Book of *in situ* data requirements V1

Julia Marshall (DLR)

coco2-project.eu

Co-ordinated by ECMWF

D7.1 Book of *in situ* data requirements V1

Dissemination Level:	Public
Author(s):	Julia Marshall (DLR)
Date:	17/01/2022
Version:	1.0
Contractual Delivery Date:	31/12/2021
Work Package/ Task:	WP7/ T7.1
Document Owner:	DLR
Contributors:	ICOS-ERIC, input from across the project
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 deliverable reports on the first survey of the *in situ* and ancillary data requirements of the CoCO2 project from tasks across work packages (WPs) 2 through 6. These requirements need to be documented to transition the project from a scientific exercise to an operational capacity, so that data dependencies and potential weak links in the future provision of timely flux estimates can be identified. The data needs were collected with an online survey, and responses were received from 16 participants across nine different tasks. Based on these responses, the identified data needs across 13 types of *in situ* measurements and 8 types of ancillary data products have been documented. Particular attention was paid to the timeliness requirements of the data, in addition to other specifications such as spatial coverage, temporal resolution, and where the project participants are currently accessing the various data streams. While it is clear that much of the data are needed in a timely manner, only one respondent reported a current need for near-real-time observations. However, because the data needs are still evolving over the first year of the project, this report reflects only a snapshot of data usage at this interim stage. This is a living book of requirements and will be updated annually over the three years of the project.

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 emissions from satellite measurements provided by the constellation of CO_2 sensors that will make up the planned CO2M mission. 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, an integrated system will also require extensive *in situ* and ancillary observations to achieve its proposed objectives. 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. Note that in the context of the Copernicus Programme, *in situ* data refers to measurements collected by ground-based, seaborne or airborne sensors, including remote sensing sensors, as well as reference and ancillary data.

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, and their report (informally referred to as "the Green Report") has guided the work reported here. Their report identified the following general areas in *which in situ* measurements will be required:

- Calibrating and validating the space component of the MVS capacity,
- Assimilating the data into models and integrating information in the core MVS capacity,
- Validating and further improving physical models that govern the evolution of CO_2 in computer simulations, and
- Evaluating the output generated by the MVS capacity for its end users.

The importance of these data streams is illustrated in the overview diagram of the CoCO2 project, which outlines the structure of the MVS capacity, found in Figure 1. The left pillar of the diagram contains all the observational requirements of the system. The spaceborne measurements are the purview of the space agencies, while the meteorological observations and the assimilation thereof are taken care of by ECMWF. The observational needs that are being documented in this report belong to the other two categories, namely “surface and airborne observations” and “auxiliary observations”.

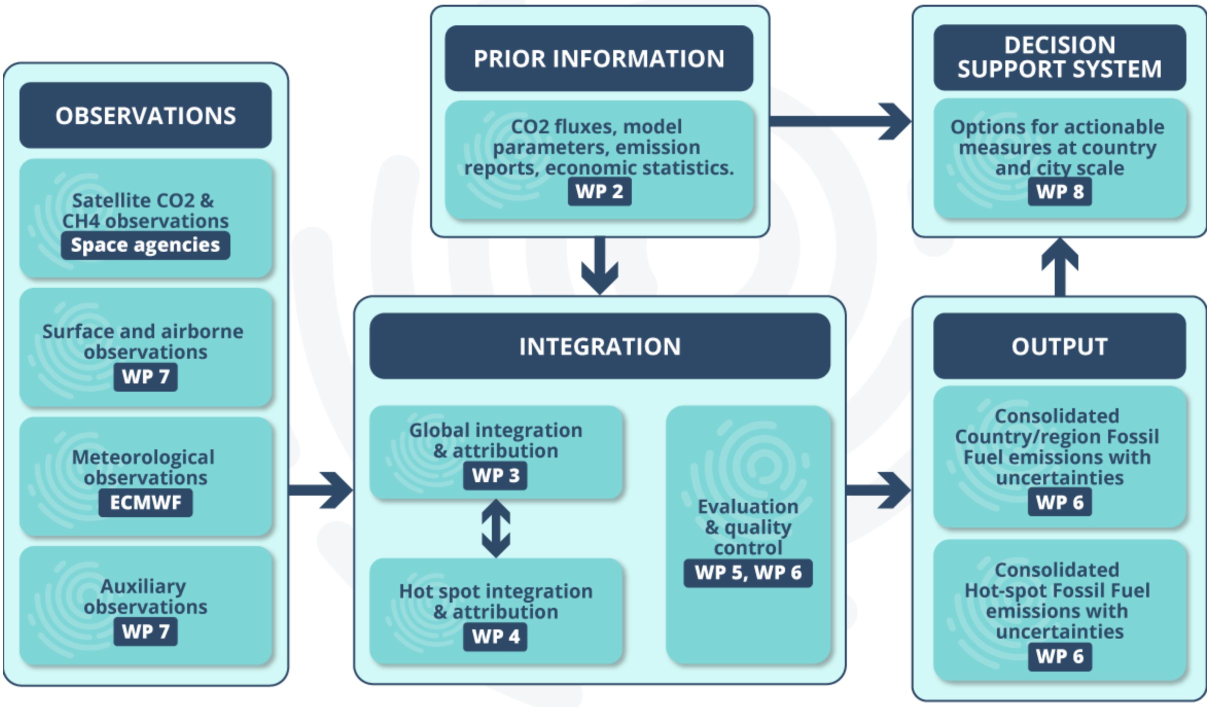

Figure 1: Schematic overview of the CoCO2 project structure, reflecting the structure of the MVS capacity itself. (Source: https://coco2-project.eu/structure)

2.2 Scope of this deliverable

2.2.1 Objectives of this deliverable

This deliverable aims to document the *in situ* and ancillary data requirements across the CoCO2 project, from WP2 through WP6. The documentation of these data needs is critical in order to move the work from a scientific exercise to an operational capacity. Through the documentation of these needs, the dependencies will become clear and potential weak links in the provision of timely emissions estimates can be identified.

To this end, the data standards (e.g. methodologies, accuracy) and specifications (e.g. spatiotemporal resolution) for measurements and auxiliary information should be collected and described. This report should be updated annually over the course of the project as the prototype evolves (with Deliverable 7.2 in month 24 and Deliverable 7.3 in month 36), and provide guidance for programmatic decisions regarding *in situ* measurement networks.

This deliverable and its successors will serve as a basis for the identification of data providers in Task 7.2, which will then be reported upon in Deliverables 7.4 and 7.5. It will also 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

In order to get feedback from across the project, an online survey was developed. The goal of the survey was to be as inclusive as possible, casting a wide net, and the possibility existed for respondents to add data streams that had been overlooked. One-on-one alpha-testing with one project participant was carried using a first draft of the survey, which resulted in a restructuring of the survey to make the overview clearer and make it easier to skip over irrelevant sections. Feedback was also collected from CoCO2 Steering Committee member Erik Andersson. This was followed by a beta-testing phase for participants in a CoCO2 land surface workshop from March, 2021. The survey was distributed to the full project in June. Video walk-throughs of the survey being completed from the perspective of a biospheric flux modeller and from the perspective of an atmospheric inverse modeller were recorded and posted, with links provided on the related confluence page.

