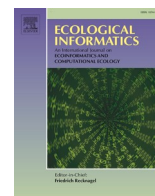


Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecoinf

A reporting format for leaf-level gas exchange data and metadata

Kim S. Ely^{a,*}, Alistair Rogers^a, Deborah A. Agarwal^b, Elizabeth A. Ainsworth^c, Loren P. Albert^d, Ashehad Ali^e, Jeremiah Anderson^a, Michael J. Aspinwall^f, Chandra Bellasio^g, Carl Bernacchi^c, Steve Bonnage^h, Thomas N. Buckleyⁱ, James Bunce^j, Angela C. Burnett^a, Florian A. Busch^k, Amanda Cavanagh^l, Lucas A. Cernusak^m, Robert Crystal-Ornelasⁿ, Joan Damerowⁿ, Kenneth J. Davidson^a, Martin G. De Kauwe^o, Michael C. Dietze^p, Tomas F. Domingues^q, Mirindi Eric Dusenge^r, David S. Ellsworth^s, John R. Evans^t, Paul P.G. Gauthier^u, Bruno O. Gimenez^v, Elizabeth P. Gordon^w, Christopher M. Gough^x, Aud H. Halbritter^y, David T. Hanson^z, Mary Heskel^{aa}, J. Aaron Hogan^{ab}, Jason R. Hupp^w, Kolby Jardine^{ac}, Jens Kattge^{ad}, Trevor Keenan^{n,ae}, Johannes Kromdijk^{af}, Dushan P. Kumarathunge^{ag}, Julien Lamour^a, Andrew D.B. Leakey^{ah}, David S. LeBauer^{ai}, Qianyu Li^a, Marjorie R. Lundgren^{aj}, Nate McDowell^{ak}, Katherine Meacham-Hensold^{al}, Belinda E. Medlyn^s, David J.P. Moore^{am}, Robinson Negrón-Juárez^b, Ülo Niinemets^{an}, Colin P. Osborne^{ao}, Alexandria L. Pivovarov^{ap}, Hendrik Poorter^{aq}, Sasha C. Reed^{ar}, Youngryel Ryu^{as}, Alvaro Sanz-Saez^{at}, Stephanie C. Schmiege^{au}, Shawn P. Serbin^a, Thomas D. Sharkey^{av}, Martijn Slot^v, Nicholas G. Smith^{aw}, Balasaheb V. Sonawane^{ax}, Paul F. South^{ay}, Daisy C. Souza^{az}, Joseph Ronald Stinziano^{ba}, Ellen Stuart-Haëntjens^{bb}, Samuel H. Taylor^{aj}, Mauricio D. Tejera^{bc}, Johan Uddling^r, Vigdis Vandvik^y, Charuleka Varadharajanⁿ, Anthony P. Walker^{bd}, Berkley J. Walker^{av,be}, Jeffrey M. Warren^{bd}, Danielle A. Way^{bf}, Brett T. Wolfe^{bg}, Jin Wu^{bh}, Stan D. Wullschlegel^{bi}, Chonggang Xu^{bj}, Zhengbing Yan^{bh}, Dedi Yang^a

^a Department of Environmental and Climate Sciences, Brookhaven National Laboratory, Upton, NY 11973, USA^b Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA^c USDA ARS GCPRU, 1201 W. Gregory Drive, Urbana, IL 61801, USA^d Department of Biology, West Virginia University, Morgantown, WV 26506, USA^e Department of Bioclimatology, University of Goettingen, Goettingen, Germany^f School of Forestry and Wildlife Sciences, Auburn University, Auburn, AL 36849, USA^g Research group biology of plants under Mediterranean conditions, Department of Biology, University of Balearic Islands, Cra. de Valldemossa Km 7.5, Palma 07122, Illes Balears, Spain^h ADC Bioscientific Ltd, Global House, Geddings Road, Hoddesdon, Herts EN11 0NT, UKⁱ Department of Plant Sciences, UC Davis, Davis, CA 95616, USA^j USDA-ARS, Adaptive Cropping Systems Laboratory, Beltsville, MD 20705, USA^k School of Biosciences, University of Birmingham, Birmingham B15 2TT, UK^l School of Life Sciences, University of Essex, Colchester CO4 3SQ, UK^m College of Science and Engineering, James Cook University, Cairns, QLD 4879, Australiaⁿ Earth and Environmental Sciences Area, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA^o ARC Centre of Excellence for Climate Extremes, Sydney, NSW 2052, Australia; Climate Change Research Centre, University of New South Wales, Evolution & Ecology Research Centre, University of New South Wales, Sydney, NSW 2052, Australia^p Department of Earth & Environment, Boston University, Boston, MA 02215, USA^q Department of Biology, FFCLRP, University of São Paulo, Av. Bandeirantes 3900, 14040-901 Ribeirão Preto, Brazil^r Department of Biological and Environmental Sciences, University of Gothenburg, P.O. Box 461, SE-405 30 Gothenburg, Sweden^s Hawkesbury Institute for the Environment, Western Sydney University, Locked Bag 1797, Penrith, NSW 2751, Australia^t Research School of Biology, The Australian National University, Canberra, ACT, 2601, Australia^u NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, NJ 08540, USA

* Corresponding author.

<https://doi.org/10.1016/j.ecoinf.2021.101232>

Available online 24 January 2021

1574-9541/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

- ^v Smithsonian Tropical Research Institute, Apartado Postal, 0843-03092, Panama City, Panama
- ^w LI-COR Biosciences, 4647 Superior St Lincoln, NE 68504, USA
- ^x Virginia Commonwealth University, Department of Biology, Box 842012, 1000 W. Cary Street, Richmond, VA 23284, USA
- ^y Department of Biological Sciences & Bjerknes Centre for Climate Research, University of Bergen, PO Box 7801, N-5020 Bergen, Norway
- ^z UNM Biology, MSC03-2020, 219 Yale Blvd, Albuquerque, NM 87131, USA
- ^{aa} Department of Biology, Macalester College, 1600 Grand Ave, St Paul, MN 55105, USA
- ^{ab} Department of Biological Sciences, Florida International University, Miami, FL 33199, USA
- ^{ac} Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
- ^{ad} Max Planck Institute for Biogeochemistry, Germany; German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Hans Knöll Str. 10, 07745 Jena Deutscher Platz 5e, 04103 Leipzig, Germany
- ^{ae} Department of Environmental Science, Policy and Management, UC Berkeley, Berkeley, CA 94720, USA
- ^{af} Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB23EA, UK
- ^{ag} Plant Physiology Division, Coconut Research Institute of Sri Lanka, Lunuwila 61150, Sri Lanka
- ^{ah} Departments of Plant Biology and Crop Sciences, University of Illinois at Urbana-Champaign, 1206 W Gregory Dr, Urbana, IL 61801, USA
- ^{ai} College of Agriculture and Life Sciences, University of Arizona, Tucson, AZ 85721, USA
- ^{aj} Lancaster Environment Centre, Library Avenue, Lancaster University, Lancaster LA1 4YQ, UK
- ^{ak} Pacific Northwest National Laboratory, PO Box 999, Richland, WA 99352, USA
- ^{al} Carl R. Woese Institute of Genomic Biology, University of Illinois at Urbana-Champaign, 1206 W. Gregory Drive, Urbana, IL 61801, USA
- ^{am} School of Natural Resources and the Environment, University of Arizona, Tucson, AZ 85721, USA
- ^{an} Estonian University of Life Sciences, Kreutzwaldi 1, 51006 Tartu, Estonia
- ^{ao} Department of Animal and Plant Sciences, University of Sheffield, Sheffield S7 2NF, UK
- ^{ap} Atmospheric Science and Global Change Division, Pacific Northwest National Laboratory, Richland, WA 99352, USA
- ^{aq} Plant Sciences (IBG2), Forschungszentrum Juelich GmbH, Juelich, Germany; Department of Biological Sciences, Macquarie University, North Ryde, NSW 2109, Australia
- ^{ar} US Geological Survey, Southwest Biological Science Center, 2290 S. West Resource Blvd, Moab, UT 84532, USA
- ^{as} Department of Landscape Architecture and Rural Systems Engineering, Seoul National University, Seoul, South Korea
- ^{at} Department of Crop, Soil and Environmental Sciences, Auburn University, Auburn, AL 36849, USA
- ^{au} Department of Ecology, Evolution and Environmental Biology, Columbia University, New York, NY 10027, USA
- ^{av} MSU-DOE Plant Research Laboratory, Michigan State University, 612 Wilson Rd, East Lansing, MI 48824, USA
- ^{aw} Department of Biological Sciences, Texas Tech University, Lubbock, TX 79409, USA
- ^{ax} School of Biological Sciences, Washington State University, Pullman, WA 99164, USA
- ^{ay} Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, USA
- ^{az} Forest Management Laboratory, National Institute for Amazon Research, Manaus, Amazonas, Brazil
- ^{ba} Department of Biology, University of New Mexico, Albuquerque, NM 87131, USA
- ^{bb} US Geological Survey, California Water Science Center, 6000 J St, Sacramento, CA 95819, USA
- ^{bc} Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI, 48824, USA
- ^{bd} Environmental Sciences Division and Climate Change Science Institute, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
- ^{be} Department of Plant Biology, Michigan State University, East Lansing, MI 48824, USA
- ^{bf} Department of Biology, University of Western Ontario, London, ON N6A 5B7, Canada
- ^{bg} School of Renewable Natural Resources, Louisiana State University Agricultural Center, Baton Rouge, LA 70803, USA
- ^{bh} School of Biological Sciences, The University of Hong Kong, Pokfulam Road, Hong Kong
- ^{bi} Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
- ^{bj} Earth and Environmental Sciences Division, Los Alamos National Laboratory, NM 87544, USA

