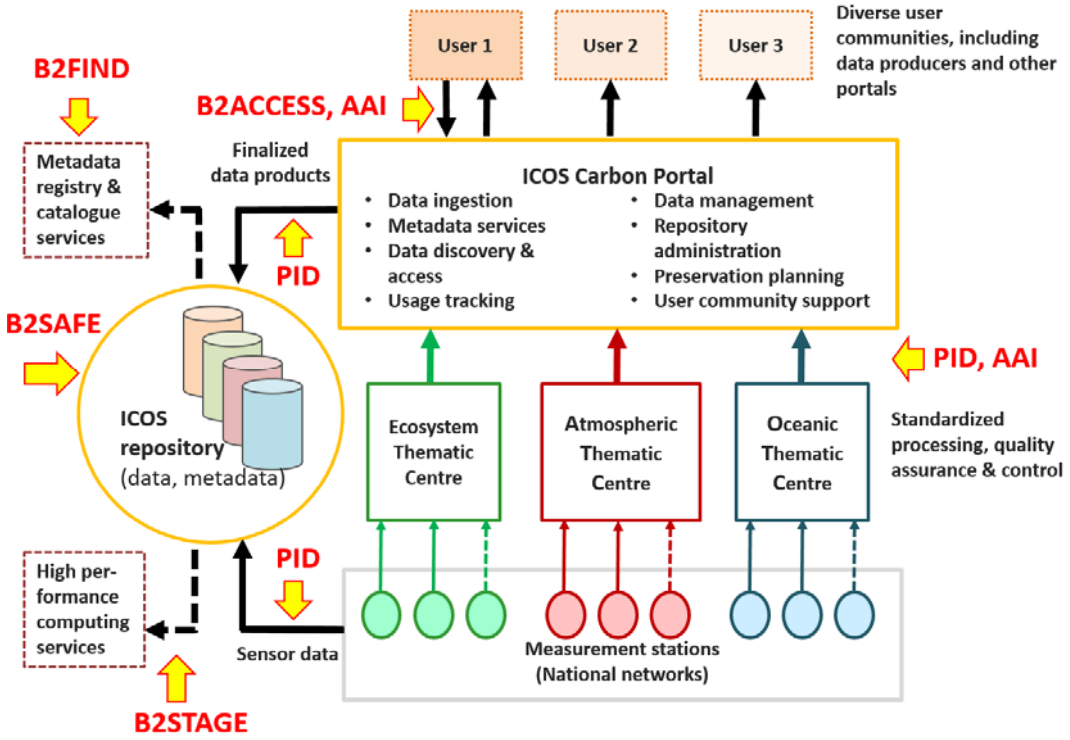


Figure 3 Data flow in ICOS RI and the roles of Carbon Portal. The yellow arrows indicate data flow where EUDAT2020 services are identified to provide interoperable functionality

The use of PID's is extended by the digital object identifiers, DOI's. These are widely used for citation of literature and data. DOI's are just a special kind of PID's that are handled by the IDF (Internaional DOI Foundation) that extends the functionality of the handle system with a strict naming convention, application profiles and services targeted at the citation functionality. Published ICOS data sets will be minted a DOI by the Carbon Portal.

Landing pages can be human readable web-pages in HTML and/or return data that is easier to read for machines (e.g. based on an ontology in OWL, JSON, XML format).

PID's can then be used to link metadata to the data objects and in all transformations and transfers the PID can be used to identify the original data object and to link the newer or enhanced metadata describing the actions to the original metadata.

Also when the data is transferred the PID can be used to inform the user of the presence of the data object (during searches) and its metadata. The metadata should in the case

of ICOS for example contain the origin of the data (station(s)), data provider(s) and a suggested citation (including the PID/DOI) when the data is used in a publication for attribution. Through the PID/DOI the Carbon Portal can also implement the usage tracking and gather citation statistics, per station, country, data provider etc.

In Figure 3 the data flow through ICOS is presented. The yellow arrows indicate data flows (essentialy all) where PID's play a vital role and also where services from the project EUDAT (see chapter: related project work) can be used in combination with the ICIS ePIC PID handles.

Station metadata entry

Using the CP metadata storage system described above the ICOS station database has been created that holds all the information needed at Head-Office and Thematic Centers for taking up the stations from the national networks into ICOS. The national contact point have provided this information in the form of a spreadsheet to ICOS HO. Using a simple ontology this spreadsheet is converted into the station metadata database. Only ICOS HO is allowed to modify this database, add countries, stations etc.

The interface to modify or view the database is completely generated automatically on the basis of the ontology and the database information. The database is also available for read-only SPARQL queries, so that the actual state can be used in any website to display up to date tables or maps of the ICOS station network. This has been implemented at the Carbon Portal website for demonstration. The viewing/editing interface and an example map of stations can be seen in Fig. 4.

