



CoCo2

Prototype system for a
Copernicus CO₂ service

Demonstrator of the updated observational FAIR data pipeline

Alex Vermeulen

Ute Karstens

Oleg Mirzov

Dario Papale

Leo Rivier

Elena Saltikoff



Co-ordinated by



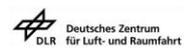


CoCO2

Prototype system for a
Copernicus CO₂ service

D7.8 Demonstrator of the updated observational FAIR data pipeline

Dissemination Level:	Public
Author(s):	Alex Vermeulen (ICOS ERIC) Ute Karstens (ULUND) Oleg Mirzov (ULUND) Dario Papale (CMCC) Leo Rivier (LSCE) Elena Saltikoff (ICOS ERIC)
Date:	22/12/2023
Version:	1.0
Contractual Delivery Date:	31/12/2023
Work Package/ Task:	WP7/ T7.4
Document Owner:	ICOS ERIC
Contributors:	ULUND, CMCC, LSCE
Status:	Final



CoCO2: Prototype system for a Copernicus CO₂ service

Coordination and Support Action (CSA)
H2020-IBA-SPACE-CHE2-2019 Copernicus evolution –
Research activities in support of a European operational
monitoring support capacity for fossil CO₂ emissions

Project Coordinator: Dr Richard Engelen (ECMWF)
Project Start Date: 01/01/2021
Project Duration: 36 months

Published by the CoCO2 Consortium

Contact:
ECMWF, Shinfield Park, Reading, RG2 9AX,
richard.engelen@ecmwf.int



The CoCO2 project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958927.



Table of Contents

1	Executive Summary	9
2	Introduction	9
3	Principles of data ingestion and storage in ICOS	9
3.1	ICOS Data and its data level definition	9
3.2	FAIR and data object granularity	10
3.3	Data ingestion	10
3.4	Data upload	11
3.5	Data license	11
3.6	Timeliness	11
3.7	Data versioning and collections	11
3.8	Datacite DOI minting	12
3.9	Data upload façade for daily packages of half-hourly files	12
4	Data access and data processing integration	12
4.1	Data access through the ICOS Data Portal graphical user interface	12
4.2	Data access through standard internet protocols	13
4.3	Data access through python	13
4.4	Data processing workflows	13
5	New data pipelines for CoCO2	14
5.1	Requirements for new data pipelines	14
5.2	Description of the new data pipelines in CoCO2	15
5.2.1	CoCO2 campaign data pipelines	15
5.3	New data pipelines in ICOS Cities/PAUL	20
5.4	Other data streams	24
5.4.1	Atmospheric GHG data products	24
5.4.2	¹⁴ CO ₂ historic data Obspack collection	24
5.4.3	Netcdf spatiotemporal data	24
5.4.4	Generation of FastTrack NRT Obspack data products	24
5.5	Progress in implementing the new data pipelines	25
6	Conclusion	25
7	References	25

Figures

Figure 1	Overview of the ICOS RI data flow. Arrows indicate the transfer of data and metadata objects between or within sub-communities of the ICOS -RI. Black arrows indicate both data and describing metadata, red indicates only data and broken line arrows only	14
Figure 2	Location of the two FORTH eddy covariance flux towers in the town of Heraklion, Crete	15

Figure 3 Surroundings of the Krakow PL-Krk CO ₂ eddy flux site and flux average footprint projected on the map of Krakow. The measurement mast is visible on the rooftop of building at the bottom left corner.	17
Figure 4 Photographs of balloon and drone launch for the greenhouse gas profile measurements in Krakow.	18
Figure 5 Example air core vertical profile at 26 April 2021 of CO ₂ and CH ₄ mixing ratios at Sodankylä (blue lines) compared with simultaneous TCCON data.	19
Figure 6 Preview at Carbon Portal of the Sodankylä AirCore vertical profile for CO ₂ for 1 Aug 2018.	19
Figure 7 Schematic overview of measurements in ICOS Cities. Each city will have its own constellation of observations using one or more of the types shown here and of each deployed type one or more instances.	20
Figure 8 Time schedule for the implementation of the different measurement systems in the ICOS Cities project.	21
Figure 9 Preview of CO ₂ observations from low cost sensors in the first half of 2023 at 6 stations in Zürich, as available from the ICOS Cities data portal.	23
Figure 10 Overview of stations in Zürich (red dots) with low-cost CO ₂ and meteorological data available at the ICOS Cities portal.	23
Figure 11 Locations of the 5 Total Column EM27 measurements in München and an example time series plot of the XCO ₂ observations for 1 October 2023 for them Northern- and Southernmost stations, showing a clear gradient of about 1-2 ppm developing over the city over the course of the day.	23
Figure 12 Preview of ¹⁴ CO ₂ historic time series from Niwot Ridge.	24

Tables

Table 1 Overview of the implementation progress as of February 2023.	25
---	----

Glossary

ATC	Atmosphere Thematic Centre
CAL	Calibration Laboratory
CC0	Public domain data licence, meaning data is free to use, copy and distribute
CC4BYPID	Creative Commons Attribution 4.0 International data licence, meaning that data is free to use and distribute, but requires that attribution to the data provider is required, and that the licence should be passed on at redistribution or derived products
CP	Carbon Portal
CRDS	Cavity Ring Down Spectroscopy
Datacite	Provider of DOI registration for research data, extending the DOI metadata with research data specific additional metadata
DOI	Data Object Identifier, a special PID that requires a specific set of metadata to go with the data object, describing provenance and attribution
ECMWF	European Centre for Medium-range Weather Forecasts
Eddy flux	Eddy-covariance (also known as eddy correlation and eddy flux) is a key atmospheric measurement technique to measure and calculate vertical turbulent fluxes within atmospheric boundary layers
ETC	Ecosystem Thematic Centre
GUI	Graphical User Interface
Handle	Provider of 'bare-bone' PIDs with minimal metadata requirement, basis of the DOI and ICOS PIDs
ICOS	Integrated Carbon Observations System
JSON(-LD)	JavaScript Object Notation (for Linked Data)
Jupyter	Project to develop open-source software, open standards, and services for interactive computing across multiple programming language
ICOS Data Levels	
Level 0 (L0)	Raw data as delivered by the instrument, uncalibrated and not necessarily in useful units (e.g. mV, mA, μ S)
Level 1 (L1)	Near real time data, automatically quality controlled, reduced (averaged) and calibrated
Level 2 (L2)	Final quality controlled data, including manual control and data reduction
Level 3 (L3)	Elaborated data, where additional information has been added to ICOS data, using a model and/or other observations, often gridded data
netCDF	Network Common Data Form, set of software libraries and self-describing, machine-independent data formats that support the creation, access, and sharing of array-oriented scientific data
Ontology	Way of showing the properties of a subject area and how they are related, by defining a set of concepts and categories that represent the subject. Used mainly to organize and represent knowledge in a machine-readable way.
OTC	Ocean Thematic Centre
PID	Persistent Identifier
QA	Quality assurance, measures taken to make sure the data has the proper quality

QC	Quality control, checks performed on the data to see whether the desired quality level has been reached
RDF	Resource Description Framework
RI	Research Infrastructure
Sparql	Query language for triple stores
SSL	Secure socket layer, basic protocol for safe encrypted one-to-one data transfer over the internet
TC	Thematic Center (ATC, OTC, ETC)
TCCON	Total Column CO2 Observing Network
Triple	A semantic triple (or RDF triple) is the atomic data entity in the RDF data model. A triple is a set of three entities that codifies a statement about semantic data in the form of subject–predicate–object expressions (e.g., "Bob is 35", or "Bob knows John")
Triple store	Database for storing and retrieval of triples
Turtle	Terse RDF Triple Language
WMO	World Meteorological Organisation
XML	eXtended Markup Language

1 Executive Summary

This report, deliverable of the CoCO2 project, funded by the European Union's Horizon 2020 research and innovation program, focuses on integrating new data streams from various tracers and instruments into the existing ICOS (Integrated Carbon Observations System) data infrastructure. It elaborates on the principles of data ingestion, storage, access, and processing in ICOS, describing various data levels from raw data (Level 0) to elaborated data (Level 3). The report details the development of new data pipelines for CoCO2, including specific case studies like the Heraklion eddy flux towers and the Krakow eddy flux tower. It also covers the data processing workflows and integration with ICOS Cities/PAUL project, which aims to enhance greenhouse gas monitoring in urban environments. The document concludes by highlighting the adaptability of the ICOS-CP process to incorporate diverse data streams while maintaining high data quality standards.