The survey was split into two main sections, one focussing on *in situ* measurements, and one on ancillary and auxiliary data. Metadata about the respondent and their role in the project was collected in a separate introductory segment. The sections that were included are listed in Table 1. If a respondent indicated that he or she made use of data from a given category, they were directed to follow-up questions about how they were using the data, as well as their requirements for the data, namely: the pre-processing, measurement precision, spatial distribution, access, and timeliness. The full survey is still available to view (and even complete) here.

The *in situ* data needs are described in Section 3 of this deliverable, while the ancillary and auxiliary data needs are documented in Section 4. A list of the project tasks from which responses were collected, along with the task description and the number of respondents, is given in Table 2.

Table 1: List of the data streams that were included in the survey.

In situ data needs	
Q1	Eddy covariance flux data
Q2	*In situ* CO_2 measurements
Q3	*In situ* CH_4 measurements
Q4	*In situ* measurements of co-emitted species
Q5	Measurements from urban networks
Q6	Ocean fluxes/partial pressures
Q7	Radiocarbon
Q8	Other tracers (e.g. radon, OCS)
Q9	Ground-based remote sensing (e.g. TCCON)
Q10	Site-level ecosystem parameters
Q11	Site-level information on management and/or lateral fluxes
Q12	*In situ* soil moisture
Q13	*In situ* meteorological measurements
Q14	Anything else?
Ancillary/auxiliary data needs	
Q1	Meteorological fields
Q2	Nightlights
Q3	Activity data
Q4	Satellite-based indices
Q5	Satellite-based measurements of SIF
Q6	Other satellite-based measurements
Q7	Landcover maps
Q8	Concentration fields from a global model
Q9	Anything else?

Table 2: List of tasks from which responses were received, with the number of respondents.

Task	Description	# of respondents
2.1	Prior Emission Dataset	1
3.1	Forward modelling and data assimilation developments for operational global prototype	1
3.2	Fossil fuel emission modelling and parameter estimation	1
3.3	Community land-surface modelling for vegetation carbon exchange fluxes: ECLAND	2
4.1	Local scale model performance assessment and improvement	2
4.3	Local inversion approaches using atmospheric transport models	2
4.4	National scale inversions	6
5.2	Assessing and quantifying errors of biogenic CO_2 fluxes	1
5.6	Assessment of uncertainties in European inversion of CO_2 and CH_4	1

2.2.3 Deviations and counter measures

No significant deviations arose, and no countermeasures were undertaken. However, the response rate to the survey was somewhat underwhelming, with only 16 unique respondents, reflecting the data usage from 9 different tasks, as summarized in Table 2. This is less than had been foreseen: we had hoped to receive responses from each task in Work Packages 2 through 6 (for a total of 26 tasks). Based on this experience, the survey will be replaced with (online) interviews for the follow-up versions of this document to ensure a more complete picture of the *in situ* data needs within the project.

3 *In situ* data needs

The first part of the survey focussed on the use of *in situ* measurements throughout the project. The responses for each type of data are summarized below. For each data type there is a summary of how many respondents reported using the data stream and in which tasks. The timeliness needs of the data users is explicitly described, as this information is particularly important for the design of the operational data pipeline in Task 7.4.

3.1 Eddy covariance flux data

Half of the respondents (8) indicated that they were using measurements from eddy covariance flux data in some way. Of these, half were involved in WP3, while the others were spread across WPs 2, 4, and 5. For these data we distinguished which specific measurements were used: only CO_2 flux measurements, or also moisture and energy fluxes? Was the direct measurement of net CO_2 fluxes used (NEE, or net ecosystem exchange), or rather the modelled partitioning into gross primary production (GPP) and ecosystem respiration? Flux tower data are characterized by gaps due to instrument problems and meteorological conditions under which the technique does not work well, e.g. stability. Consequently, gap-filled data sets are often used rather than raw measurements; the survey attempted to capture this use of processed vs raw data.

3.1.1 Near-real-time data needs

Only **Task 3.1** indicated that near-real-time eddy covariance flux data were needed. This task focusses on the evaluation of modelling and data assimilation developments in the global CO_2 MVS. Fluxes of CO_2, site-level meteorology, energy, and moisture fluxes are used. The need was for quality-controlled, gap-filled data with partitioning of the measured CO_2 fluxes into GPP and ecosystem respiration. The sigma values given by the data providers is used to estimate the measurement uncertainty. Data from all stations are considered, as they are working on a global scale. For evaluation purposes, data at different time scales are used (e.g. hourly, daily, weekly), depending on which scale of flux variability (diurnal, synoptic, seasonal) is being evaluated. For near-real-time data, the ICOS Carbon Portal was listed as a current source of measurements. The biggest concern with the use of the measurements was unreliable data quality (e.g. negative values for GPP), limited global coverage, and temporal delays.

3.1.2 Needed for use in year of simulation

One respondent taking part in **Task 4.4** (national-scale inversions) indicated that they would be using flux measurements from the simulation year directly. However, further discussion indicated that this was rather for the purpose of evaluating the modelled fluxes rather than direct assimilation, and as such is documented in the following sub-section.

3.1.3 Needed only afterwards for model evaluation

One of the national-scale flux inversion modelling groups from **Task 4.4** indicated that they would be using quality-controlled total CO_2 flux measurements to evaluate the NEE fluxes in their model. All stations within their (regional) modelling domain were considered for this purpose at a temporal resolution of one hour, potentially aggregating in time after sampling to model to ensure consistency in the approach. At present, this group was getting the measurements directly from the station PIs.

Two respondents from WP5 indicated that they would be using eddy covariance flux data for model evaluation, specifically **Task 5.2**, which focusses on site-level simulations of land surface models, and **Task 5.6**, which assesses inverse modelling results. While Task 5.6 was only using CO_2 flux measurements (quality-controlled, gap-filled, partitioned into GPP and ecosystem respiration), Task 5.2 was also considering energy and moisture fluxes in addition to CO_2 fluxes (quality-controlled and partitioned, but without gap filling). Both groups were making use of the quality flags provided with the data, and were using the data at half-hourly or hourly resolution (as well as aggregated scales). The respondent from Task 5.2 mentioned concerns about the quality of the data that were readily available, and as such preferred to use only data from stations that they knew to be reliable. Issues related to Intellectual Property Rights when using others' data was also mentioned as a concern. In contrast, the respondent from Task 5.6 was happy to make use of all data available, and was currently accessing data through ICOS and/or FLUXNET. They considered the representativity of the available datasets to be the biggest concern.

A respondent from **Task 3.2** reported that the use of urban flux tower measurements was planned, specifically for the evaluation of anthropogenic emission inventories from WP2. To this end, the use of raw flux measurements from urban and residential sites was foreseen at temporal resolutions as high as that of the model timestep (10 to 20 minutes, though hourly or 3-hourly data were preferred). However, these data were not yet being used within the task, and their availability and mode of access were unclear.