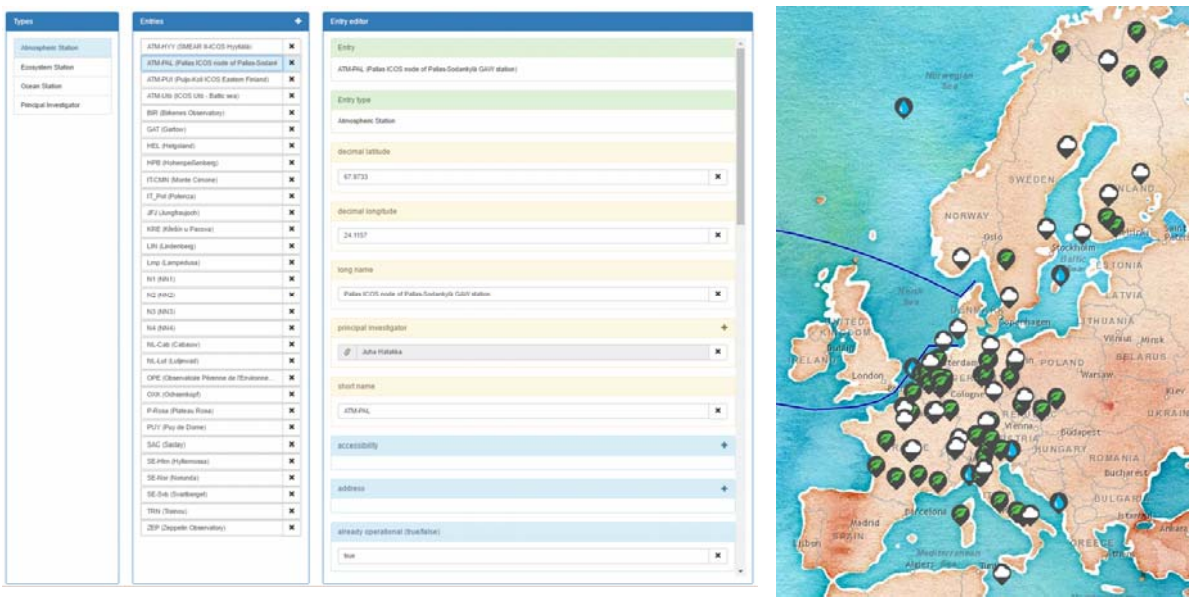


Figure 4

Station Labelling

A similar setup as for the station entry has been chosen for the assistance at CP for the ICOS station labelling process. Here PI's can insert all information requested by the Thematic Centres needed to enter step 1 of the labelling process. Also the TC's can update here the progress of the labelling process. All information is open and this

allows for anyone to monitor the progress of the process and see how many and which stations have applied and how far their application is, the data that has been submitted and the results of the evaluation steps.

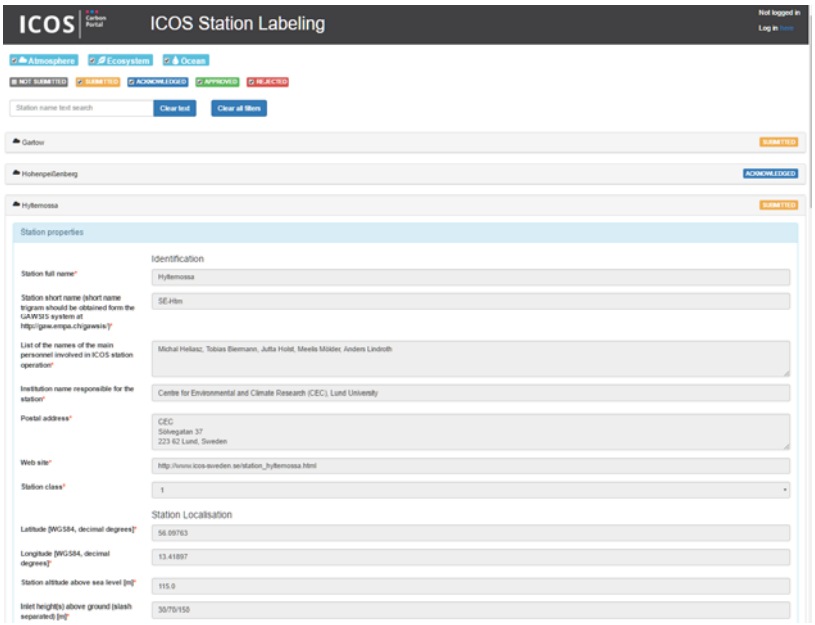


Figure 5

Website

For the website a Content Management System had to be chosen, to allow for easy maintenance and editing of websites even by non-technical staff. After some initial tests it was decided to use Drupal version 7, based on its user base and existing experience with the system in the team. Together with HO the most essential elements of the original Idean design for the ICOS Brand Visual Design was ported to Drupal. This design is a so called mobile-first visual design based on the Bootstrap Framework, supporting CSS3 and HTML5.

Figure 6 ICOS CP homepage

Currently the CP server hosts the websites for Carbon Portal (https://www.icos-cp.eu) and ICOS RI (https://www.icos-ri.eu), both in separate Dockers. It is possible to provide template Dockers for use by Central facilities or national networks and to host their website through the Carbon Portal server. The website domains for ICOS ATC and ETC are redirected to dedicated servers at those thematic centres. Work has started with ATC and OTC to use the ICOS Drupal dockers to implement (part) of their websites.

End 2015 the port to Drupal version 8 was started for the CP website, as this new version of Drupal was now finally released. Drupal 8 has major advantages over Drupal 7, being more flexible, efficient and easier to use for non-technical users with better support for dynamic content.

Document management system

The open-source package Alfresco Community 5 has been chosen as the Document Management System (DMS) for ICOS. The ICOS Alfresco (https://docs.icos-cp.eu) is installed in a Docker under 64-bit Linux, and is backed up to an external location nightly. The system administration (managing users and folders) is mainly performed using scripts. The ICOS installation currently features a Document Library, a Calendar, an Imager Gallery and a Discussion Forum – all accessible via a Dashboard system. At the end of 2015, there were ca 100 individual user accounts and a handful of "generic" accounts (the latter offering read-only access to specific parts).

