Earth Syst. Sci. Data Discuss., 7, 495–519, 2014
www.earth-syst-sci-data-discuss.net/7/495/2014/
doi:10.5194/essdd-7-495-2014

Open Access Earth System Science Data Discussions

ObsPack: a framework for the preparation, delivery, and attribution of atmospheric greenhouse gas data

K. A. Masarie[1], W. Peters[2], A. R. Jacobson[1,3], and P. P. Tans[1]

[1]NOAA Earth System Research Laboratory, Boulder, Colorado, USA
[2]Department of Meteorology and Air Quality, Wageningen University, Wageningen, the Netherlands
[3]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado, USA

Received: 30 July 2014 – Accepted: 15 August 2014 – Published: 9 September 2014

Correspondence to: K. A. Masarie (kenneth.masarie@noaa.gov)

Published by Copernicus Publications.

Abstract

Observation Package (ObsPack) 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 (called "ObsPack products") are intended to support carbon cycle modeling studies and represent the next generation of value-added greenhouse gas observation products modeled after the cooperative GLOBALVIEW products introduced in 1996. Depending on intended use, ObsPack products may include data in their original form reformatted using the ObsPack framework or may contain derived data consisting of averages, subsets or smoothed representations of original data. All products include extensive ancillary information (metadata) intended to help ensure the data are used appropriately, their calibration and quality assurance history are clearly described, and that individuals responsible for the measurements (data providers or principal investigators (PIs)) are properly acknowledged for their work. ObsPack products are made freely available using a distribution strategy designed to improve communication between data providers and product users. The strategy includes a data usage policy that requires users to directly communicate with data providers and an automated e-mail notification system triggered when a product is accessed. ObsPack products will be assigned a unique Digital Object Identifier (DOI) to ensure each product can be unambiguously identified in scientific literature. Here we describe the ObsPack framework and its potential role in supporting the evolving needs of both data providers and product users.

1 Introduction

Long-term monitoring of carbon dioxide (CO_2) in the atmosphere has a two-fold purpose. First, it documents the gradual global build up of this important greenhouse gas

ESSDD
7, 495–519, 2014

due to human activities. To this end, one or two remote measurement sites would be sufficient. Second, measurements of CO_2 are used to facilitate a better understanding of the processes that control the global carbon cycle. For this purpose, researchers utilize the detailed temporal and spatial patterns in the concentration of CO_2 from observations at many locations. These patterns reflect the location and strength of major sources and sinks as modified by the action of transport and mixing in the atmosphere. An essential component of this research is the use of three-dimensional models of atmospheric transport and mixing to translate the observed CO_2 patterns into surface sources and sinks. An ongoing challenge to this approach is the sparseness of observations even when we limit questions to spatial scales as large as the entire expanse of the boreal forest ecosystems. A global CO_2 monitoring system with unprecedented spatial resolution and temporal continuity is needed. Since both the costs and logistics of operating such a global system prove too formidable a task for a single organization, a coordinated international effort is required.

In 1996, the National Oceanic and Atmospheric Administration (NOAA), Earth System Research Laboratory (ESRL), Global Monitoring Division (formerly Climate Monitoring and Diagnostics Laboratory) coordinated an international effort to bring together as many measurements as possible from different laboratories with careful attention to direct compatibility with respect to the calibration and methodology. The resulting GLOBALVIEW data products (http://www.esrl.noaa.gov/gmd/ccgg/globalview/) were introduced at a time when carbon cycle data assimilation systems could not yet accommodate data with temporal gaps and spatial irregularities. To address this limitation, GLOBALVIEW products included "extended" atmospheric CO_2 records that were synchronized, smoothed, and gap-free (Masarie and Tans, 1995). The annually-updated GLOBALVIEW products have been used in many carbon cycle modelling studies since first introduced in 1996 (e.g., Maksyutov et al., 2013; Niwa et al., 2012; Bruhwiler et al., 2011; Nevison et al., 2008; Gurney et al., 2002; Gloor et al., 2000).

In 2007, demand for GLOBALVIEW started to wane as modelers began using new assimilation strategies that could utilize actual observations and accommodate the tem-

ESSDD
7, 495–519, 2014

poral and spatial irregularities often found in measurement records. Modelers were now in need of new products that included actual measurements. Since none yet existed, many began collecting data directly from individual laboratories and data centers. At the same time, measurement PIs found themselves receiving regular requests by a growing number of modelers for access to more up-to-date data and metadata. Thus, there emerged a demand for a new generation of data products that could meet the needs of both data users and data providers.

The work presented here builds upon the existing international effort to prepare cooperative data products. In addition to improving data coverage, accessibility, and usability, ObsPack development also includes the goal of improving communication between the large group of researchers using data products and the much smaller group of researchers contributing original data to these products. This is accomplished in several ways. First, the ObsPack framework presents measurements or data derived from these original measurements along with extensive metadata. Metadata are ancillary information about the measurements which give users a better understanding and appreciation of the data. Metadata include estimates of measurement uncertainty and representativeness, a summary of calibration and quality assurance history, references to relevant literature, and detailed contact information for those laboratories and individuals responsible for creating and providing the original data. Second, the ObsPack framework introduces a data usage policy that requires product users to directly contact data providers to discuss the nature of the work and the appropriate level of acknowledgement. Lastly, the product delivery system employs an automated e-mail notification strategy whereby a user (via e-mail) is "introduced", albeit indirectly, to providers when an ObsPack product is downloaded.

