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

Authors
Ely, Kim S
Rogers, Alistair
Agarwal, Deborah A
et al.

Publication Date
2021-03-01

DOI
10.1016/j.ecoinf.2021.101232

Peer reviewed

Instructions on how to contribute

- Contributions from all authors can be made using comments. To do this, highlight a section of text and click the “Add a comment” icon that appears on the right margin.
- Please add your contact details to the author list. See below for details.

NOTE V1 is now closed for comments. Please go to Version 2.

Title

“A metadata and data standard for archive of leaf-level gas exchange data”

Authors

If you have contributed to development of the data standard, or this paper, please add your name to the author list. Note that this list will also be used as a contact list for updates on the development of the standard and the progress of this paper. Inclusion on the list is not binding; we will finalize the author list and seek approval from all co-authors before submission. If you have required acknowledgements, please add them as a comment in the Acknowledgement section below.

Link to author list

Target journal

Ecological Informatics, special issue, "Integrating long-tail data: how far are we?"

Submission deadline 31 August 2020.

https://www.journals.elsevier.com/ecological-informatics/call-for-papers/integrating-long-tail-data

Note: Editors have been contacted and have approved submission of this paper for this special issue.

“Papers are invited that describe how data integration has been made more efficient through advances in: … standards developments”.

Highlights 3-5 points, < 85 characters

Abstract < 400 words

Leaf-level gas exchange data provide mechanistic understanding of plant and ecosystem fluxes of carbon and water. These data yield important parameterizations for terrestrial biosphere models and are necessary to understand the response of plants to global change. Collection of these data is both specialist and time consuming, and individual studies generally focus on limited species or restricted geographic regions. The high value of these data is recognized as evidenced by many publications that reuse and synthesize gas exchange data, however the lack of data and metadata standards make enhanced use of gas exchange data challenging. We have developed a standard reporting format for leaf-level gas exchange data and metadata to provide guidance to data contributors on how to store data in data repositories to maximize the value of that data and facilitate efficient data re-use. For data users, the standard will expand the capacity of data repositories to optimise data search and extraction, and more readily integrate similar data into synthesis products. The standard comprises metadata elements, standard vocabularies and required variables for survey measurements, dark respiration, CO_2 and light response curves, and parameters derived from those measurements. A crosswalk across the outputs of common instruments was developed to enable accurate data compilation. A process of extensive consultation with data collectors, data users and data scientists was undertaken to ensure that the standard would meet community needs. The standard presented here is intended to form a foundation for future development that will incorporate additional measurement types and variables. Access to the standard documentation, and future additions, will be enabled by hosting the standard on an open source version control system.

Keywords 4-6 keywords

Photosynthesis, stomatal conductance, carbon dioxide, irradiance

Main text < 7000 words, maximum 10000 words.

1. Introduction

The interface between plant and ecological sciences and research data infrastructure is rapidly evolving, with greater expectation for data preservation, reproducible and open research, and the potential to incorporate data into synthesis products. Moreover, publicly accessible data archiving is increasingly required by funding providers and publishers. Numerous databases and data repositories have been developed to fulfill these needs e.g. TRY (Kattge *et al.*, 2020), Environmental Data Initiative (environmentaldatainitiative.org), Dryad (datadryad.org) and figshare (figshare.com), yet the re-use of these data resources remains hampered by the difficulty of locating, harmonizing and assessing the quality of disparate 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 is the measurement of the flux of carbon dioxide and water vapor into and out of a leaf. Typically collected with portable infrared gas analyzers, these data are used to calculate a wide range of physiological traits, principally the rates of CO_2 assimilation, respiration and the stomatal conductance of water vapor. Gas exchange data are used to answer a wide range of scientific questions regarding plant function and their response to environmental change (Long *et al.*, 1996; Long & Bernacchi, 2003). They are also the basis of estimating and scaling photosynthesis from the leaf to canopy (Yang *et al.*, 2020), and are used to parameterize global biogeochemical models (Rogers *et al.*, 2017). The products of photosynthesis are critical to society, as they provide food, fuel and fibre, all of which are at the core of modern society (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; 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 carbon dioxide concentration and in turn the rate of global change (Walker *et al.*, 2020). Sensitivity 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; Sargsyan *et al.*, 2014; Ricciuto *et al.*, 2018). 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 specific experiments, instruments or environments. Thus resulting data products are typical long tail data, i.e. data are low volume and have diverse and heterogeneous content and context (Palmer *et al.*, 2007; Wallis *et al.*, 2013). Currently, most research data infrastructure focuses on generic metadata types, which limits the use of services such as search and data discovery for long tail data types (Limani *et al.*, 2019). Our review of existing data portals 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 re-use that data are often missing. The need for specialist data standards for specific 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 (Kissling *et al.*, 2018; Schneider *et al.*, 2019a; Gallagher *et al.*, 2020), but there has yet to be such a standard developed for leaf-level gas exchange data types. This is in spite of a recent increase in large compendia of such data (Lin *et al.*, 2015; Kumarathunge *et al.*, 2019; Smith *et al.*, 2019).