ARTICLE INFO

Keywords:

Photosynthesis
Carbon dioxide
Irradiance
Data reporting format
Metadata
Data standard

ABSTRACT

Leaf-level gas exchange data support the mechanistic understanding of plant fluxes of carbon and water. These fluxes inform our understanding of ecosystem function, are an important constraint on parameterization of terrestrial biosphere models, are necessary to understand the response of plants to global environmental change, and are integral to efforts to improve crop production. Collection of these data using gas analyzers can be both technically challenging and time consuming, and individual studies generally focus on a small range of species, restricted time periods, or limited geographic regions. The high value of these data is exemplified by the many publications that reuse and synthesize gas exchange data, however the lack of metadata and data reporting conventions make full and efficient use of these data difficult. Here we propose a reporting format for leaf-level gas exchange data and metadata to provide guidance to data contributors on how to store data in repositories to maximize their discoverability, facilitate their efficient reuse, and add value to individual datasets. For data users, the reporting format will better allow data repositories to optimize data search and extraction, and more readily integrate similar data into harmonized synthesis products. The reporting format specifies data table variable naming and unit conventions, as well as metadata characterizing experimental conditions and protocols. For common data types that were the focus of this initial version of the reporting format, i.e., survey measurements, dark respiration, carbon dioxide and light response curves, and parameters derived from those measurements, we took a further step of defining required additional data and metadata that would maximize the potential reuse of those data types. To aid data contributors and the development of data ingest tools by data repositories we provided a translation table comparing the outputs of common gas exchange instruments. Extensive consultation with data collectors, data users, instrument manufacturers, and data scientists was undertaken in order to ensure that the reporting format met community needs. The reporting format presented here is intended to form a foundation for future development that will incorporate additional data types and variables as gas exchange systems and measurement approaches advance in the future. The reporting format is published in the U.S. Department of Energy's ESS-DIVE data repository, with documentation and future development efforts being maintained in a version control system.

1. Introduction

The interface between plant and ecological sciences and research data infrastructure is rapidly evolving, with greater expectations for data preservation, reproducible and open research, and the potential to synthesize data across different studies maximizing investments in research. Moreover, publicly accessible data archiving is increasingly required by funding bodies and publishers. Numerous databases and repositories, and other data infrastructure, have been developed to fulfill these needs, including TRY (Kattge et al., 2020), Environmental Data Initiative (environmentaldatainitiative.org), Dryad (datadryad.org), figshare (figshare.com) and DataOne (dataone.org). Yet the reuse of these data resources remains hampered by the difficulty of locating, unifying and assessing the quality of data, and the absence of important metadata needed for inter-site comparison or synthesis. The challenges that must be addressed for data managers to best support scientific discoveries are summarized by the FAIR principles, a call to improve Findability, Accessibility, Interoperability and Reusability of data (Wilkinson et al., 2016).

Leaf-level gas exchange measurements quantify the flux of carbon dioxide (CO₂) and water vapor into and out of a leaf. Typically collected with infrared gas analyzer instruments, these measurements are used to determine a range of physiologically important fluxes and traits, principally the rates of net CO₂ assimilation, respiration and stomatal conductance. Gas exchange data are used to answer a wide variety of scientific questions regarding plant function and response to environmental change (Long et al., 1996; Long and Bernacchi, 2003). They are the basis for estimating and scaling photosynthesis from the leaf to canopy (Yang et al., 2020), and are used to parameterize global biogeochemical models (Kattge et al., 2009). The products of photosynthesis are critical to society, as they provide renewable supplies of food, fuel, medicine, and fiber (Vitousek et al., 1986). Understanding and improving photosynthesis, and water- and nutrient-use efficiencies are currently considered to be key targets to improve the resilience of crops to global change (Ainsworth et al., 2008; Leakey et al., 2019; López-Calcano et al., 2020; Ort et al., 2015; Simkin et al., 2019). Furthermore, plants play a critical and unique role in determining the response of the terrestrial biosphere to rising CO₂ concentration and in turn influence the rate of global change (Walker et al., 2020). Analyses have also shown that terrestrial biosphere model outputs are particularly sensitive to parameters derived from gas exchange data (Bonan et al., 2011; Booth et al., 2012; LeBauer et al., 2013; Ricciuto et al., 2018; Sargsyan et al., 2014) and that the use of derived parameters from gas exchange data can effectively constrain uncertainty in model simulations (Dietze, 2014). In short, gas exchange data are central to understanding, improving and modelling the response of plants to global and environmental change.

However, collection of these data requires specialist training, is time consuming, can involve elaborate logistics (Ellsworth et al., 2012; Weerasinghe et al., 2014), and often utilizes techniques adapted to particular experiments, instruments and environments. Thus, resulting data products are typical long-tail data, i.e. data are low volume, and have diverse and heterogeneous content, and are thus not easily shared (Heidorn, 2008; Wallis et al., 2013). Currently, most data repositories that store diverse data typically focus on describing generic package-level metadata, and not metadata specific to the data type, which limits the use, search and data discovery services for long-tail data types (Limani et al., 2019). Our review of existing data repositories and plant trait databases revealed that where leaf-level gas exchange data are available, the data provided are limited and metadata required to properly interpret and reuse those data are often missing. The need for specialist data standards for disciplines is well recognized (Bruneau et al., 2019; Limani et al., 2019), and the importance of developing standards for the collection and storage of plant trait data has been the subject of several recent studies (Gallagher et al., 2020; Kissling et al., 2018; Schneider et al., 2019). Despite recent increases in compendia of

gas exchange data (Lin et al., 2015; Keenan and Niinemets, 2016; Kumarathunge et al., 2019; Smith et al., 2019; Niinemets et al., 2015; De Kauwe et al., 2016; Ali et al., 2015), and previous calls for standard archiving (Dietze, 2014), there is no standardized reporting format that enables syntheses of these data.

Data archiving is only the first step towards maximizing the value of data. In order to be reused or incorporated into models or synthesis products, data must be both findable and accessible; these characteristics are optimized by appropriate, machine-readable search terms and persistent dataset identifiers. Currently, gas exchange instruments do not share a common output format, i.e., file structure, variable names, and units, and include column headers that are not machine readable. Additionally, reuse is enabled by including sufficient metadata to correctly interpret the data, and use of common formats and terms that allow processing multiple studies from different sites, with various measurement methods (Christianson et al., 2017). Currently, metadata associated with gas exchange data collections are largely limited to location and species (Kattge et al., 2020). Lack of documentation and metadata are recognized as data archiving risk factors (Mayernik et al., 2020), with the implication that without adequate metadata, data cannot be interpreted or used correctly. To reuse data, researchers often have to refer to original publications to access essential metadata or other key information (e.g. leaf temperature), which can be a prohibitively resource-intensive process, or, especially in the case of older work, impossible because information is unavailable. Also, as research data infrastructure moves towards advanced capabilities such as application programming interfaces (APIs) to facilitate data upload and download, or support for data visualization and analytics, standardization of data and metadata in machine readable formats will become increasingly essential (Bruneau et al., 2019). For example, the Darwin Core standard for biodiversity data has enabled the global integration of hundreds of millions of species occurrence records through the Global Biodiversity Information Facility (GBIF, www.gbif.org), and has facilitated reuse of these data in countless studies (Ball-Damerow et al., 2019). However, specific guidance for leaf-level gas exchange data and metadata is lacking.

Here we present a new data and metadata reporting format for common types of leaf-level gas exchange data, reaching consensus among over 80 researchers in the field. We describe the process of development of these guidelines, for which the aim was to find the balance between maximizing the usefulness of the reporting format to the research community with ease of compliance when a data provider is preparing a new dataset. A key aspect has been engaging the community of leaf-level gas exchange experts in the development of this reporting format, with a concerted effort to reach as many potential data contributors and users as possible. Our goal with this initial focused effort on a leaf-level gas exchange reporting format was to develop a solid foundation for further development that could include a wider range of data types. An important component of this proposed reporting format is the public archive of complete instrument outputs. While we cannot foresee all future data uses or different processing methods, the preservation of the unprocessed instrument output is a way of future-proofing rare and valuable leaf-level gas exchange data sets (Rogers et al., 2017).

