11

CoCO2

Prototype system for a Copernicus CO_2 service

D7.5: Report on data providers and long-term data availability V2

Ville Kasurinen, ICOS ERIC

coco2-project.eu

Co-ordinated by ECMWF

CoCO2

Prototype system for a Copernicus CO_2 service

D7.5 Report on data providers and long-term data availability

Dissemination Level: Public

Author(s): Ville Kasurinen, Elena Saltikoff, Sindu Raj Parampil (ICOS ERIC)

Date: 29/09/2023
Version: 1.1
Contractual Delivery Date: 30/09/2023
Work Package/ Task: WP7/ T7.2
Document Owner: ICOS ERIC
Contributors: ICOS ERIC, FORTH, FMI, CMCC
Status: Final

CoCO2: Prototype system for a Copernicus CO_2 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

Figures

Tables

1 Executive Summary

This report is the second version of a catalogue of providers of *in situ* observations and ancillary data needed for the prototype system for a Copernicus CO_2 monitoring service. It follows the structure of D7.1, outlining the current data needs identified by (inverse) modellers in that deliverable, but providing a more inclusive view. Data needs are classified by timeliness (latency), which can vary significantly, as some data are currently used for parameter estimation and can be several years old, while assimilation needs near-real time data as well as evaluation data from the past year. It is to be expected that the needs will evolve over time depending on new user requirements.

In situ observations discussed in this report can be coarsely classified as either fluxes between atmosphere and ecosystems or oceans, or concentrations in the atmosphere. In addition, we briefly cover auxiliary data from satellites and meteorological models. The largest section discusses ecosystem fluxes, which are used by several modelling groups and available from several, geographically overlapping databases, each of which has its own data policy and metadata model. In terms of atmospheric data, WMO has developed the OSCAR database which attempts to collect information about stations. However, the WMO member countries are usually represented by the national meteorological services, while the GHG-related measurements are often run by universities and research institutes focussing on agriculture, forestry and health. This disconnect is apparent in the level of inclusiveness of the OSCAR database.

This is the updated version of the previous deliverable D7.4. In the updated version we have added some remarks and new information gained during the project, which is related to aspects that data users should consider when compiling and combining data from different data sources.

2 Introduction

2.1 Background

The prototype CO_2 Monitoring & Verification Support (MVS) capacity being developed within the CoCO2 project aims to extract information about anthropogenic greenhouse gas (GHG) emissions from satellite measurements provided by the planned CO2M constellation. These satellites will provide imager-type column-integrated measurements of atmospheric CO_2, CH_4, and NO_2 at ~2 km x 2 km resolution with a swath ~250 km wide, enabling the imaging of emission plumes from point sources and hot spots associated with anthropogenic activities, and global coverage to constrain emissions on national scales.

While these satellites are being developed with this application in mind, it is clear that such an integrated system will require extensive *in situ* and ancillary observations in order to achieve its proposed objectives. Note that in the context of Copernicus, *in situ* data refers to measurements collected by ground-based, seaborne or airborne sensors, including remote sensing measurements, as well as reference and ancillary data. Multiple data streams of in situ measurements will play a role here, including, but not limited to, measurements of greenhouse gas fluxes and atmospheric mixing ratios of greenhouse gases. These data can be used for a variety of applications within the MVS prototype, each of which comes with a different set of requirements in terms of timeliness, coverage, and precision. In Pinty et al. (2019), the CO_2 Monitoring Task Force convened by the European Commission documented the needs and high-level requirements for *in situ* measurements that are foreseen in the MVS capacity, which have guided the work reported here and earlier in CoCO2 deliverable D7.1. *Book of in-situ data requirements V1.*

Table 1: Summary of the timeliness requirements by task, as collected by Task 7.1. The colours indicate the timeliness requirements of the different data streams by task.

	Task 2.1	Task 3.1	Task 3.2	Task 3.3	Task 4.1	Task 4.3	Task 4.4	Task 5.2	Task 5.6	responses
eddy covariance flux data	green	red	yellow	yellow			orange	yellow	yellow	8
in situ CO2 mixing ratios		red			yellow	orange	orange		yellow	10
in situ CH4 mixing ratios		red				orange	orange		yellow	8
in situ co-emitted species		red			yellow		yellow			5
urban air quality networks			orange		yellow					3
ocean fluxes/partial pressures							orange			1
radiocarbon						orange	orange			2
other atmsopheric species		red			yellow		yellow		yellow	4
ground-based remote sensing		red	yellow		yellow		yellow		yellow	7
site-level ecosystem parameters								yellow		1
site-level management, lateral fluxes										0
in situ soil moisture										0
in situ meteorological data	green		orange		yellow	orange	orange		yellow	7
meteorological model fields	orange		orange	yellow	orange	orange	orange	orange		13
nightlights			orange							1
activity data			yellow							1
satellite-based indices	orange			orange		orange	orange			4
solar-induced fluorescence								orange		1
other satllite-based measurements					yellow					1
landcover maps	green	orange	green	orange	yellow		green	green		10
global concentration fields		red			orange	orange	orange			9
number of responses	1	1	1	2	2	2	6	1	1	

- red: near real time
- orange: use in the currently simulated year
- yellow: use in evaluation (one year delay OK)
- green: use in parameter estimation (any year)

In this report, we have tried to elaborate upon each of the data requirements identified in D7.1, and confront them to the current spatial and temporal coverage and timeliness of the available measurements (Table 1). While the D7.1 was limited to data already used by European groups participating in the project, some of which were working on regional scales only, we put here more emphasis in finding datasets outside of Europe. ICOS (Integrated Carbon Observation System) is the European Research Infrastructure that provides high-quality in situ data and elaborated data products for science on the carbon cycle and for quantifying greenhouse gas emissions and sinks across Europe, and its data releases are described here as the primary European data source.

2.2 Scope of this deliverable

2.2.1 Objectives of this deliverable

This deliverable aims to document the available *in situ* and ancillary data corresponding to needs expressed across the CoCO2 project, from WP2 through WP6. For each of the data requirements identified, it provides the current spatial and temporal coverage, and timeliness of the available measurements with a link to the data providers. The documentation of these data is critical in order to move the work from a scientific exercise to an operational capacity. Most of the needs have been expressed by CoCO2 users in the perspective of the CoCO2 project itself (for the period of the project), i.e., as part of one of the steps towards the operational capacity

The needs have been collected in Task 7.1 and reported in Deliverable 7.1. This deliverable and its successors will serve as a basis for the gap analysis in Task 7.3, to be reported in D7.6 which is called in the proposal "*Gap analysis report of the current in situ measurement capacity"*.

This report is based mainly on internet research and contact with experts in the domain. More interaction with data users and data providers has taken place during a side event of ICOS Science conference in September 2022. Furthermore, the needs of data users will be updated in future versions of the Deliverable 7.1 report, as the project progresses.

In this work we talk about "available" observations, and that covers more than just making the measurements. An operational system has also requirements for timeliness of data processing, bandwidth of data infrastructure, homogeneity of metadata, accessibility of the data, and openness of data policy. This deliverable and its successors will provide guidance for the development of the prototype of the operational data pipeline in Task 7.4, resulting in Deliverable 7.8.

2.2.2 Work performed in this deliverable

The following people outside of this project have been interviewed:

- Jörg Klausen, MeteoSwiss about WMO OSCAR
- Pia Anttila, FMI about Air Quality data
- Ute Karstens, Lund, about radon
- Martijn Palland, Max Planck Institute, Arctic Eddy Covariance sites
- Frank Hase at KIT about COCCON
- George Burba, LICOR

We have also updated the review of scientific publications considering recent studies related to eddy covariance data usage, quality assurance, gap-filling, and post-processing.

2.2.3 Deviations and counter measures

According to the Grant Agreement, this deliverable should also discuss the uncertainty of available datasets. We have not been able to carry out a detailed analysis related to uncertainty of available data sets but some recent publications (Walthers et al 2023, Jung et al 2023) do report domain specific problems related to available data sources. Existing gaps in current data streams are more deeply discussed in Deliverable 7.6.

3 In situ data

3.1 Eddy covariance flux data of CO_2 and CH_4

The Eddy Covariance (EC) method has become the dominant technique to measure ecosystem-scale fluxes during the past three decades. Task 3.1 (Forward modelling and data assimilation developments for operational global prototype) reported a need for near-real-time observations of CO_2, CH_4, water vapour as well as latent heat and sensible heat exchange. Other users contributing to D7.1 stated that they can use data made available up to a year after it was measured, and even from any year for parameter estimation.

3.1.1 Current spatial and temporal coverage

The first EC measurement sites started to operate in early 1990's and some of these are still operational and continuously produce data. Measurements are typically made at a frequency of 10 or 20 Hz and aggregated to 30 minutes or 1 hour mean values. The measurement networks typically focussing on specific geographical areas (e.g., Europe, North and South America, Australia and Asia) have developed partly into more consolidated research infrastructures (RIs) during the past two decade. At global scale, FLUXNET, a voluntary initiative out of the scientific community has been an active player in encouraging different networks to submit EC data to a global collection that can be distributed to researchers.

The first FLUXNET workshop was held in 1998. At this event, scientists discussed flux comparisons between different biomes and initiated the idea to compile ecosystem flux data from several sites. Since then, FLUXNET has compiled measurements from several existing and historical regional networks and several FLUXNET data sets have been released. The latest of them, FLUXNET 2015[1] contains more than 2000 site years of data from over 200 sites. This data set contains data from different geographical regions, for example North and South America, Asia, Europe and Australia. FLUXNET has recently suggested a common postprocessing (OneFlux) and a common data format to the contributing stations, networks and RIs that supports data integration and comparison.

The following analysis lists the most important networks and RIs. Most of them provide their data through own portals but also have contributed to FLUXNET data sets. Thus, it should be noted that a measurement site can be listed in multiple databases (e.g the same site can be a part of AmeriFlux and FLUXNET data collection). It should also be noted that this analysis does not cover EC sites that have not been registered to any network.

The Integrated Carbon Observation System (ICOS) is a European Research Infrastructure coordinated by ICOS ERIC and releasing its data through the the ICOS Carbon Portal (CP). At the time of writing, 85 ecosystem stations, mainly in Europe, were members of the ICOS measurement network. Data derived from ICOS stations are freely available in FLUXNET format under Creative Commons Attribution 4.0 International Licence (CCBY 4.0). The ICOS Carbon Portal has also served as a platform where ICOS driven initiatives like Drought 2018[2] (52 sites) and Warm Winter 2020[3] (73 sites) data sets have been released in FLUXNET format.