The Document Library is built up around a catalogue tree where the access to each subfolder is separately defined. This allows "private" file spaces to be set up for ICOS entities (like the HO, CP, and the CFs) or cross-cutting working groups, while at the same time allowing all logged-in users to read, add or modify content in common subfolders. Access rights are defined for each registered user via memberships in appropriate user groups (managers, collaborators, contributors and consumers).

Important to ICOS is that Alfresco offers both content versioning (allowing the complete history of changes to documents to be tracked), as well as document workflows (supporting collaborative editing and deadlines). In addition to using the web-based interface, all content in the Document Library can also be accessed by external clients, including the ICOS Drupal system, via the CMIS (Content Management Information Services) 1.1 standard.

Web services

Time series

Early work at the CP concentrated on testing fast storage of time series like the NRT or Level 2 data products and high performance plotting of these time series. Adequate solutions have been identified and demonstrations have been shown on early versions of the CP website.

As the actual official ICOS data products can only be expected when stations have been labelled (soon after the summer of 2016) this work will picked up again in 2016. Currently the routines are further developed tested using the large complete data series from the WMO World Data Center of Greenhouse Gases (WDCGG), containing 4000 data sets, and sometimes containing 40 years of hourly data per set.

Spatial data

Many of the elaborated products and the input files required for the models providing these data are spatial data. There are many (evolving) standards for these data and CP will make use of the standards as developed by the OGC (Open Geospatial Consortium). OGC is also the origin of the ISO19115 standard for the metadata of geospatial data. Some of these standards are WMS (Web Map Service) and WCS (Web Coverage Service). At CP we developed both services and clients that allow users to display and retrieve spatial data on the basis of the OGC standards and also services that access for example netCDF files directly for more control and performance (Fig 7).

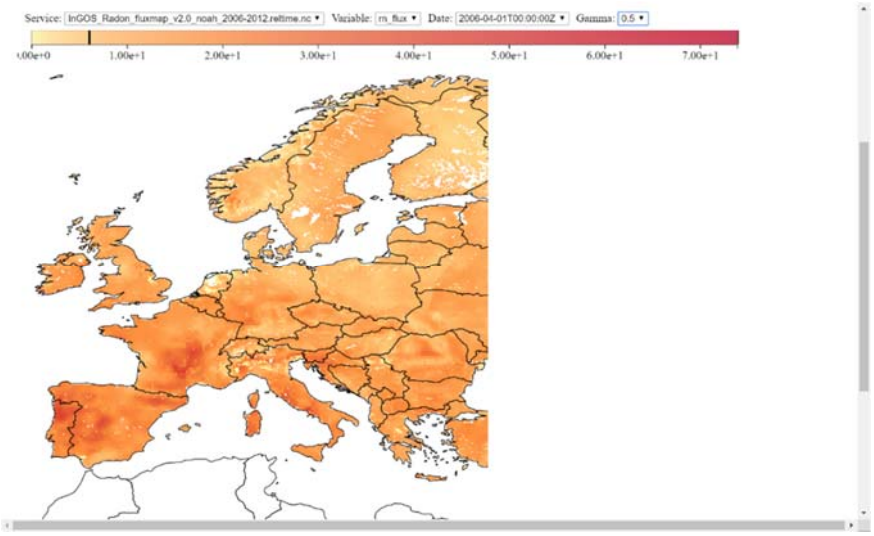


Figure 7 Illustration of the CP service providing direct visualization of spatial data in a netCDF file.

We also deploy the THREDDS server software that provides cataloguing (based on the metadata of the data objects) and WCS and WMS services (See Fig. 8 for a CP client that displays a catalog of data at the THREDDS server and maps data from a selected dataset and variable in that dataset).

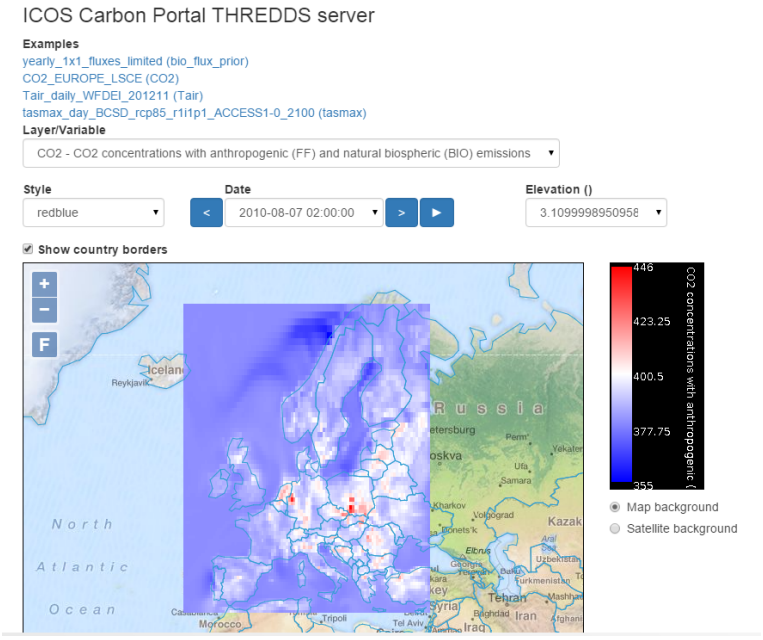


Figure 8

Metadata services

Metadata is data about data. Metadata is needed to make data interpretable, discoverable and searchable: the users need to be able to find datasets that meet their qualifications. It gives information on the characteristics of the data: e.g. quality (limitations and possibilities).