The creation of this reporting format for leaf-level gas exchange data was initiated by a call for community accepted data formats for the U.S. Department of Energy's (DOE) Environmental Systems Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE) data repository (Varadharajan et al., 2019). Accordingly, the reporting format described here is known as the 'ESS-DIVE reporting format for leaf-level gas exchange data and metadata', and referred to in this paper as 'the reporting format'. However, development of the format and documentation has considered global needs for these data. This paper presents a reporting format for leaf-level gas exchange data designed for wide adoption across ecological data repositories, and thus does not describe implementation in a specific repository or database. The reporting

format is designed to complement, and not duplicate, other metadata requirements for sites, samples and additional relevant information, and when possible this format should be used in combination with such requirements. For example, in the ESS-DIVE data repository a data submission must also include package-level metadata (e.g. authors, keywords, publication date, spatial and temporal coverage) and sample-level metadata (e.g. sample material, latitude, longitude, elevation, biome). It is encouraged that, where available, this reporting format is used in conjunction with established ontologies, such as Darwin Core (Wieczorek et al., 2012), the Plant Ontology (Cooper et al., 2013) and the Environment Ontology (Buttigieg et al., 2013).

The scope of this reporting format for leaf-level gas exchange focused on defining data and metadata variables to describe the most common type of measurements and those that have been the focus of recent synthesis efforts, i.e. survey style measurements; the response of photosynthesis to CO₂ and irradiance; and parameters derived from these relationships. In this paper we 1) describe the process of developing the format, including review of existing standards and conventions, and community consultation; 2) provide details of the components of the reporting format, including the guidance for data and metadata fields, vocabularies, units and definitions; and 3) discuss challenges to reaching consensus, the future potential to include additional measurement types and the use of this reporting format as a basis for the development of data management tools.

2. Methods

2.1. Review of existing standards and conventions

2.1.1. Search for published standards

Literature and web resources were searched to identify any published standards guiding best practice for the archive of leaf-level gas exchange data. A list of ecological trait databases was assembled, based on web searches, and a comprehensive table published by Schneider et al. (2019). Of these, databases and repositories identified as containing plant trait data were reviewed to determine if they included leaf-level gas exchange data, and if submission of data required adherence to any standards or formats (Supplementary Tables 1 & 2). A catalog of over 1400 data standards, including 17 categorized as concerning physiology, available at FAIRsharing (Sansone et al., 2019), was searched for standards that define variable names and metadata terms required to describe leaf-level gas exchange data.

2.1.2. Variable names and definitions

Existing data repositories, databases and synthesized datasets were reviewed, and the most commonly used variable terms and definitions were adopted into this reporting format (Supplementary Tables 1–3). The TRY plant trait database (Kattge et al., 2020) was identified as the most extensive publicly available plant trait database that contains leaf-level gas exchange data. Variable definitions in TRY are adopted from TOP, a thesaurus of plant characteristics (Garnier et al., 2017). Several relevant variables are included in BETYdb, the biofuel ecophysiological traits and yields database (LeBauer et al., 2018). Another resource for measurement variable definitions are several published guides to standard measurement protocols, including ClimEx (Halbritter et al., 2020), the Plant Handbook (Pérez-Harguindeguy et al., 2013) and PrometheusWiki (Evans and Santiago, 2014; Sack et al., 2010). The use of variable names in the large datasets including GlobResp (Atkin et al., 2015) and GlobAmx (Maire et al., 2015) were also considered. The default output from ten commercially available gas exchange instruments (manufactured by ADC Bioscientific, CID-Bioscience, LI-COR Biosciences, PP Systems and Walz; Supplementary Table 4) was assembled into a translation table to allow comparison and identify commonalities. Each variable was defined with a name, unit and description, and in some cases, an expected value range.

2.1.3. Metadata requirements

Many data repositories have existing metadata requirements to cover general experimental and sample parameters, such as characteristics of the location where measurements were conducted. Here we identified specific metadata parameters that would allow users of leaf-level gas exchange data to discriminate between data types, experimental protocols, and sample characteristics. We chose variables based on our collective expertise of conducting gas exchange measurements across diverse ecosystems and experimental designs, and of using those data in syntheses and meta-analysis. Our goal was to include metadata requirements for variables that would be most relevant for synthesis activities, including variables to distinguish data obtained from natural or cultivated plants, and to differentiate between common experimental manipulations and leaf sampling techniques. Controlled vocabularies—lists of preferred terms—were developed for each metadata variable to allow consistent metadata reporting. The selection of required metadata variables sought to find a balance between optimizing data discoverability and usability, while at the same time not placing undue burden on data contributors.

2.2. Community consultation

The draft reporting format was made available to the community of leaf-level gas exchange experts for suggestions and comment. Input was sought from data contributors, data scientists, data users, and instrument manufacturers. The invitation to participate was sent via direct email to 120 contacts, and reached a wider (unquantified) audience through encouraged sharing of the invitation to participate and social media. The development of this data reporting format was very well received by the community; eighty individuals contributed to the reporting format documentation, and thus this paper (Supplementary Fig. S1).

An introduction to the purpose, structure and components of the reporting format was presented as a publicly accessible webinar hosted by ESS-DIVE in July 2020, followed by a month-long period of feedback and discussion. Follow up video conferences were scheduled to discuss refinements and solutions. Feedback was gathered in an open manner, with comments and suggestions available to view by all on a collaborative online document. The reporting format documentation was then migrated to a public GitHub repository, where additions and refinements can continue to be made, and version controlled releases will be freely available for use.

3. Results

There are a number of common conventions in use for reporting of leaf-level gas exchange data, however they are not universal, and our search did not discover any published standards for data reporting. This directed our efforts into the development of a new metadata and data reporting format to enable diverse data contributors to use unified terminologies and formats when publishing data, thus lowering the barrier for data reuse and harmonization. The range of measurements that can be made with gas exchange instruments is broad. We have developed the foundation for a common reporting format by reaching consensus on a list of standard variable names and units for data tables, and metadata elements specific to leaf-level gas exchange measurements. Further guidance for data and metadata content is also proposed for data types that are commonly measured: survey style gas exchange measurements, dark adapted respiration measurements, CO₂ and light response curves, and parameters derived from those response curves.

The reporting format documentation comprises a number of elements relevant to all types of leaf-level gas exchange measurements; a list of variable names and units (Section 3.1) that should be used in data tables, a translation table of data outputs from commercial gas exchange instruments (Section 3.1.1) and comprehensive metadata requirements with controlled vocabularies (Section 3.2). For selected data types

(Section 3.3), the reporting format also specifies the minimum required variables to be included in the data table, and a list of details to be included in the measurement protocol description. Each of these elements is described in more detail in the sections below. The reporting format templates and complete documentation is available for download from the ESS-DIVE repository (Ely et al., 2020) and an example data package following the reporting format guidelines is hosted on the NGEA-Arctic data archive (Rogers et al., 2019).

Here we use the term ‘data package’ to refer to a collection of data and metadata files to be published together in a data repository (Christianson et al., 2017). A data package of gas exchange data should contain formatted data tables, metadata tables and the complete instrument output. Any data package may also include additional data types and variables not yet covered by the reporting format. Data packages should also include general metadata as required by the hosting data repository or database (e.g. author list and other citation information, data licensing terms). Fig. 1 shows the relationship between the components of a data package and definition tables included in the reporting format documentation.

The documentation includes user guides and templates to present the methods metadata and instrument details. Data and metadata tables should be in comma separated value (.csv) format; additional materials can be also included as text if appropriate. All components of a data package should be in English language; other language translations can be included as an additional resource.

3.1. Variable names and unit specifications

Consistent use of variable names (also known as field names or headers) in data tables is a key element of generating standardized datasets that can be readily combined or imported into a database. A list of variables, including measured and calculated instrument outputs, calculated parameters, and constants, were designated as *variableName*, *variableUnit* and *variableDefinition* in the defined variables table. Note that the *camelCase* naming used here indicates variables defined in the reporting format documentation. These conventions for variables were reached based on the most common usage in existing publications, databases and instrument outputs. In cases where common usage had not

already been established, *variableNames* were selected to be human and machine readable, and with no recognized conflicts with other uses. The units for each variable are listed separately, and are not included as part of the *variableName*. The measurement quantities are described in each *variableDefinition*, thus *variableUnits* are presented without information about the quantity, following NIST guidelines (Thompson and Taylor, 2008). For many *variableNames*, the reporting format also specifies an expected range of values resulting from common measurement approaches; these limits can be used to guide quality checking of data.

3.1.1. Instrument output translation table

A translation table of 23 measured and calculated output variables from ten commercially available gas exchange instruments (Supplementary Table 4) was compiled to assess the most common variable names and measurement units, and provided input into the process of defining the variables for this reporting format. It was found that there is some variation among the output of different instruments, in both variable names and units used, and thus the default instrument outputs are not always exactly aligned with the proposed reporting format. For example, measurement of photosynthetic flux density (PPFD) incident on the leaf is variously labeled as Q, Qleaf, Qin, PAR, PARI or PARtop across the different instruments. The instrument output translation table provides a guide for conversion of results to standard *variableNames* and *variableUnits*, can assist data users to understand instrument output from unfamiliar instruments, and be used for future advances such as the development of automated tools for data upload. This compilation was based on current instrument manuals and software versions; users should note that future instrument and software updates may change the outputs. It is emphasized that re-labeling and conversion of variable names and units to match the format is not a requirement for the complete instrument output files.