The European Eddy Fluxes Database Cluster[4] is an initiative to improve standardization, integration and collaboration between databases that are part of European research projects. It has been created with the aim to host flux measurements between ecosystems and the atmosphere in a single infrastructure and to provide standard and high-quality data processing and data sharing tools. The database contains data from past EU funded research projects

[1] https://fluxnet.org/data/fluxnet2015-dataset/
[2] https://www.icos-cp.eu/data-products/YVR0-4898
[3] https://www.icos-cp.eu/data-products/2G60-ZHAK
[4] http://www.europe-fluxdata.eu/home

like Carbo Africa, Carbon Extreme and GHG-Europe. The European Fluxes Database Cluster contains at the moment 426 sites that are mainly located in Europe, Africa, Russia, Greenland and North and South America.

AmeriFlux[5] is a network of PI-managed sites measuring ecosystem CO_2, water, and energy fluxes in North, Central and South America. The network was launched in 1996, after an international workshop on flux measurements in La Thuile, Italy, in 1995, where some of the first year-long flux measurements were presented. Since its establishment, the AmeriFlux network has been continuously expanding. In February 2022 the database contained 562 registered sites and 2842 downloadable sites years of data. A subset of AmeriFlux sites are permanently funded through a core project and can be considered as RI.

NEON[6] (National Ecological Observatory Network) is an RI that collects Eddy Covariance flux data using sensors mounted on towers at terrestrial field sites across the US. Currently NEON has 47 Eddy Covariance sites that provide turbulent exchange of CO_2, latent and sensible heat, storage of CO_2 Net Ecosystem exchange and energy balance data. NEON is also monitoring 34 fresh water sites like lakes and rivers. NEON provides its data through the AmeriFlux data base. Most of the AmeriFlux and the NEON data are available in FLUXNET format.

OZ Flux[7] is an Australian and New Zealand Flux Research and Monitoring program and some of those sites in the network are operated together with TERN (Terrestrial Ecosystem Research Network) research infrastructure. TERN is Australia's land ecosystem observatory, which has 12 super sites, 3 affiliated super sites, and 1 inactive super site. Super sites are part of the TERN Ecosystem Processes Platform that monitors the environment at a high level of details for small number of representative sites. In super sites monitoring is done using instrument or sensor measurements complemented with classical surveys and remote sensing activities, and the sites have been located in significant Australian biomes. A part of the OZ Flux and TERN data have been included in previous FLUXNET releases. TERN is preparing for annual releases in the FLUXNET format.

AsiaFlux[8] is a regional research network bringing together scientists from universities and institutions in Asia to study the exchanges of carbon dioxide, water vapour and energy between terrestrial ecosystems and the atmosphere across daily to inter-annually time scales. The AsiaFlux database currently provides data from 117 sites (December 2021). A subset of the AsiaFlux database has been included in the FLUXNET 2015 synthesis dataset.

ChinaFlux[9] is an observation and research network that applies EC and chamber methods to measure the terrestrial CO_2, water vapor and energy fluxes in China. Many of the sites are part of the Chinese Ecosystem Research Network (CERN) research infrastructure. At the moment the ChinaFlux data base covers in total 71 sites including 17 Farmlands, 15 grasslands, 20 forests, 15 wetlands, 1 urban, 2 desserts and 1 lake site. At the time of writing this report, availability of ChinaFlux data is unclear.

The South African research infrastructure EFTEON/SAEON runs a network of currently six EC stations covering all major biomes of South Africa.

In this report we cannot compile a comprehensive list of available site years in each network but reported in the first version of the deliverable 1699 Eddy Covariance measurement sites (Figure 1). There are some publications available focussing on EC data availability in specific regions. The recent review of Palland et al (2022) is dealing with Arctic Eddy Covariance data availability and comprises 120 EC sites[10] of which 83 are listed active and 25 of these active

[5] https://ameriflux.lbl.gov/about/network-at-a-glance/
[6] https://www.neonscience.org/data-collection/flux-tower-measurements
[7] https://www.ozflux.org.au/index.html
[8] https://www.asiaflux.net
[9] http://www.chinaflux.org/enn/index.aspx
[10] https://cosima.nceas.ucsb.edu/carbon-flux-sites/

sites remain operational throughout the winter. Some site data can be downloaded directly and some are available upon request through the site PI's.

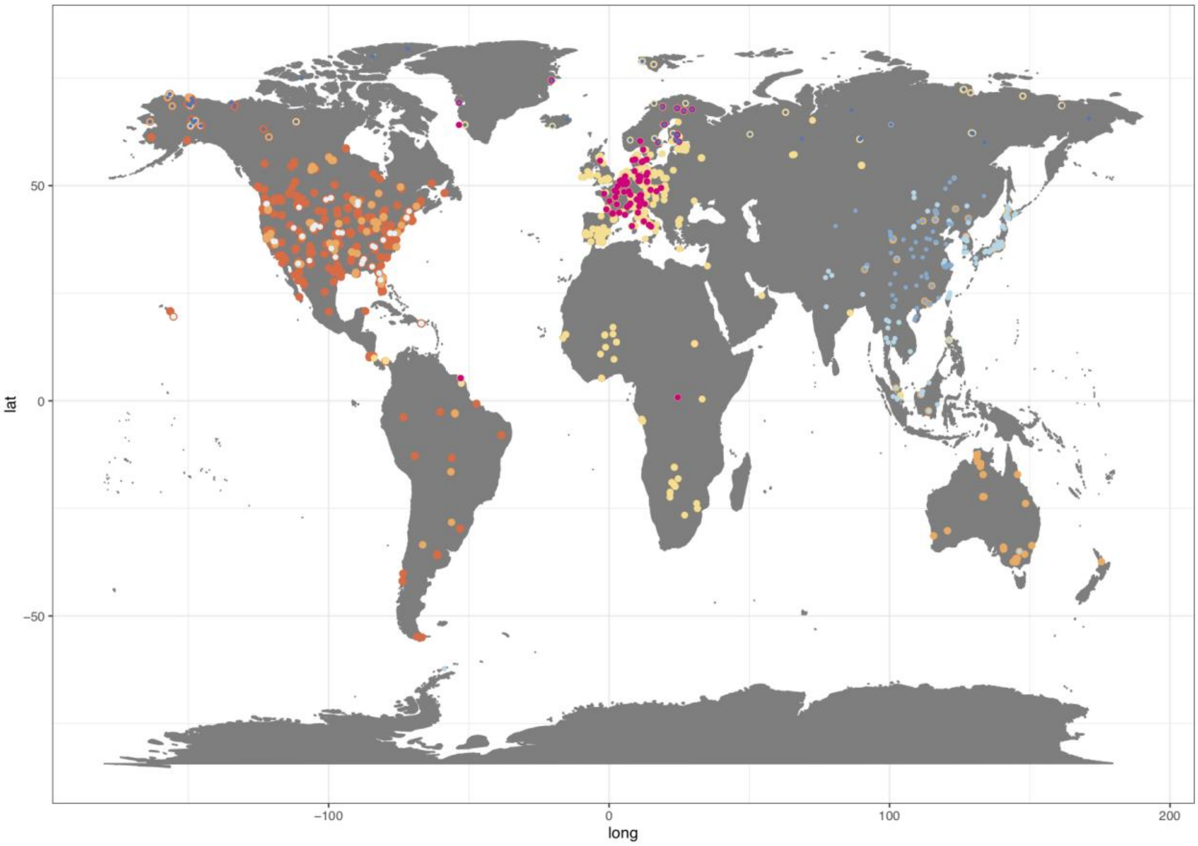

Figure 1. Map of global distribution of 1699 Eddy Covariance measurement sites reporting CO_2 and Energy Fluxes. Eddy Covariance measurement sites based on following databases: ●AmeriFlux (556 sites), ● FluxNet (267) ● European Flux Database (425), ●AsiaFlux (113), ●ChinaFlux (75), ● ICOS (86), ●NEON (47) and ● Palland et al. 2021 (120 sites). Sites from different databases may overlap and data from one site can be a part of several databases.

Based on abstracts published during the ICOS Science conference in 2022, there are +2100 locations where eddy covariance flux measurements of CO_2 and H_2O have been done in the past and where they are being done now (Figure 2, Burba & Daugherty 2022). According to Burbra & Daugherty (2022) there are roughly +400 sites that our previous deliverable D7.4 was not able to report. Additional sites have been found based on questioners targeted to site PI networks for which PI's are not willing to share data or share the exact tower location before data has been published in peer reviewed journals.

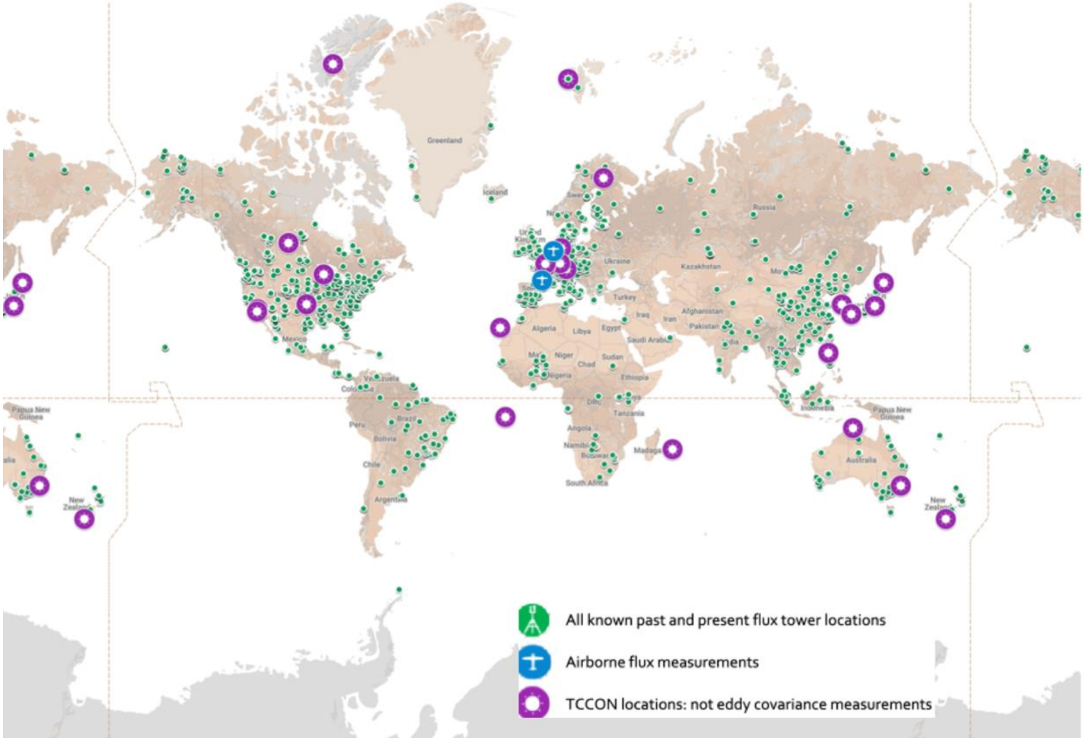

Figure 2. Illustrative Maps of Past and Present Eddy Covariance Measurement Locations (Burba 2019 & Burba & Daugherty R., B. (2022)) reporting +2100 EC sites

3.1.2 Data access

While ICOS data are fully open under CC-BY-4.0 license, access to EC data from other sources is often restricted by data policies and can divert even within networks. For example, some AmeriFlux data are licenced with the CC-BY-4.0 license, similarly to ICOS data. However, other AmeriFlux data are distributed according to the AmeriFlux Legacy policy, which means that data contributors need to be contacted directly so that they have the opportunity to contribute substantively and become a co-author. NEON data are released to the public domain under the Creative Commons CC0-1.0 license. The FLUXNET 2015 data set is mainly under CC-BY-4.0 with few sites following a more restrictive policy (Tier 2).