The ObsPack framework including strategies for data delivery, attribution, and registration is fully developed and operational. Several ObsPack products are already freely available at http://www.esrl.noaa.gov/gmd/ccgg/obspack/. Ongoing feedback from product users and data providers helps us identify what is working well and what requires improvement. We believe the distribution of data and metadata using the Ob-

ESSDD
7, 495–519, 2014

sPack framework is a considerable improvement over methods currently in use at ESRL and elsewhere. ESRL is exploring the possible use of the ObsPack framework for distribution of all its greenhouse gas measurement records.

2 Product description

ObsPack itself is not a product. A product prepared using the ObsPack framework is called an ObsPack product. The first ObsPack product (called PROTOTYPE) was released in 2012. The PROTOTYPE product includes actual data as well as derived data (averages) specifically prepared for the CarbonTracker CO_2 data assimilation system (Peters et al., 2007, with updates described at http://carbontracker.noaa.gov). Each product includes a collection of data sets (defined in Sect. 2.1) and a set of product summary files. The current PROTOTYPE product includes 190 data sets with contributions from 20 laboratories. Also in 2012, we updated the long time cooperative data product GLOBALVIEW-CO_2 using the ObsPack framework. The current release of GLOBALVIEW-CO_2 includes 336 data sets from 23 laboratories. GLOBALVIEW products consist of smooth representations of original data and not the original data themselves.

While some ObsPack products may contain actual data, these products are not the primary source of original data. Rather, the data provider remains the primary source. Furthermore, while each data set includes extensive metadata, the included metadata are not comprehensive. Original data and complete documentation should be obtained directly from the data provider using the contact information included with each data set. Original data, often reformatted, may also be available from the World Meteorological Organization (WMO) World Data Center for Greenhouse Gases (WDCGG) or from other data centers.

ESSDD
7, 495–519, 2014

2.1 Data sets

The primary component of every ObsPack product is the collection of data sets. A data set contains original or derived mole fraction (a less precise word is "concentration") data from one or more providers for a single trace gas species (e.g., CO_2 or methane (CH_4)), sample location, sampling project, and selection strategy. Sampling project describes the sampling platform (e.g., surface, aircraft, tower, or ship) and sampling strategy (e.g., flask or in situ) separated by a hyphen (e.g., surface-flask and aircraft-insitu). The selection strategy indicates how the original data have been prepared. Original data are selected to be consistent with the product's overall intended use. For example, each data set in the PROTOTYPE product has been selected to be as representative as possible of large well-mixed air masses to give CarbonTracker the best opportunity to derive estimates of CO_2 surface fluxes. Selection constraints may be based on time-of-day, sampling intake height, origin of sampled air, and measurement variability. Typically, PIs apply selection criteria to the original data and identify the outcome by including a flag with each original datum. If needed, individuals preparing the ObsPack data sets will apply additional selection criteria to the original data to ensure the derived data support the product's intended use. Each data set is uniquely named using the following structure:

<trace gas identifier>_<site code>_<project>_<lab number>_<selection tag>

where lab number (http://www.esrl.noaa.gov/gmd/ccgg/obspack/labinfo.html) identifies the laboratory providing the original data.

Each ObsPack product is prepared using a set of configuration files that inform the ObsPack software on how to read original data and prepare data sets for each sampling project. Configuration files do not contain executable code but instructions on which codes should be called and how. There is typically one configuration file for each contributing laboratory but sometimes several files for a single laboratory. ESRL, for example, operates several different sampling projects (e.g., surface-insitu, tower-insitu,

ESSDD
7, 495–519, 2014

surface-flask, and aircraft-flask) and each requires a separate configuration file. Table 1 shows an excerpt from an ESRL tower-insitu project configuration file. This excerpt instructs the software to prepare three tower data sets at day resolution using original data. The Argyle, Maine (AMT) data set is prepared using afternoon hourly average values from the 107 m above ground level (m a.g.l.) intake height beginning in 2003. Two Walnut Grove, California (WGC) data sets are prepared; one using afternoon hours from the 483 m a.g.l. intake height and one using nighttime hourly values from the same intake height. Each data set is presented as a single netCDF file (an ASCII text file is an optional output format). The three data sets from Table 1 will have the netCDF file names co2_amt_tower-insitu_1_afternoon-107magl.nc, co2_wgc_tower-insitu_1_afternoon-483magl.nc, and co2_wgc_tower-insitu_1_nighttime-483magl.nc.

Each data set also includes extensive metadata describing the source of the original data, how the original or derived data were prepared, and general information about the product in which the data set resides. Metadata describing the data set include the location of the sampling site, sampling strategy, selection strategy, calibration and quality assurance history, contributing laboratories, programs, partners, and principal investigators (data providers). Additional metadata fields describe the overall ObsPack product including the product name, description, creation date, caveats, fair use policy, required citation, and contact information for those who prepared the data set. A partial list of these metadata fields is shown in Table 2.