Finally, **Task 3.3** will be using CO_2, moisture and energy fluxes (quality-controlled, gap-filled, partitioned data) in order to evaluate improvements in land use/land cover and prognostic leaf area index (LAI) mapping. In this case, the use of all sites is foreseen, addressing a range of temporal scales, from diurnal to seasonal to interannual. Data access is expected to be through ICOS and FLUXNET.

3.1.4 Data from other years can be used

Five of the eight respondents indicated that they were using measurements from eddy covariance flux towers for the purpose of parameter estimation, such that measurements from previous years could be used. (This usage was sometimes in addition to other timeliness needs, as in Tasks 3.1, 3.2 and 3.3 – more than one answer was possible per response.) Flux tower data was used for parameter estimation in **Tasks 2.1, 3.1, 3.2**, and **3.3**.

In Task 2.1, the gap-filled, quality-controlled partitioned CO_2 flux measurements from sites representative of certain land cover types are used to estimate model parameters. In this case, sites from the domain of interest (Europe) are used from other years, relying on the half-hourly fluxes from the FLUXNET 2015 release. The main concern was that some land cover types are poorly covered, but the respondent noted that this was a bigger concern for other areas of the world (like the Tropics).

The responses from Tasks 3.1 and 3.2 are discussed in more detail above in Section 3.1.1 and 3.1.3, respectively. In addition to these near-real-time and evaluation applications, these tasks are employing (or will be employing) flux measurements for parameter estimation as well.

There were two respondents from Task 3.3, one of which is also described in more detail in Section 3.1.3. Both of these respondents are working on updates and improvements to vegetation and land cover modelling (one specifying that this is in CHTESSEL), and the impact that these changes have on the biogenic fluxes. Here, hourly measurements of carbon, energy, and moisture fluxes were employed, using quality-controlled data and partitioned carbon fluxes. The uncertainties provided with the data were not considered. Currently the data were accessed from FLUXNET and/or ICOS. Concerns about the data included the format, quality, and the limited coverage of the measurements.

3.2 *In situ* atmospheric mixing ratios of CO_2

Of the 16 respondents to the survey, 10 reported making use of *in situ* measurements of the atmospheric mixing ratio of CO_2. Of these, almost all were making use of quality-controlled measurements (nine of ten) while one reported using raw measurements. These data found the most use in WP4, which is focussed on regional-scale simulations, from plumes to national-scale inversions. Besides WP4, the data were also used in WP5 (for evaluation of inversion results) and WP3.

3.2.1 Near-real-time data needs

Only **Task 3.1** reported needing these data in near-real-time, for the evaluation of the modelling and data assimilation developments in the global CO2MVS at ECMWF. In this case, quality-controlled measurements are needed, and the measurement uncertainty provided with the data is considered. Hourly data are used, taking advantage of data from all available sites. At present these data are provided in near-real-time from the ICOS Atmospheric Thematic Centre (ICOS-ATC), and with delay from NOAA's ObsPack. The biggest concerns with the use of these data are the limited coverage and temporal delays, although these delays (particularly with ObsPack) have decreased in recent years as the demand for timely data has increased.

3.2.2 Needed for use in year of simulation

Six respondents reported needing *in situ* measurements of atmospheric CO_2 mixing ratios for use in **Task 4.4**, for direct assimilation in the national-scale inversions. Two respondents also reported using them for **Task 4.3**, which is focussing on city-scale inversions. All but one reported using measurement uncertainties (precision/accuracy) that were reported with the product when assimilating these. Most users reported making use of all measurements available within their domain, with only one reporting a selection based on how well the model can represent a given measurement site. Data were generally assimilated at either native temporal resolution or at hourly resolution.

The access to these data with the timeliness required seems to be a limiting factor. Some reported various efforts to gather databases, such as the ICOS 2018 drought task force (which fast-tracked the availability of some measurements within Europe) or the accelerated ObsPack releases. ICOS was named as their main access to European measurements by five respondents, and NOAA/ObsPack, the World Data Centre for Greenhouse Gases (WDCGG) and direct contact to measurement PIs were also mentioned. Specific mention was made of measurements from the UK and from stations operated by the Max Planck Institute for Biogeochemistry, which are not currently distributed through ICOS or other operationalized channels. One user, simulating Krakow within Task 4.3, mentioned specific non-public data that are being collected within CoCO2, and potential issues related to the inhomogeneity of the timing of the campaign days. General concerns listed by the data users were limited coverage and temporal delays. With respect to coverage, the Iberian Peninsula was mentioned explicitly as a region in Europe with few measurements.

3.2.3 Needed only afterwards for model evaluation

Task 5.6 reported using *in situ* measurements of atmospheric mixing ratios of CO_2 for evaluation purposes. For this application they are using all the quality-controlled data they can find at the native temporal resolution, taking the measurement uncertainty into account. These datasets were being sourced from ICOS, NOAA, and WDCGG.

There were also two respondents from **Task 4.1** who were using *in situ* measurements of CO_2 to evaluate plume simulations carried out by different models. The answers corresponded to two of the cases that will be simulated, namely the Randstad region and Paris. For these cases specific *in situ* measurements were considered, including urban and peri-urban measurements. It was also the only explicit mention of aircraft-based measurements, though this may have been implicit in the other responses. As before, quality-controlled data are used, with temporal resolution up to minutes considered (where available) for these high-resolution simulations. Unlike the other users of this data stream, these respondents were more reliant on local PIs rather than operational data centres, reflecting the plume-scale focus of the simulations. Concerns listed were the (in)homogeneity of the available measurements and poorly quantified uncertainties.

One respondent from **Task 4.4** also explicitly mentioned using *in situ* measurements of CO_2 to evaluate the national-scale inversion, presumably through comparison of optimized concentrations to *in situ* measurements that were not assimilated.

3.2.4 Data from other years can be used

One of the respondents from **Task 4.1** also identified that the plume simulations in this task would be used for parameter estimation, to help optimize model settings for simulations at these high spatial scales. All other aspects of their response are captured in the description in Section 3.2.3.

One respondent from **Task 4.4** also described the use of these measurements from other years to determine optimal model settings.

3.3 *In situ* atmospheric mixing ratios of CH_4

Of the 16 respondents to the survey, 8 reported making use of *in situ* measurements of the atmospheric mixing ratio of methane. The respondents represent a subset of the people who reported using *in situ* measurements of atmospheric CO_2, described in Section 3.2, and some of the answers overlap as a result. Of these, all but one were making use of quality-controlled measurements (seven of eight) while one reported using raw measurements. These data are used widely in WP4 inversions on national and city scale (Tasks 4.4 and 4.3, respectively). Besides WP4, the data are also applied in WP5 (for the evaluation of inversion results) and WP3.

3.3.1 Near-real-time data needs

As was the case for CO_2, only **Task 3.1** reported needing these data in near-real-time for the evaluation of the modelling and data assimilation developments in the global CO2MVS at ECMWF. For a more complete description of this data use, refer to Section 3.2.1.