As already described above, the technology chosen by CP for managing the metadata is the Semantic Web/Linked Data. Metadata should include the provenance or 'history' of the data: how was it processed, gap filled, used instruments etc. Metadata should be defined so interoperability is possible; exchange/connect the data to other (large) datasets. Last but not least metadata should make it possible to make the dataset citable and keep track of the use of the datasets (following e.g. the standards given by DataCite).

Metadata is defined in a profile which is a metadata scheme consisting of elements required to describe the data and which following a given standard. For ICOS data the metadata profile will follow the Infrastructure for Spatial Information in the European Community – INSPIRE directive. The metadata standard of INSPIRE is again in turn based on the ISO 19115 standard. This standard is especially relevant for spatially explicit data. The final metadata profiles at CP will however incorporate different standards to fulfil the requirements as given above.

Figure 9 INSPIRE logo.

The metadata profile is being developed using Web Ontology Language (OWL), modelled from the dataset and site descriptions, as well as from the ISO19100 series standards. The first metadata description is derived from the site labelling formulation and the next step is including the dataset description supplied by the 4 TC's.

Based on the metadata profile developed so far, and using the metadata storage system developed at CP (see Ontology/Linked Data above), several metadata services are already run in production and available online.

Generic metadata editing application

This application is developed to allow rapid creation of custom (meta)data-entry web applications. The software code is agnostic of the concrete knowledge model (ICOS activities in our case) and is configured by OWL ontologies. Can be used for rapid formulation of specialized interfaces giving users metadata editing capabilities,

focusing on specific narrow parts of the comprehensive ICOS metadata graph. One example of such interface (already being actively used by ICOS Head Office) is the provisional station information editing app, available online: **https://meta.icos-cp.eu/edit/stationentry/** (described in the previous chapter)

Provisional station information

The metadata edited by the provisional station editing app mentioned above is immediately available in our custom-made metadata-driven visualizations:

Provisional map of ICOS station candidates: https://static.icos-cp.eu/share/stations/

Candidate station info (tabular summary): https://static.icos-cp.eu/share/stations/table.html

SPARQL endpoint and SPARQL query web application

In addition to the aforementioned specialized visualizations of the provisional station info, CP also provides a SPARQL endpoint – a generic metadata service where the metadata can be examined using custom-made queries. It is standards based (see Ontology/Linked data section), available via URL **https://meta.icos-cp.eu/sparql** and can be used to embed ICOS metadata info into arbitrary web pages, even outside of ICOS domains. However, this approach is suitable only for website developers, and to extend the support to the users of intermediate to low technical levels, CP has developed a user-friendly web application for executing SPARQL queries. It is available via **https://meta.icos-cp.eu/sparqlclient/** . In addition to query-editing, it also offers a library of predefined queries (rather short at the time of writing, but planned to expand considerably), making the service of interest even to the less-technical users.

Station labelling

The station labelling web application has been described in the previous chapter, but it is worth mentioning as it is a metadata service and relies heavily on the SPARQL endpoint. In fact, all the CP metadata services make use of it.

Data Ingestion Service

Much of the preparatory work in the Carbon Portal comes together in the design for the Data Ingestion Service. This is foreseen to be the central entry point for all ICOS data and data products. Making use of the EUDAT and EPIC services all data submitted will be receiving a persistent identifier. The service also requires that all relevant metadata is provided according to the metadata profile of the data type. The CP will take care that the metadata for the data object will be enriched using the 'static' metadata that is provided by for example the Thematic Centers. All metadata will be stored in the CP metadata store and of course be linked to the PID of the data object. During the upload of the data object this can also (depending on the metadata profile) be replicated in the B2SAFE store or in the CP data store for fast access for example in the visualization.

Through the PID that will be returned to the data provider the data objects (if the data license allows and this is accepted by the downloader) and the metadata can be accessed through the so called landing page, to which the PID resolves through the handle system.

Also through the metadata search at CP (or from any other data portal linked to it) people can find the data and through the retrieved PID and the resolving landing page get access to the data and/or use the visualization service to browse the data.

A first working prototype of the data ingestion service has been tested in November 2015. In 2016 the metadata profiles will be defined in cooperation with the Thematic Centers and the work will start to enable the enriching of metadata by integrating the Thematic Center metadata systems with the CP metadata store.