Additional metadata define the variables reported with each datum. These metadata fields include sample collection time and position, reported value and uncertainty, intake height above ground, selection flag, and unique ObsPack number and ID. The ObsPack number is a unique index number assigned to each datum across all data sets in the ObsPack product. The ObsPack ID is a unique identification string that distinguishes a datum from all other data across all ObsPack data products (see Table 3). All data sets include a common set of reported variables. The list of variables depends on the product itself. Table 3 shows the list of common variables from the current PRO-

ESSDD
7, 495–519, 2014

TOTYPE product. The graph in Fig. 1 includes data and metadata from the single data set, co2_wgc_tower-insitu_1_afternoon-483magl.

Occasionally, data sets may include additional variables not available for all data sets in the product. For example, the Comprehensive Observation Network for Trace gases by AIrLiner (CONTRAIL) sampling program (http://www.cger.nies.go.jp/contrail) operated jointly by the National Institute for Environmental Studies (NIES) and the Meteorological Research Institute (MRI) in Japan additionally includes with each reported value, the ambient pressure, which represents the vertical coordinate of the sample collection position. The ObsPack framework supports inclusion of additional metadata fields as they are needed or become available.

ObsPack metadata are not yet fully compliant with the International Organization for Standardization (ISO; http://www.iso.org/iso/) because many ObsPack attributes have not yet been defined in those standards. Because ISO-compliant metadata improves both data discovery and usability, we continue to work towards full ISO-compliance as it becomes clear how to match ObsPack metadata fields with the recommended standards.

2.2 Ancillary files

Each ObsPack product includes several additional files that provide a quick overview of its contents. This summary information includes a listing of all data sets, an e-mail distribution list of all product contributors, an HTML file listing all data providers and their affiliations, the required citation, and the set of configuration files describing which original data are used and how they are prepared for each laboratory's sampling project.

ESSDD
7, 495–519, 2014

2.3 Product name

Each ObsPack product has a unique name using the following structure:

obspack_<trace gas identifier>_<lab number>_<product name>_
<product version number>_<preparation date>.

Here, lab number identifies the laboratory that prepared the overall ObsPack product. The version numbering scheme is major.minor[.minor] where a major release is indicated by the first number in the sequence and minor revisions are indicated by the second and third (optional) numbers in the sequence. The latest minor revision of a major release includes all changes included in intermediate minor revisions if they exist. For example, obspack_co2_1_PROTOTYPE_v1.0.0_2012-11-05 was the first major release of the PROTOTYPE product. Obspack_co2_1_PROTOTYPE_v1.0.1_2012-12-10 was a minor revision made available on 10 December 2012.

2.4 Programming languages and tools

Several programming languages support the ObsPack framework. Python (http://www.python.org) is used to read original data, prepare data sets, create summary files, and package the product. ObsPack metadata are managed using a MySQL database (http://www.mysql.com). These two open-source languages suffice for the creation of ObsPack products on any platform. ESRL additionally uses IDL (http://www.exelisvis.com) for product quality control, and Perl (http://www.perl.org) to interface with the database and compile product usage statistics. Data sets are presented using the netCDF4 data file protocol (http://www.unidata.ucar.edu). ASCII text files are optionally available and generated directly from the netCDF files. The list of variables included in text files is fixed to ensure consistency and readability, and as a result, may not include all variables available in the netCDF files. All files are packed using both the zip and tar.gz protocols.

ESSDD
7, 495–519, 2014

3 Product distribution

ESRL ObsPack products are available and distributed from a single web portal (http://www.esrl.noaa.gov/gmd/ccgg/obspack). Users are required to provide their name, e-mail address, affiliation, and intended use of the data before accessing a product. Requiring registration is a departure from the earlier GLOBALVIEW product distribution strategy, which was anonymous. Still, ObsPack products are available freely and without delay. Upon product download, automated e-mails are sent to the user and to the data providers. E-mail sent to the user includes the data usage statement, required citation, the e-mail list of all contributors to the product, and the URL link to the requested product. E-mail to data providers summarizes the user registration and product download information. The automated e-mails establish indirect communication between user and providers, and give providers the ability to track use of their data from ObsPack products. Registration information is saved to an ObsPack database so that users can be notified when products are revised or updated.

4 Product DOI

Recent ObsPack products prepared by ESRL have been assigned an internationally-recognized digital object identifier (DOI) through DataCite (http://www.datacite.org). A DOI is a sequence of characters that uniquely identify an electronic document, which in this context, is an ObsPack data product. Using DOIs has several important benefits. By including the DOI in an ObsPack product citation, researchers provide an unambiguous link to the version of the product used in their study. Citation services (e.g., Thomson Reuters' Web of Science) that provide statistics on published paper citations are now adapting their tools to include data and data products. DataCite member organizations responsible for administering DOIs typically require the submitting authors to provide metadata that meets ISO standards and to archive the product at a recognized

ESSDD
7, 495–519, 2014

data center. These requirements ensure that products are preserved, accessible, and usable for all time. The use of DOIs complements the ObsPack distribution strategy.