3.2. Metadata

All published data packages that use this reporting format should include metadata to ensure that data are adequately described, in order to allow users to fully understand how the data were generated, and maximize findability of data with certain characteristics. The reporting

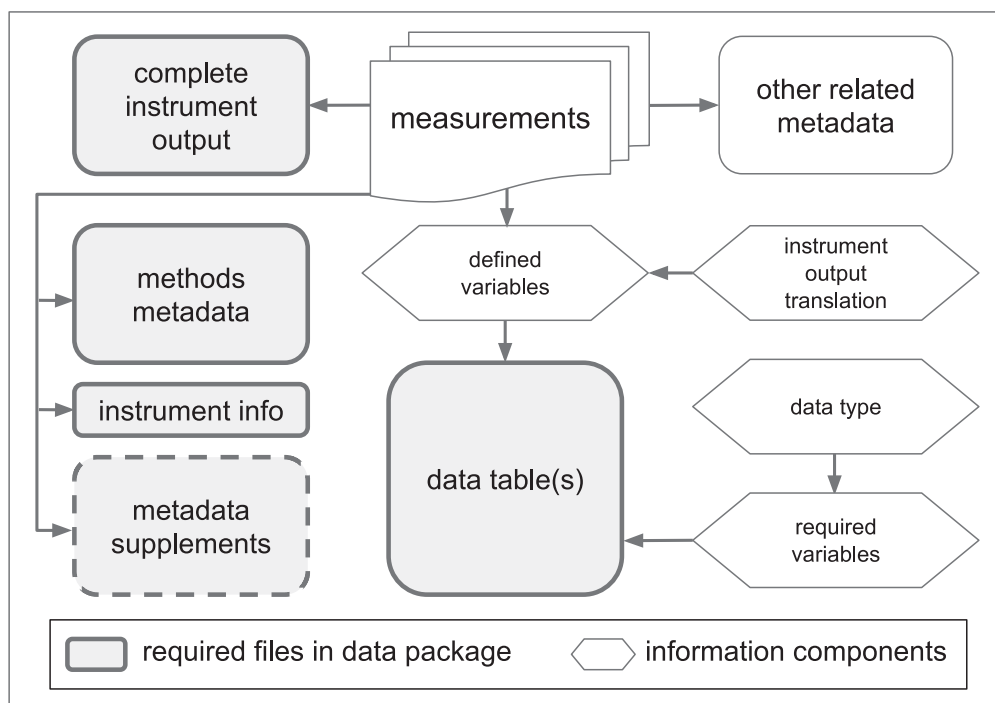


Fig. 1. The relationships between gas exchange measurements and the components of this data and metadata reporting format. The characteristics of the ‘measurements’ (experimental design and recording of information) inform the content of metadata and data tables. Components shown in boxes should be included in a data package. Data tables, methods metadata, instrument details and the complete instrument output (if available) are required to be included in a data package (grey boxes with solid borders). Inclusion of metadata supplement tables will be dependent on the experiment (grey box with dashed border). The requirements for other related metadata (white box) could be set by a data repository, or be mandated by other specialist data standards or conventions; those details are not covered by this reporting format. Information components (hexagons) are reference tables to guide the format and content requirements of the submitted data. Refer to the reporting format documentation for a complete list of variables, definitions and controlled vocabularies (Ely et al., 2020).

format provides controlled vocabularies and template files for the required methods metadata (Section 3.2.1) and instrument details (Section 3.2.2). The inclusion of methods supplement tables (Section 3.2.3) and other related metadata (Section 3.2.4) will depend on the design of individual experiments; thus for these items the reporting format provides guidelines and recommendations only.

3.2.1. Methods metadata

The methods metadata is a record of data types, measurement protocols, experimental and sample characteristics, and details of data processing and calculation approaches, summarized in a single file. The reporting format includes a template file and controlled vocabularies to simplify metadata creation and ensure consistency across datasets. However, the diversity of experimental approaches is recognized, and flexibility is accommodated by allowing use of free text for many variables if the controlled vocabulary is not adequate.

Development of the methods metadata focused on important search filters for data users such as the growth conditions and treatments of the plants on which measurements were made; these should be indicated using the *growthEnvironment* and *experimentalManipulation* variables. For example, the *growthEnvironment* variable captures if the plants were grown in natural or controlled environments, while *experimentalManipulation* can be employed by data users to include or exclude common treatments such as atmospheric, water or nutrient manipulation. Further categorization is enabled by use of variables such as *canopyPosition*, *lightExposure*, *leafAge* and *plantAge*.

3.2.2. Instrumentation details

This data reporting format provides a template to record details of the instruments used for data collection, including model information, software version, type of chamber used, and a statement of instrument calibration. This will enable users of data to understand the data provenance, and achieve data equivalency in synthesis products.

3.2.3. Methods supplement tables

Leaf-level gas exchange data are often measured with the purpose of comparing between sample types or treatments; these discriminators are commonly included in data tables as codes to represent species, treatments, plots or other characteristics. The methods supplement tables component of this reporting format demonstrates how the explanation of these descriptors should be included in a data package with a range of examples. Inclusion of metadata supplements in a data package is highly dependent on the nature of the experiment, and as such, examples are provided as guidelines only and are not required by the reporting format.

3.2.4. Other related data and metadata

Gas exchange data are frequently associated with other measurements, e.g. leaf nitrogen content. We strongly encourage the use of unique, persistent sample identifiers to link gas exchange data and other data and metadata associated with the same sample. The unique sample identifier should be a column in the data file. Also, in simple cases associated data could be included as additional variables (i.e. columns) in a gas exchange data table and where data are collected in-line with gas exchange, e.g. fluorescence, logically they should be included in the same file. In cases where a *variableName* is not defined by this reporting format, data providers should follow other appropriate standards or conventions. Similarly, for experimental data not covered by the methods metadata variables, such as reporting of environmental, landscape, or climatic characteristics, or genotype variation, data providers should utilize published standards or formats for that data type. Metadata associated with the sample collection (e.g. location information, sample description) can be provided separately using a file that conforms to recognized sample reporting formats (e.g. [Damerow et al., 2020](#)).

3.3. Specific requirements for selected data types

Additional reporting guidelines are provided for seven data types identified as common gas exchange measurements (e.g. photosynthetic CO₂ response curves) and analytical approaches (e.g. one-point method) (Table 1). For each of these data types the reporting format includes a detailed description of the data type, a list of elements required in the protocol description, and the minimum required variables to include in the data table. For data types not described here, data creators should use a protocol and minimum variable requirements as judged appropriate for their data.

For each data type, a list of 5–8 required variables was developed in order to capture the result variable (e.g. V_{cmax}) and covariates required to interpret that result in context. Of the existing standards and databases reviewed, only the BETYdb specifies any required or optional covariates ([LeBauer et al., 2018](#)). Thus the minimum required variables presented in this reporting format are the result of an iterative feedback process involving both domain expert contributors and users. Data contributors may also include any other variables, using the *variableNames* defined in this reporting format. The data table should also include the sample identifier, and other sample variables (e.g. species, treatment) as required.

Table 1

The data types for which this reporting format makes specific recommendations for variables required in the data table, and protocol descriptions. Refer to the reporting format documentation for a detailed description of each data type.

| Data type | Description |
|--|---|
| Survey | Single point measurement of leaf gas exchange. |
| Response of photosynthesis to intercellular CO ₂ concentration (AC_i curves) | Sequential measurements on the same leaf material of photosynthetic rate with varying CO ₂ concentration. |
| Photosynthetic parameters derived from AC_i curves | Results of fitting photosynthetic CO ₂ response curves to derive parameters, e.g. apparent V_{cmax} , J_{max} , TPU. |
| V_{cmax} from one-point | Apparent V_{cmax} calculated from A_{sat} measurements using the one-point method. |
| Response of photosynthesis to irradiance (AQ curves) | Sequential measurements on the same leaf material of photosynthetic rate with varying irradiance. |
| Photosynthetic parameters derived from AQ curves | Results of fitting light response curves to derive parameters, e.g. quantum yield of CO ₂ fixation. |
| Dark adapted respiration | Respiration rate of a dark adapted leaf. |

3.4. Inclusion of instrument output data

The methods metadata and required variables are designed to capture adequate information to allow proper interpretation of datasets. However, not all possible data reuse can be foreseen. The inclusion of the complete instrument output (commonly referred to as ‘raw data’) in a data package is seen as the ultimate future-proofing for a dataset ([Fig. 1](#)). Archiving of raw gas exchange data is recognized as good science practice and has been highlighted as important for the preservation and reuse of data ([Dietze, 2014](#); [Rogers et al., 2017](#)). Ideally we would like to mandate archiving of quality controlled complete instrument output to allow reanalysis of highly valuable datasets as new knowledge, analytical approaches or data corrections are developed. The term ‘complete instrument output’ is used here to recognize that instrument data files with some quality control applied, such as correction of user input errors, are generally more valuable to data users than true raw data. However, this ideal has to be balanced by the need to ensure we do not create a barrier for data submission, particularly for older data sets where complete instrument output may no longer be available, or for data collected with custom built gas exchange systems.