A group of RIs working on terrestrial ecosystems (namely ICOS, TERN, CERN, NEON, EFTEON and eLTER) have recently started a coordinated activity called ‘Global Ecosystem Research Infrastructure’ that aims to support FLUXNET in providing an annual data release from those sites that follow open and FAIR data policies.

The European Fluxes Database Cluster provides an option for Public Data Access and restricted data access. The OzFlux data portal allows users to download site data after registering and accepting the data policy. Data registered to the AsiaFlux and ChinaFlux databases require a direct contact and data request to the station PI. The review of Palland et al (2022) reports the PI and data source for each site introduced.

In addition to CO_2, CH_4 measurements are also becoming available. In the AmeriFlux network 74 sites are measuring CH_4 fluxes using either EC technique or chambers (updated January 2019). ICOS Class 1 Ecosystem stations are also measuring CH_4 flux using the EC method. A first CH_4 inventory has been done in the FLUXNET-CH_4 initiative that includes half-hourly and daily gap-filled and non-gap-filled aggregated CH_4 fluxes and meteorological data from 79 sites globally: 42 freshwater wetlands, 6 brackish and saline wetlands, 7 formerly drained ecosystems, 7 rice paddy sites, 2 lakes, and 15 uplands (Delwiche et al 2021). There data are shared through the FLUXNET database.

3.1.3 Timeliness

RIs and regional networks release data using different approaches. Usually, there is a delay of a few months to a few years before new measurements are released in a regional data collection. ICOS ecosystem data are typically released twice per year by the ICOS Carbon Portal. ICOS furthermore allows to download near-real time data with lower data quality.

AmeriFlux, NEON and TERN provide data through their data portal and data availability is varying between sites in the network.

From some networks like AsiaFlux and ChinaFlux data should be requested directly from the site PI.

The FLUXNET 2015 data collection is not updated frequently. Thus, timeliness of global data remains an unsolved problem. A concept to solve it has been provided by Papale (2020) who suggested that regional measurement networks or RIs would compile a data global set to be directly shared in a machine-readable way through their databases instead of creating periodic global data collections. The proposed concept is called ‘FLUXNET shuttle’ and would aim to provide continuous updates to the FLUXNET data collection.

3.1.4 Uncertainty

A major source of uncertainty in Ecosystem ecosystem flux data is that they are always gap-filled and usually the energy balance closure in measurement sites is incomplete (Foken et al 2006). The gap-filled data can be up to 80% of the provided data. For example, in the global FLUXNET2015 data set, with 1532 site-years of data, on average 68% of the half-hourly CO_2 fluxes are missing. Even if the site-years that have gaps longer than two months are excluded, the mean observation coverage is 40%. Only 50 site-years have a coverage higher than 60% and only 5 site-years a coverage higher than 70%. The gap-filling methods are not always applicable to different environments, as seen in a recent article by Vekuri et al. (2023). They showed that the commonly used gap-filling method MDS causes significant carbon balance errors for northern ecosystems (latitude > 60 °N) sites. MDS systematically overestimates the CO_2 emissions through respiration and underestimates the sequestration of carbon by the biospheric sink. They developed methods to substantially reduce the northern site bias.

Gap-filling is not the only methodological approach that is influencing the uncertainty of the final eddy covariance flux product that is provided to the data user. Eddy Covariance measurements and final post-processed data that are typically submitted to a data portal are sensitive to the measurement method used on the site. Previous literature has reported significant differences between open path and closed path systems as well as between devices using different laser techniques (Papale et al. 2020). These differences in situ measurements are also influencing to the final data products that are typically used by the data user as defined in this deliverable.

The nature of Eddy Covariance measurements carried out in 10-20 Hz frequency is such that decisions made during the post-processing of high frequency data into 0.5-1h values are influencing to final data (Mammarella et al. 2020). The ICOS measurement protocol in Nemitz et al. (2018) provide further guidelines that are implemented in ICOS sites. In ICOS network post-processing of raw data files is centralized in ETC.

According to Mammarella et al. 2020 at least following methodological aspects are influencing the final EC flux product used by the data user:

1) Data acquisition and synchronization.

2) Overflow inlet systems.

3) Raw data de-spiking.

4) Time lag estimation.

5) Spectral correction approach for high frequency loss.

6) Friction velocity filtering.

7) Gap-filling.

8) Additional guidance on the conditions under which site conditions fluxes of CO_2 is sensible to measure with EC approaches.

From the data user perspective, it is important to acknowledge that gap-filling and post-processing of raw data files are influencing the final EC flux data product:

1) Different networks might have carried out post-processing of raw data using different methods, while some data collections like FLUXNET has been processed using the same post-processing software like OneFlux (Pastorello et al 2020).

2) There might be differences in post-processing even between sites of the same network.

3) Differences between same site and between years do exist (Jung et al 2023).

4) Different releases or versions of the same data might be available.

5) Several different gap-filling methods can be implemented to fill missing observations (Vekuri et al 2023), which might be reflected also in the final product.

As reported by Jung et al. (2023), the FLUXNET community has been developing a series of standardized tools for post-processing, quality controlling and quality assurance and gap-filling. Some part of these tools are also organized in a set of routines (like ONEFlux Pastorello et al. (2020)) that have been used in the FLUXNET2015 collection and continental networks releases (e.g., AmeriFlux FLUXNET product, ICOS Level2 data, Drought2018) and WarmWinter2020 collections).

It has been acknowledged that good quality flags are important for the data user in order to define difference between good quality and bad quality measurements and gap-filled data. Data providers making their data openly accessible should make sure that all variables have appropriate quality flags clearly identifying gap-filled and measured data. The influence of post-processing and gap-filling methodologies EC fluxes are discussed in Deliverable 7.5.

3.1.5 Data providers

The following list provides and overview of the described data providers. Note also that the underlined words here (and also in previous paragraphs) are links to websites, which give additional information such as data policy of the discussed database.

1) ICOS Carbon portal: https://www.icos-cp.eu/

2) FLUXNET: https://fluxnet.org

3) AmeriFlux: https://ameriflux.lbl.gov

4) NEON: https://www.neonscience.org

5) European Flux Data Base: http://www.europe-fluxdata.eu

6) TERN & Oz Flux : https://www.tern.org.au & https://www.ozflux.org.au

7) FluxAsia: https://www.asiaflux.net

8) ChinaFlux: http://www.chinaflux.org/enn/

9) Palland et al 2021 data set: https://cosima.nceas.ucsb.edu/carbon-flux-sites/

3.2 In-situ atmospheric mixing ratios of CO_2

Task 3.1 (Forward modelling and data assimilation developments for operational global prototype) reported a need for near-real-time observations of atmospheric mixing ratios of CO_2. Tasks 4.3 and 4.4 use data from currently simulated year. Other users contributing to D7.1 stated that they can use data made available up to a year after it was measured (for evaluation).

3.2.1 Current spatial and temporal coverage

The WMO OSCAR database reports 186 stations measuring CO_2 globally. Their distribution is shown in table 2. It is however well known that this does not cover all the stations. OSCAR/Surface is the World Meteorological Organization's official repository of WIGOS metadata for all surface-based observing stations and platforms. It contains metadata such as station location and supervising organization, as well as affiliation in programs and networks such as WMO GAW, but usually no direct link to the data.

Table 2 Geographic distribution of stations measuring CO_2 according to WMO OSCAR database. February 2022

Area / date	Africa	Antarctica	Asia	Europe	North + Central America, Caribbean	South America	SW Pacific	Total
Feb 2022	13	9	27	64	45	7	18	186
May 2023	13	10	35	75	68	18	8	227

The World Data Centre for Greenhouse Gases (WDCGG) is a World Data Centre (WDC) operated by the Japan Meteorological Agency (JMA) under the Global Atmosphere Watch (GAW) programme of the World Meteorological Organization (WMO). WDCGG collects, archives and distributes data provided by contributors on greenhouse gases (such as CO_2, CH_4, CFCs, N_2O) and related gases (such as CO) in the atmosphere and elsewhere. WDCGG publishes roughly annually data summaries which cover observational data collected at surface stations and on certain ships for the period from 1968 to the previous year based on monthly mean data submitted to WDCGG – e.g. the summary 45 published in September 2021 covers data up to September 2020[1]. The data summaries consist of observational data and the related analysis. The list of stations seems to include more units than the summary in table 2, but sometimes the same station is listed under two organizations.

Data from partially overlapping subgroups of these stations is available via different channels. Observation Package (ObsPack) operated by NOAA and described in detail in Masarie et al.(2014), is a framework designed to bring together atmospheric greenhouse gas observations from a variety of sampling platforms, prepare them with specific applications in mind, and package and distribute them in a self-consistent and well-documented product. Data products created using the ObsPack framework are called "ObsPack products".

NOAA is preparing ObsPack data products in consultation with data providers. The ObsPack globalView+ product includes 524 atmospheric carbon dioxide concentration time series

[1] https://gaw.kishou.go.jp/static/publications/summary/sum45/sum45.pdf

derived from observations made by 70 laboratories from 21 countries. Data for the period 1957-2020 (where available) are included.

Masarie et al. (2015) suggested an approach where ObsPack data would be distributed as complementary products via distributed or central access. These sub-products would be responsibilities of national or regional networks such as CSIRO in Canada or ICOS in Europe.