Data archiving is only the first step towards maximizing the value of data. In order to be re-used or incorporated into models or synthesis products, the data must be findable and accessible; these characteristics are optimized by appropriate, machine-readable search terms and persistent dataset identifiers. A key to interoperability and reusability is having sufficient metadata to correctly interpret the data, and process comparably across multiple studies from different sites, with various measurement methods (Christianson *et al.*, 2017). A lack of documentation and metadata is recognized as a data archiving risk factor (Mayernik *et al.*, 2020), with the implication that, without adequate metadata, data cannot be interpreted or used correctly. In order to reuse data, researchers often have to refer to original publications to access essential metadata, which can be a prohibitively resource intensive process, and still yield inconsistent results. Also, as research data infrastructure moves towards using application programming interfaces (APIs) to facilitate data upload and download, standardization of data and metadata, in machine readable formats, will become increasingly essential (Bruneau *et al.*, 2019). These needs are met in part by initiatives such as Darwin Core (Darwin Core Task Group, 2009) that provide a glossary of defined terms to describe biological data, yet specific guidance for leaf-level gas exchange data is lacking.

Here we present a data and metadata standard for the archive of a number of types of leaf-level gas exchange measurements, and describe the process of development of this standard. The approach taken here is to find the balance between maximizing the usefulness of the standard to the research community with ease of compliance. A key aspect has been engaging the community in the development of this data reporting standard, with a concerted effort to reach as many potential users as possible, seeking contributions and feedback. Our goal with this initial tightly focused effort on a leaf gas exchange standard is to develop a standard with broad consensus that provides a solid foundation for further development by the community. It is expected that future development of this standard will encompass a wider range of measurement types (e.g. fluorescence). An important key component of this proposed standard 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 standard for leaf-level gas exchange data reporting was initiated by a call for community accepted data standards 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 standard described here is known as the 'ESS-DIVE standard for data and metadata reporting of leaf-level gas exchange data', and referred to in this paper as 'the standard'. However, development of the standard and documentation has considered global needs for these data, and it will be available for implementation in other data repositories and databases. The standard is designed to be complementary, and not duplicate, existing metadata requirements, and 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 and sample-level metadata. It is encouraged that, where available, this standard be used in conjunction with established ontologies, such as Darwin Core (Darin Core Team, 2009), the Plant Ontology (Cooper *et al.*, 2013) and the Environment Ontology (Buttigieg *et al.*, 2013).

The scope of this standard for archive of leaf-level gas exchange data is focused on survey style measurements, and the response of photosynthesis to carbon dioxide (CO_2) and irradiance, and parameters derived from these relationships. In this paper we, 1) describe the process of developing the standard, including review of existing standards and community consultation; 2) provide details of the components of the standard, including the guidance for data and metadata fields, vocabularies and definitions; 3) demonstrate implementation of the standard in ESS-DIVE; and 4) discuss how this standard may be extended to cover a broader range of leaf-level gas exchange measurements.

2. Methods

2.1 Review of existing standards

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 (Schneider *et al.*, 2019b). Of these, databases and data 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 (Supplementary Tables 1 & 2). A catalogue of over 1400 data standards, including 17 categorised as concerning physiology, available at FAIRsharing (Sansone *et al.*, 2019), was searched for applicable standards.

2.1.2 Variables and definitions

Existing data repositories, databases and synthesized datasets were reviewed, and the most commonly used variable terms and definitions were adopted into this standard (Supplementary Tables 2 & 3). The TRY plant trait database (Kattge *et al.*, 2020) was identified as the only publicly available plant traits 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). It is noted that BETYdb specifies required and recommended covariates for several variables of interest. 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 (Sack *et al.*, 2010; Evans & Santiago, 2014). The use of variable names in the large datasets GlobResp (Atkin *et al.*, 2015) and GlobAmax (Maire *et al.*, 2015) were also considered. Default variable outputs, and their definitions, from eight commercially available portable gas analyzer instruments were compiled.

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. Here we identified metadata parameters that would allow users of leaf-level gas exchange data to discriminate between specific data types, experimental protocols and sample characteristics. The variables were chosen, and defined vocabularies were established, via input from the domain experts that contributed to the development of this standard. Our goal was to include metadata requirements and controlled vocabularies for variables that would be most relevant for synthesis activities, including variables to distinguish data obtained from natural or cultivated plants, and differentiate between common experimental manipulations and leaf sampling techniques.

2.2 Community consultation

The draft standard was made available for review and comment. Review was sought from data collectors, data scientists, data users (empiricists and modelers) and instrument manufacturers. The opportunity to participate was advertised via direct email to over eighty researchers identified as working in this field, and through social media. An introduction to the purpose, structure and components of the standard was presented as a free public webinar hosted by ESS-DIVE on 28 July 2020, followed by a month long period of feedback and discussion. For the purpose of review and feedback, the draft standard was made available for comment in a Google Sheets format. Review was conducted in an open manner, with comments and suggestions available to view by all reviewers. Feedback was received from over 50 contributors, and included over 130 separate comment threads on the standard draft document. Follow up video conferences were scheduled to discuss refinements and solutions. Suggestions for improvements made in initial rounds of review were considered and changes made to the standard and explanatory materials prior to the next round of review. The standard was then migrated to a public Github repository, where additions and refinements can continue to be made, and version controlled releases will be available for use, published in a Gitbooks format, along with templates to guide metadata compilation.