2 Introduction

This demonstrator describes the implementation of the data streams from additional tracers and new instrumentation as described in CoCO2 deliverable 7.7. The foundation for the demonstrator is the existing data infrastructure developed for ICOS at the ICOS Carbon Portal. For clarity sake and to enable stand-alone reading of this document, some of the information in this deliverable 7.7 is repeated here.

The ICOS research infrastructure generates and processes many data streams that have a dedicated pipeline that involves one or more parties next to ICOS Carbon Portal (ICOS-CP) central data repository, but all data in the end is published by ICOS-CP together with rich metadata. Most of the data streams curated by ICOS-CP are operational ICOS data and have their own specific requirements for processing and usage. In the ICOS data streams we differentiate between observational data streams and elaborated data streams, the latter are mostly based upon ICOS observational data and combine these with other observational or modelled data. Elaborated data are mostly contributed by other parties or the ICOS community.

New observational data streams can be incorporated, and this will require in most cases definition of new stations and instruments, new data types for raw and processed data and the associated descriptive metadata, and setting up a processing pipeline to process the raw data and creating the needed account and right authorisations for upload of the data.

For this provisions needs to be made to develop and test the new data pipelines without disturbing the operational data provision and in the early development phase before regular data provision for this temporary solutions have been developed based on file sharing, e.g. through Nextcloud and/or sftp.

3 Principles of data ingestion and storage in ICOS

3.1 ICOS Data and its data level definition

ICOS data is data that has been generated by the ICOS infrastructure. Data is always associated with a defined project, ICOS is just one of the projects. At ICOS-CP we differentiate between three major data levels for all data:

- L0: Level 0 or raw data. This is untouched data as received by the instrument or data logger. Data could be compressed by a loss-less algorithm like zip or gzip for the sake of efficiency. L0 data is in general not usable for normal users, as units are not always meaningful and information on what the values are representing is lacking from the data object itself.

- L1: Level 1 or near-real time data. Data produced from L0 data and that has been converted to meaningful units, has been automatically quality controlled and calibrated, and in most cases reduced due to averaging and other data reduction techniques. L1 data is used extensively by station PIs and thematic centres for quality control and monitoring the instrumentation but can also be used by external users for analysis for specific applications with (near-)real time requirements. L1 is usually a time series.
- L2: Level 2 or final quality-controlled data. These are the main data product of ICOS and fulfil all the quality control requirements of ICOS, including the manual quality control and data flagging by the PI and thematic centre. L2 data is usually a time series.
- L3: Level 3 or elaborated data. This is data that has been generated in most cases using L1 or L2 data but then combined with other (model) data. Examples are spatially extrapolated data (SOCAT, FLUXCOM) or inverse model estimated priors or posterior flux data. L3 products are mostly contributed by the ICOS community or external parties and are usually spatiotemporal data, often formatted in cf-netcdf.

3.2 FAIR and data object granularity

The ICOS data system is designed from the ground up to be fully FAIR, open and transparent. The data lifecycle is in detail explained in the ICOS Data Improved Lifecycle (2023 update, <https://doi.org/10.18160/D2JV-KB6B>).

Granularity of the data objects (binary blobs represented by files) is decided by the uploading user and in case of operational ICOS observational data it is mostly as daily files for L0+L1 and annual releases for L2 (see ICOS data types in chapter 3.2).

3.3 Data ingestion

Ingestion and storage of any observational data streams by ICOS-CP follows the following principles:

- Data ingestion is a completely machine to machine operation that does not require any human intervention
- Data is accepted only for known and registered data types, each data type requires a specific authorisation for upload and for data levels > 0 the variables that the data contains are defined in rich metadata
- Data is stored without any modification, including their original filename
- Data should be accompanied by a descriptive metadata package, that among others contains the SHA256 checksum, the data type and detailed provenance information, and links to, where applicable, stations, instruments and persons, protocols and descriptive documents or related scientific publications
- Only data is registered that has been completely transferred and fulfils all requirements with regards to metadata, passed the validation checks (when applicable), and passed the checksum calculation
- The data object receives a unique Handle Persistent identifier (PID) that is linked to the ICOS-CP database and when resolved results into a machine and human readable landing page. The landing page provides access to the data through qualified and persistent link and lists the metadata linked to the ICOS-CP ontology. PID resolving is completely dynamic and the landing pages are generated on the fly based on the current state of the metadata database. The ICOS PID contains a character encoded part of the data object checksum so it is uniquely linked to the data object.

- Data objects are stored permanently and can be always retrieved through their PID together with all the rich metadata.
- The data object is stored at the ICOS-CP and a copy is simultaneously stored at the EUDAT CDI B2SAFE services at CSC (Finland).

3.4 Data upload

Data ingestion is technically described at the ICOS-CP Github repository (<https://github.com/ICOS-Carbon-Portal/meta#upload-instructions-scripting>, <https://github.com/ICOS-Carbon-Portal/data#instruction-for-uploading-icos-data-objects>).

There is also a user friendly GUI to assist in the manual upload (<https://meta.icos-cp.eu/uploadgui/>), or update metadata elements. All upload requires login and proper authorization for the respective data type to be uploaded.

Individual (one-off) and/or irregular publication of data will have to be done by manual curation by ICOS-CP staff, that serve as Data Steward, to inform and assist users on the (meta)data requirements and to perform the quality control and submission of metadata and data.

3.5 Data license

All ICOS data is licensed under the Creative Commons Attribution 4.0 International (CCBY 4.0, <https://creativecommons.org/licenses/by/4.0/>). All metadata is licensed as CC0.

The ICOS-CP also supports other data licenses, most notably Public Domain (CC0, <https://creativecommons.org/share-your-work/public-domain/cc0/>) for some of the L3 data. Other standard and internationally accepted data licenses can be added when needed.

As often projects or scientific journals require that data publication is postponed until a certain release date, ICOS-CP supports that at ingestion an optional moratorium date is associated with the data object, which makes that the data object is only visible and accessible after that date.

3.6 Timeliness

All ingested data from L0 to L3 is available immediately at the ICOS-CP to all users without restriction, unless a moratorium has been specified in the ingestion metadata. Another exception is raw data from the atmosphere domain that is only available on request.

Observational data are the core of ICOS data and each of the domains provides in principle the raw data provided by the instruments as prompt as possible (within 24 hours) to ICOS-CP. This raw data can be sent to ICOS-CP directly by the instruments or data loggers, for later pick up and further processing, or after collection at one of the thematic centres. In case no internet connection is available, like in the case of ships of opportunity or buoys without satellite data connection in the ocean domain, data is transferred as soon as the ship enters the harbour, or when the station is visited for maintenance.

ICOS L1 data is provided usually within 24 hours after the generation of the underlying L0 data. L1 data always covers the period starting from the end of the latest L2 release to the current day in so called growing time series.

ICOS L2 data is generated at least once per year as major release in the first half of the new year. In case of the ecosystem a second annual release is provided at the end of the growing season of the northern hemisphere. L2 data always covers the complete period that the station provided ICOS data. From the ICOS-CP side there is nothing preventing from more frequent L2 releases.

3.7 Data versioning and collections

At upload, data objects can be marked in the metadata as new versions of previous data objects by linking to their PID. The older version will normally get a deprecated state in the database and by default won't be visible any more in the ICOS-CP data portal for search

(unless the option to show deprecated data is enabled). However, older version PID(s) are permanent and remain valid and landing pages will be generated when called. Both the landing page of the latest and the older versions of the data will point to the respectively previous and newer version of the data, forming a linked chain up to the earliest data version.

Next to L0 to L3 data objects it is possible to generate so-called collections of multiple data objects. This done by minting a Handle PID for a list of PIDs that all point to data objects. All landing pages of data objects in a collection will refer to the collection(s) to which they belong. Collections can also be versioned. Collections can be downloaded and then the system will on the fly compose a zip file that contains all the data objects in the collection, including an index of all included data objects and their PID.

3.8 Datacite DOI minting

Datacite DOIs (Data Object Identifier) can be minted for data objects or collections. Handle PIDs prescribe only a minimal amount of metadata to be registered. ICOS adds to this its own rich metadata model that varies with data type to the PID associated metadata. Datacite DOIs like ICOS use Handle PIDs but then use another simpler metadata model that mainly takes care of the attribution and some provenance of the data. The metadata schema currently is version 4.4 (<https://schema.datacite.org/meta/kernel-4.4/>).