4. Discussion

We have developed a reporting format for leaf-level gas exchange data and metadata with the goal of improving findability and the capacity to reuse these valuable data (Ely et al., 2020). This reporting format will be adopted by the ESS-DIVE data repository and be freely available to the community. We encourage its use by all data producers, repositories and databases. More than eighty data contributors, data users, manufacturers and data scientists contributed to the development of this effort, which we hope will form a foundation for future development by the community. The reporting format aims to provide a resource for data contributors to enhance the value of their data, reduce the overheads to re-using and synthesizing data, and provide prescribed metadata that will simplify parsing of data for analysis and synthesis (Fig. 2).

4.1. Development of a community standard

Given the importance of gas exchange data, the effort taken to collect it, and the widespread use of gas exchange data in synthesis activities and model parameterization, it was surprising that a data standard did not yet exist. However, the need and desire for the development of a common reporting format was readily apparent. Both data contributors and data users were very supportive of the effort, were quick to contribute, and provided valuable input.

Data reporting formats, and the mandate by funding agencies to use them, burden the data contributor with the task of preparing and uploading their data. In contrast, the data user is hopefully relieved of the burden associated with finding and harmonizing datasets prior to analysis. Therefore there is a perception that data reporting formats and preservation of data in a repository offer little direct return for the contributor's effort. However, whilst not readily tangible, there are several benefits to contributors. These include the provision of a formal way to meet mandates for data preservation, defined data descriptions and units, and tools for data quality control (e.g. expected data ranges). Widespread adoption of a data reporting format will also accelerate the development of data ingest tools that will benefit the contributor. Furthermore, sharing of data in an accessible and searchable format increases the impact of their data collection, jump-starts collaborations, and, with conscientious data users, can lead to invitations to co-author novel data syntheses where data contributors can share knowledge of their data and also gain additional insight from their collaborators (Allen and Mehler, 2019; Cheruvilil and Soranno, 2018). One issue that remains challenging for the field is formal recognition of datasets through citations and ensuring the continued recognition of a given contribution. For example, if the original dataset is combined with other data into a larger dataset the original association with the data contributor can be lost.

While a formal data format had not existed before we started this work, the vocabulary of leaf-level gas exchange was well established and

was fairly similar across instruments. Therefore, incorporating many variables and definitions that are already in widespread use resulted in large parts of the reporting format being readily accepted by the community. Most feedback was focused on additional components, and fine tuning of definitions, rather than large changes to the first draft proposal. It was necessary to provide precise descriptive information to clearly communicate our goal of developing a data reporting format; in some instances this goal was conflated with documentation of measurement protocols, defining a gold standard method or building a database. The data reporting format does not attempt to constrain method choice by data contributors but is intended to be inclusive of all approaches and methodologies. However, there were several issues that garnered significant commentary and these are discussed further below.

4.2. Decisions and compromises

As expected there was a necessary compromise between the desire for additional metadata detail and the need for a relatively simple and manageable reporting format for data contributors. Many of the requests for increased metadata would increase the effort, and therefore the barrier, to format data for some contributors whilst providing only limited value for most data users. Experimental and sample details that are not covered by the methods metadata variables may be included in protocol descriptions or methods supplement tables. While not yet providing specific formats for data, methods metadata variables have been included to indicate the inclusion of canopy height information and additional data collected in-line with gas exchange. There are no restrictions preventing conscientious data contributors from including more metadata detail or data types. Similarly, when developing the required variables for each data type we resisted adding requirements for variables that are not essential to effectively reuse the data. We hope that by strongly encouraging (and perhaps, in time, mandating) the submission of complete instrument output we will preserve all data fields for the specialist data user, and for future, currently unanticipated uses of the data.

There were several comments about missing measurement types; in many cases this reflected the desire to expand the reporting format to cover more data types, e.g. temperature and vapor pressure deficit response curves, or porometer measurements of stomatal conductance. The combination of fluorescence with gas exchange data is very powerful and many instruments allow simultaneous collection of both data types. Whilst we recognize the value of including fluorescence data, developing common reporting formats for these data would have significantly expanded the scope of this initial effort. Development of a common data reporting format for fluorescence data presents some additional challenges, since these data are not always associated with gas exchange data, they can be collected with a wider variety of instruments, and the vocabulary and protocols used are not as constrained as for gas exchange measurements (Baker, 2008; Maxwell and Johnson, 2000; Murchie and Lawson, 2013).

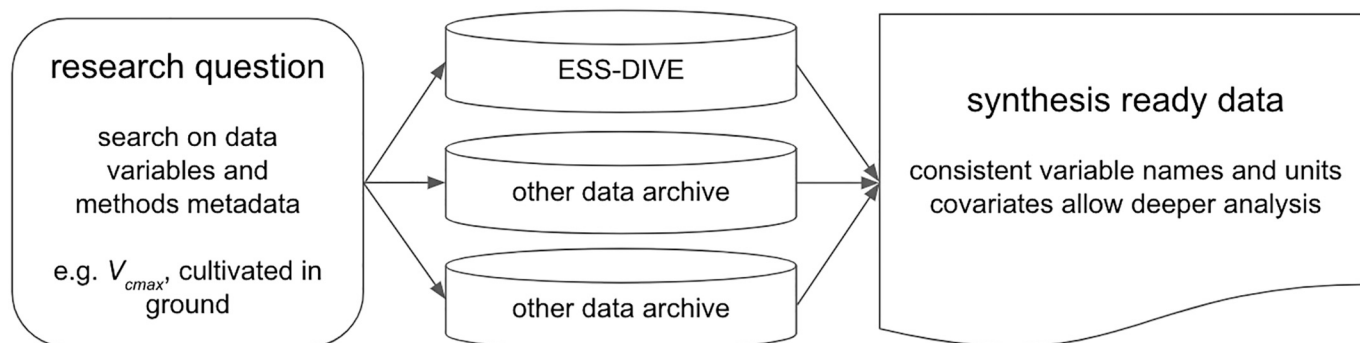


Fig. 2. Schematic showing how the implementation of this data reporting format across data archives will facilitate data discovery and reuse.

Estimates of photosynthetic parameters derived from the response of photosynthesis to intercellular CO₂ concentration (C_i) provide apparent estimates of those parameters, i.e. the estimate assumes an infinite mesophyll conductance (g_m) and C_i is assumed to be equal to the CO₂ concentration inside the chloroplast (C_c)—the site of carboxylation. Whilst g_m and hence C_c can be estimated from gas exchange data (Ethier and Livingston, 2004; Sharkey et al., 2007), the most robust approaches require in-line measurements of fluorescence or isotopic discrimination (Bongi and Loreto, 1989; Busch et al., 2020; Evans et al., 1986; Harley et al., 1992; Loreto et al., 1992; von Caemmerer and Evans, 1991). Estimates of photosynthetic parameters based on C_c are different from those that do not account for g_m , and the data can be used in different ways, so it is important to distinguish which data (C_i or C_c), were used to calculate the derived parameters. Additionally, for the specialist data user, knowledge of additional fluorescence or isotopic discrimination data collected in parallel with gas exchange data would be valuable. Therefore we added methods metadata requirements for photosynthetic CO₂ response data to capture assumptions about g_m and indicate the presence of additional data in the data package.

Specialist approaches of gas exchange measurements mean that equivalence cannot be assumed between different studies, even within the same lab, as protocols are adjusted for individual experiments, depending on species measured, ambient environmental conditions, and the experimental goals. The methods metadata categories have been defined to allow equivalency between data sets to be recognized, and provide the required information to recalculate if necessary. Similarly, calculations of parameters such as maximum carboxylation capacity (V_{cmax}) are dependent on fitting approaches (Bernacchi et al., 2013; Gu et al., 2010; Sharkey et al., 2007), the choice of kinetic constants (Rogers et al., 2017), inclusion of mesophyll conductance (Ethier and Livingston, 2004; Warren, 2006) and whether and how investigators applied corrections for gasket diffusion leaks (Flexas et al., 2007; Rodeghiero et al., 2007). In some cases, capturing these metadata can enable data users to recalculate derived parameters using a common approach (e.g., Niinemets et al., 2015). Ideally data users should recalculate derived parameters from the underlying data.

4.3. Future developments

Development of this data reporting format highlighted the strong desire by the community for additional functionality to be added to repositories to aid data ingestion, search and compilation. For example, tools that would reduce the burden of curating instrument output files, enable data validation during upload, and enable full search and compilation of all measured and calculated instrument output variables. Whilst the creation of such tools is out of the scope of developing a data reporting format we considered these needs in our decision making processes. For example, we have provided an instrument output translation table that provides a column by column comparison of default output from commercially available gas exchange systems, have created and defined a machine readable vocabulary, defined units, and provided expected ranges for commonly measured variables that can aid validation and curation. Furthermore, we have mandated that data are published using a single non-proprietary file format (.csv) further reducing the challenge for long-term archival and cyberinfrastructure tool development.