3.3.2 Needed for use in year of simulation

As was the case for CO_2, the direct use of these measurements through assimilation was dominated by WP4, specifically **Task 4.4** (four respondents) and **Task 4.3** (two respondents). Most of the responses are similar to those for CO_2, found in Section 3.2.2. There were some differences in terms of access to the data: one user mentioned that they relied upon project-specific efforts to gather datasets (e.g. in VERIFY), while another mentioned that some sites were still not reporting data as long as two years after the measurements were collected.

3.3.3 Needed only afterwards for model evaluation

As previously described for CO_2 in Section 3.2.3, **Task 5.6** is using *in situ* measurements of atmospheric mixing ratios of methane for the purpose of evaluating inverse modelling results. For this application they are using all the quality-controlled data they can find at the native temporal resolution, taking the measurement uncertainty into account. These datasets were being sourced from ICOS, NOAA, and WDCGG.

One respondent from **Task 4.4** also explicitly mentioned using *in situ* measurements of methane to evaluate their national-scale inversion results, presumably through comparison of optimized concentrations to *in situ* measurements that were not assimilated.

3.3.4 Data from other years can be used

As in Section 3.2.4 for CO_2, one respondent from **Task 4.4** described the use of *in situ* methane measurements from other years to determine optimal model settings.

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

Five respondents from WP3 and WP4 reported making use of *in situ* measurements of co-emitted species. In all cases, CO and NO_x/NO_2 were named as the species of interest, and in all cases quality-controlled data were used.

3.4.1 Near-real-time data needs

As was the case for *in situ* measurements of atmospheric CO_2 and CH_4, only **Task 3.1** indicated a need for these measurements in near-real-time. Here all available data were used at hourly spatial resolution, and currently data were being sourced from ICOS-ATC and air quality networks. Measurement uncertainty provided by the data providers was used. Limited coverage and temporal delays were identified as limiting factors in the use of these measurements.

3.4.2 Needed for use in year of simulation

None of the respondents from WP4 reported using assimilating these measurements directly.

3.4.3 Needed only afterwards for model evaluation

All four respondents from WP4 reported using these measurements after the fact for validation, with two from **Task 4.1** (plume simulations) and two from **Task 4.4** (national-scale inversions). In this case, the measurements were sourced from the EEA, air quality databases, or local sources, with CO from ICOS and ICOS-like GHG measurement sites being used. Both respondents from Task 4.4 reported using only sites representing regional rather than local signals, such as rural background sites. One of the respondents from Task 4.1 preferred using measurements when CO_2 was co-sampled, but would use everything for evaluation purposes. Temporal scales from minutes to hourly were used. One user reported that for air quality data the formats are not always consistent, and time was needed when converting to a new format.

3.4.4 Data from other years can be used

One of the respondents from **Task 4.1** (plume simulations) expressed that data from other years could be used, as they were focussing on optimizing chemistry and transport settings.

3.5 Measurements from urban air quality networks

Three respondents reported making use of measurements from urban air quality networks, specifically in **Task 3.2** (for evaluating anthropogenic emission inventories) and **Task 4.1** (plume modelling). Within Task 4.1, atmospheric abundances of CO_2, CO, NO_2, NO, and (for one respondent) ozone were of interest, and quality-controlled data were used. Temporal resolutions from minutes to hours were found to be useful, and any measurements within the regional domain of interest (specifically Rotterdam, Amsterdam, Paris, Berlin, or at stack level) were sought. The data were being sourced through station PIs, national organisations, air quality databases, or through colleagues. One concern with the available data was the lack of co-located meteorological measurements.

In Task 3.2, the interest was rather in raw urban flux measurements of CO_2 (and NO_x, if available), as well as any meteorological measurements. In theory all data were welcome, at a temporal resolution ranging from tens of minutes to hours, but these measurements had not yet been accessed.

3.5.1 Near-real-time data needs

No users expressed a need for urban air quality networks in near-real time.

3.5.2 Needed for use in year of simulation

Task 3.2 expressed the need for these (flux) measurements specific to the year being simulated.

3.5.3 Needed only afterwards for model evaluation

All three respondents indicated that these measurements would be used for evaluation purposes, and thus would still be of use after the fact. (The quantitative representation of air quality tracers within urban environments is quite challenging, but a qualitative comparison and evaluation could still be useful.)

3.5.4 Data from other years can be used

One of the respondents from **Task 4.1** and the respondent from **Task 3.2** indicated that data from other (previous) years could still also be used for e.g. parameter estimation or model optimization.

3.6 Ocean fluxes/partial pressures

At this stage of the project, only one user from **Task 4.4** indicated an interest in using ocean flux measurements in national-scale inversions. At this point the potential use was rather exploratory in nature, but they were considering using quality-controlled measurements for areas near Europe. The data were not in use yet, but an application of these measurements for parameter estimation, model evaluation, and direct assimilation were all being considered.

Based on the proposal these measurements may be used in the creation of an ocean flux product in the scope of Task 2.1, but no one responded to the survey in relation to this activity, and the measurement-based product may only become available in the next year of the project. This will be further pursued in follow-up surveys and versions of this report.

3.6.1 Near-real-time data needs

No near-real-time data needs were reported for measurements of ocean fluxes or partial pressures.

3.6.2 Needed for use in year of simulation

The respondent from Task 4.4 indicated that direct assimilation of ocean flux and/or partial pressure measurements from the year of simulation was being considered in an exploratory manner.

3.6.3 Needed only afterwards for model evaluation

The respondent from Task 4.4 reported that measurements of ocean flux and/or partial pressure might be used for model evaluation.

3.6.4 Data from other years can be used

Depending on what comes of the exploratory use of these measurements, the respondent from Task 4.4 considered that these measurements might be used for parameter estimation.

3.7 Radiocarbon in CO_2

Two respondents, one for **Task 4.3** and **4.4** and one only for **Task 4.4**, indicated that they would be using measurements of radiocarbon in CO_2 for the purpose of source attribution. Both users wanted to make use of quality-controlled measurements, and would make use of measurement uncertainty if available. All measurements within the domain of interest (Europe and/or Krakow) were being considered.

Because radiocarbon measurements are sometimes integrated over a longer measurement period (e.g. two weeks at ICOS Class 1 stations) 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.

Regarding the source of the measurements, for the Krakow domain the campaign data from WP7 is being exploited, in addition to data provided directly from measurement PIs and ICOS. The other respondent indicated using ICOS and NOAA measurements. Both respondents indicated the availability of these measurements was the most important limitation at present, both spatially and in terms of temporal coverage.

3.7.1 Near-real-time data needs

No users reported a need for near-real-time measurements of radiocarbon.

3.7.2 Needed for use in year of simulation

Both respondents indicated that they would directly assimilate the measurements in their activities, and as such required measurements from the year of simulation.

3.7.3 Needed only afterwards for model evaluation

Radiocarbon measurements were not identified for use in model evaluation.

3.7.4 Data from other years can be used

No users reported using radiocarbon measurements from other years, e.g. for parameter estimation.