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 formal published data standards. This directed our efforts into development of a standard to meet this need. The range of measurements that can be made with portable gas exchange instruments is vast; the scope of this initial standard development was refined to some specific data types that have been the focus of recent synthesis activities; survey style gas exchange measurements, dark adapted respiration measurements, and CO_2 and light response curves, and parameters derived from those response curves. The standard comprises a number of components, being, a description of the data types, comprehensive metadata requirements with defined vocabularies, a list of standardized variable field names and definitions, required variables for each data type, and a cross walk of data outputs from common portable gas analyzer instruments (Figure 1). Each of these components is described in more detail in sections below; note that *camelCase* naming conventions used here follow the terminology used in the full standard documentation, available from ESS-DIVE (REF, link). 'Data package' is the term used to describe the collection of data and metadata files to be submitted to a data archive (Christianson *et al.*, 2017). Decisions were made as to how prescriptive the standard should be in order to maximize compliance by data contributors, and the benefit for data users. This standard specifies the minimum requirements for a data package that includes the data types described here; a data package may also include additional data types and variables not yet covered by the standard. A data package should also include general metadata as required by the hosting data archive or database (e.g. latitude and longitude).

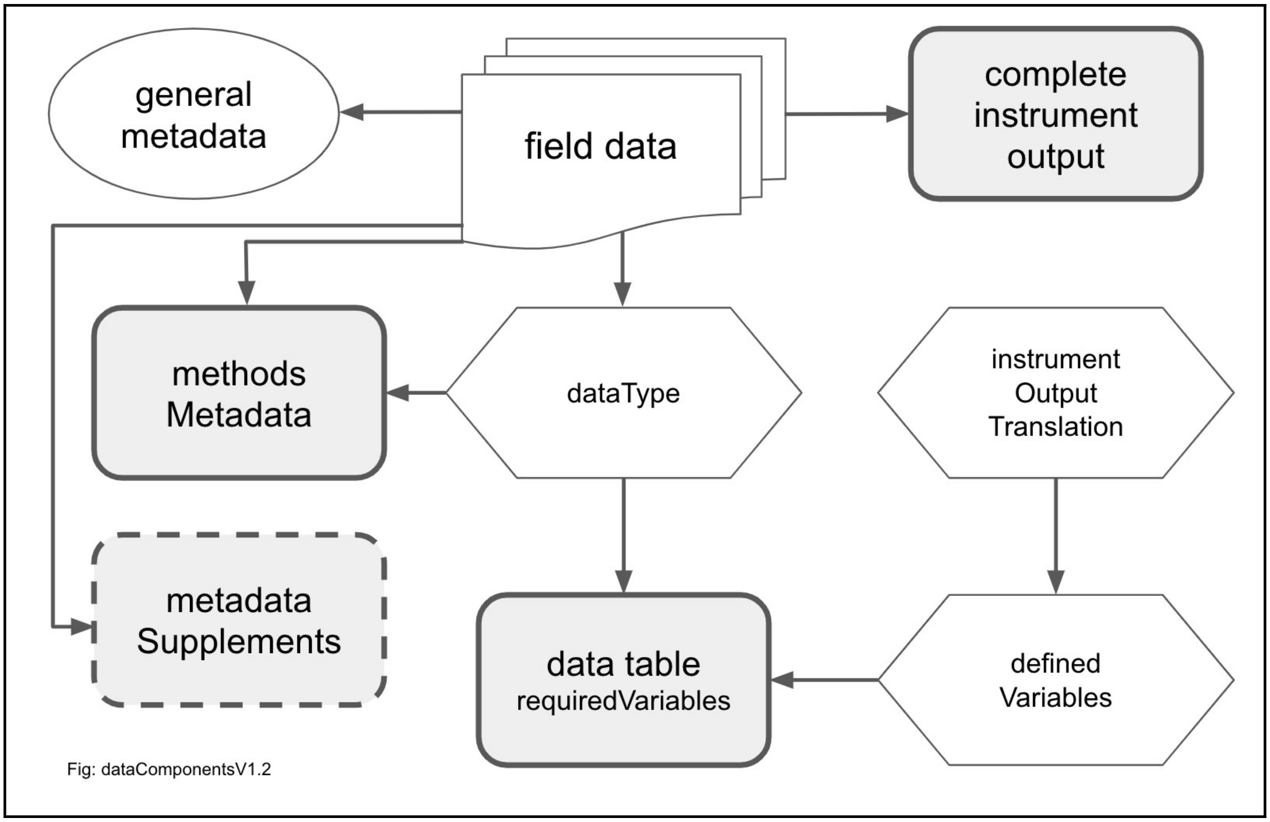

Figure 1. The relationships between field data and components of this standard for leaf-level gas exchange data. Grey boxes with solid borders indicate components required to be included in a data package. Use of the *metadataSupplements* component will be dependent on the experiment (grey box with dashed border). The requirements for general metadata (ellipse) will be set by each data archive; those details are not covered by this standard. Other components (hexagons) are informational, and guide the format and content requirements of the submitted data.

3.1 Data types

The *dataTypes* component provides detailed description and definition of the leaf-level gas exchange measurements in the scope of this standard. Seven *dataTypes* are defined (Table 1), and metadata requirements to describe the *measurementProtocols* used for each *dataType* are specified. Described *dataTypes* encompass common gas exchange measurements (e.g. photosynthetic CO_2 response curves), and analysis approaches (e.g. One point method).

3.2 Methods metadata

Comprehensive metadata provision is a key element in meeting FAIR principles, and enabling maximum re-use of data. While most data repositories have standard metadata requirements, these metadata cover the generic aspects of the data, such as authors, dates and locations. Systematic recording of detailed information about experimental conditions and protocols for leaf-level gas exchange is lacking. The *methodsMetadata* establishes a framework to ensure that data is well described. The provision of defined vocabularies for many variables will simplify metadata creation and the resulting consistency across datasets will enable more accurate search outcomes for data users. However, the diversity of experimental variables is recognised, and flexibility is allowed by the inclusion of free text options for many variables if the defined vocabulary is not adequate. The *methodsMetadata* captures a summary of the dataset and measurement protocols employed, and captures *dataTypes*, experimental and samples characteristics, and details of data processing and calculation approaches.