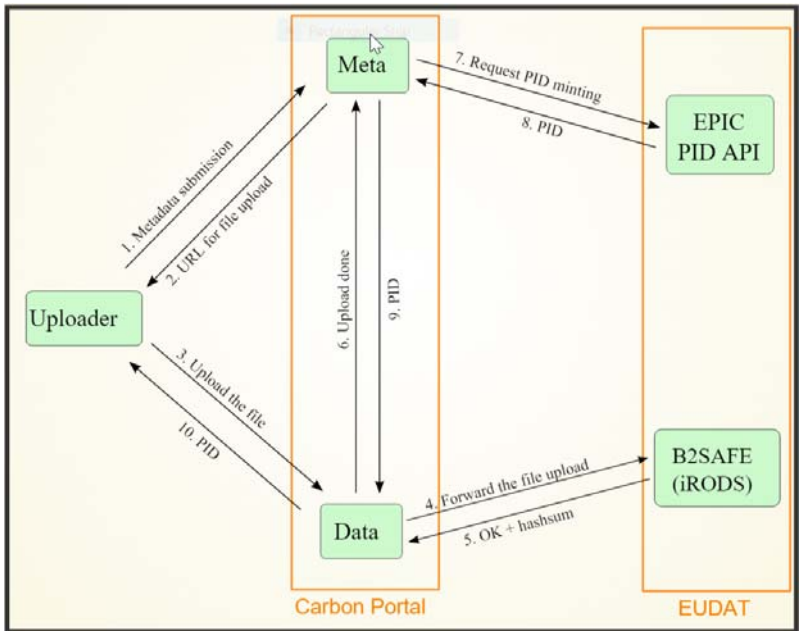


Figure 10 Design for the data flow of the CP data ingestion service

Elaborated products

One of the roles of the CP is to coordinate, facilitate and ensure production of elaborated products based on ICOS data in direct contact with the modelling community. Global as well as regional flux and emission datasets are collected and will be analysed and displayed at the CP.

The work on elaborated products are split in products with a global or regional focus. For the global focus, a start has been made with

1) the synthesis of three inverse model results from European research groups, included in the Global Carbon Project (GCP),
2) the setup of a demonstrator inverse modelling interface, and
3) ObsPack data products (see also next section).

In 2015 we worked on the annual update of an example of a global inverse model (CarbonTracker Europe), of which the results have been included in several (high-impact) publications.

Regional inversion systems (see Fig. 7) with a focus on Europe are currently being developed at several institutes and the resulting flux estimates are not yet operationally updated. Therefore, contacts to the individual PIs of the regional inversion systems are established.

An intercomparison project to harmonize regional CO_2 inversions with respect to time periods or general set-up components (e.g. use of ICOS data) will be initiated in collaboration with the French-Swedish project EUROCOM. This allows a direct comparison and synthesis of the estimated fluxes together with an estimate of associated uncertainties.

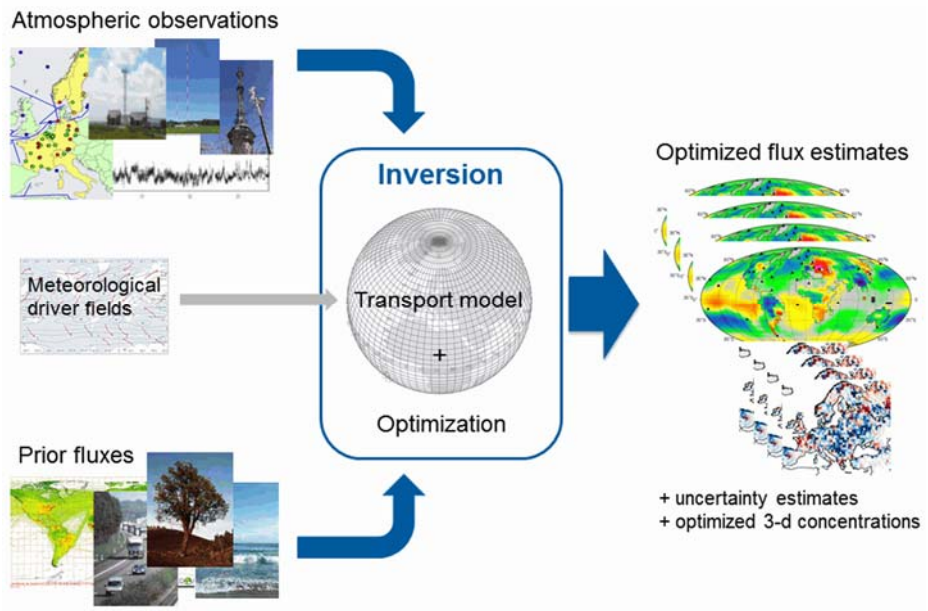


Figure 11 Schematic of atmospheric transport models combined with data assimilation of ICOS observations to improve emission estimates.

The contributions of natural fluxes and anthropogenic emissions to the atmospheric CO_2 concentrations at currently existing and planned ICOS sites are analyzed separately based on model simulations, thus providing a very simplified version of a tool for the evaluation of measurement strategies.

The STILT (Stochastic Time Inverted Lagrangian Transport) model framework provided by C. Gerbig (MPI-BGC) is used to simulate time series at the measurement stations separating the contributions from natural fluxes and anthropogenic emissions (See Fig. 8). In the STILT model framework so-called footprints are generated, which provide spatially resolved, quantitative information linking upstream surface fluxes with the concentration at the measurement site. These footprints are then coupled to surface flux maps to derive tracer concentration changes.

In the current set-up anthropogenic emissions are based on a preliminary version of the EDGARv4.3 inventory plus additional information on the fuel mix per category. Temporal variations (diurnal, weekly and seasonal cycles) of the emissions are included based on time profiles compiled in the TROTREP/POET project in order to provide hourly emission estimates. Additional temporal extrapolation is possible by using fuel consumption data at the national level (British Petroleum statistics).

Biosphere fluxes are based on the diagnostic biosphere model VPRM (Vegetation Photosynthesis and Respiration Model), implemented in STILT. The STILT model framework is installed at the CP and will in future allow for an online computation of the time series based on pre-computed footprints and the different flux fields. The

footprint database and the emissions are continuously updated and extended to other flux datasets and tracers.