3.8 Atmospheric mixing ratios of other species

Four respondents from WPs 3, 4, and 5 reported using radon measurements, if available. All users indicated that they wanted quality-controlled data from as many sites as possible, with temporal resolution from minutes to hours. Currently, three of the respondents indicated that they accessed these data directly from the station PIs, through personal contacts, or through national networks (in one case). Only one respondent indicated accessing these data through ICOS-ATC, and this was for near-real-time application. All users reported that the limited coverage of these measurements was the most serious limitation to their use.

3.8.1 Near-real-time data needs

Task 3.1 was the only place in the project in which the near-real-time application of radon measurements was identified. In this case, data access through ICOS-ATC was reported.

3.8.2 Needed for use in year of simulation

No users reported assimilating these data directly for the year in question, but in a non-near-real-time capacity.

3.8.3 Needed only afterwards for model evaluation

Respondents from **Task 4.1** (plume simulations), **Task 4.4** (national-scale inversions) and **Task 5.6** (evaluation of inverse modelling results) reported using these data for evaluation purposes, for which a delayed delivery was not critical.

3.8.4 Data from other years can be used

The respondents from Task 4.1 reported that data from other (previous) years could also be used, e.g. for parameter estimation or model optimization.

3.9 Ground-based remote sensing measurements of atmospheric composition

Seven respondents from across WPs 3, 4, and 5 reported using or planning to use ground-based remote sensing measurements of atmospheric trace gases within the project. All planned on making use of XCO_2 (total column atmospheric CO_2) measurements, four planned on using XCH_4 measurements as well, and one reported using column-integrated carbon monoxide measurements (XCO) in addition to the other two species.

3.9.1 Near-real-time data needs

Task 3.1 indicated needing XCO_2, XCH_4, and XCO measurements in near-real time in order to evaluate the modelling and data assimilation developments in the global CO2MVS. For this purpose, hourly data were sufficient, and currently were accessed directly from the TCCON database, taking the reported measurement precision into account while using them. Both limited coverage and temporal delays were identified by this respondent as significant limitations in the use of this data stream.

3.9.2 Needed for use in year of simulation

No users reported using these data for direct assimilation in (non-IFS) simulations.

3.9.3 Needed only afterwards for model evaluation

All other respondents indicated that the measurements were used primarily for evaluation purposes, and as such a delay in their availability is less critical. This included one respondent from **Task 3.2** (implementing and evaluating anthropogenic emission inventories), two from **Task 4.1** (plume simulations), two from **Task 4.4** (national-scale inversions), and one from **Task 5.6** (evaluation of inverse model results). Most of these users reported accessing the measurements through the TCCON data archive. In addition to this, the respondents from Task 4.1 indicated that they also contacted station PIs directly in order to access EM27 measurements that are not part of TCCON. Two respondents indicated that limited data coverage in some areas was a shortcoming of this data stream.

3.9.4 Data from other years can be used

No respondents indicated that ground-based remote sensing data from other years could be used for their activities within the project.

3.10 Measurements of site-level ecosystem parameters

Only one respondent from **Task 5.2** reported that site-level measurements of ecosystem parameters, specifically leaf area index (LAI), could be useful in their site-level simulations. In this case, they would be using quality-controlled measurements on daily to monthly temporal resolution, specifically for the Fr-Tou flux tower site in Toulouse. Sampling time was considered to be the most critical issue with this data stream.

3.10.1 Near-real-time data needs

No near-real-time need for this data stream was reported within the project.

3.10.2 Needed for use in year of simulation

No respondents reported directly assimilating site-level measurements of ecosystem parameters.

3.10.3 Needed only afterwards for model evaluation

The respondent indicated that these data would be used for model evaluation purposes, and a temporal delay was not critical.

3.10.4 Data from other years can be used

No use of site-level measurements of ecosystem parameters for parameter estimation was reported.

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.

3.13 *In situ* measurements of meteorological parameters

Seven users from across WPs 2, 3, 4, and 5 reported directly using *in situ* measurements of meteorological parameters. Most users reported using quality-controlled measurements, one from **Task 4.3** reported preferring raw wind speed measurements and one from **Task 4.4** reported using both raw and quality-controlled data.

3.13.1 Near-real-time data needs

None of the respondents reported needing *in situ* meteorological measurements in near-real time.

3.13.2 Needed for use in year of simulation

The respondent from **Task 3.2** indicated that meteorological measurements from the current year would be needed for the implementation and evaluation of anthropogenic emission inventories. For this task, only urban measurements of relevant NWP variables (e.g. 2-m temperature, 10-m wind speed) would be used, with a temporal resolution ranging from model timestep (10-20 minutes) to 3-hourly. These data were not yet actively in use and as such no information about how they were accessed was available.

One respondent from **Task 4.3** (city-scale inversions) and one from **Task 4.4** (national-scale inversions) reported directly assimilating *in situ* measurements of wind speed and direction within their modelling domain. For this, a temporal resolution of hourly (or native) was foreseen. For the national-scale inversions the user reported retrieving the data from the ICOS Carbon Portal, and found that mast corrections and data availability to be the most challenging aspects related to its use.

3.13.3 Needed only afterwards for model evaluation

In **Task 5.6** direct measurements of planetary boundary layer height are being used for the purpose of model evaluation. In this case, the data are being collected directly from station PIs and applied at native temporal resolution. The most limiting aspect of this data stream is reported to be the correct application of the measurement uncertainties.

Both respondents from **Task 4.1** reported that *in situ* measurements of wind and temperature were being used for model evaluation for the plume simulations, and one further mentioned the use of moisture, boundary layer height, surface heat flux, moisture flux, and ground heat flux measurements. All available measurements with temporal resolutions from minutes to hourly within the domain of interest were being considered. The data were sourced from either national networks or meteorological databases. The lack of co-located meteorological and composition measurements was named as a limiting factor in making optimal use of these data.

The respondent from Task 4.4 indicated that wind measurements would be used for model evaluation as well.

3.13.4 Data from other years can be used

A respondent from **Task 2.1** reported on the use of *in situ* meteorological measurements from flux tower sites for the purpose of parameter estimation in the context of ecosystem modelling. In this case, incoming shortwave radiation and air temperature measured at eddy covariance flux tower sites representative of specific plant functional types were used. Half-hourly data were used, and the data were taken from the FLUXNET 2015 release.

The respondents from Task 3.2, 4.4 and one respondent from Task 4.1 (described in more detail in Sections 3.13.2 and 3.13.3) also characterized their use of *in situ* meteorological measurements in terms of parameter estimation or model optimization, for which measurements from other years could be used.

4 Ancillary/Auxiliary data needs

Of the 16 respondents, all reported using at least some ancillary or auxiliary data in their activities within CoCO2. Their data needs are reported here, following a similar format to that for *in situ* measurements in Section 3.