The DOI string includes a prefix and a suffix separated by a forward slash (/). The prefix identifies the organization registering the document. The suffix identifies the registered document. The suffix may appear as a random sequence of characters or may be readable and provide some indication of the document's content. The DOI is often embedded in a URL to create a direct link to the product via the service providing support for the DOI (e.g., doi:10.3334/OBSPACK/1001). The link loads the landing page which includes product documentation and direct access to the product. The URL may change over time; the DOI is permanent.

Each ObsPack product includes a citation which is required when used in any publication or presentation. The citation includes the product name, version, publication date and DOI. When a revision to an existing product is made available, the publication date will reflect the update but the original DOI will be preserved. We cannot rule out the possibility that a revision may require a new DOI, but this has not yet occurred.

5 Product fair use statement

Each product includes the ObsPack Fair Use Statement summarizing the user's responsibilities to data providers. First, the user is required to contact all providers contributing to the requested product. This requirement ensures that the user discusses with data providers, in advance, how measurements are to be acknowledged and what level of collaboration is appropriate. In addition, this requirement stimulates communication between product user and data provider, which is beneficial to both. Users gain a better understanding of the data and their limitations, and providers learn how their data are being used and what additional information may be helpful to improve usability. Second, the user must reference the product in publications and presentations using the provided required citation, which includes the DOI. Again, the DOI ensures that the product can be unambiguously identified in the literature. To help users meet these

ESSDD
7, 495–519, 2014

requirements, users receive upon download, an automated e-mail which includes the Fair Use Statement, the citation, and the list of e-mail addresses of all providers.

The Fair Use Statement is:

> These cooperative data products are made freely available to the scientific community and are intended to stimulate and support carbon cycle modeling studies. We rely on the ethics and integrity of the user to assure that each contributing national and university laboratory receives fair credit for their work. Fair credit will depend on the nature of the work and the requirements of the institutions involved. Your use of an ObsPack data product implies an agreement to contact each contributing laboratory to discuss the nature of the work and the appropriate level of acknowledgement. If an ObsPack data product is essential to the work, or if an important result or conclusion depends on an ObsPack product, co-authorship may be appropriate. This should be discussed with each data provider at an early stage in the work. Contacting the data providers is not optional; if you use an ObsPack data product, you must contact the data providers. To help you meet your obligation, each data product includes an e-mail distribution list of all data providers. ObsPack data products must be obtained directly from the ObsPack Data Portal at www.esrl.noaa.gov/gmd/ccgg/obspack/ and may not be re-distributed.
>
> Beginning November 2013, all new ObsPack data products have a unique Digital Object Identifier (DOI) registered with the International DOI Foundation. In addition to the conditions of fair use as stated above, users must also include the ObsPack product citation in any publication or presentation using the product. The required citation is included in every data product and in the automated e-mail sent to the user during product download.

The above usage statement is considerably different from that of the earlier GLOBALVIEW data products, which only required the user to cite the product itself when

ESSDD
7, 495–519, 2014

used in publications and presentations. Several users have provided feedback on the new ObsPack distribution strategy and attribution requirements. Some have already acknowledged that the policy does, in fact, improve communication and collaboration with data providers. Others have found the attribution requirements confusing. Users are discovering that data usage policies differ among the contributing laboratories, and the expectation of co-authorship varies widely. Since there is often no clear way to assess the relative contribution of one data set over another in a study, the list of co-authors providing observations may appear somewhat arbitrary. Our aim is to arrive at a fair use policy that is easy to implement and benefits users and providers alike. We expect refinements to the policy as both users and providers gain experience with these new requirements.

6 Product updates

For nearly 20 years, ESRL has coordinated the effort to produce annually-updated multi-laboratory cooperative data products. The first product, GLOBALVIEW-CO2, was introduced in 1996. In 2012, we updated GLOBALVIEW-CO2 and introduced the PROTOTYPE product both using the ObsPack framework. From 1996 to 2012, GLOBALVIEW-CO_2 was downloaded ~ 17 000 times by users from more than 70 countries. Since 2012, ObsPack products have been downloaded ~ 500 times. ESRL is currently developing additional ObsPack products for other (non-)greenhouse gas measurements including methane, carbon monoxide, sulfur hexafluoride, nitrous oxide, and the stable isotopes of CO_2.

ObsPack products are generally updated in the fall and derived using data through December of the preceding year. Several months are required to acquire new measurement records, receive updates to existing records, verify metadata, and prepare and quality control the product before release. Occasionally, products need to be revised after initial release. While we intend to keep revisions to a minimum, the larger number

ESSDD
7, 495–519, 2014

of contributors and data sets increases the likelihood of finding errors. A revision may correct data, metadata, or both.

To prepare new products and product updates, ESRL first surveys the WMO WDCGG for up-to-date data before sending an invitation to participate directly to data providers. Direct communication between ESRL and the participating laboratories has helped foster many productive and long-term collaborations. Laboratories, including ESRL, recognize the value of including their measurements in these types of cooperative data products as it increases visibility, use, and often leads to fruitful collaboration with product users.