Datacite DOI minting is in most cases a manually curated process that at ICOS-CP is assisted by the upload GUI and the DOI GUI (<https://doi.icos-cp.eu/>). All major ICOS products such as L2 data collections and L3 products are minted DOIs. Data objects or collections with a DOI show in their landing page information the reference to their DOI and show the citation string connected to the DOI.

Users can pre-mint Datacite DOIs through the ICOS-CP DOI GUI, for example for a L3 product they want to publish through ICOS-CP. After curation by the ICOS-CP and ingestion of the data this DOI will then be published to and indexed by Datacite. These pre-minted DOIs are often required by publishers to include in a draft scientific paper to refer to supplementary data sets.

3.9 Data upload façade for daily packages of half-hourly files

An exception in the standard workflow is made for the ingestion of the raw data from ecosystem eddy covariance towers and phenocams, to accommodate for real time, half hourly transfer of data by instrumentation in the field that is not able to transfer data in the most up to date SSL protocol and does not have the computational power to calculate SHA256 checksums.

For this ICOS-CP provides the so called façade API for data upload (<https://github.com/ICOS-Carbon-Portal/data#simplified-etc-specific-facade-api-for-data-uploads>). Data are then merged into daily data files that are submitted by the façade on behalf of the station as daily L0 data through the standard ingestion routine of the ICOS-CP.

4 Data access and data processing integration

4.1 Data access through the ICOS Data Portal graphical user interface

At <https://data.icos-cp.eu/portal> a multi-faceted search function allows the user to drill down through the over 1 million data objects stored at the portal, to the data objects needed. Data objects can be selected for example by any combination of project(s), station(s), data level(s), variable name(s), domain(s), keyword(s), acquisition or submission interval, and/or through a spatial selection on the map. From the list of data objects the user can select all or some and download them directly, or add them to a virtual shopping cart for later download together with data from other queries. For most data types one or more data objects can be selected for preview as time series or spatiotemporal data on the map. The data object list of course also contains the PIDs as links that provide access to the data object landing pages (also machine

readable by content negotiation and also encoded in the schema.org scheme) that contain the rich metadata. From each landing page the data can be downloaded directly or if available also previewed. Downloading requires the user to accept the ICOS data licence.

To present all the results the portal GUI is using the same linked open data technology that is used in the back-end. The actual (SPARQL) query resulting from the GUI settings in the faceted search can be viewed and copied for execution in an automated machine to machine operation. These queries and all data and metadata access uses standard and simple open web technology without requiring login or authentication (see section 4.2). It is possible to register an account and login, which will enable the user to bypass the licence acceptance page through a one-time acceptance of the data licence in the account profile. In the user profile also a record of previous downloads and all individual preview settings are stored for convenience.

4.2 Data access through standard internet protocols

Just like upload, also access to download of data and metadata through PID landing pages is done through basic https GET calls. Metadata exchange can take place in a variety of metadata profiles (ICOS JSON, rdf/xml, rdf/turtle, iso-19115 XML) through content negotiation. For download of ICOS data the download requires authentication through a token, available at the user profile page or using the procedure described at <https://github.com/ICOS-Carbon-Portal/meta/#authentication>. Authentication is required to assure that the ICOS data licence is accepted and to allow usage tracking and service abuse prevention.

4.3 Data access through python

ICOS provides a python library for easy access to the ICOS data and metadata. The library can be installed through pip. Detailed information is available at <https://pypi.org/project/icoscp/> and <https://icos-carbon-portal.github.io/pylib/>. When using the ICOS Jupyterhub this library is already available, and all data will be available directly without need for download to local storage. In this Jupyterhub dedicated project groups can be defined that have access to a common shared storage area for exchanging data and code and that can also have access to dedicated associated group disks at the ICOS Fileshare, for example to access non-public measurement data from campaigns of data pipelines in the test phase and results from prototype data pipelines.

4.4 Data processing workflows

All three ICOS thematic centers (TCs) and the central labs have developed automated and standardized workflows that process raw data to daily NRT L1 data and regular final quality controlled L2 data releases. These workflows are following detailed specifications that are based on the community guidelines for high precision and long term consistent and highest quality data, following and in many cases even leading the WMO, FLUXNET recommendations. The general dataflow is shown in Figure 1

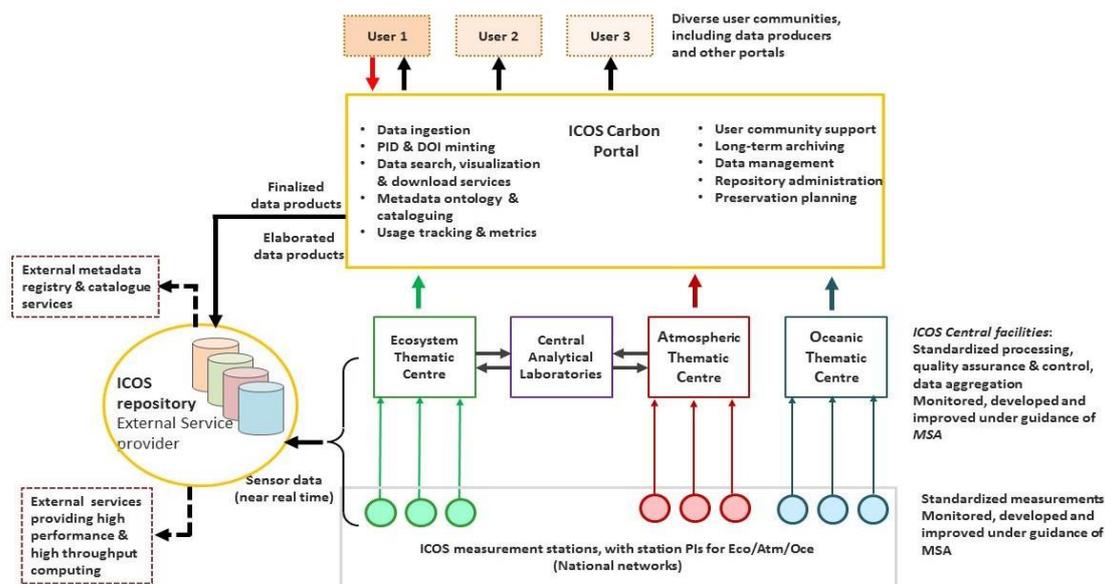


Figure 1 Overview of the ICOS RI data flow. Arrows indicate the transfer of data and metadata objects between or within sub-communities of the ICOS -RI. Black arrows indicate both data and describing metadata, red indicates only data and broken line arrows only

The Atmosphere TC (ATC) provides the station PIs with daily detailed overviews of the processed data, including statistics on uncertainty and calibration quality. Both the OTC and ATC provide GUI applications for the PIs to flag and assess the quality of the measurements and provide important metadata on the configuration and state of the measurements. In the ETC all metadata is exchanged with PIs using the BADM standard (<https://ameriflux.lbl.gov/data/badm/>). The daily metadata exchange between thematic centers is processed by the ICOS-CP into a unified and harmonized metadata model that describes the stations, instruments, measured quantities, methods, provenance, persons and organisations and their relations.

5 New data pipelines for CoCO2

5.1 Requirements for new data pipelines

New observations that cannot use any of the existing pipelines in ICOS, for example because the instruments are of a new type or because new variables not yet described are measured or through methods that do not follow the ICOS protocols, will in many cases require the formulation of a new data type. If the data will not be ICOS data one needs to define an associated project, which can be one of the existing projects, e.g. *Miscellaneous* or a newly defined project, for example named CoCO2. Only the ICOS-CP can create new projects in the ICOS metadata, upon request by email to info@icos-cp.eu.

Each data type is connected to a single project. For the project “Miscellaneous” there were before CoCO2 three data types available that are available for new observational data. These are the data types “Fluxnet Product”, “Atmospheric Measurements Results Archive” and “NetCDF spatial data”. The disadvantage of these data types is that they are either very specific or very unspecific respectively. Data type can also connect to a value type and a variable specification. Again data types can only be added by the ICOS-CP and decisions on new data types will be made through consultation with the data provider during the first data curation steps.