In 2022, ICOS CP is preparing for “ObsPack Europe”, a dataset in ObsPack format covering both ICOS and non-ICOS stations from Europe. The latest release is covering years 2013-2022 and has been released early 2023.

As of 2021, ICOS data is available from the 33 atmospheric stations. Its processing chain is significantly faster than that of full ObsPack release, but the metadata format is different.

Figure 3 Global distribution of sites measuring atmospheric CO2 from the WDCGG

3.2.2 Timeliness

ICOS is publishing near-real time data daily. These NRT time series are generated within 24 hours after measurement, using only completely automated quality control procedures. The set-up of the data flow and the automated quality control was supported by the Copernicus Atmosphere Monitoring Service (CAMS) to address the specific operational data requirements of the Copernicus programme.

The final completely quality-controlled and flagged ICOS data is released with a delay between 6-12 months. It includes all corrections and maximum completion of missing data. In ICOS terminology this is called “Level 2 data”.

NOAA ObsPack is released as near-real time data and annually, but in NOAA terminology, the NRT product is a dataset released approximately every 3 months following the annual release of the GLOBALVIEW+ product.

The WDCCC data is updated from time to time as contributors submit data. The frequency of data submission varies depending on contributor. Some contributors submit their data once a year, others submit once a month.

3.2.3 Data providers

1) ICOS Carbon portal: https://www.icos-cp.eu/data-products/atmosphere-release (Europe)

2) NOAA Obspack: https://gml.noaa.gov/ccgg/obspack/ (Global)

3) WDCGG: https://gaw.kishou.go.jp/ (Global)

3.3 In-situ atmospheric mixing ratios of CH_4

Task 3.1 (Forward modelling and data assimilation developments for operational global prototype) reported a need for near-real-time observations. Tasks 4.3 and 4.4 use data from currently simulated year, and other users contributing to D7.1 stated that they can use data made available up to a year after it was measured (for evaluation use).

3.3.1 Current spatial and temporal coverage

WMO OSCAR database reports globally 177 stations measuring CH_4. Their distribution is shown in table 3. It is however well known that this does not cover all the stations.

Table 3 Geographical distribution of stations measuring CH_4 according to WMO OSCAR database, Feb 2022

	Africa	Antarctica	Asia	Europe	North + Central America, Caribbean	South America	SW Pacific	Total
Feb 22	11	10	21	61	49	6	18	177
May 23	11	12	24	68	69	6	19	209

The NOAA ObsPack product includes 362 atmospheric methane concentration timeseries derived from observations made by 46 laboratories. Data for the period 1983-2020 (where available) are included. As of 2021, ICOS data release includes methane data from 26 stations, 67 levels.

Figure 4. Global distribution of Methane from WDCGG

3.3.2 Timeliness

ICOS is publishing near-real time CH_4 data daily. These NRT time series are generated within 24 hours after measurement, using only completely automated quality control procedures. The set-up of the data flow and the automated quality control was supported by the Copernicus Atmosphere Monitoring Service (CAMS) to address the specific operational data requirements of the Copernicus programme.

The final completely quality controlled and flagged ICOS data is released with a delay between 6-12 months. It includes all corrections and maximum completion of missing data. In ICOS terminology this is called “Level 2 data”.

Since 2020, the NOAA ObsPack methane product is also released as near-real time data and annually, but in NOAA terminology, the NRT product is a dataset released approximately every 3 months.

The WDCGG data is updated from time to time as contributors submit data. The frequency of data submission varies depending on contributor. Some contributors submit their data once a year, others submit once a month.

3.3.3 Data providers

1) https://www.icos-cp.eu/data-products/atmosphere-release (Europe)

2) NOAA Obspack: https://gml.noaa.gov/ccgg/obspack/ (Global)

3) WDCGG : https://gaw.kishou.go.jp/ (Global)

3.4 In-situ atmospheric mixing ratios of co-emitted species

Task 3.1 (Forward modelling and data assimilation developments for operational global prototype) reported a need for near-real-time observations. There is also need for data from the simulated year.

3.4.1 Current spatial and temporal coverage

The ObsPack CO product includes 249 atmospheric carbon monoxide datasets from 50 countries.

ICOS has included measurements of atmospheric NO_2 and CO mixing ratios as part of its atmosphere release. Currently this includes measurements of NO_2 from 13 stations, at 44 levels. CO data are provided from 26 stations and 67 vertical levels.

More information is provided in Chapter 3.5 related to air quality measurements, as the World Air Quality Index (WAQI) also contains information about NO_2 and CO.

3.4.2 Timeliness

ICOS is publishing near-real time data daily. These NRT time series are generated within 24 hours after measurement, using only completely automated quality control procedures.

The final completely quality controlled and flagged ICOS data is released with a delay between 6-12 months. It includes all corrections and maximum completion of missing data. In ICOS terminology this is called “Level 2 data”.

3.4.3 Data providers

1) ICOS: https://www.icos-cp.eu/data-products/atmosphere-release

2) WAQI portal http://waqi.info/

EEA, Air Quality databases (covered in next section)

3.5 Measurements from urban air quality networks

In the CoCO2 Project, no task reported a need for near-real-time observations of air quality related trace gases. There is however need for data from the simulated year, and data with delay of a year (for evaluation use).

3.5.1 Current spatial and temporal coverage

Nitrogen dioxide (NO_2), ground-level ozone (O_3) and particulate matter (PM_{10} and $PM_{2.5}$) are the three main urban air pollutants. Also, CO and SO_2 are often measured. In many countries, the measurements are responsibility of the municipalities or regional health authorities.

In Europe, EEA has information about European networks, altogether 62 000 datasets from 6150 stations.

In North America, data is available on AirNow site reporting air quality using the official U.S. Air Quality Index (AQI), based on ozone, PM_{10}, and $PM_{2.5}$. They have data from 500 U.S. cities, Canada and Mexico and a number of U.S. embassies. The total number of stations is over 2000.

The World Health Organization WHO has an air quality database, but it only contains PM data as yearly averages. While there is no global organization delivering data at higher temporal resolution, researchers are often relying on informal sources, easily leading to scattered and heterogenous data sources. The *World Air Quality Index* project is a non-profit project started in 2007. The team is based in Beijing, China. The project is providing air quality information for more than 130 countries, covering more than 30,000 stations in 2000 major cities, via its websites. There are also commercial companies which share data from their customers, such as IQ Air.

For other types of urban stations, European Fluxes Database has listed 20 urban flux stations in Europe. The data are half-hourly or hourly.

The International Association for Urban Climate has also had a site of micrometeorological towers in Urban environment, but the website is maybe not entirely up-to-date.

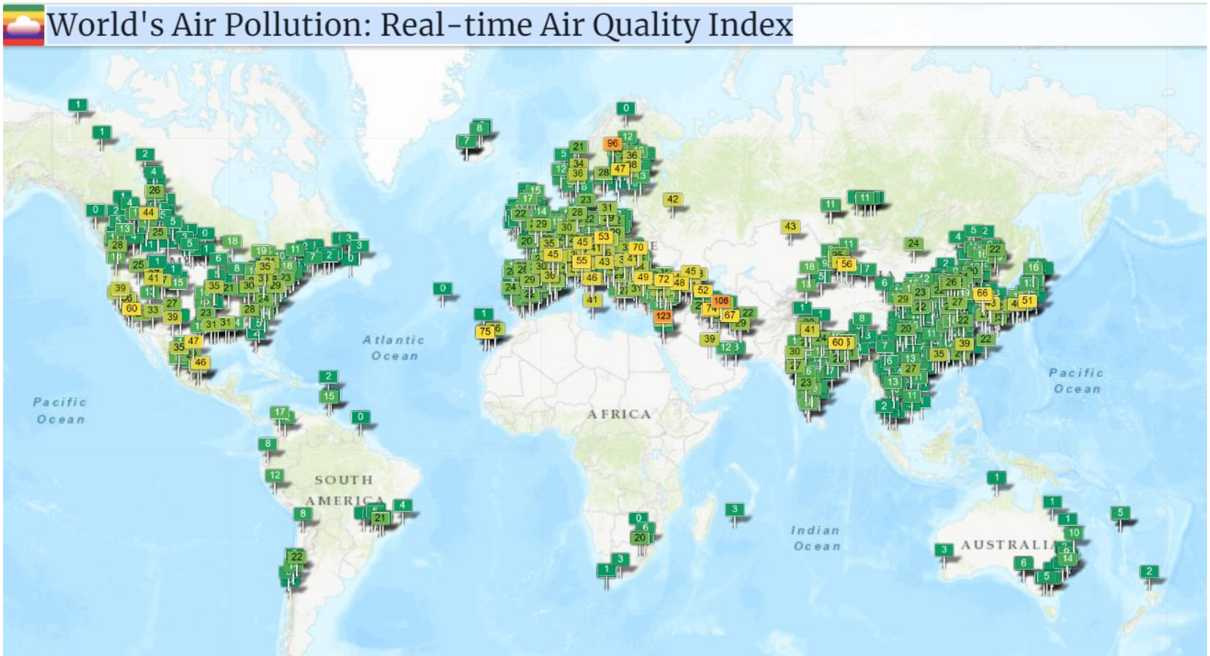

Figure 5. WAQI coverage of near-real-time air quality data

3.5.2 Timeliness

In EEA, preliminary data on hourly basis from most of the member countries is updated every night. Once a year (in September) a quality-controlled official dataset which covers the year before the delivery is uploaded by the member states, and all the preliminary data until that date is deleted. In AirNow, most data arrive by half-past the previous hour and are quality assured and released by the end of the hour. WHO database consists of annual averages.

WAQI provides near-real-time data based on hourly readings: for a value reported at 8AM the measurement was done from 7AM to 8AM.

3.5.3 Data providers

1) European Environmental Agency EEA: https://discomap.eea.europa.eu/map/fme/AirQualityExport.htm

2) North America: https://www.airnow.gov/

3) World Health Organization: https://www.who.int/data/gho/data/themes/air-pollution/who-air-quality-database

4) Industry portal: https://www.iqair.com/

5) WAQI portal http://waqi.info/

6) European Fluxes Database (www.europe-fluxdata.eu)

7) International Association for Urban Climate https://ibis.geog.ubc.ca/urbanflux/

3.6 Ocean fluxes/partial pressures

Note that ocean fluxes are not part of the CoCO$_2$ prototype.

The participants in survey of task 7.1 expressed only need for data from the simulated year.

3.6.1 Current spatial and temporal coverage

The Ocean component of ICOS consists of two types of stations: Ships Of OPportunity (SOOP) and Fixed Ocean Stations (FOS) which are moored to one location. ICOS release 2020-1 of the final quality-controlled Level 2 data from the ICOS Ocean network. This collection of 108 files was produced by labelled ICOS stations and processed following the ICOS Ocean data protocols using the QUINCE software tool (https://otc.icos-cp.eu/data-processing).