To aid long-term development, the reporting format will be a dynamic document hosted as a public repository on GitHub, a version control platform (<https://github.com/ess-dive-community/essdive-leaf-gas-exchange>). This platform will allow the user community to flag issues, make suggestions, discuss amendments, and prioritize development of the reporting format, all in the open so the community can understand the motivation behind development and contribute to decision making. The published reporting format can be revised with minor edits, ensuring users can easily access the latest update. Contributors on the GitHub platform could also facilitate more substantial

changes, such as the addition of new data types, leading to publication of a new version of this reporting format in the future.

We hope that widespread adoption of this first data reporting format for leaf-level gas exchange data will increase the preservation and reuse of these valuable, and hard won, data sets and elevate the importance of data storage in the mindset of data contributors. We also hope that this work will form the cornerstone for a more comprehensive effort by the community to expand and develop the reporting format, including expansion to include full consideration of additional data types that were beyond the initial scope of this effort.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

KSE and AR were supported by Environmental Systems Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE) community funds, through the Office of Biological and Environmental Research in the Department of Energy, Office of Science. Further support was received from several additional projects funded by the Office of Biological and Environmental Research, including the Next Generation Ecosystem Experiments -Arctic (NGEE-Arctic) and -Tropics (NGEE-Tropics) projects and through the United States Department of Energy contracts No. DE-SC0012704, DE-AC02-05CH11231, and DE-AC05-00OR22725 to Brookhaven National Laboratory, Lawrence Berkeley National Laboratory, and Oak Ridge National Laboratory. In addition, ASS was supported by Alabama Agricultural Experiment Station and the Hatch program of the National Institute of Food and Agriculture, U.S. Department of Agriculture; BJW was supported by the Division of Chemical Sciences, Geosciences and Biosciences, Office of Basic Energy Sciences of the United States Department of Energy [Grant DE-FG02-91ER20021]; JK was supported by the Biotechnology and Biological Sciences Research Council UK, grant reference BB/T007583/1; MJA was supported by the U.S. Department of Agriculture - National Institute of Food and Agriculture (Award number: 2019-67013-29161). SCR was supported by US Department of Defense (RC18-1322) and the USGS Ecosystems Mission Area. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Appendix A. Supplementary Data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ecoinf.2021.101232>.

References

- Ainsworth, Elizabeth A., Rogers, Alistair, Leakey, Andrew D.B., 2008. Targets for crop biotechnology in a future high-CO₂ and high-O₃ world. *Plant Physiol.* <https://doi.org/10.1104/pp.108.117101>.
- Ali, Ashehad A., Xu, Chonggang, Rogers, Alistair, McDowell, Nathan G., Medlyn, Belinda E., Fisher, Rosie A., Stan D. Wullschlegel, et al., 2015. Global-scale environmental control of plant photosynthetic capacity. *Ecol. Appl.* <https://doi.org/10.1890/14-2111.1>.
- Allen, Christopher, Mehler, David M.A., 2019. Open Science challenges, benefits and tips in early career and beyond. *PLoS Biol.* 17 (5), e3000246 <https://doi.org/10.1371/journal.pbio.3000246>.
- Atkin, Owen K., Bloomfield, Keith J., Reich, Peter B., Tjoelker, Mark G., Asner, Gregory P., Bonal, Damien, Bönisch, Gerhard, et al., 2015. Global variability in leaf respiration in relation to climate, plant functional types and leaf traits. *The New Phytol.* 206 (2), 614–636. <https://doi.org/10.1111/nph.13253>.
- Baker, Neil R., 2008. Chlorophyll Fluorescence: A Probe of Photosynthesis. In *Vivo Ann. Rev. Plant Biol.* 59 (1), 89–113. <https://doi.org/10.1146/annurev-arplant.59.032607.092759>.
- Ball-Damerow, Joan E., Brenskelle, Laura, Barve, Narayani, Soltis, Pamela S., Sierwald, Petra, Bieler, Rüdiger, LaFrance, Raphael, Ariño, Arturo H., Guralnick, Robert P., 2019. Research applications of primary biodiversity databases