Data objects of a certain data type can be associated with a station and sampling height, and one or more instruments. Stations must exist in the ICOS-CP database before they can be referenced as associated to the data object. The same applies to instruments. Stations

minimally need metadata with coordinates in longitude, latitude, and elevation (WGS84), but can be also associated with ICOS, WMO, organisations, personnel, and their roles. In principle for ICOS data this metadata is provided by the thematic centre of the respective domain of the station. But in case of no association with a Thematic Centre, when new measurement data are added through a new data type, responsibility for the provision of metadata needs to be assigned to the data provider or an associate so that the ICOS-CP can create in the metadata database the associated stations, instruments, organisations, and persons. In the upload metadata these need to be referenced using the URL's that the ICOS-CP will provide. In case the data processing pipeline will be managed by an existing thematic centre this will be arranged between the TC and the ICOS-CP based on input from the measurement PI, otherwise by direct contact between the data processor and the ICOS-CP.

In case a new project is planned involving a central data processing unit that will handle the metadata and data pipeline, access can be given to a specific part of the metadata that as a start can be managed directly in the ICOS-CP metadata database and ontology through the manual admin interface, this is also the way that the OTC interfaces with the ICOS-CP. A relatively easy to use metadata editor can be made available to provide metadata, an example is the OTC entry editor at <https://meta.icos-cp.eu/edit/otcentry/>.

5.2 Description of the new data pipelines in CoCO2

5.2.1 CoCO2 campaign data pipelines

Three partners deliver new campaign-wise observations in the CoCO2 project. All data files resulting from the CoCO2 project will become findable through the ICOS Data Portal through the keyword CoCO2. An example search URL to achieve this is: <https://data.icos-cp.eu/portal/#%7B%22filterKeywords%22%3A%5B%22CoCO2%22%5D%7D>

5.2.1.1 FORTH's Heraklion eddy flux towers for urban CO₂ fluxes

RSLab, FORTH provides, for two locations in Heraklion, Crete urban eddy covariance flux observations of CO₂. Detailed metadata on the flux sites can be found at <http://rslab.gr/fluxtowers.html>. HECKOR (since 2016) is located in the city centre and HECMAS (since 2021) in at a residential area, Mastambas south of the centre. Furthermore, FORTH has established an online interface for both flux towers displaying near-real-time data (http://rslab.gr/heraklion_eddy.html).

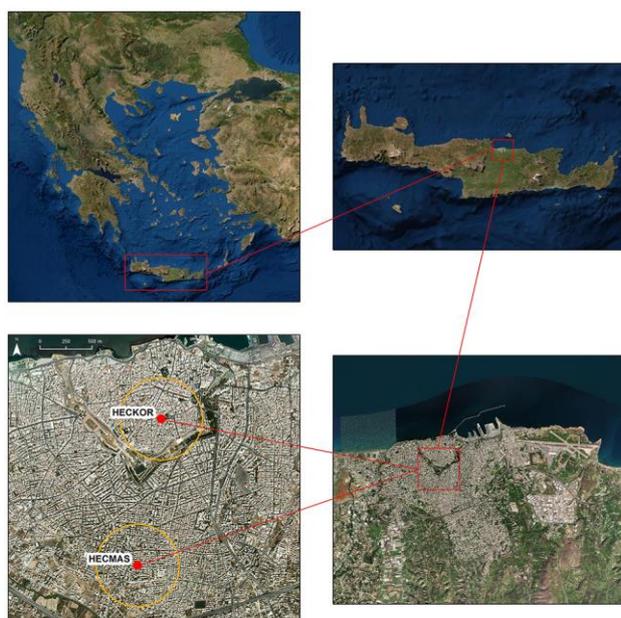


Figure 2 Location of the two FORTH eddy covariance flux towers in the town of Heraklion, Crete

Both towers deliver flux data using ICOS compliant hardware and their data streams can be integrated using the ETC data pipeline, using the platinum option, both towers delivering about 40 MB of raw data per day that can be ingested using the ICOS-CP ETC Façade interface (Section 3.7). From there the data can be picked up by the ETC for daily processing in the cloud, resulting in daily NRT-24h half-hourly automatically quality controlled data flux product and an annual Level 2 final quality controlled half-hourly data product. Published through the ICOS-CP together with the regular ICOS flux data products.

However, as the fluxes concern an urban area with no ecosystem, the station definition is different and also the standard gap-filling techniques used in the processing that take into account vegetation dynamics do not apply, so that some experimenting on choosing working approaches is required, also considering the challenging conditions like sea breeze and high sensible heat fluxes.

As urban eddy covariance flux measurements are still in the experimental phase, and the eddy flux measurements at Krakow and Heraklion are not part of the ICOS Cities project, the processing for both Krakow and Heraklion have been processed by the PI. Data submission has been coordinated with respect to data format and metadata and data is expected to arrive in January 2024.

HECKOR measurements have shown the impact to the CO₂ emissions due to an intervention started in January 2018 by the municipality of Heraklion, introducing traffic regulations and rebuilding initiatives (road closures, pedestrianizations, minor planting), and further reductions during the COVID lock-downs in 2020 and 2021.

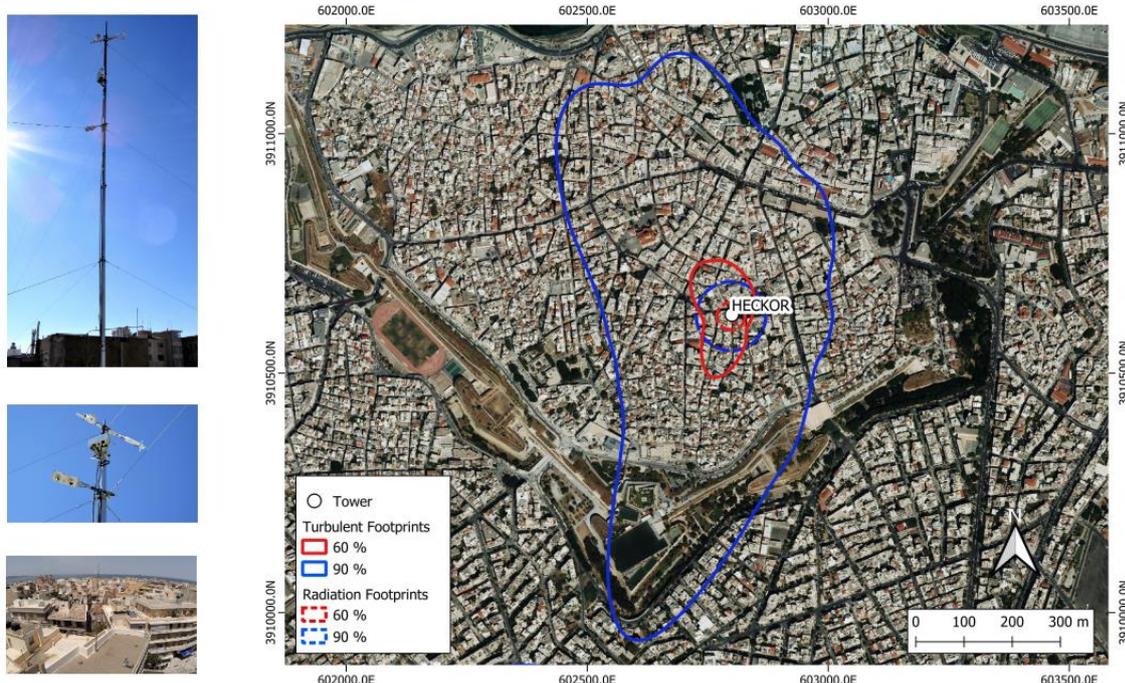


Figure 4 On the left, surroundings of HECKOR, on the right CO₂ eddy flux site and flux average (1-year) footprint projected on the map of Heraklion city center.