A significant discriminator for data users is the growth condition of the plants on which measurements were made. At the highest level, categorization of the *growthEnvironment* is an important discriminator. Details of *experimentalTreatment* can be employed by data users to include or exclude common treatments as appropriate. Further categorization is enabled by specification of *canopyPosition*, *lightExposure*, *leafAge* and *plantAge*. Refer to the *methodsMetadata* documentation for a complete list of variables, definitions and controlled vocabularies (link again?)

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 *methodsMetadata* 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 ($V_{c,max}$) are dependent on fitting approaches (Sharkey *et al.*, 2007; Gu *et al.*, 2010; Bernacchi *et al.*, 2013) the choice of kinetic constants (Rogers *et al.*, 2017), inclusion of mesophyll conductance (Ethier & 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) but ideally data users should return to the underlying data, i.e. the instrument output.

Leaf-level gas exchange data is often measured with the purpose of comparing between sample types or treatments; these discriminators are often included in data tables as codes to represent species, treatments, plots or other characteristics. The *methodsSupplements* component of this standard demonstrates how explanation of these descriptors should be included in a data package. Inclusion of *metadataSupplements* in a data package is highly dependent on the nature of the experiment, and as such, these examples are provided as guidelines only and are not required components. However, in the interest of achieving data equivalency in synthesis products, inclusion of the *instrumentation* metadata class, including a statement of instrument calibration, is highly recommended.

3.3 Inclusion of instrument output data

The *methodsMetadata* and *requiredVariables* (Section 3.5) are designed to capture adequate information to allow proper interpretation of datasets. However, not all use scenarios can be foreseen. The inclusion of the complete instrument outputs (commonly referred to as ‘raw data’) in a data package is seen as the ultimate future-proofing for a dataset. Archiving of raw data is recognised as good science practice and has been highlighted as important for the preservation and reuse of data (Dietze *et al.*, 2013; 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 recognise that instrument data with some quality control applied is 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, non-commercial gas exchange systems. Furthermore, it is recognized that full standardization of these data can require considerable effort, so in order to reduce this burden and allow researchers to upload the underlying instrument output the standard specifies three quality control options. 0 = ‘Data not available’. 1 = ‘Full instrument output with minimal quality control’. 2 = ‘Full instrument output with complete quality control, files are verified to only contain valid data, measurement values are reasonable and within expected range, area corrections are made where required’. Suggested quality control measures include correction of user input errors, removal of non-data rows such as test logs, and data points recognised as invalid. Measured values can be verified to fall within the expected range (see Section 3.5 for range values). The instrument output data should include all output variables (columns), for all valid data points (rows).

3.4 Variable names, definitions and units

The most common data field names (also known as headers) were designated a *variableName*, *variableUnit* and *variableDefinition*. These standards were developed based on the most common usage in existing databases and instrument outputs. In cases where common usage has not already been established, field names were selected to be human and machine readable, and with no recognised conflicts with other uses. Units for each variable are part of the variable definition, thus are not required to be part of the *variableName*. For *variableNames* that are *requiredVariables* (see Section 3.5), the standard also specifies an expected range of values; these limits can be used during data processing as part of a quality checking workflow.

3.5 Required variables for different data types

The list of minimum *requiredVariables* for each *dataType* was developed in order to capture the result variable (e.g. $V_{c,max}$) 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 *requiredVariables* presented in this standard are the result of an iterative feedback process of domain expert contributors. Again, a balance between the ideal dataset and the ease of compliance was considered during the development of these requirements. It should be noted that these lists are intended to be essential requirements, and the standard allows data contributors to include additional variables, using standard *variableNames*.

3.6 Instrument output translation table

The default output from eight 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 crosswalk to assess commonalities and provide a tool for data contributors and users. Instrument output varies in the types of variables, naming and units, and includes measured and calculated variables. The *instrumentOutputTranslation* table includes variables that are common across most instruments considered, and relevant to the *dataTypes* considered here. Twenty-four variables were compiled and cross referenced to the list of *variableNames* developed in this standard. The crosswalk is intended to act as a guide to users to translate their results to standard variables and units. This standard does not require translation of field headers for full instrument data included in a data package. The crosswalk may also assist data users to understand instrument output from unfamiliar instruments.

4. Discussion

We have developed a gas exchange data standard for the ESS-DIVE data repository that is available to the community through (link). Over 50 data contributors, data users, manufacturers and data scientists contributed to the development of this initial standard which we hope will form a foundation for future development by the community. The standard 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 (Figure 2). This has not been afforded by the recent set of gas exchange data compendia (Ali *et al.*, 2015; Lin *et al.*, 2015; De Kauwe *et al.*, 2016; Kumarathunge *et al.*, 2019; Smith *et al.*, 2019; Kattge *et al.*, 2020). The standard represents a compromise between data contributors and users that we hope reflects the consensus of the community and provides a readily usable but valuable contribution.