ESRL is now receiving regular requests for more frequent product updates that would include more up-to-date observations. We expect this trend to continue. For some laboratories, making available near real-time measurements that have been properly quality-controlled by principal investigators is not feasible given existing resources. Other laboratories already make near real-time data available but with the caveat that data are provisional and will likely be revised. Thus, more frequent product updates would likely include some data sets using preliminary measurements with increased uncertainty and some data sets that are unchanged.

6.1 Complementary ObsPack products

While preparing the CT2013 update of CarbonTracker CO_2 (http://www.esrl.noaa.gov/gmd/ccgg/carbontracker/), ESRL developed the ability to utilize multiple ObsPack products during assimilation. This capability enables us to think of ObsPack products in a new way. Instead of creating a single product that includes data from many laboratories, we can now consider a suite of complementary ObsPack products each containing similarly prepared data from a single laboratory or subset of laboratories. Further, laboratories other than ESRL could prepare complementary ObsPack products provided they adhere to the standards of the ObsPack framework. The ObsPack software is freely available to assist laboratories with this process (http://www.esrl.noaa.gov/gmd/ccgg/obspack/). Smaller complementary products would likely improve our ability to

ESSDD
7, 495–519, 2014

create product revisions and update products using near real-time data (where available) without impacting data sets or products whose original data remain unchanged.

A collection of ObsPack products can only be complementary if data are prepared in a consistent manner, and content and structure are fully compatible. ESRL has configured a development server with versioning and revision control software to promote community development of the ObsPack framework. In this way, improvements to the ObsPack framework can be under version control and easily shared with ObsPack product users, data contributors, and developers. Complementarity is a critical issue that must be formalized by the communities generating ObsPack products and developing the framework.

It is worth noting that the product distribution strategy described here is used for products created by ESRL. Other labs may choose to distribute products themselves using a similar or different strategy.

7 Future developments

ESRL will continue to develop the ObsPack framework and refine the data delivery and attribution strategies. We are currently working with partner laboratories to convey in more consistent terms the meaning of the reported estimated uncertainty of original and derived data and the description of what measured air samples represent.

Uncertainty in original data includes measurement repeatability, systematic and random uncertainties, and in some instances, a component of atmospheric variability. Derived data may have an additional uncertainty term depending on how the original data are prepared. We would like to include these estimates of component uncertainty along with overall uncertainty for each reported data value included in a data set.

Measurement PIs often consider factors such as wind direction, measurement variability, and measurements of other atmospheric constituents when identifying a measurement as being representative of certain conditions. For instance, reported values can be characterized as representing marine or continental air masses, as having

ESSDD
7, 495–519, 2014

hemispheric, continental, or regional origin, or as being representative of a local CO_2 source. These categories vary by measurement site and laboratory, and we are working to develop a consistent way for data providers to communicate this information to users.

8 Conclusions

ObsPack products benefit both product users and data providers. The ObsPack framework is designed to aggregate atmospheric greenhouse gas observations from a variety of sampling platforms, prepare them with specific uses in mind, and package and distribute them in a self-consistent and well-documented product. ObsPack products are intended to stimulate and support carbon cycle modeling studies. The ObsPack product delivery and attribution strategies are designed to ensure data owners receive proper credit for their contribution of original data and to improve communication between the product user and the data providers.

Today, atmospheric greenhouse gas observations and products critical to carbon cycle science are vast, varied, and available from many different data distribution web sites.

- The WMO WDCGG in Tokyo, for example, is the designated data center for WMO Global Atmosphere Watch (GAW) high-precision atmospheric greenhouse gas measurements (http://ds.data.jma.go.jp/gmd/wdcgg/). WDCGG data holdings include contributions from 92 laboratories and 110 trace gas species.
- Several programs contributing original data to the WDCGG are large and, in practice, act as data centers themselves. ESRL, for example, makes its primary data and products available directly from the ESRL web site (http://www.esrl.noaa.gov/gmd/dv/data/).
- The Integrated Carbon Observation System (ICOS; http://www.icos-infrastructure.eu/), currently under development in Europe, is a multi-national

ESSDD
7, 495–519, 2014

program designed to make internally-consistent, high-precision atmospheric measurements from a large European observing network. ICOS will also likely act as a data center.

- Several additional programs, which produce valuable observations are not part of GAW and distribute original data from other centers. For example, the HIAPER Pole-to-Pole Observations (HIPPO; http://hippo.ornl.gov/) study has produced high-precision measurements of atmospheric greenhouse gas and related trace gas species from aircraft flying pole-to-pole over the Pacific Ocean. HIPPO data are distributed by the Carbon Dioxide Information Analysis Center (CDIAC) operated by the United States Department of Energy, Oak Ridge National Laboratory. CDIAC is also a designated World Data Center for Atmospheric Trace Gases.

- The Total Carbon Column Observing Network (TCCON; http://www.tccon.caltech.edu/), another multi-national program, makes ongoing total column measurements from a global network of ground-based Fourier Transform Spectrometers. TCCON data are distributed by the California Institute of Technology.

- Satellite retrieval products from the Greenhouse Gases Observing Satellite (GOSAT) are distributed by the National Institute for Environmental Studies (NIES) in Japan (http://www.gosat.nies.go.jp/eng/gosat/page5.htm).