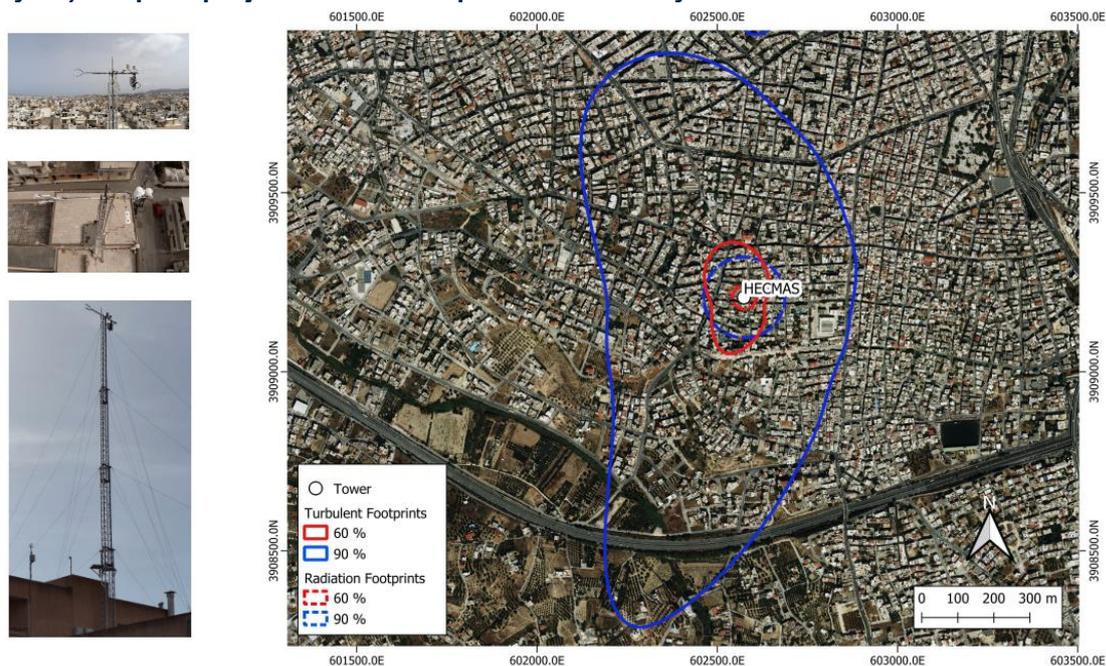


Figure 5 On the left, surroundings of HECMAS, on the right CO₂ eddy flux site and flux average (1-year) footprint projected on the map of Heraklion’s residential area, Mastambas.

5.2.1.2 Krakow (Poland) eddy flux tower PL-Krk

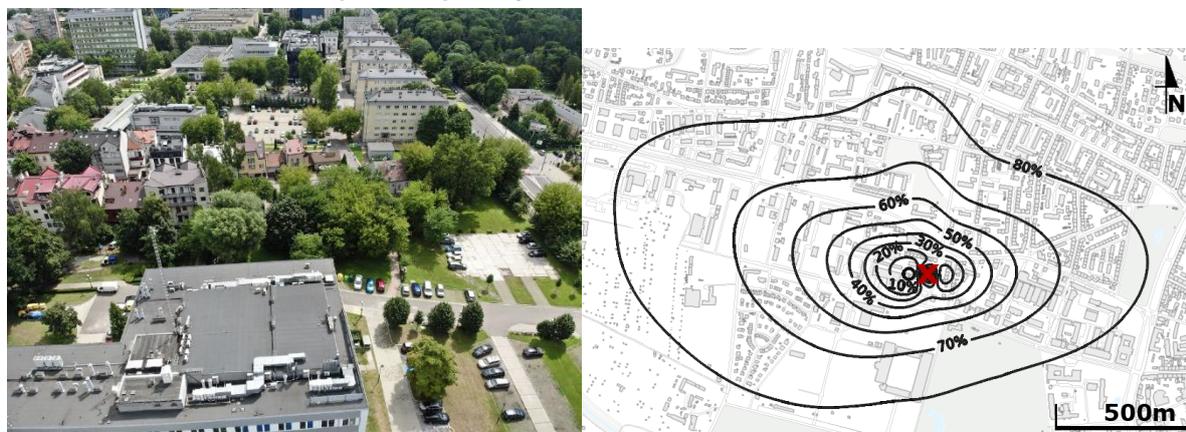


Figure 3 Surroundings of the Krakow PL-Krk CO₂ eddy flux site and flux average footprint projected on the map of Krakow. The measurement mast is visible on the rooftop of building at the bottom left corner.

The AGH University of Science and Technology (AGH UST) in Krakow provides data from a rooftop eddy covariance system set up at the rooftop of a university building in the city of Krakow, Poland. The CO₂ flux data are operational since February 2021 and is using an ICOS data pipeline compatible setup using a Smartflux station. This data stream can be integrated using the ETC data pipeline, delivering about 40 MB of raw data per day that can be ingested using the ICOS-CP ETC Façade interface and further analyses as described under 5.2.1.1 for the Heraklion towers, thus using the platinum option. This has been fully implemented for the ICOS Cities project.

As urban eddy covariance flux measurements are still in the experimental phase, and the eddy flux measurements at Krakow and Heraklion are not part of the ICOS Cities project, the processing for both Krakow and Heraklion have been processed by the PI and final processed

Level 2 data for Krakow was submitted as single data file with PID <https://hdl.handle.net/11676/JLEi-8twlcdbRH0YqOuElu-J>, that is part of the complete data collection of CoCO2 data for Krakow as Zimnoch et al (2023) with DOI <https://doi.org/10.18160/8DSK-R4JS>.

5.2.1.3 Krakow vertical gradients using balloon and drone lift

From March 2021 to February 2022, measurement campaigns were performed at Krakow, close to the eddy covariance site, each consisted of a number of flights. Air pressure, temperature, and humidity were measured as well as wind speed and CO₂ and CH₄ concentration during each flight. Each measurement campaign lasted at least from the afternoon until the next morning. At first, instruments travelled up and down inside the basket of an air balloon. This was later change by sampling from a line lifted through a drone. In total 269 vertical profiles have been obtained over the measurement period. Profiles consist of data on air temperature, pressure, humidity, wind speed, CO₂ and CH₄ mixing ratio.

Data will be provided according to the bronze option (Section 5.1.5) as L2 data files in the WMO time series CSV format, similar to these as used by ICOS ATC, one file for each campaign. ICOS-CP gathered the individual files in a zip file, and assigned Handle PID <https://hdl.handle.net/11676/qpMqcmY2cnf18KwLMqYP4wV>. This dataset was added to the Krakow collection of CoCO2 data with its corresponding DOI for the complete dataset <https://doi.org/10.18160/8DSK-R4JS>. The ICOS-CP assisted in this process of data upload in establishing the correct data format and metadata.

Total data volume of the vertical profile data in the zip archive is about 12 MB.



Figure 4 Photographs of balloon and drone launch for the greenhouse gas profile measurements in Krakow.

5.2.1.4 Krakow ¹³CO₂ and ¹⁴CO₂ isotopic mixing ratios

In 12 campaigns lasting each 4 hours, flask grab samples have been collected at the AGH UST university campus in Krakow that are analysed for CO₂, ¹⁴CO₂ and ¹³CO₂ mixing ratios. Furthermore, monthly average samples were collected at the campus and at the Kasprowy Wierch mountain station, that were analysed for the same components.

Data has been provided according to the bronze option as L2 data files in the WMO time series CSV format, similar to those used by ICOS ATC, one file for campaigns and one for the monthly samples respectively. The ICOS-CP assembled all files into a zipped file that was minted Handle PID <https://hdl.handle.net/11676/w6Xyqv9hBU1B-7UhwDKjLpDi> and added to the Krakow CoCO2 data collection and corresponding DOI for the complete dataset collection <https://doi.org/10.18160/8DSK-R4JS>.

5.2.1.5 Sodankylä AirCore vertical profiles by FMI

FMI has measured at the Sodankylä site vertical profiles using the AirCore technique. The instrument used in Sodankylä is built as a stainless-steel tube of a length of about 100m. The

instrument is flown using a meteorological balloon and typical flight altitudes are 30-35 km. During the balloon ascent the AirCore tube gets evacuated, while during the descent the tube fills with air and in that way stores a continuous record of air samples from top to ground. After retrieval of the AirCore it is analysed by evacuating the tube and analysing for CO₂ and CH₄ with a CRDS analyser. The AirCore instrument also records pressure, air temperature and humidity during the flight. In the future also N₂O profiles will be measured using the AirCore instrument.

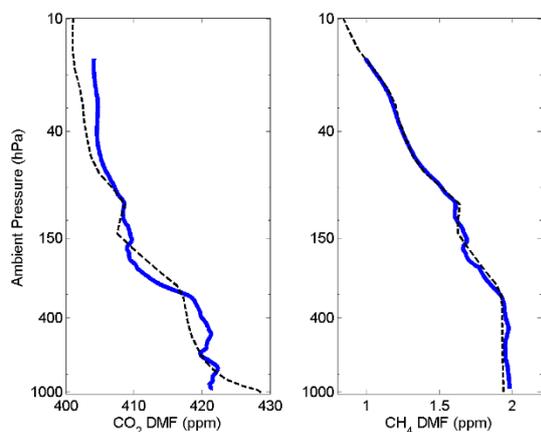


Figure 5 Example air core vertical profile at 26 April 2021 of CO₂ and CH₄ mixing ratios at Sodankylä (blue lines) compared with simultaneous TCCON data

The final analysis results consist of vertical profile data on air temperature, pressure, humidity, CO₂ and CH₄ mixing ratio.