4.1 Meteorological model fields

Thirteen respondents reported using fields from meteorological models within their activities in the project, from WPs 2 through 5. Different products were used, mostly from ECMWF but also Meteo-France. Respondents reported accessing meteorological fields through MARS or ERA5 fields through the Climate Data Store (CDS) or local data pools (such as at the German supercomputing centre DKRZ). There were some concerns about the difficulty of access, with the CDS reported as being very slow, and MARS access being restricted, such that colleagues with full access had to be asked to retrieve data. One user reported having to log in to ECMWF weekly to maintain access. Two users reported concerns about dealing with the data volume, and one commented that the GRIB1 and GRIB2 file formats were somewhat complicated to deal with. Only one user, from **Task 3.2**, reported using the IFS-ensemble to estimate uncertainties in the provided fields. The users were split about the spatial resolution used, with seven reporting using ERA5 (at ~0.25° hourly resolution) and seven reporting using short-term forecast products with higher spatial resolution but 3-hourly temporal resolution. (One respondent reported using both.)

4.1.1 Near-real-time data needs

No users reported needing meteorological model fields in near-real time.

4.1.2 Needed for use in year of simulation

Most users reported needing these data for the year that they were simulating, including the only respondent from **Task 2.1** (for running a biospheric flux model), one respondent from **Task 3.2** (for implementing and evaluating online anthropogenic flux models), both respondents from **Task 4.1** (for plume modelling), one respondent from **Task 4.3** (city-scale inversions), all five respondents from **Task 4.4** (national-scale inversions) and the respondent from **Task 5.2** (site-level simulations).

4.1.3 Needed only afterwards for model evaluation

Both respondents from **Task 3.3** indicated that they used meteorological fields for model evaluation after adapting their landcover schemes. The respondent from **Task 5.2** reported similar usage for evaluation of site-level simulations. The respondents looking at plume-modelling case studies (Task 4.1) indicated that they were using the data for model evaluation as well, and as such had less stringent timeliness requirements.

4.1.4 Data from other years can be used

Both respondents from **Task 3.3** reported using meteorological fields for model parameter estimation, for which data from other years could be used. The respondent from **Task 3.2** and one respondent from **Task 4.4** reported similar usage, for model optimization.

4.2 Nightlights

Only one user reported using (or planning to use) nightlight data, for the purposes of implementing and evaluating online anthropogenic emission inventories in **Task 3.2**. This activity had not yet started in earnest however, and as such only rough information could be provided: using any available data at any resolution available.

4.2.1 Near-real-time data needs

A near-real-time need for this data stream was not yet reported.

4.2.2 Needed for use in year of simulation

The respondent from Task 3.2 planned on using these data for direct assimilation, and as such would need them for the year currently being simulated.

4.2.3 Needed only afterwards for model evaluation

No one reported needing these data *a posteriori* for model evaluation.

4.2.4 Data from other years can be used

The respondent from Task 3.2 also reported potentially using these data for parameter estimation, for which information from other years could be used.

4.3 Activity data

One user from **Task 3.2** reported that they were planning to use activity data, for the purposes of evaluating anthropogenic emission inventories produced in WP2. As this activity had not yet started, no concrete specifications regarding the data were provided.

4.3.1 Near-real-time data needs

No near-real-time data needs were reported.

4.3.2 Needed for use in year of simulation

The respondent did foresee the need to use data from the current year.

4.3.3 Needed only afterwards for model evaluation

It was not reported that these data would be used for model evaluation or validation.

4.3.4 Data from other years can be used

The respondent reported that data from other years could presumably be used to evaluate the upscaling approach used in emission inventories.

4.4 Satellite-based indices

Four respondents reported using satellite-based measurements of ecosystem properties. Two users, one from **Task 2.1** and one from **Task 4.4** reported using MODIS reflectance data to calculate EVI (enhanced vegetation index) and LSWI (land surface water index). In both cases they were using measurements with a spatial resolution from 500 m to 1 km at 8-day resolution and accessed the data from a NASA ftp server. Both users were using the indices as input for the diagnostic biospheric flux model VPRM (Vegetation Photosynthesis Respiration Model, Mahadevan et al., 2008), one offline for the production of fluxes in Task 2.1 and one online as part of their regional-scale inversion framework in Task 4.4. One of the users identified the pre-processing of the data as a potential limitation, requiring poorly documented software with a remapping library that is no longer maintained.

Two other project members reported using the Copernicus Global Land Service (CGLS) Leaf Area Index (LAI) product CGLS LAI v2, based on a daily synthesis of top-of-canopy (TOC) reflectance data from the PROBA-V satellite and the SPOT-VEGETATION Programme. This was for the purpose of improving land use and land cover mapping, prognostic LAI modelling, and evaluating the impact of land cover updates on the fluxes in CHTESSEL within **Task 3.3**. The data are used with a spatial resolution of 300 m and 1 km and with a temporal resolution from 3 to 10 days. The data were available through the CGLS and C3S portals. The most limiting factor in using the product was temporal delays.

4.4.1 Near-real-time data needs

No respondents reported needing these data in near-real time.

4.4.2 Needed for use in year of simulation

All users reported needing the satellite-derived measurements for direct use in the year currently being simulated.

4.4.3 Needed only afterwards for model evaluation

One respondent from Task 3.3 reported that the data were also used for validation and evaluation purposes.

4.4.4 Data from other years can be used

The respondent from Task 3.3 also reported that data from other years could be used for testing and improving the land cover model.

4.5 Satellite measurements of SIF

At this stage in the project, only one user from **Task 5.2** reported using spaceborne sun-induced fluorescence (SIF) measurements within the project, for the purpose of site-level simulations. The work was still at a planning stage however, and specific details regarding the data stream could not be provided.

4.5.1 Near-real-time data needs

No near-real-time data needs were reported.

4.5.2 Needed for use in year of simulation

The respondent who planned on using these data in Task 5.2 foresaw needing the measurements for the year being simulated.

4.5.3 Needed only afterwards for model evaluation

The use of these measurements for model evaluation was also foreseen.

4.5.4 Data from other years can be used

No users reported using these data for parameter estimation or similar within the project.

4.6 Other satellite-based measurements

Here, the question specified if other satellite-based measurements (other than atmospheric mixing ratios) were used. One user reported on their use of other satellite-based measurements of atmospheric composition nonetheless (GOSAT, IASI, OCO-2, GOME-2, TROPOMI), which is not included in the current report. One other user reported the use of LAI data, and this response has been included in Section 4.4. Finally, one user reported general plans to use satellite measurements to characterize the land surface for plume simulations in **Task 4.1**. This application appeared to be still at the planning stage, and only general specifications of the data need were provided: a spatial resolution on the order of meters was sought, with the best temporal resolution possible.

4.6.1 Near-real-time data needs

No near-real-time data needs were reported.

4.6.2 Needed for use in year of simulation

As the plume simulations are somewhat decoupled from the operational chain and are simulating scenes from the past, the data were not considered to be needed for the currently simulated year.

4.6.3 Needed only afterwards for model evaluation

The use of delayed data for model evaluation was foreseen.

4.6.4 Data from other years can be used

As the plume simulations focus on past years, this data use falls into this category.