While we may wish that these diverse greenhouse gas data holdings could be consistently formatted and accessed from a single portal, it is not likely to happen nor is it truly necessary. Well-documented data adhering to ISO metadata standards make a distributed data system viable. ISO-compliant metadata improves both data discovery and usability. Improved data discovery makes it possible for many web sites to provide what appears to a user as one-stop-shopping for atmospheric greenhouse gas observations but what is, in fact, a portal to distributed data archives. In the near future, our best strategy for improving visibility and usability of greenhouse gas observations

ESSDD
7, 495–519, 2014

may be to encourage designated world data centers and institutions acting as data centers to continue their efforts to serve metadata that are fully ISO-compliant.

Acknowledgements. The authors are indebted to the many individuals who have contributed original data to the ObsPack project. They are P. Steele, P. Krummel, R. Langenfelds, and M. van der Schoot (Commonwealth Scientific and Industrial Research Organization (CSIRO), Australia); L. Gatti (Instituto de Pesquisas Energéticas e Nucleares (IPEN), Brazil); E. Gloor (Leeds University, UK); D. Worthy (Environment Canada (EC), Canada); J. Hatakka and T. Aalto (Finnish Meteorological Institute (FMI), Finland); M. Ramonet, M. Delmotte, M. Schmidt, and P. Ciais (Laboratoire des Sciences du Climat et de l'Environnement – UMR8212 CEA-CNRS-UVSQ (LSCE), France); I. Levin and S. Hammer (University of Heidelberg, Institut für Umweltphysik (UHEI-IUP), Germany); L. Haszpra (Hungarian Meteorological Service (HMS), Hungary); A. di Diodato, M. Alemanno, L. Lauria, and A. Proietti (Italian Air Force Meteorological Service (IAFMS), Italy); K. Tsuboi, H. Koide, and M. Takahashi (Japan Meteorological Agency (JMA), Japan); T. Machida and H. Mukai (National Institute for Environmental Studies (NIES), Japan); H. Matsueda and Y. Sawa (Meteorological Research Institute (MRI), Japan); S. Morimoto, T. Nakazawa, and S. Aoki (National Institute of Polar Research (NIPR) and Tohoku University, Japan); H. Meijer (University of Groningen, Centre for Isotope Research (RUG), the Netherlands); L. Zhou and S. Fang (Chinese Academy of Meteorological Sciences (CMA), Peoples Republic of China); K. R. Kim, M. Park, S. Park, and J. Kim (Seoul National University/School of Earth and Environmental Sciences (SNU), Republic of Korea); E. Brunke and C. Labushagne (South African Weather Service (SAWS), South Africa); A. Gomez-Pelaez, R. Ramos, V. Gomez-Trueba, and R. Campo-Hernandez (Izaña Atmospheric Research Center, Meteorological State Agency of Spain (AEMET), Spain); M. Steinbacher (Swiss Federal Laboratories for Materials Science and Technology (EMPA), Switzerland); M. Leuenberger (University of Bern, Physics Institute, Climate and Environmental Physics (KUP), Switzerland); S. Wofsy (Harvard University (HU), United States); S. Biraud and M. Torn (Lawrence Berkeley National Laboratory (LBNL), United States); A. Andrews, E. Dlugokencky, J. B. Miller, C. Sweeney, and K. Thoning (Global Monitoring Division (NOAA), United States); B. Stephens (National Center For Atmospheric Research (NCAR), United States); R. Keeling (Scripps Institution of Oceanography (SIO), United States).

ESSDD
7, 495–519, 2014

References

Bruhwiler, L. M. P., Michalak, A. M., and Tans, P. P.: Spatial and temporal resolution of carbon flux estimates for 1983–2002, Biogeosciences, 8, 1309–1331, doi:10.5194/bg-8-1309-2011, 2011.

Gloor, M., Fan, S.-M., Pacala, S., and Sarmiento, J.: Optimal sampling of the atmosphere for purpose of inverse modeling: a model study, Global Biogeochem. Cy., 14, 407–428, doi:10.1029/1999GB900052, 2000.

Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L. M. P., Chen, Y.-H., Ciais, P., Fan, S., Fung, I. Y., Gloor, M., Heimann, M., Higuchi, K., John, J., Maki, T., Maksyutov, S., Masarie, K. A., Peylin, P., Prather, M., Pak, B. C., Rangerson, J., Sarmiento, J., Taguchi, S.,Takahashi, T., and Yuen, C.-W.: Towards robust regional estimates of CO2 sources and sinks using atmospheric transport models, Nature, 415, 626–630, doi:10.1038/415626a, 2002.

Maksyutov, S., Takagi, H., Valsala, V. K., Saito, M., Oda, T., Saeki, T., Belikov, D. A., Saito, R., Ito, A., Yoshida, Y., Morino, I., Uchino, O., Andres, R. J., and Yokota, T.: Regional CO_2 flux estimates for 2009–2010 based on GOSAT and ground-based CO_2 observations, Atmos. Chem. Phys., 13, 9351–9373, doi:10.5194/acp-13-9351-2013, 2013.

Masarie, K. A. and Tans, P. P.: Extension and integration of atmospheric carbon dioxide data into a globally consistent measurement record, J. Geophys. Res.-Atmos., 100, 11593–11610, 1995.