4.1 Development of a community standard

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

The development of the standard aimed to balance requirements for an unburdensome standard that would be readily adopted by data contributors with the desire for detail by data users. We have strived to develop a standard that requires the essential information that will enable efficient search and reuse of data. For these data types we have provided controlled vocabulary, a definition and units. We recognize that other detail is often desired and therefore we enable data contributors to describe unique measurement conditions, variables or approaches and provide more detail if desired. Importantly the community recognized and highlighted the desire to preserve the original instrument output. We strongly encourage the preservation of the underlying instrument output but recognize that this is not always possible, may involve inconsistencies based on the instrument used, and in some cases a major burden.

While a formal standard had not existed before we started this work, the vocabulary of leaf level gas exchange was well established and very similar between instrumentation. Therefore, incorporating many variables and definitions that are already in widespread use resulted in large parts of the standard 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. Some feedback conflated the goal of developing a data standard with documentation of measurement protocols or defining a gold standard method. The data standard does not attempt to constrain method choice by data contributors but be inclusive of all approaches and methodology. However, there were several issues that garnered lots of comments that are worth discussing further.

4.2 Decisions and compromises

As expected there was a necessary compromise between the desire for additional metadata detail and the desire for a simple and manageable standard for data contributors. Many of the requests for increased metadata would increase the effort, and therefore the barrier, to uploading data from some contributors whilst providing only limited value for most data users. Our selected metadata and co-variable requirements aim to ensure that the minimum information is collected using controlled vocabulary, definitions and units. We have resisted adding requirements for variables that “would be nice to have” or that are not absolutely required to reuse the data. There is no restriction preventing conscientious data contributors adding more detail and 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. Furthermore, one benefit of a living standard is that the community can comment upon and help evolve and expand the standard with time.

There were several comments about missing measurements, in many cases this just reflected the desire to expand the standard to cover more measurement types e.g. temperature and vapor pressure deficit response curves. The combination of fluorescence with gas exchange data is very powerful and for many instruments is standard. Whilst we recognise the value of including fluorescence data, doing so would have significantly expanded the scope of the standard. Early on we recognized that inclusion of fluorescence data would be out of reach for the initial development of the standard, particularly since these data can be collected with many more instruments which are often not associated with coincident gas exchange. In addition, the vocabulary and protocols are also not as well constrained as gas exchange measurements.

Estimates of photosynthetic parameters from photosynthesis and intercellular CO_2 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 CO_2 concentration in the chloroplast (C_c) - the site of carboxylation. Whilst g_m and hence C_c can be estimated from gas exchange data (Ethier & Livingston, 2004; Sharkey *et al.*, 2007). The most robust approaches require in-line measurements of fluorescence or isotopic descrimination (Evans *et al.*, 1986; Bongi & Loreto, 1989; Caemmerer & Evans, 1991; Harley *et al.*, 1992; Loreto *et al.*, 1992). Estimates of photosynthetic parameters based on C_i are lower than those that account for g_m and so it is important to distinguish what data (C_i or C_c) were used to calculate the derived parameters, and 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 have added metadata requirements for photosynthetic CO_2 response data to capture assumptions about g_m and flag the existence of additional data.

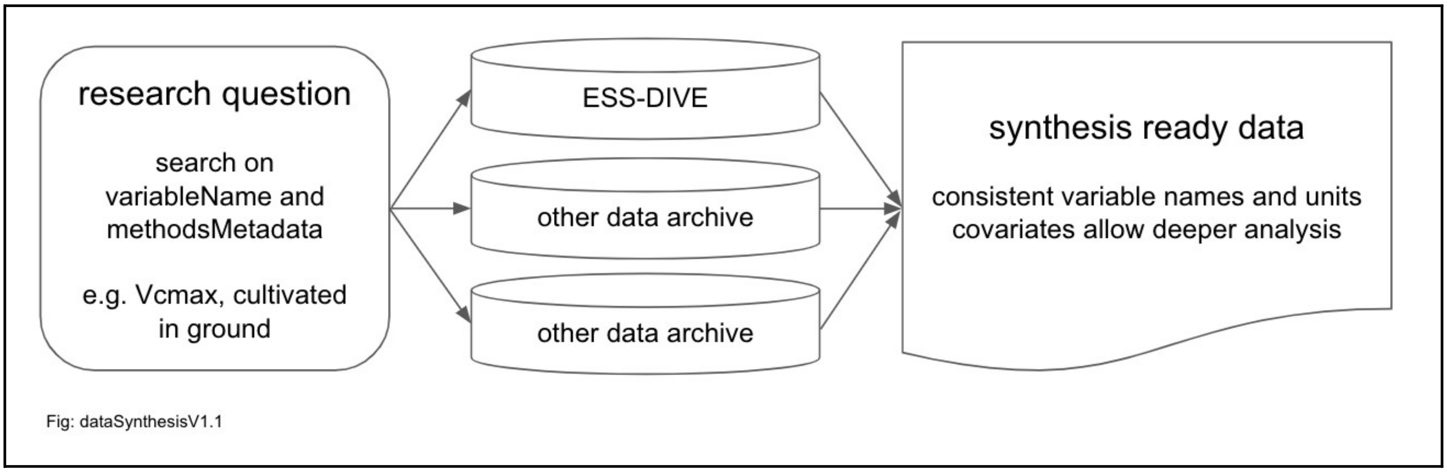

Figure 2. Schematic showing how the implementation of this data standard across data archives will facilitate data discovery and re-use.