The Surface Ocean CO_2 Atlas (SOCAT) is a synthesis activity for quality-controlled, surface ocean fCO_2 (fugacity of carbon dioxide) observations by the international marine carbon research community (>100 contributors).

3.6.2 Timeliness

Typically, both ICOS and SOCAT data releases are published once per year. Some ICOS Ocean stations provide NRT data on daily basis.

3.6.3 Data providers

1) ICOS Carbon portal : https://www.icos-cp.eu/data-products

2) SOCAT: https://www.socat.info/

3.7 Radiocarbon in CO2

In D7.1 the modelers expressed only need for data from the simulated year. However, the answers from CoCO2 users in 7.1 may be not representative regarding all the requirements for $^{14}CO_2$, and this is one area we will focus on in next version of the report.

Radiocarbon is measured at all ICOS Class1 stations as integrated 14-day samples. According to modeler's experiences, in areas of low anthropogenic/biological CO_2 ratios such as Scandinavia, the integrated data are of lesser value, as they often represent emissions transported for long distances. Hence radiocarbon measurements from both time scales: integrated and instantaneous, are needed. The project CORSO (CO2MVS Research on Supplementary Observations, Grant agreement ID: 101082194 until 31 December 2025) is temporarily increasing the sampling rate by a factor of 5 at ICOS stations.

3.7.1 Current spatial and temporal coverage

In Europe, 18 ICOS Atmosphere stations measure radiocarbon since 2021.

Because radiocarbon measurements are sometimes integrated over a longer measurement period (e.g. two weeks at ICOS sites) and sometimes are instantaneous, as in campaign-based flask measurements, both temporal scales need to be considered. Both users indicated that they would use both types of measurements, based on availability.

For background "clean" sites, long time series have been collected from Point Barrow, Alaska, La Jolla, California, Mauna Loa, Hawaii Kumukahi, Hawaii, Cape Matatula, Samoa and the South Pole, Antarctica

A network of 15 stations in China has been reported by Zhou et al, 2020.

WMO OSCAR database reports globally 17 stations measuring radiocarbon; these are not the only stations.

Table 4 Geographical distribution of stations measuring radiocarbon according to WMO OSCAR database, Feb 2022

Africa	Antarctica	Asia	Europe	North America,	South America	SW Pacific
2	4	-	4	6	-	1

3.7.2 Uncertainty and timeliness

Depending on the data source.

3.7.3 Data providers

ICOS Carbon portal

Authors of papers in the "references" section: Graven, Zhou, Levin.

3.8 Atmospheric mixing ratios of radon

Task 3.1 (Forward modelling and data assimilation developments for operational global prototype) reported a need for near-real-time observations of radon concentrations. There is also need for data with delay of a year (for evaluation use) as well as older data for parameter development.

The general problem with radon is that different measurement principles are used. So, combining data from different instruments requires careful inter-comparisons and eventually application of corrections.

Radon measurements in the radiological networks are installed for radiation protection purposes and are much less sensitive than what is required for atmospheric transport studies. Radon variability that we need for the transport analysis is much smaller than what is required to detect hot spots or incidents. Radon has a role in developing atmospheric transport models but no other operational use as such in GHG MVS.

Timeliness and access of radon data has improved recently, as NRT radon data timeseries are now available from 6 ICOS stations in France, the Netherlands, Switzerland, and the UK via the ICOS Carbon Portal. The radon data still need full QA/QC and proper calibration, which was started in the EMPIR project 19ENV01 traceRadon. Limited coverage of these measurements remains the most serious limitation to their use.The radon data still need full QA/QC and proper calibration, which was started in the EMPIR project 19ENV01 traceRadon.

3.8.1 Current spatial and temporal coverage

WMO OSCAR lists 27 stations (search Atmosphere > Radionuclide) globally.

Table 5 Geographical distribution of stations measuring radon according to WMO OSCAR database, Feb 2022

Africa	Antarctica	Asia	Europe	Canada	South America	SW Pacific
2	1	0	16	6	0	2

Most of the stations listed in OSCAR are ICOS stations and radon is/will be available at CP.

As radon is only a recommended parameter for ICOS atmospheric stations, not all stations have implemented or will implement these measurements.

3.8.2 Data providers

As long as the radon data pipeline to ICOS CP is not fully established, contacting station PIs is the only way to get access to atmospheric radon data. For European stations a shortcut could be to directly contact ICOS ATC. For data outside of Europe, the best shortcut is probably ANSTO.

Scott Chambers: https://www.ansto.gov.au/people/dr-scott-chambers
Michel Ramonet: ICOS ATC https://www.icos-cp.eu/

3.9 Ground-based remote sensing measurements of atmospheric composition

Task 3.1 (Forward modelling and data assimilation developments for operational global prototype) reported a need for near-real-time observations. There is also need for data with delay of a year (for evaluation use).

3.9.1 Current spatial and temporal coverage

TCCON Network provides column-averaged abundances of CO_2, CH_4, N_2O, HF, CO, H_2O, and HDO at 28 locations around the world (see map). The global network is managed by CalTech.

COCCON- Collaborative Carbon Column Observing Network uses EM27/SUN spectrometers from KIT to provide provides column-averaged abundances of CO_2, CH_4 and CO. 40 devices at 18 locations or supersites are operated around the globe by working groups in Germany, USA, UK, India, Namibia, Japan, China and Mexico.

NDACC Network for the Detection of Atmospheric Composition Change (NDACC) is composed of more than 70 globally distributed, ground-based, remote-sensing research stations. 26 of these operate FTIR Spectrometers, and are also members of one or both of the above mentioned (TCCON or COCCON) networks.

AirCore is a research instrument, but the campaign data may be useful for evaluation use. It is unique profiles up to stratosphere where we do not have much more, or it can be used as a total column reference.

The National Oceanic and Atmospheric Administration (NOAA) / Earth System Research Laboratory's (ESRL's) airborne flask contains now 14 active sites in North America. Their sites on other continents have been discontinued

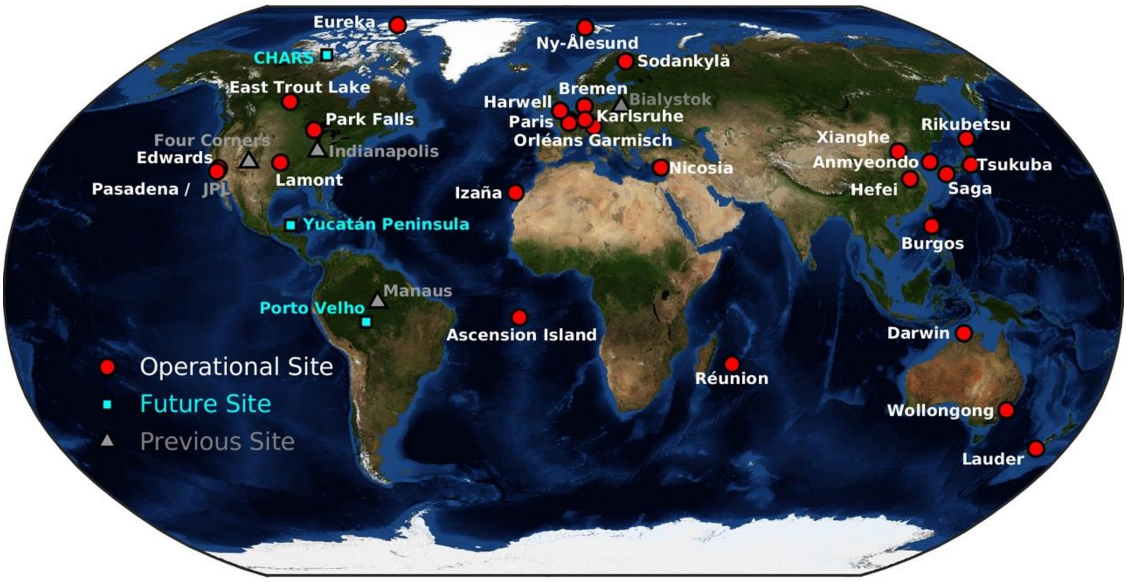

Figure 6. TCCON Network

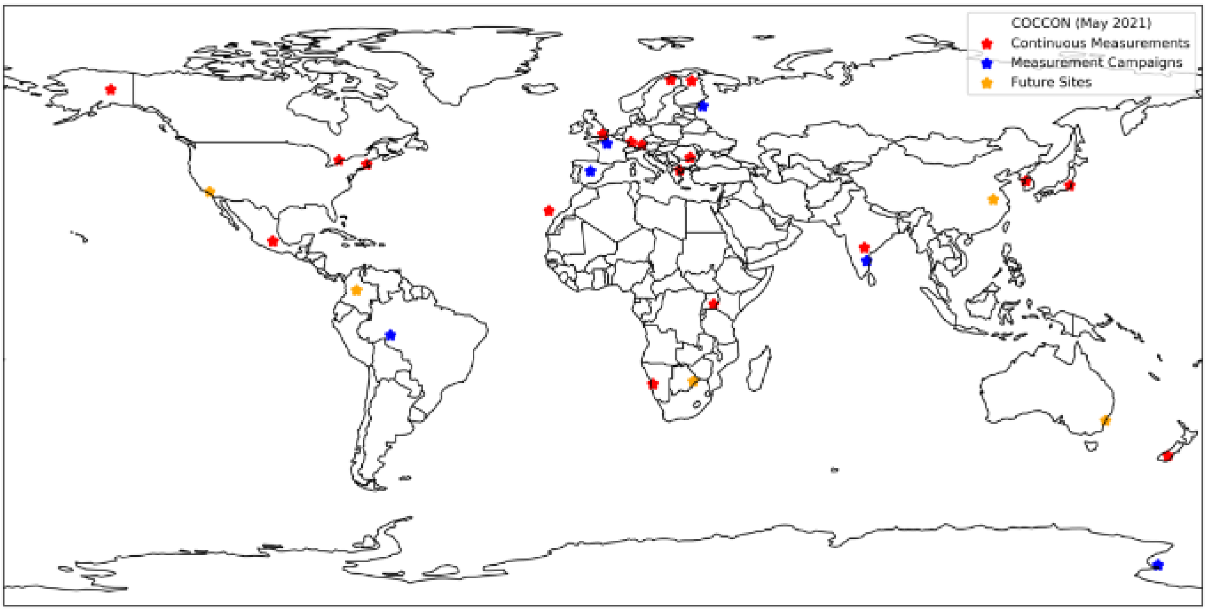

Figure 7. COCCON network (May 2021)