- in the digital age. *PLoS One* 14 (9), e0215794. <https://doi.org/10.1371/journal.pone.0215794>.
- Bernacchi, Carl J., Bagley, Justin E., Serbin, Shawn P., Ruiz-Vera, Ursula M., Rosenthal, David M., Vanloocke, Andy, 2013. Modelling C₃ photosynthesis from the chloroplast to the ecosystem. *Plant Cell Environ.* 36 (9), 1641–1657. <https://doi.org/10.1111/pce.12118>.
- Bonan, Gordon B., Lawrence, Peter J., Oleson, Keith W., Levis, Samuel, Jung, Martin, Reichstein, Markus, Lawrence, David M., Swenson, Sean C., 2011. Improving canopy processes in the community land model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data. *J. Geophys. Res.* <https://doi.org/10.1029/2010jg001593>.
- Bongi, G., Loreto, F., 1989. Gas-exchange properties of salt-stressed olive (*Olea Europea L.*) leaves. *Plant Physiol.* 90 (4), 1408–1416. <https://doi.org/10.1104/pp.90.4.1408>.
- Booth, Ben B.B., Jones, Chris D., Collins, Mat, Totterdell, Ian J., Cox, Peter M., Sitch, Stephen, Huntingford, Chris, Betts, Richard A., Harris, Glen R., Lloyd, Jon, 2012. High sensitivity of future global warming to land carbon cycle processes. *Environ. Res. Lett.* 7 (2), 024002 <https://doi.org/10.1088/1748-9326/7/2/024002>.
- Bruneau, Anne, Borges, Leonardo M., Allkin, Robert, Egan, Ashley N., de la Estrella, Manuel, Javadi, Firouzeh, Klitgaard, Bente, et al., 2019. Towards a new online species-information system for legumes. *Aust. Syst. Bot.* 53, 1. <https://doi.org/10.1071/SB19025>.
- Busch, Florian A., Holloway-Phillips, Meisha, Stuart-Williams, Hilary, Farquhar, Graham D., 2020. Revisiting carbon isotope discrimination in C plants shows respiration rules when photosynthesis is low. *Nature Plants* 6 (3), 245–258. <https://doi.org/10.1038/s41477-020-0606-6>.
- Buttigieg, Pier Luigi, Morrison, Norman, Smith, Barry, Mungall, Christopher J., Lewis, Suzanna E., ENVO Consortium, 2013. The environment ontology: Contextualising biological and biomedical entities. *J. Biomed. Semant.* 4 (1), 43. <https://doi.org/10.1186/2041-1480-4-43>.
- Cheruvilil, Kendra Spence, Soranno, Patricia A., 2018. Data-intensive ecological research is catalyzed by Open Science and team science. *Bioscience* 68 (10), 813–822. <https://doi.org/10.1093/biosci/biy097>.
- Christianson, Danielle S., Varadarajan, Charuleka, Christoffersen, Bradley, Detto, Matteo, Faybishenko, Boris, Gimenez, Bruno O., Hendrix, Val, et al., 2017. A metadata reporting framework (FRAMES) for synthesis of Ecohydrological observations. *Ecol. Informat.* 42 (November), 148–158. <https://doi.org/10.1016/j.ecoinf.2017.06.002>.
- Cooper, Laurel, Walls, Ramona L., Elser, Justin, Gandolfo, Maria A., Stevenson, Dennis W., Smith, Barry, Preece, Justin, et al., 2013. The plant ontology as a tool for comparative plant anatomy and genomic analyses. *Plant Cell Physiol.* 54 (2), e1 <https://doi.org/10.1093/pcp/pcs163>.
- Damerow J., Varadarajan C., Boye K., Brodie E., Chadwick D., Cholia S., Elbashedy H., Ely K., Goldman A., Hendrix V., Jones C., Jones M., Kakalia Z., Kemner K., Kersting A., Maher K., Merino N., O'Brien F., Perzan Z., Robles E., Snavely C., Sorensen P., Stegen J., Weisenhorn P., Whitenack K., Zavarin M., Agarwal D., 2020. Sample identifiers and metadata reporting format for environmental systems science. *Environ. Syst. Sci. Data Infrastructure for a Virtual Ecosystem (ESS-DIVE)*. <https://doi.org/10.15485/1660470>.
- Dietze, Michael C., 2014. Gaps in knowledge and data driving uncertainty in models of photosynthesis. *Photosynth. Res.* 119 (1–2), 3–14. <https://doi.org/10.1007/s11120-013-9836-z>.
- Ellsworth, David S., Thomas, Richard, Crous, Kristine Y., Palmroth, Sari, Ward, Eric, Maier, Chris, DeLucia, Evan, Oren, Ram, 2012. Elevated CO₂ affects photosynthetic responses in canopy pine and subcanopy deciduous trees over 10 years: a Synthesis from Duke FACE. *Glob. Chang. Biol.* <https://doi.org/10.1111/j.1365-2486.2011.02505.x>.
- Ely, K.S., Rogers, A., Crystal-Ornelas, R., 2020. ESS-DIVE reporting format for leaf-level gas exchange data and metadata. *Environmental Systems Science Data Infrastructure for a Virtual Ecosystem*. <https://doi.org/10.15485/1659484>.
- Ethier, G.J., Livingston, N.J., 2004. On the need to incorporate sensitivity to CO₂ transfer conductance into the Farquhar-von Caemmerer-Berry leaf photosynthesis model. *Plant Cell Environ.* <https://doi.org/10.1111/j.1365-3040.2004.01140.x>.
- Evans, John R., Santiago, Louis S., 2014. PrometheusWiki gold leaf protocol: gas exchange using LI-COR 6400. *Funct. Plant Biol.* 41 (3), 223–226. <https://doi.org/10.1071/FP10900>.
- Evans, J.R., Sharkey, T.D., Berry, J.A., Farquhar, G.D., 1986. Carbon isotope discrimination measured concurrently with gas exchange to investigate CO₂ diffusion in leaves of higher plants. *Funct. Plant Biol.* 13 (2), 281. <https://doi.org/10.1071/PP9860281>.
- Flexas, J., Díaz-Espejo, A., Berry, J.A., Cifre, J., Galmés, J., Kaldenhoff, R., Medrano, H., Ribas-Carbó, M., 2007. Analysis of leakage in IRGA's leaf chambers of open gas exchange systems: quantification and its effects in photosynthesis parameterization. *J. Exp. Bot.* 58 (6), 1533–1543. <https://doi.org/10.1093/jxb/erm027>.
- Gallagher, Rachael V., Falster, Daniel S., Maitner, Brian S., Salguero-Gómez, Roberto, Vandvik, Vigdis, Pearse, William D., Florian D. Schneider, et al., 2020. Open Science principles for accelerating trait-based science across the tree of life. *Nat. Ecol. Evolut.* 4 (3), 294–303. <https://doi.org/10.1038/s41559-020-1109-6>.
- Garnier, Eric, Stahl, Ulrike, Laporte, Marie-Angélique, Kattge, Jens, Mougouet, Isabelle, Kühn, Ingolf, Laporte, Baptiste, et al., 2017. Towards a thesaurus of plant characteristics: an ecological contribution. *The J. Ecol.* 105 (2), 298–309. <https://doi.org/10.1111/1365-2745.12698>.
- Gu, Lianhong, Pallardy, Stephen G., Kevin, Tu, Law, Beverly E., Wullschleger, Stan D., 2010. Reliable estimation of biochemical parameters from C₃ leaf photosynthesis-intercellular carbon dioxide response curves. *Plant Cell Environ.* <https://doi.org/10.1111/j.1365-3040.2010.02192.x>.
- Halbritter, Aud H., De Boeck, Hans J., Eycott, Amy E., Reinsch, Sabine, Robinson, David A., Vicca, Sara, Berauer, Bernd, et al., 2020. The handbook for standardized field and laboratory measurements in terrestrial climate change experiments and observational studies (ClimEx). *Methods Ecol. Evolut.* 11 (1), 22–37. <https://doi.org/10.1111/2041-210X.13331>.
- Harley, P.C., Loreto, F., Di Marco, G., Sharkey, T.D., 1992. Theoretical considerations when estimating the mesophyll conductance to CO₂ flux by analysis of the response of photosynthesis to CO₂. *Plant Physiol.* 98 (4), 1429–1436. <https://doi.org/10.1104/pp.98.4.1429>.
- Heidorn, P. Bryan, 2008. Shedding light on the dark data in the Long tail of science. *Libr. Trends* 57 (2), 280–299. <https://doi.org/10.1353/lib.0.0036>.
- Kattge, Jens, Knorr, Wolfgang, Raddatz, Thomas, Wirth, Christian, 2009. Quantifying photosynthetic capacity and its relationship to leaf nitrogen content for global-scale terrestrial biosphere models. *Glob. Chang. Biol.* <https://doi.org/10.1111/j.1365-2486.2008.01744.x>.
- Kattge, Jens, Bönsch, Gerhard, Díaz, Sandra, Lavorel, Sandra, Prentice, Iain Colin, Leadley, Paul, Tautenhahn, Susanne, et al., 2020. TRY plant trait database - enhanced coverage and open access. *Glob. Chang. Biol.* 26 (1), 119–188. <https://doi.org/10.1111/gcb.14904>.
- Kauwe, De, Martin, G., Lin, Yan-Shih, Wright, Ian J., Medlyn, Belinda E., Crous, Kristine Y., Ellsworth, David S., Maire, Vincent, et al., 2016. A test of the 'one-point method' for estimating maximum carboxylation capacity from field-measured, light-saturated photosynthesis. *The New Phytol.* 210 (3), 1130–1144. <https://doi.org/10.1111/nph.13815>.
- Keenan, Trevor F., Niinemets, Ulo, 2016. The canopy trait plasticity (CANTRIP) database v1.0.0 (version V1.0.0) [data set]. *Zenodo*. <https://doi.org/10.5281/zenodo.164883>.
- Kissling, W. Daniel, Walls, Ramona, Bowser, Anne, Jones, Matthew O., Kattge, Jens, Agosti, Donat, Amengual, Josep, et al., 2018. Towards global data products of essential biodiversity variables on species traits. *Nat. Ecol. Evolut.* 2 (10), 1531–1540. <https://doi.org/10.1038/s41559-018-0667-3>.
- Kumarathunge, Dushan P., Medlyn, Belinda E., Drake, John E., Tjoelker, Mark G., Aspinwall, Michael J., Battaglia, Michael, Francisco J. Cano, et al., 2019. Acclimation and adaptation components of the temperature dependence of plant photosynthesis at the global scale. *New Phytol.* <https://doi.org/10.1111/nph.15668>.
- Leakey, Andrew D.B., Ferguson, John N., Pignone, Charles P., Wu, Alex, Jin, Zhenong, Hammer, Graeme L., Lobell, David B., 2019. Water use efficiency as a constraint and target for improving the resilience and productivity of C₃ and C₄ crops. *Annu. Rev. Plant Biol.* <https://doi.org/10.1146/annurev-arplant-042817-040305>.
- LeBauer, David S., Wang, Dan, Richter, Katherine T., Davidson, Carl C., Dietze, Michael C., 2013. Facilitating feedbacks between field measurements and ecosystem models. *Ecol. Monogr.* <https://doi.org/10.1890/12-0137.1>.
- LeBauer, David, Kooper, Rob, Mulrooney, Patrick, Rohde, Scott, Wang, Dan, Long, Stephen P., Dietze, Michael C., 2018. BETYdb: a yield, trait, and ecosystem service database applied to second-generation bioenergy feedstock production. *GCB Bioenergy* 10 (1), 61–71. <https://doi.org/10.1111/gcbb.12420>.
- Limani, Fidan, Latif, Atif, Borst, Timo, Tochtermann, Klaus, 2019. Metadata challenges for Long tail research data infrastructures. *Bibliothek Forschung Und Praxis*. <https://doi.org/10.1515/bfp-2019-2014>.
- Lin, Yan-Shih, Medlyn, Belinda E., Duursma, Remko A., Colin Prentice, I., Wang, Han, Baig, Sofia, Eamus, Derek, et al., 2015. Optimal stomatal behaviour around the world. *Nat. Clim. Chang.* <https://doi.org/10.1038/nclimate2550>.
- Long, S.P., Bernacchi, C.J., 2003. Gas exchange measurements, what can they tell us about the underlying limitations to photosynthesis? Procedures and sources of error. *J. Exp. Bot.* 54 (392), 2393–2401. <https://doi.org/10.1093/jxb/erg262>.
- Long, S.P., Farage, P.K., Garcia, R.L., 1996. Measurement of leaf and canopy photosynthetic CO₂ exchange in the field. *J. Exp. Bot.* <https://doi.org/10.1093/jxb/47.11.1629>.
- López-Calcagno, Patricia E., Brown, Kenny L., Simkin, Andrew J., Fisk, Stuart J., Viallet-Chabrand, Silvere, Lawson, Tracy, Raines, Christine A., 2020. Stimulating photosynthetic processes increases productivity and water-use efficiency in the field. *Nature Plants* 6 (8), 1054–1063. <https://doi.org/10.1038/s41477-020-0740-1>.
- Loreto, Francesco, Harley, Peter C., Di Marco, Giorgio, Sharkey, Thomas D., 1992. Estimation of mesophyll conductance to CO₂ flux by three different methods. *Plant Physiol.* 98 (4), 1437–1443. <https://doi.org/10.1104/pp.98.4.1437>.
- Maire, Vincent, Wright, Ian J., Colin Prentice, I., Batjes, Niels H., Bhaskar, Radika, van Bodegom, Peter M., Will K. Cornwell, et al., 2015. Global effects of soil and climate on leaf photosynthetic traits and rates: effects of soil and climate on photosynthetic traits. *Glob. Ecol. Biogeogr.* 24 (6), 706–717. <https://doi.org/10.1111/gcb.12296>.
- Maxwell, Kate, Johnson, Giles N., 2000. Chlorophyll fluorescence—a practical guide. *J. Exp. Bot.* 51 (345), 659–668. <https://doi.org/10.1093/jxb/51.345.659>.
- Mayernik, Matthew S., Breseman, Kelsey, Downs, Robert R., Duerr, Ruth, Garretson, Alexis, Hou, Chung-Yi, 2020. Risk assessment for scientific data. *Data Sci. J.* <https://doi.org/10.5334/dsj-2020-010>.
- Murchie, E.H., Lawson, T., 2013. Chlorophyll fluorescence analysis: a guide to good practice and understanding some new applications. *J. Exp. Bot.* 64 (13), 3983–3998. <https://doi.org/10.1093/jxb/ert208>.
- Niinemets, Ülo, Keenan, Trevor F., Hallik, Lea, 2015. A worldwide analysis of within-canopy variations in leaf structural, chemical and physiological traits across plant functional types. *The New Phytol.* 205 (3), 973–993. <https://doi.org/10.1111/nph.13096>.
- Ort, Donald R., Merchant, Sabeeha S., Alric, Jean, Barkan, Alice, Blankenship, Robert E., Bock, Ralph, Croce, Roberta, et al., 2015. Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *Proc. Natl. Acad. Sci.* <https://doi.org/10.1073/pnas.1424031112>.
- Pérez-Harguindeguy, N., Díaz, S., Garnier, E., Lavorel, S., Poorter, H., Jaureguierry, P., Bret-Harte, M.S., et al., 2013. New handbook for standardised measurement of plant