Some example files have been provided in their NASA-Ames format. This format is not supported by Carbon Portal and has been transferred to the ATC tab separated CSV format. ICOS-CP assisted in this conversion and the process of data upload to illustrate the integration with the Carbon Portal repository. One demonstration data file is available with Handle PID: <https://hdl.handle.net/11676/eHlx18iMyNaCYQZ7UJ6hcqAD>. Figure 6 shows the preview of the vertical profile of CO₂ from this example file.

Showing the vertical profile in the conventional way with measurement height or potential height in hPa/mbar at the X- or Y-axis is unfortunately not working yet, at the next update of the front-end this will be corrected.

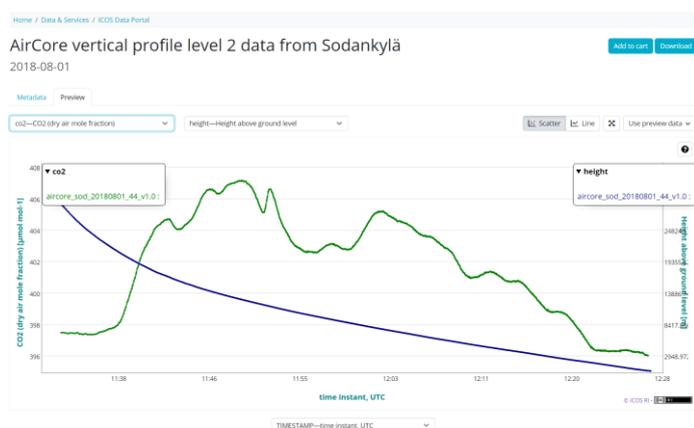


Figure 6 Preview at Carbon Portal of the Sodankylä AirCore vertical profile for CO₂ for 1 Aug 2018

More AirCore profile data are available by contacting instrument PI's (Rigel Kivi at the FMI and Huilin Chen at the Groningen University). Publishing more data through the Carbon Portal is under negotiation.

5.3 New data pipelines in ICOS Cities/PAUL

PAUL is short for Pilot Application in Urban Landscapes towards integrated city observatories for greenhouse gases. This project runs for 4 years, 2021-2025. PAUL was the name of the proposal, but externally it is now called ICOS Cities.

This project supports the European Green Deal by creating capabilities to observe and verify greenhouse gas emissions from densely populated urban areas across Europe. Cities are recognized as important anthropogenic greenhouse gas emission hotspots and therefore play a significant role in any emission reduction efforts. The PAUL project aims to increase our understanding of specific needs of greenhouse gas emission assessment in urban environments; it compares available and novel observational approaches and implements an integrated concept for a city observatory, providing unique data sets that feed diverse modelling approaches, scientific studies and will be the base of services towards the city administrations. Overarching goals of PAUL are to:

1. implement elements of a pilot city observatory in a large (Paris), a medium (Munich) and a small (Zurich) European city,
2. collaborate with city stakeholders and engage citizens in co-designing services that are required for GHG monitoring in order to validate the implementation of Paris Agreement, and
3. increase our understanding of specific needs of GHG assessment in urban environments and create a service portfolio for setting up an urban greenhouse gas observatory.

ICOS Cities data will be complementing the CO₂ MVS system by zooming in on the hot spots that cities are for the fossil fuel CO₂ emissions. In turn the analysis of these high resolution city data will require information from the Copernicus MVS to build its urban emission inversion systems on. For a large part ICOS Cities targets the same stakeholders as the Copernicus MVS and will build on the model and data developments from the CoCO2 project.

In Figure 7 an overview is given of the different measurements to be used in the ICOS Cities project.

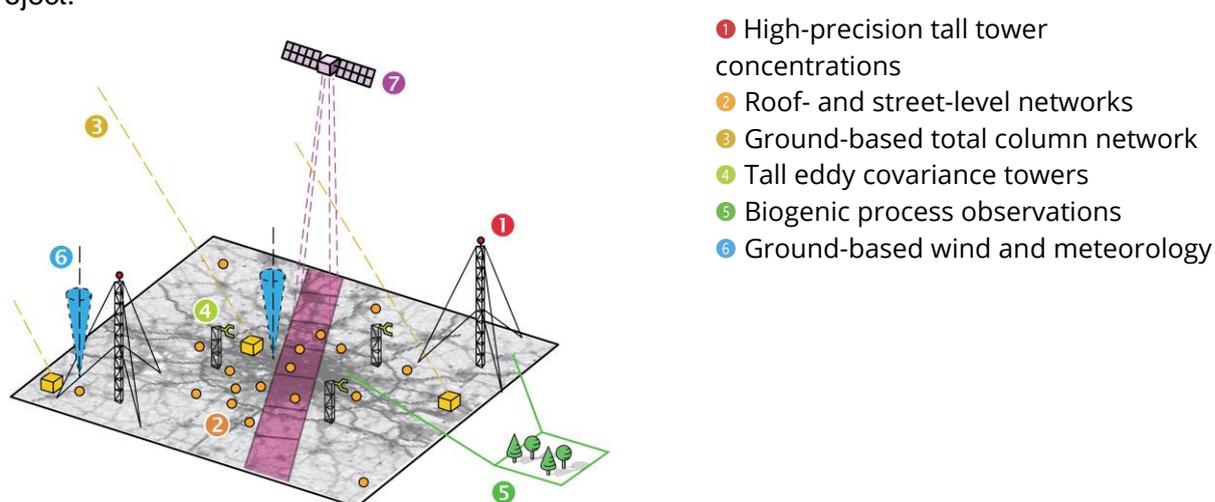


Figure 7 Schematic overview of measurements in ICOS Cities. Each city will have its own constellation of observations using one or more of the types shown here and of each deployed type one or more instances

In Figure 8 the time schedule for the implementation of the different measurements foreseen in ICOS Cities is shown. This includes many new measurements types and deployment of new instrumentation and existing measurements in different configurations. It is clear that together with rolling out the measurements also a lot of work is required for implementing new measurement data processing pipelines. The year 2023 is the year in which most of the instrumentation becomes operational and data will become available. 2023 will also be the year in which the new and updated data pipelines then should become available. This work is done in the framework of the ICOS Cities project but will be very relevant for the future CO₂ MVS.

Potential dates Project Month	Year 1				Year 2				Year 3				Year 4			
	Januar 22 1-3	April 22 4-6	Juli 22 7-9	Oktober 22 10-12	Januar 23 13-15	April 23 16-18	Juli 23 19-21	Oktober 23 22-24	Januar 24 25-27	April 24 28-30	Juli 24 31-33	Oktober 24 34-36	Januar 25 37-39	April 25 40-42	Juli 25 43-45	Oktober 25 46-48
Paris																
Task 3.1	Preparation			10 tall tower, high-quality CRDS sites with CO ₂ / co-species												migration into long-term local observatory
Task 3.2	Preparation			30 roof-level, mid-cost atmospheric CO ₂ / co-species sensor network												migration into long-term local observatory
Task 3.3	Preparation			3 x EM27 / TCCON total column measurements of CO ₂ and NO _x												
Task 3.4	Preparation			1 x tall tower eddy covariance of CO ₂ fluxes												
Task 3.4	Preparation			2 x short tower eddy covariance of CO ₂ fluxes												
Task 3.4					Preparation			1 x co-spec. eddy covariance (6 mo)								
Task 3.4								Potentially 1 x 14C REA								
Task 3.5	Preparation							6 x sap flow, field measurements								
Task 3.5	Preparation							1 x phenocam, 1 x PAR, 6 x soil stations								
Task 3.6	Preparation							2 x scanning Doppler wind LIDARS								
Task 3.6	Preparation							1 x mini wind LIDAR (6 months)								
Zurich																
Task 3.2	Preparation			20 roof-level, mid-cost atmospheric CO ₂ / co-species sensor network												migration into long-term local observatory
Task 3.2	Preparation			60 street-level, low-cost atmospheric CO ₂ sensor network												migration into long-term local observatory
Task 3.4	Preparation			1 x tall tower eddy covariance of CO ₂ fluxes												migration into long-term local observatory
Task 3.4	Preparation	Preparation		1 x co-spec. eddy covariance (6 mo)												
Task 3.4	Preparation	Preparation		1 x 14C REA												
Task 3.4	Preparation	Preparation		1 x COS EC												
Task 3.5	Preparation			6 x sap flow, 6 x soil stations, field measurements												
Task 3.5	Preparation			1 x phenocam, 1 x PAR												
Task 3.6	Preparation			1 x scanning Doppler wind LIDAR (6 months)												
Task 3.6	Preparation			1 x mini wind LIDAR (6 months)												
Munich																
Task 3.2	Preparation			20 roof-level, mid-cost atmospheric CO ₂ / co-species sensor network												
Task 3.2	Preparation			100 street-level, low-cost atmospheric CO ₂ sensor network												
Task 3.3	Preparation			5 x EM27 / TCCON total column measurements of CO ₂ and NO _x												
Task 3.4	Preparation		Preparation	1 x tall tower eddy covariance of CO ₂ fluxes												migration into long-term local observatory
Task 3.4								Preparation								
Task 3.4								1 x co-spec. eddy covariance (6 mo)								
Task 3.5	Preparation							Potentially 1 x 14C REA								
Task 3.5	Preparation							6 x sap flow, field measurements								
Task 3.5	Preparation							1 x phenocam, 1 x PAR, 6 x soil stations								
Task 3.6	Preparation							2 x Doppler wind LIDARS								
Task 3.6	Preparation							1 x mini wind LIDAR (6 months)								migration into long-term observatory