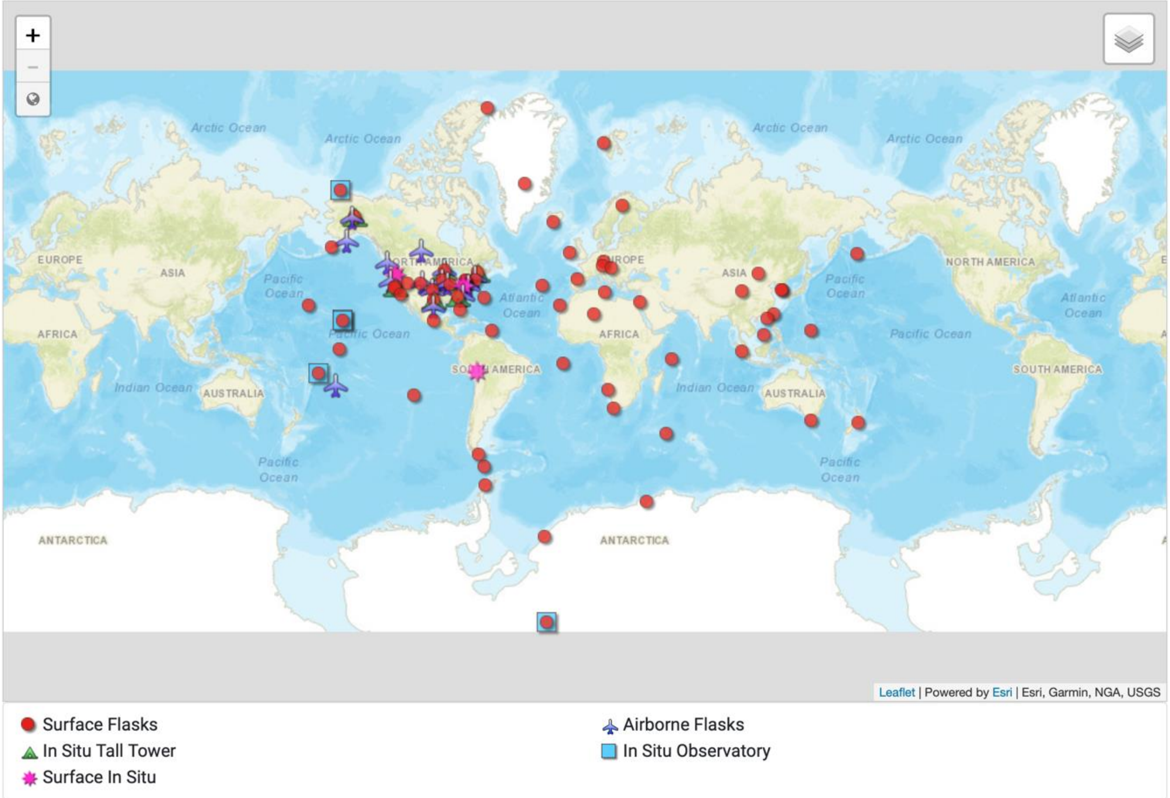

Figure 8. Air flask observation sites according to Global Monitoring Laboratory (https://gml.noaa.gov/dv/site/gmdsites.php?program=ccgg&projtable=1&active=1) accessed 14.9.2023)

3.9.2 Timeliness

TCCON data are publicly available no later than one year after the spectra are recorded; many sites release their data earlier. Timeliness varies due to site specific constraints, available manpower, raw data transfer, contractual issues, etc. For TCCON, the fastest sites are sorted in the 1-2 month category, for slow sites, data delivery might be a year or more. For NDACC, half a year to a 1.5 year lag also is a typical range, while a subset of sites is supporting a rapid delivery project for certain species (which then are delivered within a month) through a CAMS support contract.

For COCCON, the latest data on the EVDC portal are a few months old. For a selection of sites, the data delivery could in principle be speeded up significantly (mostly European and US sites) to provide data within two weeks or so.

AirCore data is campaign-based, not operational.

3.9.3 Data providers

1) TCCON data from CALTECH: https://tccondata.org/

2) COCCON data from ESA EVDC: https://evdc.esa.int

4) AIRCORE: https://gml.noaa.gov/ccgg/aircore/

AirCore NOAA: contact person Colm Sweeney. In Europe, AirCore campaigns have been executed in Finland (contact Rigel Kivi), France and Germany (Andreas Engel) and in Sweden (contact Cyril Crevoisier for Kiruna).

3.10 Measurements of site-level ecosystem parameters

There is only need for data with delay of a year (for evaluation use).

3.10.1 Current spatial and temporal coverage

In Europe, ICOS Ecosystem stations execute a set of non-continuous measurements related to vegetation: Green Area Index, aboveground biomass, litter biomass, basal area distribution. For details, see Gielen et al (2018).

3.10.2 Timeliness

Typically, once per year.

3.10.3 Data providers

ICOS Ecosystem Thematic Centre

3.11 (Information about site-level management and/or lateral fluxes)

No respondents indicated that they were utilizing site-level information about management (e.g. ploughing, harvest, fertilizer) or related lateral fluxes (e.g. wood harvest in forests, organic manure in crop fields and grasslands).

3.12 (In-situ soil moisture measurements)

No respondents reported using in situ measurements of soil moisture within this task. However, soil moisture measurements might be useful for various of applications. For example, soil water index is implemented in many weather forecasting systems and used to produce information related forest fire warning. Soil water index is also recognized as a Essential Climate Variable by the Global Climate Observing System (GCOS).

3.13 In-situ measurements of meteorological parameters

No need for near-real-time observations was expressed. There is however need for data from the simulated year, and data with delay of a year (for evaluation use) and even from any year for parameter estimation. Typically, the meteorological parameters are measured at same locations and intervals as GHG parameters, and available from the same sources.

4 Ancillary/Auxiliary data needs

4.1 Meteorological model fields

ECMWF Reanalysis v5 (ERA5) is the fifth generation ECMWF atmospheric reanalysis of the global climate. ERA5-Land[1] is a high-resolution product, which provides land surface variables at hourly resolution since 1950 to present. ERA5 is produced by the Copernicus Climate Change Service (C3S) at ECMWF. ERA5-Land provides land surface variables at higher resolution.

The Global Forecast System (GFS) is a weather forecast model from the National Center for Environmental Prediction (NCEP), that generates data for atmospheric and land-soil variables, like temperatures, winds, precipitation, soil moisture, and atmospheric ozone concentration. The system couples four separate models (atmosphere, ocean model, land/soil model, and sea ice) that work together to accurately depict weather conditions.

[1] https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land?tab=overview

4.1.1 Current spatial and temporal coverage

ERA5 provides hourly estimates of a large number of atmospheric, land and oceanic climate variables from January 1950 to present. It has a 30 km grid and resolves the atmosphere using 137 levels from the surface up to a height of 80 km.

GFS is a global model with a base horizontal resolution of 28 km between grid points. Temporal resolution covers analysis and forecasts out to 16 days. Horizontal resolution drops to 70 km between grid points for forecasts between one week and two weeks.

4.1.2 Uncertainty and timeliness

ERA5 includes information about uncertainties for all variables at reduced spatial and temporal resolutions

4.1.3 Data providers

ERA5 are provided by Copernicus and available from the ECMWF website. Link here.

ERA5-Land are provided by Copernicus and available from the ECMWF website. link here

GFS is available from the NCEP website. Link here.

4.2 Nightlights

A nightlight product is produced using measurements from the Visible Infrared Imaging Radiometer Suite Day-Night Band (VIIRS DNB) and the Defense Meteorological Satellite Program Operational Linescan System (DMSP-OLS). The underlying data are sourced from the NOAA National Centers for Environmental Information (NCEI) archive. Additional processing by the University of Michigan enables access in Cloud Optimized GeoTIFF format (COG) and search capabilities using the Spatial Temporal Asset Catalog (STAC) standard. The data are published and openly available under the terms of the World Bank's open data license, in the registry on Amazon Web Services (AWS).

4.2.1 Current spatial and temporal coverage

VIIRS DNB is available from 2012-2020 and DMSP-OLS from 1992-2013.

4.2.2 Uncertainty and timeliness

Quarterly (typically four times per year)

4.2.3 Data providers

Trevor Monroe tmonroe@worldbank.org;
Benjamin P. Stewart bstewart@worldbankgroup.org;
Brian Min brianmin@umich.edu;
Kim Baugh kim.baugh@noaa.gov

Documentation: https://worldbank.github.io/OpenNightLights/wb-light-every-night-readme.html

How to cite: World Bank - Light Every Night was accessed on DATE from https://registry.opendata.aws/wb-light-every-night.

4.3 (Activity data)

(The D7.1 has a section of activity data needs, the topic is excluded from this version of Data report which concentrates in observations.)

4.4 Satellite-based indices

4.4.1 Current spatial and temporal coverage

Surface reflectance from satellites are available from several satellites, including the LANDSAT series, MODIS, VIIRS and, more recently, Sentinel-2 & 3. Each of these sensors offers different spatial and spectral resolution and temporal coverage. All are affected by cloud cover and as such are often aggregated into data products with a coarser temporal resolution than the overpass repeat time. Often these reflectances are post-processed to yield higher-level indices, such as NDVI, EVI, and LSWI for use in biosphere modelling (Walther et al 2022).

Landsat reflectance data are available **globally** from the following instruments:

Landsat 9 Operational Land Imager 2 (OLI-2): October 2021 to present

Landsat 8 Operational Land Imager (OLI): April 2013 to present

Landsat 7 Enhanced Thematic Mapper Plus (ETM+): July 1999 to present

MODIS Surface reflectance is available in the following time and space resolutions

Aqua 8-Day L3 Global 500m: 2002-07-04 to Present, Multi-day resolution, Global coverage, Pixel size 500 m

Terra 8-Day L3 Global 250m: 2000-02-24 to Present. Multi-day resolution, Global coverage, pixel size 250 m

Terra Daily L2G Global 1km and 500m: 2000-02-24 to Present, daily, Global coverage, pixel size 1000 m and 500 m

Daily L2G Global 250m: Global coverage, pixel size 250 m, daily, Terra: 2000-02-24 to Present and Aqua: 2002-07-04 to Present, daily, Global coverage, pixel size 250 m

Sentinel

VGT S1 Surface Reflectance: L3 data, Global coverage, spatial resolution 1 km, daily

4.4.2 Data providers

Landsat: The data are in the public domain and more information is available at: https://www.usgs.gov/landsat-missions/landsat-surface-reflectance

This filedisplays the calibration, data processing, metadata, and product differences between Landsat Collection 1 Level-1, Level-2 U.S. Analysis Ready Data (ARD), and Landsat Collection 2 products.