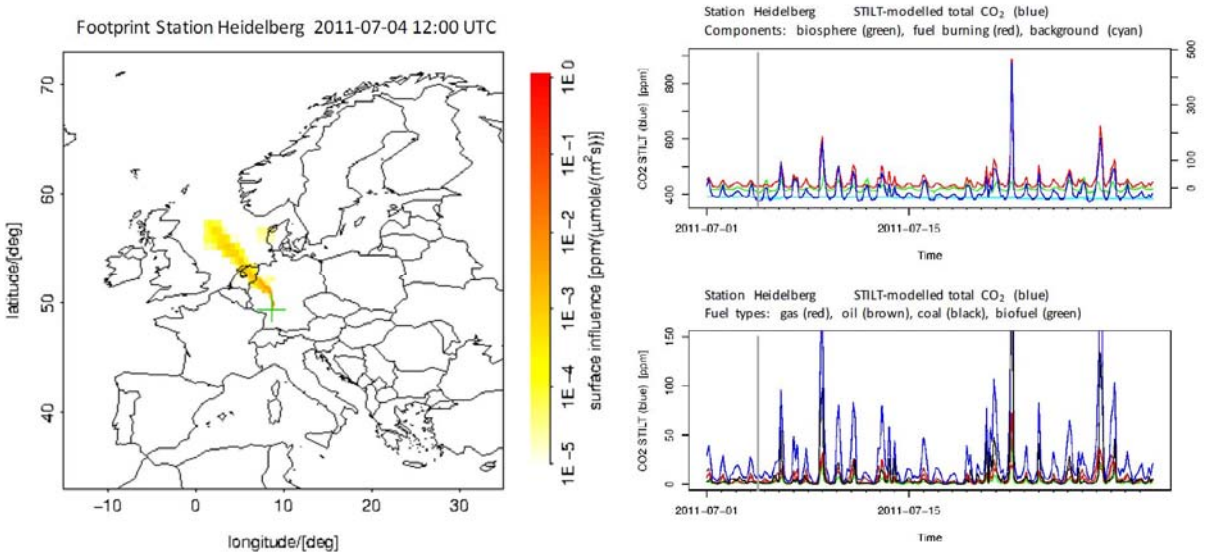


Figure 12 Example footprint (animated) figure showing contributions of different components (upper right) and anthropogenic contributions (lower right)

Data products

Obspack

ICOS CP is actively involved in the community wide Observation Package (ObsPack) products. ObsPack products are prepared using the ObsPack framework (Masarie et al., 2014) which is designed to bring together direct atmospheric greenhouse gas measurements derived from one or more national or university laboratories, prepare them with specific applications in mind, and package and distribute them in a set of self-documenting files. ObsPack products include a set of prepared data sets and metadata, a summary of included data sets, an e-mail address list of all data providers, and the complete set of configuration files used by NOAA to prepare the product.

In 2015 NOAA and ICOS released the first GLOBALVIEWplus (GV+) CO_2 ObsPack cooperative data product (v1.0). This is the first major release of the GV+ cooperative data product. GV+ replaces ObsPack products called PROTOTYPE and is the next generation of the historic GLOBALVIEW products. In contrast to the historic GLOBALVIEW products, GV+ does include actual CO_2 observations. The most significant difference between GV+ and earlier PROTOTYPE products is a nominal hourly time step for all high-frequency (quasi-continuous) measurements. Representative measurements as well as those that may be locally influenced are included and identified using the obs_flag variable.

Ken Masarie (NOAA/ESRL) has visited ICOS CP in April 2015 to work on the joint development of a new ObsPack product: ObsPack_CO2_GLOBALVIEWplus_v1.0, which was released in July 2015. This combined NOAA-ICOS product includes 205 time series with atmospheric CO_2 observations from 25 global laboratories, including 11 from Europe. ObsPack data products are free to download from http://www.esrl.noaa.gov/gmd/ccgg/obspack/data.php.

ObsPack data products are intended to stimulate and support carbon cycle modeling studies. A variety of ObsPack products are available and vary from one another depending on their intended use. GLOBALVIEWplus and its predecessor data products have been used in more than 30 peer reviewed scientific papers thus far. GV+v1.0 has been downloaded 182 times in the first 8 months after its release, mainly by carbon cycle modellers, satellite observation validators and other scientists, but also by e.g. lecturers who include the observations in their courses on the carbon cycle.

Related project work

General

Where relevant for supporting the work of the ICOS Carbon Portal we take part in Horizon2020 or related national projects. In 2014 CP was asked to join the EUDAT2020 and ENVRIplus projects. A significant contribution is delivered to these projects, and the work provided here is completely in line with the tasks of the Carbon Portal for ICOS and strengthens the positions of ICOS in the field of European Research Infrastructures in the field of Environment.

The Carbon Portal also contributes to the national eScience project eSSENCE (e@LU) and the LU internal networking program INTEGRATE ASG.

EUDAT2020

EUDAT2020 (http://eudat.eu/) is a Horizon2020-funded project aiming to set up a Collaborative Data Infrastructure to provide research data services, training and consultancy for European researchers, research communities and research infrastructures.

ICOS has, through the CP and its host Lund University, joined EUDAT as one of its "core communities", with the intention to implement and use a number of the offered services, including B2SAFE (secure long-term data storage), B2HANDLE (minting of persistent identifiers (PIDs) for ICOS data objects) and B2STAGE (data transfers to and from high-performance computing centres).

The CP is also participating in a study on optimizing the connections between EUDAT resources and computational facilities at EGI (European Grid Initiative). In addition, EUDAT is also supporting the CP in obtaining Data Seal of Approval (DSA) certification for the ICOS data repository.