Figure 8 Time schedule for the implementation of the different measurement systems in the ICOS Cities project.

Data streams fitting to existing or updated ICOS pipelines (platinum, gold):

Tall tower sites (ATC, CAL)

Picarro CRDS CO₂, CH₄, CO,

Paris: 9

Zurich: 9

Flask sampler ¹⁴CO₂

Paris: 3

Ecosystem flux urban (ETC)

Soil resp, LAI, Sap flow, TEROS, Phenocam, PAR:

Munich, Zurich (UNIBAS; platinum): 10

Eddy covariance (platinum)

Paris: 4

Zurich: 4

Munich: 1

Synops (ATC)

Paris: 10

Zurich: 10

Munich:

ANSTO Rn (ATC)

Paris: 4

Zurich : 4

New observational data pipelines (silver unless mentioned otherwise):

TCCON

Paris: 1
Zurich: 1 (TCCON)

EM27 remote sensing (platinum: ATC)

Paris: 2
Zurich: 2
München: 5

MIRO Flux/REA/ICOS Flask sampler (gold): 1 (Zurich+Paris+München respectively)

LIDAR Doppler

Paris: 2+1
München: 2+1

Aethalometer

Paris: 2

Mid cost sensors CO₂

Paris: 10+16
Zurich: 14+20
München: 20

Low cost sensors CO₂

Zurich: 60
München: 60

CAPS N500 (NO, NO₂)

Paris: 2

In addition to the observation pipelines there will be many data streams with either prior or ancillary data, like emission inventories and mapping information on land use and buildings. Most of this data will use the flex ingest (iron) option, but when the data is used to generate final model results and/or is ready for publication that data will be published through the bronze option and uploaded as elaborated data to fit with the documented modelling workflow pipelines.

At the final stages of the project (mid 2024 and further) final modelled inversion results will be uploaded and published as elaborated (level 3) data.

Almost all elaborated data will be spatio(temporal) that should be provided in cf-compliant (ATMODAT standard, <https://www.atmodat.de/atmodat-standard>) NetCDF format. A detailed protocol for ingestion at ICOS-CP of netcdf data that allows for support of discovery of variables, previews and subsetting is in preparation for summer 2023.

At this moment data pipelines have been set for:

- Ecosystem flux data for all three cities (3 stations):
 - Level 0 raw data
 - Level 1 NRT 1 day delay
 - Level 2 final quality controlled data
- Level 2 mid-cost and low cost data for CO₂ (77 stations) and meteorology (13 stations) from Zürich
- NRT vertical profiles of XCO₂ and XCO from EM27 sensors at 5 stations for München

Only the Ecosystem flux pipeline is fully operational. Operational data flows for the mid-cost sensors of all three cities will be ready in Q1 2024. An example preview plot of low-cost sensor data for Zürich for 6 stations is shown in figure 9. An overview of the stations for which data is available for Zürich is shown in figure 10.

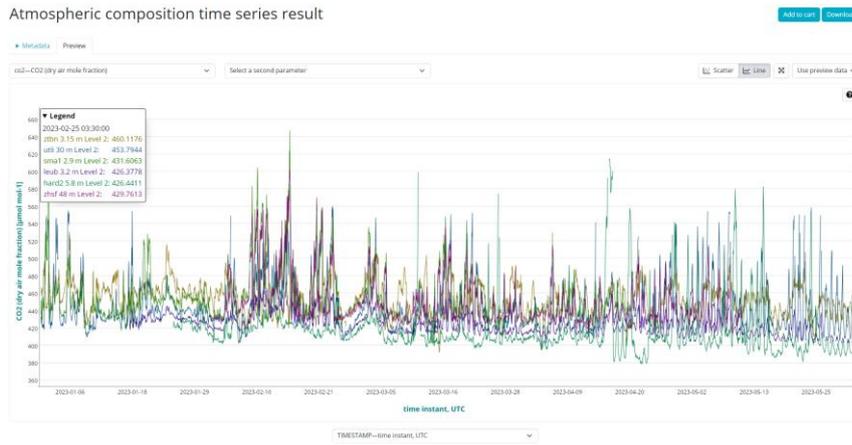


Figure 9 Preview of CO₂ observations from low cost sensors in the first half of 2023 at 6 stations in Zürich, as available from the ICOS Cities data portal

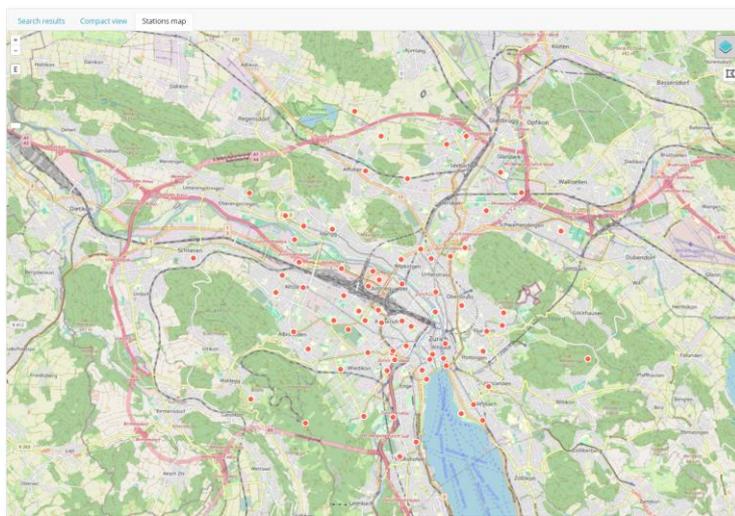


Figure 10 Overview of stations in Zürich (red dots) with low-cost CO₂ and meteorological data available at the ICOS Cities portal.

Total Column measurements from EM27 fotospectrometers are also expected to arrive in Q1 2024. As demonstration five daily files for 5 stations in München have been ingested in Carbon Portal.

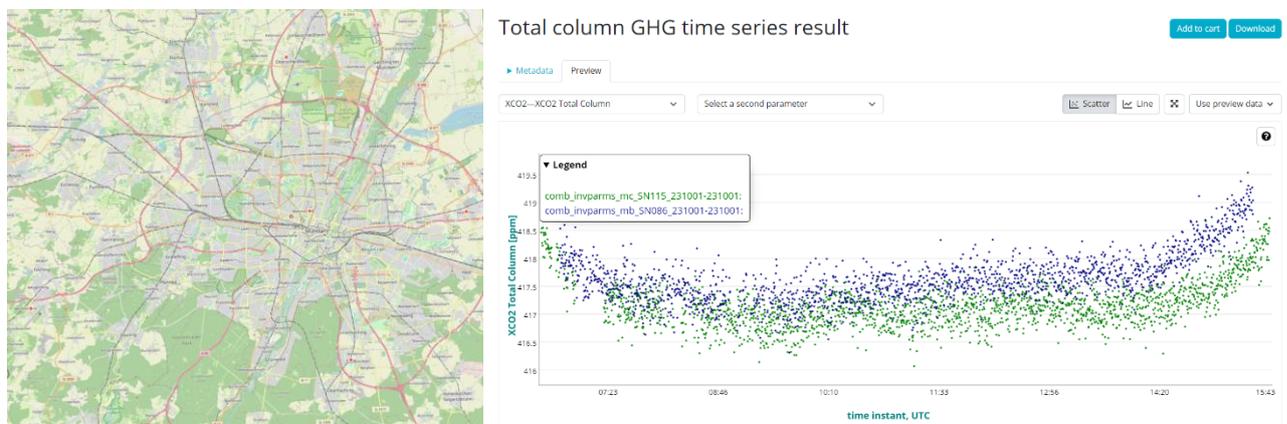


Figure 11 Locations of the 5 Total Column EM27 measurements in München and an example time series plot of the XCO₂ observations for 1 October 2023 for them Northern- and Southernmost stations, showing a clear gradient of about 1-2 ppm developing over the city over the course of the day