Nevison, C. D., Mahowald, N. M., Doney, S. C., Lima, I. D., van der Werf, G. R., Randerson, J. T., Baker, D. F., Kasibhatla, P., and McKinley, G. A.: Contribution of ocean, fossil fuel, land biosphere, and biomass burning carbon fluxes to seasonal and interannual variability in atmospheric CO_2, J. Geophys. Res., 113, G01010, doi:10.1029/2007JG000408, 2008.

Niwa, Y., Machida, T., Sawa, Y., Matsueda, H., Schuck, T. J., Brenninkmeijer, C. A. M., Imasu, R., and Satoh, M.: Imposing strong constraints on tropical terrestrial CO_2 fluxes using passenger aircraft based measurements, J. Geophys. Res., 117, D11303, doi:10.1029/2012JD017474, 2012.

Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A., Conway, T. J., Masarie, K. A., Miller, J. B., Bruhwiler, L., Petron, G., Hirsch, A., Worthy, D., van der Werf, G., Randerson, J. T., Wennberg, P., Krol, M., and Tans, P. P.: An atmospheric perspective on North

ESSDD
7, 495–519, 2014

American carbon dioxide exchange: CarbonTracker, P. Natl. Acad. Sci. USA, 104, 18925–18930, doi:10.1073/pnas.0708986104, 2007.

ESSDD
7, 495–519, 2014

Table 1. Excerpt from an ESRL tower-insitu configuration file.

Name	Value
project.lab.abbr	noaa
project.lab.comparison_activity	Routine direct air comparisons at many locations
project.project	tower-insitu
project.data.frequency	1
project.data.frequency.unit	hour
project.parse.function	noaa_tower_insitu
project.parameter.abbr	co2
project.calibration.scale	WMO (prior to 2004); WMOX2007 (2004 to present)
site.amt.1.time.window.lst	12–16
site.amt.1.intake_ht	107
site.amt.1.startdate	2003
site.amt.1.selection	afternoon hours from 107 m a.g.l. (highest) intake
site.amt.1.selection_tag	afternoon-107magl
site.wgc.1.time.window.lst	12–16
site.wgc.1.intake_ht	483
site.wgc.1.selection	afternoon hours from 483 m a.g.l. (highest) intake
site.wgc.1.selection_tag	afternoon-483magl
site.wgc.2.time.window.lst	0–4
site.wgc.2.intake_ht	483
site.wgc.2.selection	nighttime hours from 483 m a.g.l. (highest) intake
site.wgc.2.selection_tag	nighttime-483magl

ESSDD
7, 495–519, 2014

Table 2. Partial list of metadata fields included in an ObsPack data set.

Name	Description
site_ fields provide information about the sampling location.	
site_code	Site code as defined by GAWSIS (gaw.empa.ch/gawsis/)
site_name	Site name (e.g., Park Falls, Wisconsin)
site_country, site_country_flag	Country in which site is located, link to country flag image
site_latitude, site_longitude	Latitude, longitude (decimal degree) of site location
site_elevation	Ground or surface elevation at site location
site_utc2lst	Hour conversion from UTC to LST
site_map	URL link to world map highlighting site location image.
site_url	URL link to site web page
dataset_ fields provide information about the data set. Some fields inform on how to prepare the data set.	
dataset_num	Unique integer identifies the data set within the ObsPack product.
dataset_name	Unique string identifies the data set within the ObsPack product.
dataset_globalview_prefix	String equivalent of GLOBALVIEW file name prefix.
dataset_parameter	Identifies trace gas species included in data set (e.g., CO_2, CH_4)
dataset_process	Original data preparation descriptor (e.g., PassThru, TimeStepAverage)
dataset_project	Sampling platform and strategy (e.g., surface-flask, tower-insitu)
dataset_time_window_lst	Attribute set when necessary to subset original data by sample collection time (LST) (internal use only).
dataset_parse_function	Python module used to read original data (internal use only).
dataset_data_frequency	Measurement frequency of original data.
dataset_data_frequency_unit	Indicates time unit of the dataset_data_frequency attribute.
dataset_platform	Fixed or Mobile.
dataset_start_date	Date of first item in data set (ISO 8601 format).
dataset_stop_date	Data of last item in data set (ISO 8601 format).
dataset_selection	Brief description of how data have been selected by data contributor or prepared by product author.
dataset_selection_tag	Short descriptor to help convey how data have been selected by data contributor or prepared by product author. The selection tag is included in the data set name.
dataset_calibration_scale	Measurements are relative to reported calibration scale.
dataset_reference_number	Number indicating how many references to published literature to expect in this file.
dataset_reference_#_name	Reference provided by data contributor. # is a number from 1 to relative dataset_reference_number.
lab_ fields provide information about the laboratories responsible for providing the original data.	
lab_number	Number of laboratories listed in the file.
lab_#_num	Laboratory identification number
lab_#_name, lab_#_abbr	Laboratory name and abbreviation or acronym
lab_#_address, lab_#_country, lab_#_country_flag, lab_#_url, lab_#_logo	Additional laboratory information fields
lab_ongoing_atmospheric_air_comparison	If "yes", lab contributing original data participates in at least one ongoing direct atmospheric air comparison experiment.
lab_comparison_activity	Brief description of measurement comparison activities.