Finally, through EUDAT's data pilot enabling programme, the CP will also be assisting a number of other research communities interested in using EUDAT services. EUDAT2020 started in March 2015, and will end in February 2018. The CP (as Lund University) receives funding from the project for a total of 32 person months.

ENVRIplus

ENVRIplus (http://www.envriplus.eu/) is a Horizon2020-funded project bringing together Environmental and Earth System Research Infrastructures, projects and networks together with technical specialist partners to create a more coherent, interdisciplinary and interoperable cluster of Environmental Research Infrastructures across Europe.

The project is coordinated by ICOS and led from the HO in Helsinki. All in all, 37 institutions are involved in ENVRIplus, representing 21 RIs from four domains (atmosphere, biosphere, marine and solid earth) as well as EGI and EUDAT.

The CP will, through its host Lund University, mainly be engaged in the "Data for Science" theme, leading one work package on data Identification & Citation, and participating in several other tasks related to cataloguing, provenance and metadata.

In addition, the CP will also contribute to the further development of the ENVRI Reference Model. ENVRIplus started in May 2015 and will end in April 2019. The CP (as Lund University) receives funding from this project for a total of 47 person months.

Report on general activities and communication

The complete Carbon Portal team meets regularly for team meetings every two weeks. In the team meetings general and individual progress is reported and discussed. Minutes of these meetings are available through the Alfresco system. Individual feedback on progress and perspectives is gathered twice per year without formal reporting.

Once per year, the first time on 20 January 2015, the CP Management Group meets to discuss the previous year progress and financial report and the coming year working plan and budget. The Management Group is chaired by the Carbon Portal Director and consists furthermore of the ICOS Director-General, the CPD deputy, the Head of the INES Department, the CP project leader at Wageningen University and a representative of the Lund Faculty of Science.

The CPD of course took part in the ICOS IRICOM and RICOM meetings, Contractual Working Group meetings, ISIC meetings, and as guest in the General Assembly meetings.

CPD also has face-to-face meetings, in principal bi-monthly, with Head-Office in Helsinki and since 2016 weekly teleconferences with Head-Office for day-to-day business.

All Monitoring Station Assembly meetings for atmosphere, ecosystem and ocean have been attended by CP representatives. For atmosphere the representative is the CPD, for ecosystems this is Harry Lankreijer and for the Ocean this is Margareta Hellström.

Other meetings not listed here where CP has attended are listed in the following:

Meetings (co-)organized by CP

Users workshop, Lund, February 2014

ICOS MSA Atmosphere, Amsterdam, 24-26 June 2014

Network design workshop, Amsterdam, 27 June 2014

Metadata workshop, Lund, 28-29 September 2014

3rd ICOS/InGOS Summer School, Hyytiälä, 20-29 May 2015

IRIcom face-to-face meeting, Lund, 2-3 June 2015

Nordic ENVRI workshop, Lund, 27-28 October 2015

Meetings attended by CP

AGU Fall Meeting, San Francisco, 15-19 December 2014

ICOS IRIcom face-to-face, Helsinki, 10-11 February 2015

EUDAT2020 kick-off, Helsinki, 24-27 March 2015

ENVRIplus kick-off meeting, Helsinki, 11-13 May 2015

ICOS-NEON Training Workshop, Observatoire de Haute Provence, 8-12 June 2015

ICOS MSA Atmosphere, Dübendorf, 9-11 June 2015

GEO 9th European Projects Workshop, Copenhagen, 15-16 June 2015

JPI Climate workshop on GHG emission verification, Dublin, 24-24 June 2015

EUDAT Project Enabling Workshop, Amsterdam, 8-9 October 2015

ICOS Finland Science workshop, Helsinki, 26-27 October 2015

ENVRIplus week, Prague, 16-19 November 2015

Conferences attended by CP

ICOS Science Conference, Brussels, 23-26 September 2014

GGMT GHG measurements expert meeting, San Diego, 14-17 September 2015

InGOS Science Conference, Utrecht, 21-24 September 2015

COP21, Paris, 20 November-2 December

AGU Fall Meeting, San Francisco, 14-18 December 2015

Bilateral meetings with Central Facilities

ATC-CP, Paris, 18 November 2014

ETC-CP, Antwerpen, 3 March 2015

ATC-CP, Paris, 30 Juni-1 July 2015

CP-OTC, Lund, 29 September 2015

Other

Obspack: Visit Ken Masarie, Lund, 13-17 April, 2015

References

"ICOS Stakeholder Handbook 2013", prepared by the ICOS Preparatory Phase coordination office in March 2013. Available on-line at http://www.nateko.lu.se/icos/docus/Stakeholder_Handbook_2013.pdf

"Technical and Scientific Description of ICOS Research Infrastructure (ICOS RI)" by the ICOS Head Office, and approved by the ICOS Stakeholders' Interim Council (ISIC) in March 2014. Report submitted to the European Commission in spring 2014. Available on-line at http://www.nateko.lu.se/icos/docus/Technical_Scientific_Description_ICOS_2014-03-05.pdf

"ICOS Carbon Portal concept paper", prepared for the ICOS Head Office by a writing team led by A. Lindroth (2012). The document was approved by the ICOS Stakeholders' Interim Council (ISIC) in autumn 2012. Available on-line at http://www.nateko.lu.se/icos/docus/Carbon_Portal_Concept_Paper.pdf

"Sweden's application to host the ICOS Carbon Portal central facility" by A. Lindroth et al. Submitted to the ICOS Stakeholders' Interim Council (ISIC) by the Swedish Research Council in January 2013. Available on-line at http://www.nateko.lu.se/icos/docus/Carbon_Portal_Application_SE-NL_2013-01-23.pdf

"ICOS Carbon Portal: summary of activities 2013" by M. Hellström, H. Lankreijer and A. Lindroth (2014). Report submitted to the Swedish Research Council in April 2014. Available on request from the authors (Margareta.Hellstrom@nateko.lu.se).