4.3 Useability and Future developments

The standard will be a dynamic document. It will be hosted on an open access repository,Github. Github is, more often, being used to provide transparent tracking of any changes to text-based documents like data (Bryan, 2018). When standards are updated, the changes on GitHub are pushed and rendered into a user-friendly gitbook. Github also allows the user community to flag issues, discuss amendments and prioritize development of the standard, including addition of new measurement types (e.g. fluorescence), all in the open so the community can understand the motivation behind development and contribute to decision making. Standards for additional measurement types can be developed and added as compatible modules. Whilst the scope of this initial version of standard is limited to selected data types. It is hoped that standards for other measurement types will be developed and compatible with this standard.

Acknowledgements

Authors, please add your required acknowledgements as a comment.

KSE and AR were supported by ESS-DIVE community funds, through the Office of Biological and Environmental Research in the Department of Energy, Office of Science, and through the United States Department of Energy contract No. DE-SC0012704 to Brookhaven National Laboratory.

Supplementary materials

Full documentation of standard to be posted on Github (still to come)

Supplementary Tables 1 - 4

References

Extra refs not in Paperpile

Darwin Core Task Group (2009) Darwin Core (Kampmeier G, review manager). Biodiversity Information Standards (TDWG) http://www.tdwg.org/standards/450

Paperpile

Ainsworth EA, Rogers A, Leakey ADB. **2008**. Targets for Crop Biotechnology in a Future High-CO2 and High-O3 World. *Plant Physiology* **147**: 13–19.

Ali AA, Xu C, Rogers A, McDowell NG, Medlyn BE, Fisher RA, Wullschleger SD, Reich PB, Vrugt JA, Bauerle WL, *et al.* 2015. Global-scale environmental control of plant photosynthetic capacity. *Ecological Applications* **25**: 2349–2365.

Atkin OK, Bloomfield KJ, Reich PB, Tjoelker MG, Asner GP, Bonal D, Bönisch G, Bradford MG, Cernusak LA, Cosio EG, *et al.* 2015. Global variability in leaf respiration in relation to climate, plant functional types and leaf traits. *The New phytologist* **206**: 614–636.

Bernacchi CJ, Bagley JE, Serbin SP, Ruiz-Vera UM, Rosenthal DM, Vanloocke A. **2013**. Modelling C_3 photosynthesis from the chloroplast to the ecosystem. *Plant, cell & environment* **36**: 1641–1657.

Bonan GB, Lawrence PJ, Oleson KW, Levis S, Jung M, Reichstein M, Lawrence DM, Swenson SC. **2011**. Improving canopy processes in the Community Land Model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data. *Journal of Geophysical Research* **116**.

Bongi G, Loreto F. **1989**. Gas-Exchange Properties of Salt-Stressed Olive (Olea europea L.) Leaves. *Plant physiology* **90**: 1408–1416.

Booth BBB, Jones CD, Collins M, Totterdell IJ, Cox PM, Sitch S, Huntingford C, Betts RA, Harris GR, Lloyd J. **2012**. High sensitivity of future global warming to land carbon cycle processes. *Environmental research letters: ERL [Web site]* **7**: 024002.

Bruneau A, Borges LM, Allkin R, Egan AN, de la Estrella M, Javadi F, Klitgaard B, Miller JT, Murphy DJ, Sinou C, *et al.* 2019. Towards a new online species-information system for legumes. *Australian systematic botany* **53**: 1.

Bryan J. **2018**. Excuse Me, Do You Have a Moment to Talk About Version Control? *The American Statistician* **72**: 20–27.

Buttigieg PL, Morrison N, Smith B, Mungall CJ, Lewis SE, ENVO Consortium. **2013**. The environment ontology: contextualising biological and biomedical entities. *Journal of biomedical semantics* **4**: 43.

Caemmerer SV, Evans JR. **1991**. Determination of the Average Partial Pressure of CO2 in

Chloroplasts From Leaves of Several C3 Plants. *Functional plant biology: FPB* **18**: 287.

Christianson DS, Varadharajan C, Christoffersen B, Detto M, Faybishenko B, Gimenez BO, Hendrix V, Jardine KJ, Negron-Juarez R, Pastorello GZ, *et al.* 2017. A metadata reporting framework (FRAMES) for synthesis of ecohydrological observations. *Ecological informatics* **42**: 148–158.

Cooper L, Walls RL, Elser J, Gandolfo MA, Stevenson DW, Smith B, Preece J, Athreya B, Mungall CJ, Rensing S, *et al.* 2013. The plant ontology as a tool for comparative plant anatomy and genomic analyses. *Plant & cell physiology* **54**: e1.

De Kauwe MG, Lin Y-S, Wright IJ, Medlyn BE, Crous KY, Ellsworth DS, Maire V, Prentice IC, Atkin OK, Rogers A, *et al.* 2016. A test of the 'one-point method' for estimating maximum carboxylation capacity from field-measured, light-saturated photosynthesis. *The New phytologist* **210**: 1130–1144.

Dietze MC, Lebauer DS, Kooper R. **2013**. On improving the communication between models and data. *Plant, cell & environment* **36**: 1575–1585.

Ellsworth DS, Thomas R, Crous KY, Palmroth S, Ward E, Maier C, DeLucia E, Oren R. **2012**. Elevated CO2 affects photosynthetic responses in canopy pine and subcanopy deciduous trees over 10 years: a synthesis from Duke FACE. *Global Change Biology* **18**: 223–242.