5.4 Other data streams

5.4.1 Atmospheric GHG data products

In the past many projects or researchers have used the ICOS repository to publish their results. The improvements in the data pipelines implemented through the CoCO2 project also enable better support for these data. The new data type *Atmospheric GHG Data Product* allows for flexible ingestion of many greenhouse gas related variables at a station in one single time series file in WMO/ATC style text format, including the preview and access of the data through the binary data access API and the python icoscp library. With this one data type time series data can be provided for CO₂, CH₄, N₂O, APO, O₂/N₂, ²²²Rn, ¹⁴CO₂, ¹³CO₂, ¹³CH₄, C¹⁸O₂. Extension to additional data types can be provisioned whenever needed.

An example of this is the historical time series for CO₂, CO, O₂/N₂, APO and ²²²Rn from Lutjewad at <https://meta.icos-cp.eu/objects/R36ZYPJ9H7GG8VNEQcQ1vhk>, part of a data supplement hosted by ICOS CP for Nguyen et al., 2021 with DOI <https://doi.org/10.18160/QQ7D-T060>.

5.4.2 ¹⁴CO₂ historic data Obstack collection

In the framework of the CORSO project preparations have been started to support the data ingestion of ¹⁴CO₂ observations from historic time series from global background stations and to provide this unique data as a special edition of the Obstack archive. Affirst example is historic ¹⁴CO₂ data from Niwot Ridge (USA) with an Handle PID minted as: <https://hdl.handle.net/11676/HGWlet2D2hN0ApvWaJ0Az1J1>. A preview is shown in Figure

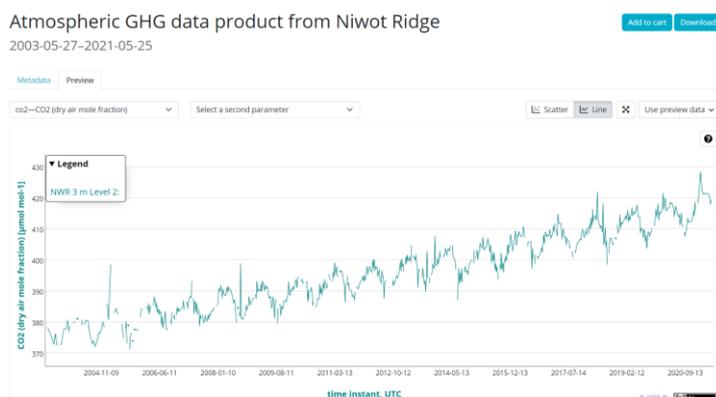


Figure 12 Preview of ¹⁴CO₂ historic time series from Niwot Ridge

The Obstack archive will become available in Q1 2024 as <https://doi.org/10.18160/RATY-OYS0> (until then this data is under embargo and unavailable).

5.4.3 Netcdf spatiotemporal data

Support has been extended for netcdf spatiotemporal data to allow previews without ingestion of the netcdf data into the binary representation. Complete ingestion of all possible configurations of variables and dimensions of cf-netcdf formatted data would be very complicated and resource intensive, but for not too complex 4-dimensional data (lat-lon projection, time dimension and multiple variables with same dimensions), an interactive preview is now supported.

5.4.4 Generation of FastTrack NRT Obstack data products

Starting early 2024, ICOS Atmosphere will start producing so called Fast Track NRT data products at regular, initially 3-months intervals. With a delay of 1-2 months PIs will perform manual control on the data from the previous quarter since the last final quality control Level 2 data release and ATC will perform a complete data processing step over all quality controlled data. Carbon portal will assemble from the historic pre-ICOS, Level 2, Fast Track NRT and

the last NRT data a Fast Track Obspack release for Europe. Initially the number of stations that volunteer to participate will be limited. After experience has been gathered and we know the effort required a roll-out to the whole net work of the Fast Track NRT production will be considered. This product will support the modelling efforts in the framework of the Global Greenhouse Gas Watch initiative.

5.5 Progress in implementing the new data pipelines

Table 1 Overview of the implementation progress as of February 2023

Project	Location	Type	Provider	Data level(s)	Pipeline	Data Processor	Status	Data Start	Data End	Pipeline Completed
CoCO2	Heraklion	Urban CO ₂ Eddy covariance (2x)	FORTH	L0, NRT, L2	platinum	ETC	●	1-2022	12-2022	3-2024
CoCO2	Krakow	Urban CO ₂ Eddy covariance	AGH	L0, NRT, L2	platinum	ETC	●	2-2021	8-2023	6-2023
CoCo2	Krakow	CO ₂ , CH ₄ profiles	AGH	L2	bronze	AGH	●	3-2021	2-2022	9-3023
CoCO2	Krakow, KAS	¹³ CO ₂ , ¹⁴ CO ₂	AGH	L2	bronze	AGH	●	2-2021	1-2022	9-2023
CoCo2	Sodankylä	Aircore vert. profiles (CO ₂ , CH ₄ , (N ₂ O))	FMI	L2	bronze	FMI	●	2018	2022	9-2023
ICOS Cities	Paris, Munchen, Zürich	Many different data types	Many	L0-L3	platinum& bronze	ATC, ETC,local	●	2021	Until 2025	6-2024

○ In progress ● Ready ● Pipeline ready, data on the way

6 Conclusion

All data streams that are not part of the current ICOS-CP process can be implemented to flow through ICOS-CP, but the management is different based on observation protocols, compliance of metadata standards, level of QA/QC and the requirements of timeliness, so the data will be ingested as project specific (non-ICOS) data, providing relatively limited metadata on provenance compared to full ICOS data.

Short campaigns (mainly useful for algorithm development purposes) can be saved as “flux ingest”, but if data is close enough to ICOS requirements, a NRT pipeline can be created (e.g. daily accumulating time series).

7 References

ICOS Data Improved Lifecycle (2023), <https://doi.org/10.18160/D2JV-KB6B>

Nguyen, L. N. T., Meijer, H. A. J., van Leeuwen, C., Kers, B. A. M., Scheeren, B. A., Jones, A. E., Brough, N., Barningham, T., Pickers, P. A., Manning, A. C., & Lujikx, I. T. (2021). Two decades of flask observations of atmospheric $\delta(O_2/N_2)$, CO₂ and APO at stations Lutjewad (the Netherlands) and Mace Head (Ireland) plus 3 years from Halley station (Antarctica). Copernicus GmbH. <https://doi.org/10.5194/essd-2021-213>

Nguyen, L.N.T., Lujikx, I., Meijer, H.A.J., 2021. Two decades of flask observations of atmospheric $\delta O_2/N_2$, CO₂, and APO at stations Lutjewad (the Netherlands) and Mace Head (Ireland) plus 3 years from Halley station (Antarctica). <https://doi.org/10.18160/QQ7D-T060>

Zimnoch, M., Sekuła, P., Jasek-Kamińska, A., Skiba, A., Gałkowski, M., Chmura, Ł., Bartyzel, J., Jagoda, P., Kud, M., Nęcki, J., 2023. Observational datasets of urban CO₂ fluxes, atmospheric vertical profiles of CO₂ and CH₄ and ¹⁴CO₂, and isotopic composition of atmospheric CO₂ at Krakow, Poland; period 2021-2023; part of the CoCO2 project . <https://doi.org/10.18160/8DSK-R4JS>

Document History

Version	Author(s)	Date	Changes
1.0	Alex Vermeulen (ICOS ERIC)	22/12/2023	

Internal Review History

Internal Reviewers	Date	Comments
Wouter Peters (WU)		
Marko Scholze (ULUND)		

Estimated Effort Contribution per Partner

Partner	Effort
ICOS ERIC	2 PM
ULUND	3 PM
FORTH	? PM
AGH	? PM
Total	5 PM

This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.