4.7 Land cover maps

Ten respondents reported using land cover maps within the project, from WPs 2 through 5. Different respondents interpreted this question slightly differently, but some general themes emerged. Five respondents (from **Task 3.1**, **Task 3.3**, **Task 4.4**, and **Task 5.2**) reported making use of the ESA-CCI land cover product. Two respondents from **Task 4.4** described the built-in land surface schemes of their models (the MODIS land cover scheme distributed with WRF and the multilayer land surface scheme TERRA of ICON). Similarly, one respondent from **Task 4.1** mentioned that a landcover map was used in their LES simulations, without further specification. One user from **Task 2.1** reported using the SYNMAP product (Jung et al., 2006) but wanting to switch to the CGLS 100-m land cover map in the near future. The user from Task 5.2 reported using ECOCLIMAP in addition to the ESA-CCI product, and the respondent from **Task 3.2** was using the ECMWF urban cover map from the Copernicus SLIM project.

The spatial resolution required by users was often tied to the spatial resolution of the model used: in Task 3.3, users reported using the full 300-m resolution of the ESA-CCI product, in Task 5.2 resolutions from 300 m to 1 km were used. In Task 2.1, land cover maps are aggregated to the 1-km resolution of the flux product, taking into account the fractional land cover at higher resolution. The respondents carrying out national-scale inversions in Task 4.4 aggregated the land cover onto the resolution of their model, from 5 km to 13 km, and for the global IFS simulations in Task 3.1 the land cover map was aggregated to 9-km or 25-km resolution for forecast and analysis respectively. The urban cover map was used at 1-km resolution.

Temporally, the responses varied between those using essentially static maps (in Task 5.2, Task 4.4, Task 3.2 and currently for Task 2.1) and those who used annually updated fields (in Task 3.1 and 3.3).

The timeliness requirements for this data stream are somewhat different from most, as it is either annually (or less frequently) updated, or essentially static in time. Two respondents (one from Task 3.3 and one from Task 3.1) indicated that they needed a landcover map specific to the year they were simulating. Three other respondents (from Task 2.1, Task 3.3, and Task 4.1) indicated that they were updating the landcover only when new products became available, but not regularly. A further five respondents (from Task 3.2, Task 4.4, and Task 5.2) reported that they were using fixed maps, and the timeliness of the product was not critical.

Data users turned to different sources for accessing these maps, whether directly through the ESA landcover CCI website, in-house at ECMWF (for Task 3.2), through Meteo-France, the CDS, or C3S. The following factors limiting the use of these datasets were identified: limited temporal resolution (for reanalyses especially), uncertainties in converting land classes to model classes, temporal delays in the provision of the maps, and inflexible tools provided along with the datasets. One user commented that they themselves needed to be more proactive in regularly checking for updates.

4.8 Concentration fields from a global model

Nine respondents reported that they were making use of concentration fields from a global simulation in their task. These responses were dominated by the regional modelling activities WP4 (two from **Task 4.1**, one from **Task 4.3**, and five from **Task 4.4**) with one positive response from **Task 3.1**. As for the species needed, all but one respondent indicated that they were using CO_2 fields, and five indicated that they were using CH_4 fields. Five respondents were using some combination of co-emitted species, such as CO, NO_2, and other chemical tracers. Most were using products retrieved from ECMWF (via MARS) or CAMS, one indicated that they calculate their own global model fields or use NOAA's CarbonTracker product. One user from Task 4.1 indicated that they were using TROPOMI data for NO_2 and CO, which is not from a global model, but might provide lateral boundary conditions for plume modelling. The spatial resolutions used ranged from "the highest possible" to 3° x 2°, with temporal resolutions from hourly to 3-hourly. No users reported using uncertainties of any sort.

4.8.1 Near-real-time data needs

Only one survey respondent reported requiring these data in near-real time, for the evaluation of the modelling and data assimilation developments in the global CO2MVS within **Task 3.1**. Within this task, 3-hourly concentration fields at 9 to 30 km spatial resolution from the MARS archive were being used. The most limiting factor was found to be their temporal coverage.

4.8.2 Needed for use in year of simulation

All users but one from WP4 reported needing these data for the year actively being simulated, presumably as initial and lateral boundary conditions for regional simulations.

4.8.3 Needed only afterwards for model evaluation

One respondent from Task 4.1 reported using these data only for evaluation purposes, but also described using TROPOMI measurements for this purpose.

4.8.4 Data from other years can be used

No respondents reported use of this data stream from previous years.

5 Conclusion

Based on the responses that were supplied, a picture of the data needs within the project has emerged. While it is clear that much of the data is needed in a timely manner, only the respondent from Task 3.1 reported a need for near-real-time observations. This may however change over the course of the project.

The results are tabulated by Task and observation type in

Table 3, where the number of respondents per data type and task are listed to the right and bottom of the table. For each Task and data type the most stringent reported timeliness requirement is colour-coded, with red representing near-real-time need, orange for use in the currently simulated year, yellow for use in model evaluation, and green for parameter estimation, for which other years of data could be used. Where no colour is shown, no respondents reported using this data stream in a particular task.

Table 3: Summary of the timeliness requirements and number of positive responses by task. The number on the right indicates the total number of respondents who indicated using this data stream, and the number of respondents per task is listed at the bottom of the table. The colours indicate the timeliness requirements of the different data streams by task, with red for near-real-time, orange for use in the currently simulated year, yellow for use in evaluation (for which a time delay of a year is considered acceptable), and green for cases where the measurements are used only for parameter estimation, and data from another year could be used. When more than one timeliness requirement was listed per task and data stream, the most stringent requirement is shown.

	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 CO_2 mixing ratios		red			yellow	orange	orange		yellow	10
in situ CH_4 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 atmospheric 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 satellite-based measurements					yellow					1
land cover 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	

5.1 What's missing?

The most obvious thing that is lacking at this point is more responses. Part of this may be the result of how early in the project the exercise began – not all activities have begun yet, and some are still determining exactly which data they will be using. When asked for feedback on the survey, some users reported being unsure about exactly what data they will need or be able to use, pointing out that they may not need everything that they listed, but may well use additional data as well. There was some ambiguity as to whether the survey should be collecting what data are (ideally) needed, or what users currently have and are using. Moving the data collection to online interviews for the next year of the project can hopefully capture some of this nuance, and provide an improved picture of the use of observational data across the project. Furthermore, the reported data needs are expected to converge to reality over future versions of this deliverable (D7.2 and D7.3) as the project matures.

When asked about which data streams respondents thought should be more explicitly considered, suggestions included site-level soil moisture measurements (which was then added) and aircraft-based observations of atmospheric composition. While the latter is implicitly included in the discussion of *in situ* measurements of atmospheric mixing ratios of

CO_2 and CH_4, airborne measurements will be addressed separately in follow-up versions of this survey.