Ethier GJ, Livingston NJ. **2004**. On the need to incorporate sensitivity to CO2 transfer conductance into the Farquhar-von Caemmerer-Berry leaf photosynthesis model. *Plant, Cell and Environment* **27**: 137–153.

Evans JR, Santiago LS. **2014**. PrometheusWiki Gold Leaf Protocol: gas exchange using LI-COR 6400. *Functional plant biology: FPB* **41**: 223–226.

Evans JR, Sharkey TD, Berry JA, Farquhar GD. **1986**. Carbon Isotope Discrimination measured Concurrently with Gas Exchange to Investigate CO2 Diffusion in Leaves of Higher Plants. *Functional plant biology: FPB* **13**: 281.

Flexas J, Díaz-Espejo A, Berry JA, 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. *Journal of experimental botany* **58**: 1533–1543.

Gallagher RV, Falster DS, Maitner BS, Salguero-Gómez R, Vandvik V, Pearse WD, Schneider FD, Kattge J, Poelen JH, Madin JS, *et al.* 2020. Open Science principles for accelerating trait-based science across the Tree of Life. *Nature ecology & evolution* **4**: 294–303.

Garnier E, Stahl U, Laporte M-A, Kattge J, Mougenot I, Kühn I, Laporte B, Amiaud B, Ahrestani FS, Bönisch G, *et al.* 2017. Towards a thesaurus of plant characteristics: an ecological contribution (P Vesk, Ed.). *The Journal of ecology* **105**: 298–309.

Gu L, Pallardy SG, Tu K, Law BE, Wullschleger SD. **2010**. Reliable estimation of biochemical parameters from C3 leaf photosynthesis-intercellular carbon dioxide response curves. *Plant, Cell & Environment* **33**: 1852–1874.

Halbritter AH, De Boeck HJ, Eycott AE, Reinsch S, Robinson DA, Vicca S, Berauer B, Christiansen CT, Estiarte M, Grünzweig JM, *et al.* 2020. The handbook for standardized field and laboratory measurements in terrestrial climate change experiments and observational studies (ClimEx) (R Freckleton, Ed.). *Methods in ecology and evolution / British Ecological Society* **11**: 22–37.

Harley PC, Loreto F, Di Marco G, Sharkey TD. **1992**. Theoretical Considerations when

Estimating the Mesophyll Conductance to CO(2) Flux by Analysis of the Response of Photosynthesis to CO(2). *Plant physiology* **98**: 1429–1436.

Kattge J, Bönisch G, Díaz S, Lavorel S, Prentice IC, Leadley P, Tautenhahn S, Werner GDA, Aakala T, Abedi M, *et al.* 2020. TRY plant trait database - enhanced coverage and open access. *Global change biology* **26**: 119–188.

Kissling WD, Walls R, Bowser A, Jones MO, Kattge J, Agosti D, Amengual J, Basset A, van Bodegom PM, Cornelissen JHC, *et al.* 2018. Towards global data products of Essential Biodiversity Variables on species traits. *Nature ecology & evolution* **2**: 1531–1540.

Kumarathunge DP, Medlyn BE, Drake JE, Tjoelker MG, Aspinwall MJ, Battaglia M, Cano FJ, Carter KR, Cavaleri MA, Cernusak LA, *et al.* 2019. Acclimation and adaptation components of the temperature dependence of plant photosynthesis at the global scale. *New Phytologist* **222**: 768–784.

LeBauer D, Kooper R, Mulrooney P, Rohde S, Wang D, Long SP, Dietze MC. **2018**. BETYdb: a yield, trait, and ecosystem service database applied to second-generation bioenergy feedstock production. *GCB Bioenergy* **10**: 61–71.

LeBauer DS, Wang D, Richter KT, Davidson CC, Dietze MC. **2013**. Facilitating feedbacks between field measurements and ecosystem models. *Ecological Monographs* **83**: 133–154.

Limani F, Latif A, Borst T, Tochtermann K. **2019**. Metadata Challenges for Long Tail Research Data Infrastructures. *Bibliothek Forschung und Praxis* **43**: 68–74.

Lin Y-S, Medlyn BE, Duursma RA, Colin Prentice I, Wang H, Baig S, Eamus D, de Dios VR, Mitchell P, Ellsworth DS, *et al.* 2015. Optimal stomatal behaviour around the world. *Nature Climate Change* **5**: 459–464.

Long SP, Bernacchi CJ. **2003**. Gas exchange measurements, what can they tell us about the underlying limitations to photosynthesis? Procedures and sources of error. *Journal of experimental botany* **54**: 2393–2401.

Long SP, Farage PK, Garcia RL. **1996**. Measurement of leaf and canopy photosynthetic CO2exchange in the field. *Journal of Experimental Botany* **47**: 1629–1642.

Loreto F, Harley PC, Di Marco G, Sharkey TD. **1992**. Estimation of Mesophyll Conductance to CO 2 Flux by Three Different Methods. *Plant physiology* **98**: 1437–1443.

Maire V, Wright IJ, Prentice IC, Batjes NH, Bhaskar R, van Bodegom PM, Cornwell WK, Ellsworth D, Niinemets Ü, Ordonez A, *et al.* 2015. Global effects of soil and climate on leaf photosynthetic traits and rates: Effects of soil and climate on photosynthetic traits. *Global ecology and biogeography: a journal of macroecology* **24**: 706–717.

Mayernik MS, Breseman K, Downs RR, Duerr R, Garretson A, Hou C-Y (sophie). **2020**. Risk Assessment for Scientific Data. *Data Science Journal* **19**.