Appendix A: Tasks and deliverables of the operational phase

With reference to the CP "roles" defined in the *CP Concept Paper*, the following tasks and associated deliverables have been identified for the operational phase of the Carbon Portal.

1. ICOS data flow coordination (Role 3)

This involves maintaining the ICOS data and metadata standards (Role 3), as well as monitoring the use of standardized data exchange protocols and data descriptions within the RI.

Deliverables include a summary report to the Head Office.

2. Data collection and the digital curation of all ICOS data (Role 1)

Ensuring long term data archiving and (back-up) storage (Role 1) of ICOS data from all levels (raw data to elaborated products); Safe storage, future long-term access as complement to TC delivered data; Storage of datasets in different versions.

Deliverables include providing PIDs to data sets made available at the web portal, and a report to HO on the status of data curation, and the long and short-term storage of ICOS data products.

3. Providing access to data with CP web platform (Roles 2, 4 and 10)

Data mining, data extraction, collocation: to provide effective tools to publish, discover, and retrieve observations according to users need; Providing web services to users, including implementation of advanced web service for presentation, visualization, retrieval and processing of data, datasets, metadata and elaborated data;

Deliverables include access to observational datasets, operational web portal (24/7 access); software tools

4. Keeping track of ICOS data usage (Roles 5, 6 and 7)

User registration and traceability. Traceability of downloaded ICOS data; Mentions and references to ICOS data products in publications, reports, media etc. (Role 6)

Deliverables: User database, usage database, bibliometrics database

5. Coordination of data related products (Roles 8, 9 and 11)

ICOS data description and releases publications (Role 8); publication of the ensemble ICOS data; Coordinate, facilitate and ensure production of elaborated products based on ICOS data (Role 9); Interface with other data portals in and outside Europe (Role 11) - Exchange of data with other data portals

Deliverables include published datasets and related publications; Written synthesis report (European greenhouse budget);

6. Running and planning CP activities

Deliverables: Annual report, budget, and operational plan…

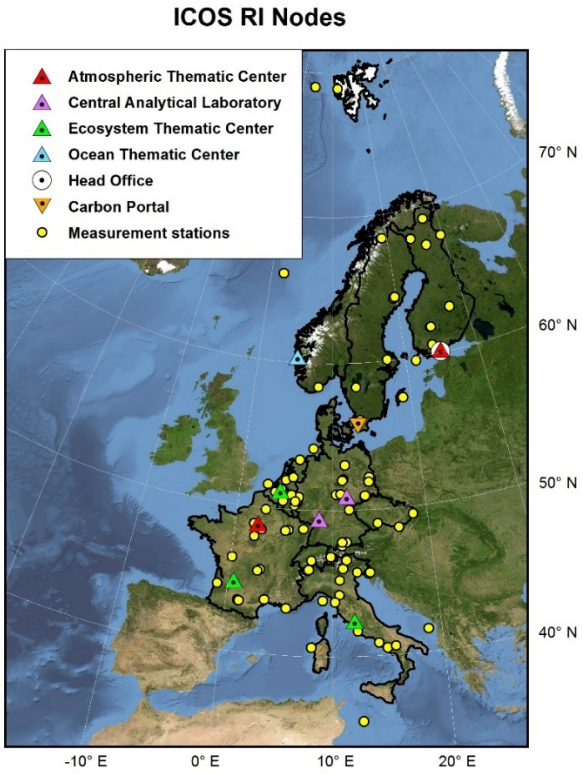
About ICOS and the ICOS Carbon Portal

What is ICOS?

ICOS RI is a pan-European Research Infrastructure for quantifying and understanding the greenhouse gas balance of the Europe and its neighbouring regions. ICOS has received the official legal status as an ERIC from the European Commission on 20 November 2015. ICOS' Head Office is located in Helsinki, Finland. Currently ICOS has 11 member countries and involves more than 100 measurement locations where greenhouse gas concentrations and fluxes are measured. ICOS provides long term and high quality observations, and will hopefully expand to many more countries and stations in the near future. Read more about ICOS...

What is the Carbon Portal?

ICOS Carbon Portal is part of ICOS ERIC and offers access to research data, as well as easily accessible and understandable science and education products. It is a joint effort from Sweden and The Netherlands. The Carbon Portal is is located in Lund and is hosted by the University of Lund (Sweden). The contribution from The Netherlands is located at Wageningen University. All measurement data available in the Carbon Portal is quality controlled through the ICOS thematic centres, divided into Ecosystem, Atmospheric and Ocean Thematic Centres and a Central Analytical laboratory. Dedicated researchers all over the world will contribute to the elaborated products catalogue.



ICOS Carbon Portal, c/o Sölvegatan 12, 223 62 Lund, Sweden Email: info@icos-cp.eu
ICOS RI website: http://www.icos-ri.eu ICOS CP website: http://www.icos-cp.eu