ESSDD
7, 495–519, 2014

Table 2. Partial list of metadata fields included in an ObsPack data set.

Name	Description
Original data may also be associated with larger program entities. program_ fields provide information about overarching programs if they exist. program_ fields are identical to lab_ fields.	
program_number	Number of programs listed in the file.
provider_ fields provide information about individuals responsible for the original data (e.g., PIs).	
provider_number	Number of providers listed in the file.
provider_#_name	Name of provider.
provider_#_address, provider_#_email	Contact information.
provider_#_organization provider_#_organization_abbr	Organization name and acronym most directly associated with provider.
Partners are individuals or agencies that provide critical infrastructure or technical support for the original data record. If a partner is an individual, the partner_ fields are identical to provider_fields otherwise they are identical to lab_fields.	
partner_number	Number of partners listed in the file.
obspack_ fields provide information about the creation of the ObsPack product.	
obspack_contact_name [_lab, _email]	Contact information of ObsPack preparer.
obspack_data_time_step	Time interval at which ObsPack data are presented (e.g., day, hour).
obspack_name	Unique ObsPack identification string.
obspack_description	Brief description of data product contents.
obspack_creation_date	Date ObsPack data product was prepared.
obspack_citation	Required ObsPack citation. This citation is in addition to the requirements of the ObsPack Fair Use statements.
obspack_warning	Important caveats of the data product.
obspack_fair_use	Complete ObsPack Fair Use Statement.

ESSDD
7, 495–519, 2014

Table 3. Partial list of variables included in an ObsPack data set.

Variable Name	Description
obs_num	Unique observation number within a single data set.
obs_id	Unique identification string that distinguishes the datum from all other data in the ObsPack data product. It includes dataset_name and obs_num.
obspack_num	Unique observation index number across all data sets in the ObsPack distribution.
obspack_id	Unique identification string that distinguishes the data item from all other data items in any ObsPack data product. It includes obspack_name, dataset_name, and obspack_num delimited by a tilde (~).
time	Air sample collection time (UTC). POSIX time (number of seconds since 1 Jan 1970 in UTC).
time_decimal	Air sample collection time (UTC) in decimal year notation (e.g., 2012.4523312).
time_components	Air sample collection time (UTC) represented as a 6-element array [year, month, day, hour, minute, second].
solartime_components	Air sample collection time (solar time) represented as a 6-element array [year, month, day, hour, minute, second]. UTC time is converted to local solar time based on longitude and day-of-year.
value	Reported mole fraction or isotope ratio. Units depend on trace gas species.
value_unc	Standard deviation of the reported mean value when nvalue is greater than 1.
nvalue	Number of individual measurements used to compute reported value.
latitude	Latitude at which air sample was collected (units: decimal degrees).
longitude	Longitude at which air sample was collected (units: decimal degrees).
altitude	Altitude (surface elevation plus sample intake height) at which air sample was collected. Units are meters above sea level (m a.s.l.).
elevation	Surface or ground elevation at which air sample was collected. Units are meters above sea level (m a.s.l.).
intake_height	Height above ground at which air sample was collected. Units are meters above ground level (m a.g.l.).
obs_flag	Representation flag indicates that reported value has large spatial scale representation (1) or is locally influenced (0). This variable is derived from information provided by the data owners and is currently being further developed.

ESSDD
7, 495–519, 2014

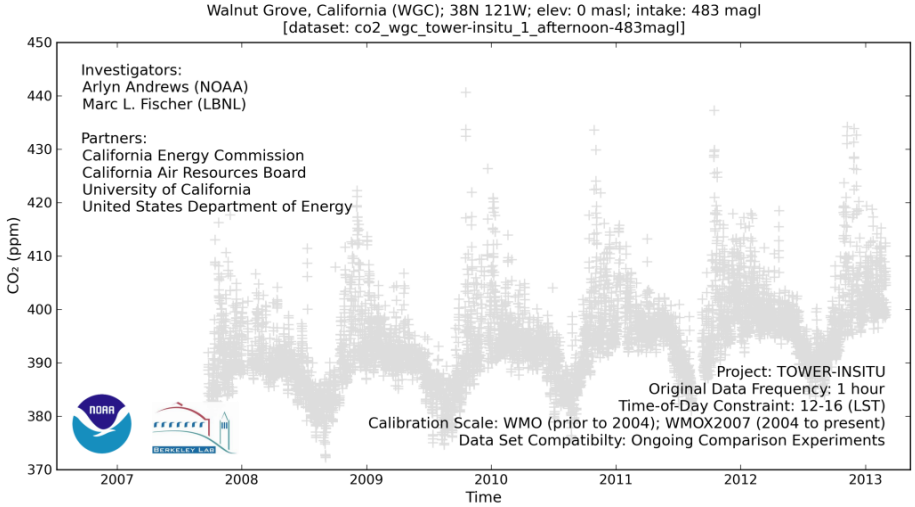

Figure 1. Graph constructed using data and metadata from a single ObsPack data set.

ESSDD
7, 495–519, 2014