One respondent described a general need for co-sampled meteorology in conjunction with atmospheric composition measurements, and suggested that this needed to be included at more sites. Furthermore, measurement campaigns were identified as being useful to test models, and a need for measurements specifically focussing on vertical profiles and free-tropospheric values was identified. Uncertainty estimates, including covariances, on biogenic and anthropogenic emissions was found to be a missing component within the project, but this does not really constitute a measurement as such. Nonetheless, estimates of these uncertainties and covariances, at least for biogenic fluxes, may emerge from the benchmarking efforts of WP5, which are based on observations.

One aspect that is difficult to capture is the need for improved network coverage: while many respondents reported that limited data coverage was one of their main struggles in using the data, there was often little concrete feedback given as to how or where measurement networks could optimally expand. This will be addressed through targeted follow-up questions in the next version of the survey, and passed on to Task 7.3, which assesses gaps in the currently available observations.

Finally, one should keep in mind the tendency for some *in situ* data streams to be employed for validation or model evaluation rather than direct evaluation due to the delayed availability of these data, resulting in a sort of Catch-22: There is no identified need for these measurements in near-real-time because no one is using them in this way, but no one is using these measurements in this way because they are not available soon enough. Ground-based remote sensing measurements, such as those from TCCON, might be an example of this phenomenon.

Referring back to the project overview diagram in Figure 1, it is clear that some observations are also used in WP2 for the production of the flux or emission products that are then used in WPs 3-6. Indeed, perhaps it would be sensible to include another arrow from the “Observations” pillar to the WP2 box. At present only one participant in WP2 has responded to the survey, which means that some of the required datasets that are needed for the production of the prior fluxes are not included in this report. Over the next year of the project we will make an effort to reach out to colleagues in WP2 to ensure that their data needs are also documented, to avoid any gaps in provision of input during the transition from a scientific endeavour to an operational service.

This deliverable and its successors will serve as a basis for the identification of data providers in Task 7.2, which will then be reported upon in Deliverables 7.4 and 7.5. It will also provide guidance for the development of the prototype of the operational data pipeline in Task 7.4, resulting in Deliverable 7.8.

6 References

Jung, M, Henkel, K, Herold, M, Churkina, G (2006). Exploiting synergies of global land cover products for carbon cycle modeling. REMOTE SENSING OF ENVIRONMENT, 101(4), 534-553.

Mahadevan, P., Wofsy, S. C., Matross, D. M., Xiao, X., Dunn, A. L., Lin, J. C., Gerbig, C., Munger, J. W., Chow, V. Y., and Gottlieb, E. W. (2008), A satellite-based biosphere parameterization for net ecosystem CO_2 exchange: Vegetation Photosynthesis and Respiration Model (VPRM), *Global Biogeochem. Cycles*, 22, GB2005, doi:10.1029/2006GB002735.

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 af, doi: 10.2760/182790, European Commission Joint Research Centre, EUR 29817 EN.

7 List of acronyms

C3S	Copernicus Climate Change Service
CAMS	Copernicus Atmosphere Monitoring Service
CDS	Climate Data Store
CGLS	Copernicus Global Land Service
CHE	CO_2 Human Emissions project
CHTESSEL	Carbon-Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land
CIF	Community Inversion Framework - A modular system, developed in the VERIFY project, for estimating GHG fluxes.
CLMS	Copernicus Land Monitoring Service
CO2M	Copernicus Carbon Dioxide Monitoring mission
CO2MVS	CO_2 Monitoring and Verification Support capacity
CoCO2	Prototype system for a Copernicus CO_2 service
DLR	Deutsches Zentrum für Luft- und Raumfahrt, German Aerospace Center
ECMWF	European Centre for Medium-Range Weather Forecasts
ECOCLIMAP	A dual database at 1 km resolution that includes an ecosystem classification and a coherent set of land surface parameters
EEA	European Environment Agency
EM27	A solar absorption spectrometer
ERA5	An ECMWF reanalysis data product from 1979 to near real time
ESA	European Space Agency
ESA-CCI	ESA Climate Change Initiative
FLUXNET	1) The data portal and 2) measurement site network.
Fr-Tou	Flux tower site in Toulouse
GOME-2	Global Ozone Monitoring Experiment–2
GOSAT	Greenhouse gases Observing SATellite
GPP	Gross Primary Production
GRIB1, GRIB2	GRidded BInary data file formats
IASI	Infrared Atmospheric Sounding Interferometer
ICON	Icosahedral Nonhydrostatic Weather and Climate Model
ICOS	Integrated Carbon Observation System
ICOS ATC	ICOS Atmospheric Thematic Centre
IFS	Integrated Forecasting System - The atmospheric model and data assimilation system at ECMWF
IPCC	Intergovernmental Panel on Climate Change
LAI	Leaf area index
LES	Large Eddy Simulation, a mathematical model for turbulence used in computational fluid dynamics
MARS	Meteorological Archival and Retrieval System
MODIS	Moderate Resolution Imaging Spectroradiometer
MVS	Monitoring & Verification Support

NEE	Net Ecosystem Exchange - NPP minus the heterotrophic respiration
NOAA	U.S. National Ocean and Atmosphere Administration
NPP	Net Primary Production - GPP minus the autotrophic respiration
NRT	Near-Real-Time
NWP	Numerical Weather Production
ObsPack	Observation Package
OCO-2	Orbiting Carbon Observatory 2
OCS	Carbonyl sulfide
PI	Principal Investigator
PROBA-V	PROBA-Vegetation, a satellite in the European Space Agency's PROBA series
SIF	Solar-Induced Fluorescence
SLIM	Surface Land Information Mapping
SPOT-VEGETATION	Satellite pour l'Observation de la Terre
SYNMAP	A global land cover product
TCCON	Total Carbon Column Observing Network
TERRA	Land surface scheme
TOC	Top Of Canopy
TROPOMI	TROPOspheric Monitoring Instrument
VERIFY	Verifying greenhouse gas emissions project
VOD	Vegetation Optical Depth
VPRM	Vegetation Photosynthesis Respiration Model
WDCGG	World Data Centre for Greenhouse Gases
WMO	World Meteorological Organisation
WP	Work Package
WRF	Weather Research and Forecasting model
XCO2, XCH4, and XCO	Total column atmospheric measurements of CO_2, CH_4, and CO

Document History

Version	Author(s)	Date	Changes
0.1	Julia Marshall (DLR)	21/12/2021	First draft complete
1.0	Julia Marshall (DLR)	17/01/2021	Incorporation of comments and suggestions described below

Internal Review History

Internal Reviewers	Date	Comments
Elena Saltikoff, Werner Kutsch (ICOS-ERIC)	30/12/2021	Addition of list of acronyms, inserted table of responding tasks, addressed network coverage in the conclusions
Dominik Brunner and Erik Koene (EMPA)	03/01/2021	Mostly minor corrections (e.g. consistency in subscripts, spacing, how "*in situ*" is written throughout)
Paul Palmer (UEDIN)	06/01/2021	Minor stylistic corrections, some clarifications
Janne-Markus Rintala (ICOS-ERIC)	09/01/2021	Improvement of Executive Summary, explanation of abbreviations

Estimated Effort Contribution per Partner

Partner	Effort
DLR	2
Total	**2**

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.