MODIS: E. Vermote. (2015). MOD09A1 MODIS Surface Reflectance 8-Day L3 Global 500m SIN Grid V006. NASA EOSDIS Land Processes DAAC.

http://doi.org/10.5067/MODIS/MOD09A1.006 (Terra)

http://doi.org/10.5067/MODIS/MYD09A1.006 (Aqua)

Sentinel: https://sentinels.copernicus.eu/web/sentinel/sentinel-data-access

4.5 Satellite measurements of SIF

Satellite Solar Induced Fluorescence (SIF) are available from NASA's Orbiting Carbon Observatory-2 (OCO-2) and ESA's TROPOMI (TROPOspheric Monitoring Instrument).

OCO-2 Level 2 (L2) Version 10 is the current version. The data is available as bias-corrected, daily files.

The TROPOSIF products are provided in self-explanatory netCDF-4 files as ungridded data (Guanter et al., 2021). Two types of data files are available:

L2 data: orbit files, all retrievals (quality flag)

L2B data: daily files, only valid retrievals

4.5.1 Current spatial and temporal coverage

OCO-2: -180 to 180. -90 to 90 and 2014-09-06 to 2021-12-01

Timeliness: Daily

4.5.2 Data providers

OCO-2 on NASA data portal and TROPOSIF is available from the data website.

4.6 Other satellite-based measurements

Table 6 Summary of available satellite data

	Platform	Sensor	Spatial reso-lution	Temporal Reso-lution	Measure-ment	Data providers (link)
1	Orbiting Carbon Observatory -2 (OCO-2)	OCO-2	2.25 km x 1.29 km	Daily	CO2	OCO-2 landing page
2	Aqua	Atmospheric Infrared Sounder (AIRS)	40.5 km at nadir	Daily Monthly	CO2, CH4, H2O	AIRS landing page
3	Sentinel-5 Precursor	Tropospheric Monitoring Instrument (TROPOMI)	5.5 km x 3.5 km	Daily, Monthly	O3, NO_2, CO, CH4	TROPOMI landing page
4	Global Observing Satellite for Greenhouse gases (GOSAT)	Thermal And Near infrared Sensor for carbon Observation - Fourier Transform Spectrometer (TANSO-FTS)	10.5 km	Daily	CO2, CH4	GOSAT landing page
5	MetOp	Infrared Atmospheric Sounding Interferometer (IASI)	1 km		Vertical profiles of atmospheric humidity and temperature, CO, O3, CH4, CO2	EUMETSAT data page
5	MetOp-A	Global Ozone Monitoring	80 km x 40 km	Daily	O3, NO2, SO2	EUMETSAT data

		Experiment-2 (GOME-2)				navigator site

4.7 Landcover maps

The Copernicus Global Land Service (CGLS) is a component of the Land service that provides a series of bio-geophysical products on the status and evolution of land surface at global scale. The Dynamic Land Cover map at 100 m resolution (CGLS-LC100) is a new product in the CGLS and delivers a global land cover map at 100 m spatial resolution. The CGLS Land Cover product provides a primary land cover scheme. The product also includes continuous field layers for all basic land cover classes that provide proportional estimates for vegetation/ground cover for the land cover types. This continuous classification scheme may depict areas of heterogeneous land cover better than the standard classification scheme. Thus can be tailored for various applications (e.g. forest monitoring, crop monitoring, biodiversity and conservation, monitoring environment and security in Africa, climate modelling, etc.).

(Other landcover maps to be discussed in next version).

4.7.1 Current spatial and temporal coverage

The Land Cover maps (v3.0.1) are provided for the period 2015-2019 over the entire Globe, derived from the PROBA-V 100 m time-series, a database of high-quality land cover training sites and several ancillary datasets, reaching an accuracy of 80% at Level1 for all years.

4.7.2 Uncertainty and timeliness

Currently, the global land cover data is available only for 2015-2019. Yearly updates from 2020 through the use of a Sentinel time-series are being planned.

4.7.3 Data providers

The Land cover maps can be viewed in the data viewer interface https://lcviewer.vito.be/2015. These maps can be downloaded as 20x20 degree tiles of global files on Zenodo or analysed in Google Earth Engine™

4.8 Concentration fields from a global model

The CAMS reanalysis is the latest global reanalysis data set of atmospheric composition (AC) produced by the Copernicus Atmosphere Monitoring Service (CAMS), consisting of 3-dimensional time-consistent AC fields, including aerosols, chemical species and greenhouse gases (GHGs) through the separate CAMS global greenhouse gas reanalysis (EGG4).

The CAMS reanalysis was produced using 4DVar data assimilation in Cycle 42r1 of ECMWF's Integrated Forecasting System (IFS), with 60 hybrid sigma/pressure (model) levels in the vertical, with the top level at 0.1 hPa. Atmospheric data are available on these levels and they are also interpolated to 25 pressure levels, 10 potential temperature levels and 1 potential vorticity level. "Surface or single level" data are also available. The model level fields are in GRIB2 format. More information here (Agusti-Panareda et al., 2017; Flemming et al., 2017).

4.8.1 Current spatial and temporal coverage

The CAMS reanalysis dataset is available for the period: January 2003 to June 2021. The data are available at a sub-daily and monthly frequency and consist of analyses and 48h forecasts, initialised daily from analyses at 00 UTC. For sub-daily CAMS reanalysis data, the analyses are available 3-hourly. The daily forecast, run from 00 UTC, has 3-hourly steps from 0 to 48 hours for the 3D model level and pressure level fields, and hourly steps from 0 to 48 hours for the surface fields.

4.8.2 Data providers

CAMS global greenhouse gas reanalysis (CAMS EGG4) data is available here

4.9 Other auxiliary data repositories

Even though these needs were not expressed by the users, we have been encouraged to include three data types in this report: forests, soil classification and lakes. There might be a link between these and use of landcover maps.

Forest resources estimates are available from two sources. The European Forest Information SCENario Model (EFISCEN) is a large-scale forest model that projects forest resource development on regional to European scale. EFISCEN provides data on basic forest inventory data (species, area, stem wood volume, increment, mortality, age-structure) as well as multiple indicators allowing projection of forest resource development for a period of 50 to 60 years.

The Carbon Budget Model of the Canadian Forest Sector (CBM-CFS3) has been adapted, tested, and applied by JRC to forests of 26 EU countries over the last years for EU policy making, and scientific research.

FAO has published soil maps and databases on global scale since 1960s.

TERN has recently published an Australia-wide, consistent and accurate map of soil types at a 90m resolution, with estimates of reliability.

Data sources containing not continuous but annual, seasonal or periodical flux data and information regarding carbon stocks in measurement sites are 1) Global Forest Ecosystem Structure and Function Data for Carbon Balance Research[1] (Luyessart et al 2009) 2) Global Forest Carbon Database ForC[2]: a global database for forest carbon stocks and fluxes (Anderson-Teixeira et al 2018) that contain 9,762 records, 10,608 plots and 1,532 distinct geographic areas 3) A Global Soil Respiration Database[3] version 5.0 (Jian et al 2021) is a near-universal compendium of published soil respiration (Rs) data. Version 5 (V5) is the compilation of 2,266 published studies with measurements taken between 1961-2017. 4) COSORE[4]: A community database for continuous soil respiration and other soil-atmosphere greenhouse gas flux data (Bond-Lamberty et al 2020).

There are over a hundred million lakes with an area of more than 0.002 km2 in the world. Lakes influence the structure of the atmospheric boundary layer by affecting the surface fluxes, influence temperature and other weather parameters. To consider lake influence, the numerical weather prediction community has developed a Global Lake Database (GLDB). It contains *in situ* information about 14 960 lakes. This third version of GLDB is a global lake depth data set with *in situ* and estimated values on the 1 km grid (Toptunova et al, 2019).

Data sources and references

1) CBM: Pilli, Roberto (2021): EFA FRA CBM Data analysis. European Commission, Joint Research Centre (JRC) [Dataset] PID: http://data.europa.eu/89h/d4be2da6-54a1-4767-a262-dcebf66bf10b
2) EFI https://efi.int/knowledge/models/efiscen
3) FAO https://www.fao.org/soils-portal/data-hub/soil-maps-and-databases/global-soil-organic-carbon-map-gsocmap/en/
4) TERN: https://www.tern.org.au/news-australian-soil-classification-map/

[1] https://daac.ornl.gov/VEGETATION/guides/forest_carbon_flux.html
[2] https://forc-db.github.io
[3] https://daac.ornl.gov/SOILS/guides/SRDB_V5.html
[4] https://onlinelibrary.wiley.com/doi/full/10.1111/gcb.15353

5) https://daac.ornl.gov/VEGETATION/guides/forest_carbon_flux.html
6) https://forc-db.github.io
7) https://daac.ornl.gov/SOILS/guides/SRDB_V5.html
8) https://onlinelibrary.wiley.com/doi/full/10.1111/gcb.15353

5 Conclusion

We expect that all the modellers interviewed for D7.1. know from where to get the data they are currently using, or have used in the past, which may result in a certain response bias. This report attempts to provide a more global view. On the atmosphere side, the WMO has developed the OSCAR database which has collected information about stations. The weak point is that the WMO member countries are usually represented by the national meteorological services, while the GHG-related measurements are often run by universities, research institutes in areas of agriculture, forestry and health. Thus, this database does not always represent the state of the art in terms of atmospheric composition and related datasets.

For ecosystem fluxes we managed to create a global map of stations in major regional networks and for this deliverable update a map reporting +2100 EC measurement locations., We also noticed that the metadata is very heterogenous, and it is not always clear which stations are still operational.

This report feeds into “D7.6: Gap analysis report of the current *in situ* measurement capacity” and “D7.7: Requirements for data streams from additional tracers and new instrumentation.” Based on our findings, we suggest that both D7.6 and D7.7 should include information about the data policy and metadata availability. Also, discussion of access to non-real-time data for parameter estimation and model evaluation.

This report will also help to define more clarifying questions for the surveys that will be carried out for D7.2, the second iteration of the Book of *in situ* requirements. An online workshop in connection with the ICOS Science conference in September should also help to fill gaps in the first versions of both of these documents.

6 References

Agusti-Panareda, A., M. Diamantakis, V. Bayona, F. Klappenbach, and A. Butz, 2017: Improving the inter-hemispheric gradient of total column atmospheric CO2 and CH4 in simulations with the ECMWF semi-Lagrangian atmospheric global model, *Geosci. Model Dev.*, 10, 1-18, https://doi.org/10.5194/gmd-10-1-2017.