Niinemets Ü, Keenan TF, Hallik L. **2015**. A worldwide analysis of within-canopy variations in leaf structural, chemical and physiological traits across plant functional types. *The New phytologist* **205**: 973–993.

Ort DR, Merchant SS, Alric J, Barkan A, Blankenship RE, Bock R, Croce R, Hanson MR, Hibberd JM, Long SP, *et al.* 2015. Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *Proceedings of the National Academy of Sciences* **112**: 8529–8536.

Palmer CL, Cragin MH, Heidorn PB, Smith LC. **2007**. Data curation for the long tail of science: The case of environmental sciences. In: Third International Digital Curation Conference.

Pérez-Harguindeguy N, Díaz S, Garnier E, Lavorel S, Poorter H, Jaureguiberry P, Bret-Harte MS, Cornwell WK, Craine JM, Gurvich DE, *et al.* 2013. New handbook for standardised measurement of plant functional traits worldwide. *Australian journal of botany* **61**: 167.

Ricciuto D, Sargsyan K, Thornton P. **2018**. The Impact of Parametric Uncertainties on Biogeochemistry in the E3SM Land Model. *Journal of Advances in Modeling Earth Systems* **10**: 297–319.

Rodeghiero M, Niinemets U, Cescatti A. **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**: 1006–1022.

Rogers A, Medlyn BE, Dukes JS, Bonan G, von Caemmerer S, Dietze MC, Kattge J, Leakey ADB, Mercado LM, Niinemets Ü, *et al.* 2017. A roadmap for improving the representation of photosynthesis in Earth system models. *The New phytologist* **213**: 22–42.

Sack L, Cornwell WK, Santiago LS, Barbour MM, Choat B, Evans JR, Munns R, Nicotra A. **2010**. A unique web resource for physiology, ecology and the environmental sciences: PrometheusWiki. *Functional plant biology: FPB* **37**: 687.

Sansone S-A, McQuilton P, Rocca-Serra P, Gonzalez-Beltran A, Izzo M, Lister AL, Thurston M, FAIRsharing Community. **2019**. FAIRsharing as a community approach to standards, repositories and policies. *Nature biotechnology* **37**: 358–367.

Sargsyan K, Safta C, Najm HN, Debusschere BJ, Ricciuto D, Thornton P. **2014**. DIMENSIONALITY REDUCTION FOR COMPLEX MODELS VIA BAYESIAN COMPRESSIVE SENSING. *International Journal for Uncertainty Quantification* **4**: 63–93.

Schneider FD, Fichtmueller D, Gossner MM, Güntsch A, Jochum M, König-Ries B, Le Provost G, Manning P, Ostrowski A, Penone C, *et al.* 2019a. Towards an ecological trait-data standard. *Methods in Ecology and Evolution* **10**: 2006–2019.

Schneider FD, Fichtmueller D, Gossner MM, Güntsch A, Jochum M, König-Ries B, Le Provost G, Manning P, Ostrowski A, Penone C, *et al.* 2019b. Towards an ecological trait-data standard (D Orme, Ed.). *Methods in ecology and evolution / British Ecological Society* **10**: 2006–2019.

Sharkey TD, Bernacchi CJ, Farquhar GD, Singsaas EL. **2007**. Fitting photosynthetic carbon dioxide response curves for C(3) leaves. *Plant, cell & environment* **30**: 1035–1040.

Simkin AJ, López-Calcagno PE, Raines CA. **2019**. Feeding the world: improving photosynthetic efficiency for sustainable crop production. *Journal of Experimental Botany* **70**: 1119–1140.

Smith NG, Keenan TF, Colin Prentice I, Wang H, Wright IJ, Niinemets Ü, Crous KY, Domingues TF, Guerrieri R, Yoko Ishida F, *et al.* 2019. Global photosynthetic capacity is optimized to the environment. *Ecology Letters* **22**: 506–517.

Varadharajan C, Cholia S, Snavely C, Hendrix V, Procopiou C, Swantek D, Riley W, Agarwal D. **2019**. Launching an Accessible Archive of Environmental Data. *Eos* **100**.

Vitousek PM, Ehrlich PR, Ehrlich AH, Matson PA. **1986**. Human Appropriation of the Products of Photosynthesis. *Bioscience* **36**: 368–373.

Walker AP, De Kauwe MG, Bastos A, Belmecheri S, Georgiou K, Keeling R, McMahon SM, Medlyn BE, Moore DJP, Norby RJ, *et al.* 2020. Integrating the evidence for a terrestrial carbon sink caused by increasing atmospheric CO 2. *The New phytologist*.

Wallis JC, Rolando E, Borgman CL. **2013**. If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. *PloS one* **8**: e67332.

Warren C. **2006**. Estimating the internal conductance to CO2 movement. *Functional Plant Biology* **33**: 431.

Weerasinghe LK, Creek D, Crous KY, Xiang S, Liddell MJ, Turnbull MH, Atkin OK. **2014**. Canopy position affects the relationships between leaf respiration and associated traits in a tropical rainforest in Far North Queensland. *Tree physiology* **34**: 564–584.

Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE, *et al.* 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data* **3**: 160018.

Yang J, Medlyn BE, De Kauwe MG, Duursma RA, Jiang M, Kumarathunge D, Crous KY, Gimeno TE, Wujeska-Klause A, Ellsworth DS. **2020**. Low sensitivity of gross primary production to elevated CO_2 in a mature eucalypt woodland.