- functional traits worldwide. *Aust. J. Bot.* 61 (3), 167. <https://doi.org/10.1071/BI12225>.
- Ricciuto, Daniel, Sargsyan, Khachik, Thornton, Peter, 2018. The impact of parametric uncertainties on biogeochemistry in the E3SM land model. *J. Adv. Model. Earth Systems*. <https://doi.org/10.1002/2017ms000962>.
- Rodeghiero, Mirco, Niinemets, Ulo, Cescatti, Alessandro, 2007. "Major diffusion leaks of clamp-on leaf cuvettes still unaccounted: how erroneous are the estimates of Farquhar et al. model parameters?" *Plant. Cell & Environment* 30 (8), 1006–1022. <https://doi.org/10.1111/j.1365-3040.2007.001689.x>.
- Rogers, Alistair, Medlyn, Belinda E., Dukes, Jeffrey S., Bonan, Gordon, von Caemmerer, Susanne, Dietze, Michael C., Kattge, Jens, et al., 2017. A roadmap for improving the representation of photosynthesis in earth system models. *The New Phytol.* 213 (1), 22–42. <https://doi.org/10.1111/nph.14283>.
- Rogers, Alistair, Ely, Kim, Serbin, Shawn, 2019. Leaf Photosynthetic Parameters: Quantum Yield, Convexity, Respiration, Gross CO₂ Assimilation Rate and Raw Gas Exchange Data, Barrow, Alaska, 2016. In: Next Generation Ecosystem Experiments Arctic Data Collection. Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee, USA. <https://doi.org/10.5440/1482338>.
- Sack, Lawren, Cornwell, Will K., Santiago, Louis S., Barbour, Margaret M., Choat, Brendan, Evans, John R., Munns, Rana, Nicotra, Adrienne, 2010. A unique web resource for physiology, ecology and the environmental sciences: PrometheusWiki. *Funct. Plant Biol.* 37 (8), 687. <https://doi.org/10.1071/FP10097>.
- Sansone, Susanna-Assunta, McQuilton, Peter, Rocca-Serra, Philippe, Gonzalez-Beltran, Alejandra, Izzo, Massimiliano, Lister, Allyson L., Thurston, Milo, FAIRsharing Community, 2019. FAIRsharing as a community approach to standards, repositories and policies. *Nat. Biotechnol.* 37 (4), 358–367. <https://doi.org/10.1038/s41587-019-0080-8>.
- Sargsyan, Khachik, Safta, Cosmin, Najm, Habib N., Debusschere, Bert J., Ricciuto, Daniel, Thornton, Peter, 2014. Dimensionality reduction for complex models via Bayesian compressive sensing. *Int. J. Uncertain. Quantif.* <https://doi.org/10.1615/int.j.uncertaintyquantification.2013006821>.
- Schneider, Florian D., Fichtmueller, David, Gossner, Martin M., Güntsch, Anton, Jochum, Malte, König-Ries, Birgitta, Le Provost, Gaëtane, et al., 2019. Towards an ecological trait-data standard. *Methods Ecol. Evol.* <https://doi.org/10.1111/2041-210x.13288>.
- Sharkey, Thomas D., Bernacchi, Carl J., Farquhar, Graham D., Singsaas, Eric L., 2007. Fitting photosynthetic carbon dioxide response curves for C(3) leaves. *Plant Cell Environ.* 30 (9), 1035–1040. <https://doi.org/10.1111/j.1365-3040.2007.01710.x>.
- Simkin, Andrew J., López-Calcagno, Patricia E., Raines, Christine A., 2019. Feeding the world: improving photosynthetic efficiency for sustainable crop production. *J. Exp. Bot.* <https://doi.org/10.1093/jxb/ery445>.
- Smith, Nicholas G., Keenan, Trevor F., Colin Prentice, I., Wang, Han, Wright, Ian J., Niinemets, Ülo, Kristine Y. Crous, et al., 2019. Global photosynthetic capacity is optimized to the environment. *Ecol. Lett.* <https://doi.org/10.1111/ele.13210>.
- Thompson, Amber, Taylor, Barry N., . Guide for the Use of the International System of Units (SI) NIST Special Publication 811, 2008 Edition (version 3.2) [Online] Available. <http://physics.nist.gov/SP811> [2020-08-28]. National Institute of Standards and Technology, Gaithersburg, MD.
- Varadharajan, Charuleka, Cholia, Shreyas, Snaveley, Cory, Hendrix, Valerie, Procopiou, Christina, Swantek, Diana, Riley, William, Agarwal, D., 2019. Launching an accessible archive of environmental data. *Eos* 100. <https://doi.org/10.1029/2019EO111263>.
- Vitousek, Peter M., Ehrlich, Paul R., Ehrlich, Anne H., Matson, Pamela A., 1986. Human appropriation of the products of photosynthesis. *Bioscience* 36 (6), 368–373. <https://doi.org/10.2307/1310258>.
- von Caemmerer, S.V., Evans, J.R., 1991. Determination of the average partial pressure of CO₂ in chloroplasts from leaves of several C3 plants. *Funct. Plant Biol.* 18 (3), 287. <https://doi.org/10.1071/PP9910287>.
- Walker, Anthony P., De Kauwe, Martin G., Bastos, Ana, Belmecheri, Soumaya, Georgiou, Katerina, Keeling, Ralph, McMahon, Sean M., et al., 2020. Integrating the evidence for a terrestrial carbon sink caused by increasing atmospheric CO₂. *The New Phytol.* <https://doi.org/10.1111/nph.16866>.
- Wallis, Jillian C., Rolando, Elizabeth, Borgman, Christine L., 2013. If we share data, Will anyone use them? Data sharing and reuse in the Long tail of science and technology. *PLoS One* 8 (7), e67332. <https://doi.org/10.1371/journal.pone.0067332>.
- Warren, Charles, 2006. Estimating the internal conductance to CO₂ movement. *Funct. Plant Biol.* <https://doi.org/10.1071/fp05298>.
- Weerasinghe, Lasantha K., Creek, Danielle, Crous, Kristine Y., Xiang, Shuang, Liddell, Michael J., Turnbull, Matthew H., Atkin, Owen K., 2014. Canopy position affects the relationships between leaf respiration and associated traits in a tropical rainforest in far North Queensland. *Tree Physiol.* 34 (6), 564–584. <https://doi.org/10.1093/treephys/tpu016>.
- Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., et al., 2012. Darwin Core: an evolving community-developed biodiversity data standard. *PLoS One* 7 (1), e29715. <https://doi.org/10.1371/journal.pone.0029715>.
- Wilkinson, Mark D., Michel, Dumontier, Aalbersberg, I. J., Brandt, Jan, Appleton, Gabrielle, Axton, Myles, Baak, Arie, Blomberg, Niklas, et al., 2016. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* 3 (March), 160018. <https://doi.org/10.1038/sdata.2016.18>.
- Yang, Jinyan, Medlyn, Belinda E., De Kauwe, Martin G., Duursma, Remko A., Jiang, Mingkai, Kumarathunge, Dushan, Crous, Kristine Y., Gimeno, Teresa E., Wujeska-Klaue, Agnieszka, Ellsworth, David S., 2020. Low Sensitivity of Gross Primary Production to Elevated CO₂ in a Mature Eucalypt Woodland. *Biogeosciences* 17, 265–279. <https://doi.org/10.5194/bg-17-265-2020>.