Anderson-Teixeira, K. J., Wang, M. M., McGarvey, J. C., Herrmann, V., Tepley, A. J., Bond-Lamberty, B., & LeBauer, D. S. (2018). ForC: a global database of forest carbon stocks and fluxes. *Ecology*, 99(6), 1507.

Bond-Lamberty, B., Christianson, D. S., Malhotra, A., Pennington, S. C., Sihi, D., AghaKouchak, A., & Zou, J. (2020). COSORE: A community database for continuous soil respiration and other soil-atmosphere greenhouse gas flux data. *Global change biology*, 26(12), 7268-7283.

Burba, G. and Daugherty R., B. (2022) ICOS Science Conference Book of Abstracts, 2100+ CO2 and H2O Flux Measurements Across the Globe: Sitting on a Golden Egg?

Burba G., 2019. Illustrative Maps of Past and Present Eddy Covariance Measurement Locations: II. High-Resolution Images. Retrieved September 7, 2023, from https://www.researchgate.net. DOI: 10.13140/RG.2.2.33191.70561, 9 pp.

https://www.researchgate.net/publication/335004533_Illustrative_Maps_of_Past_and_Present_Eddy_Covariance_Measurement_Locations_II_High-Resolution_Images

Delwiche, K. B., Knox, S. H., Malhotra, A., Fluet-Chouinard, E., McNicol, G., Feron, S., Ouyang, Z., Papale, D., Trotta, C., Canfora, E., Cheah, Y.-W., Christianson, D., Alberto, M. C. R., Alekseychik, P., Aurela, M., Baldocchi, D., Bansal, S., Billesbach, D. P., Bohrer, G., … Jackson, R. B. (2021). FLUXNET-CH4: a global, multi-ecosystem dataset and analysis of methane seasonality from freshwater wetlands. *Earth System Science Data*, 13(7), 3607–3689. doi: 10.5194/essd-13-3607-2021

Flemming, J., and Coauthors, 2017: The CAMS interim Reanalysis of Carbon Monoxide,Ozone and Aerosol for 2003–2015. *Atmos. Chem. Phys*., 17, 1945–1983, https://doi.org/10.5194/acp-17-1945-2017

Foken, T., Wimmer, F., Mauder, M., Thomas, C., & Liebethal, C. (2006). Some aspects of the energy balance closure problem. *Atmospheric Chemistry and Physics*, *6*(12), 4395-4402.

Gielen, B., Acosta, M., Altimir, N., Buchmann, N., Cescatti, A., Ceschia, E., ... & Wohljahrt, G. (2018). Ancillary vegetation measurements at ICOS ecosystem stations. International Agrophysics, 32(4), 645-664.

Graven, H. D., Guilderson, T. P., & Keeling, R. F. (2012). Observations of radiocarbon in CO2 at seven global sampling sites in the Scripps flask network: Analysis of spatial gradients and seasonal cycles. *Journal of Geophysical Research: Atmospheres*, 117(D2).

Jian, J., R. Vargas, K.J. Anderson-Teixeira, E. Stell, V. Herrmann, M. Horn, N. Kholod, J. Manzon, R. Marchesi, D. Paredes, and B.P. Bond-Lamberty. 2021. A Global Database of Soil Respiration Data, Version 5.0. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/1827

Jung, M., Nelson, J., Migliavacca, M., El-Madany, T., Papale, D., Reichstein, M., ... & Wutzler, T. (2023). Flagging inconsistencies in flux tower data. *Biogeosciences Discussions*, *2023*, 1-45. In review.

Levin, I., Hammer, S., Kromer, B., Preunkert, S., Weller, R., & Worthy, D. E. (2021). Radiocarbon in Global Tropospheric Carbon Dioxide. *Radiocarbon*, 1-11.

Luyssaert, S., I. Inglima and M. Jung. 2009. Global Forest Ecosystem Structure and Function Data for Carbon Balance Research. Data set. Available on-line [http://daac.ornl.gov/] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A. doi:10.3334/ORNLDAAC/949

Mammarella,I., Aslan, T., Burba, G., Cowan, N., Helfter, C. et al. 2022, Protocol for non-CO 2 eddy covariance measurements, QA/QC, data processingand gap-filling. Deliverable D3.5 in RINGO: Readiness of ICOS for Necessities of integrated Global Observations. https://www.icos-cp.eu/sites/default/files/2020-12/D3.5.%20Protocol%20for%20non-CO2%20eddy%20covariance%20measurements%2C%20QA-QC%2C%20data%20processing%20and%20gap-filling.pdf

Masarie, K. A., Peters, W., Jacobson, A. R., and Tans, P. P.: ObsPack: a framework for the preparation, delivery, and attribution of atmospheric greenhouse gas measurements, *Earth Syst. Sci. Data*, 6, 375–384, https://doi.org/10.5194/essd-6-375-2014, 2014.

Nemitz, E., Mammarella, I., Ibrom, A., Aurela, M., Burba, G. G., Dengel, S., ... & Zahniser, M. (2018). Standardisation of eddy-covariance flux measurements of methane and nitrous oxide. *International agrophysics*, *32*(4), 517-549.

Pallandt, M., Kumar, J., Mauritz, M., Schuur, E., Virkkala, A.-M., Celis, G., Hoffman, F., & Göckede, M. (2022). Representativeness assessment of the pan-Arctic eddy-covariance site network, and optimized future enhancements. *Biogeosciences Discussions*, 1–42. doi: 10.5194/bg-2021-133

Papale, D. (2020). Ideas and perspectives: enhancing the impact of the FLUXNET network of eddy covariance sites. *Biogeosciences*, 17(22), 5587-5598.

Papale, D., Sabbatini, S., Shaukat, S., Concept on standards for data collection, classification, description, processing and distribution and methods for data identification, traceability and sharing in FLUXNET. 2020 Deliverable 4.3 in RINGO: Readiness of ICOS for Necessities of integrated Global Observations. https://www.icos-cp.eu/sites/default/files/2021-02/D4.3.%20Reprocessed%20long%20term%20data%20series%20from%209%20Ecosystem%20stations.pdf

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y. W., ... & Law, B. (2020). The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. Nature *Scientific data*, *7*(1), 1-27.

Pinty B., P. Ciais, D. Dee, H. Dolman, M. Dowell, R. Engelen, K. Holmlund, G. Janssens-Maenhout, Y. Meijer, P. Palmer, M. Scholze, H. Denier van der Gon, M. Heimann, O. Juvyns, A. Kentarchos and H. Zunker (2019) An Operational Anthropogenic CO_2 Emissions Monitoring & Verification Support Capacity – Needs and high level requirements for in situ measurements, doi: 10.2760/182790, European Commission Joint Research Centre, EUR 29817 EN.

Toptunova, O., Choulga, M., & Kurzeneva, E. (2019). Status and progress in global lake database developments. *Advances in Science and Research*, *16*, 57-61.

Vekuri, H., Tuovinen, J. P., Kulmala, L., Papale, D., Kolari, P., Aurela, M., ... & Lohila, A. (2023). A widely-used eddy covariance gap-filling method creates systematic bias in carbon balance estimates. *Nature Scientific Reports*, *13*(1), 1720.

Walther, S., Besnard, S., Nelson, J. A., El-Madany, T. S., Migliavacca, M., Weber, U., ... & Jung, M. (2022). A view from space on global flux towers by MODIS and Landsat: the FluxnetEO data set. *Biogeosciences*, *19*(11), 2805-2840.

Zhou et al, 2020 : Fossil fuel CO2 traced by radiocarbon in fifteen Chinese cities *Science of The Total Environment* Volume 729, 10 August 2020, 138639

7 List of abbreviations

AWS	Amazon Web Services	ICOS	Integrated Carbon Observation System
CAMS	Copernicus Atmosphere Monitoring Service	IFS	Integrated Forecasting System -
CBM-CFS3	Carbon Budget Model of the Canadian Forest Sector	MVS	Monitoring & Verification Support
CO2M	Copernicus Carbon Dioxide Monitoring mission	NCEI	National Centers for Environmental Information
COCCON	Collaborative Carbon Column Observing Network	NOAA	U.S. National Ocean and Atmosphere Administration
CP	Carbon Portal	ObsPack	Observation Package
EC	Eddy covariance	OCO	Orbiting Carbon Observatory
ECMWF	European Centre for Medium-Range Weather Forecasts	OCO-2	Orbiting Carbon Observatory 2
EEA	European Environment Agency	OSCAR	Observing Systems Capability Analysis and Review Tool

EFISCEN	European Forest Information SCENario Model	PM	Particulate Matter
ERA5	An ECMWF reanalysis data product 1979 to near real time	SOCAT	Surface Ocean CO_2 Atlas
ERIC	European Research Infrastructure Consortium	TCCON	Total Carbon Column Observing Network
FLUXNET	1) The data portal 2) measurement site network.	TERN	Terrestrial Ecosystem Research Network
GAW	Global Atmospheric Watch programme	TROPOMI	TROPOspheric Monitoring Instrument
GCOS	Global Climate Observing System	VIIRS DNB	Imaging Radiometer Suite Day-Night Band
GFS	Global Forecast System	WAQI	World Air Quality Index
GHG	Greenhouse gas	WMO	World Meteorological Organisation

Document History

Version	Author(s)	Date	Changes
0.1	Saltikoff, Kasurinen, Parampil (ICOS ERIC)	22/03/2022	First version
1.0	Saltikoff, Kasurinen, Parampil (ICOS ERIC)	31/03/2022	Based on review by Broquet and Marshall
1.1	Saltikoff, Kasurinen, Parampil (ICOS ERIC)	18/05/2022	Based on review by Engelen, including some late contributions from others
2.0	Kasurinen, Saltikoff, Parampil, Kutsch (ICOS ERIC)	21/09/2023	For review
2.1	Kasurinen, Saltikoff, Parampil, Kutsch (ICOS ERIC)	29/09/2023	Based on review by Jung & Segers

Internal Review History

Internal Reviewers	Date	Comments
Gregoire Broquet (LSCE)	28/03/2022	Thorough comments, some of which may be implemented for V2 in 2023
Julia Marshall	29/03/2022	Thorough comments, some of which may be implemented for V2 in 2023
Richard Engelen ECMWF	04/05/2022	Thorough comments, linking the report not only to research but to operational phase. some changes will be implemented V2 in 2023
Martin Jung, MPI	18/9/2023	Internal review
Arjo Segers, TNO	22/9/2023	Internal review