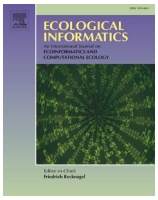
Ecological Informatics 62 (2021) 101280

Contents lists available at ScienceDirect

Ecological Informatics

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



A reporting format for field measurements of soil respiration

Ben Bond-Lamberty [a,*], Danielle S. Christianson [b], Robert Crystal-Ornelas [b], Kayla Mathes [c], Stephanie C. Pennington [a]

[a] Pacific Northwest National Laboratory, Joint Global Change Research Institute at the University of Maryland–College Park, College Park, MD 20740, USA
[b] Lawrence Berkeley National Laboratory, Berkeley, CA 94609, USA
[c] Integrated Life Sciences, Virginia Commonwealth University, Richmond, VA 23284, USA

ARTICLE INFO

Keywords:
Soil respiration
Soil gas flux
Data reporting format
Metadata
Data standard
FAIR data

ABSTRACT

Field observations of the soil-to-atmosphere CO_2 flux—soil respiration, R_S—are a prime example of 'long tail' data that historically have had neither centralized databases nor an agreed-upon reporting format. This has hindered scientific transparency, analytical reproducibility, and syntheses with respect to this globally-important component of the carbon cycle. Here we propose a new data and metadata reporting format for R_S data, based on engagement with a wide range of researchers in the earth and ecological sciences as well as expert advisory panels. Our goal was a reporting format that would be relevant and useful for synthesis activities, optimizing data discoverability and usability while not placing an undue burden on data contributors. We describe previous R_S data collection efforts, lessons learned from related databases and data-oriented networks (e.g., FLUXNET) in earth and ecological sciences, and the process of community consultation. The proposed reporting format focuses on chamber-level data and metadata, specifying measurement conditions and, for a given measurement period defined by beginning and ending timestamps, a mean R_S flux (or CO_2 concentration) and associated ancillary measurements. With input from the research community, we have also developed research data and metadata templates to support data collection adhering to the reporting format. Fundamentally, this format aims to enable findable, accessible, interoperable, and reusable data, while providing 'future-proofing' capabilities to support reanalyses using as yet unknown algorithms or approaches. This proposed R_S reporting format is openly available online and is intended to be a dynamic document, subject to further community feedback and/or change.

1. Introduction

Science is rapidly becoming more collaborative and data-intensive (Adams, 2012), and data-sharing and data-archival practices are changing as well. Journals increasingly specify and enforce data access and archival policies (Nosek et al., 2015); funding agencies now generally require detailed data management plans, open access to primary data, and use of established repositories (Borgman, 2012); and there is a growing recognition that taxpayer-funded research must be publicly available (Neylon, 2012). Encouragingly, publications with openly-available data seem to garner more citations (Dai et al., 2018; McKiernan et al., 2016). Enabling these changes is a challenge, but defining data standards and then making research data available in centralized, standardized repositories and databases is relatively straightforward for centralized, coordinated efforts such as National Ecological Observatory Network (NEON) (Schimel et al., 2007).

In many research fields, however, the majority of science is done by individual scientists leading small teams (Wu et al., 2019) producing small, heterogeneous, and ad hoc (in terms of standardization) datasets. Defining data standards and making these research data available in centralized repositories and databases is technically and culturally challenging. This 'long tail' of disparate, fragmented data almost certainly encapsulates massive amounts of scientific information (Dietze, 2014), but these datasets are difficult to access or synthesize (Wallis et al., 2013). They may be characterized by a disproportionate number of negative or non-significant results, producing a 'file drawer effect' that skews subsequent meta-analyses (Heidorn, 2008; Rosenthal, 1979). Troublingly, we know that such dispersed, unarchived data will inevitably be lost over time (Reichman et al., 2011; Vines et al., 2014). This issue is particularly relevant for research fields related to global change, as the exact same system climatic state will never recur in the future (Wolkovich et al., 2012).

* Corresponding author.
E-mail address: bondlamberty@pnnl.gov (B. Bond-Lamberty).

https://doi.org/10.1016/j.ecoinf.2021.101280
Received 16 November 2020; Received in revised form 19 February 2021; Accepted 12 March 2021
Available online 27 March 2021

In ecology and biogeosciences, a prime example of this problem revolves around soil respiration (R_S), the flux of CO_2 between soils and the atmosphere. R_S constitutes the second-largest flux in the global carbon cycle (Luo and Zhou, 2006), and its changes driven by climate, land use, and other factors portend significant climatic feedback (Bond-Lamberty et al., 2018; Liu et al., 2020). In addition, R_S data can be used as a cross-check on other components of the carbon cycle (Barba et al., 2018; Phillips et al., 2017; Wang et al., 2017). Unfortunately, soil respiration measurement instruments do not share a common machine-readable output format; this property is crucial as the scientific research infrastructure increasingly leverages application programming interfaces (APIs) for data upload and download, and scripting languages for reproducible analyses that can be scaled to larger datasets and questions. Ideally these datasets follow reporting formats that provide human-readable supporting documentation to encourage format adoption across individual scientists, teams, and international institutions. The ultimate goal is to make the resulting datasets both 'machine actionable' (sensu Wilkinson et al., 2016) and Findable, Accessible, Interoperable, and Reproducible (FAIR) (Stall et al., 2019). In fact, a precondition for machine actionable and FAIR data is adoption of data standardization (Bezuidenhout 2020; Sansone et al. 2019).

In contrast to the eddy covariance community measuring photosynthesis and land-atmosphere CO_2 exchange, there has traditionally been no centralized, standardized repository for R_S data akin to FLUXNET (Baldocchi et al., 2001) or AmeriFlux (Novick et al., 2018). R_S datasets remain widely dispersed and frequently unavailable, although efforts have been made to collect and standardize annual data in, for example, the widely-used global Soil Respiration Database or SRDB (Bond-Lamberty and Thomson, 2010; Jian et al., 2020), as well as a daily to seasonal analogue (Jian et al., 2018) More recently, Bond-Lamberty et al. (2020) presented an open database ("COSORE") for continuous R_S data. Nationally-oriented databases (e.g. Xu et al., 2015) also exist. While valuable, these efforts are individual- (as opposed to community-) driven and disparate, with no common data format linking them. This limits both the archival of FAIR data (because it takes more work for individual researchers to standardize their data) and subsequent synthesis efforts that might link and leverage multiple standardized databases (Jian et al., 2020).

Here we propose a new data and metadata reporting format for R_S data, based on engagement with a wide range of researchers in the earth and ecological sciences as well as expert advisory panels. This work was prompted 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). This paper 1) describes the development of the format, including our review of existing standards and conventions and community consultation; 2) details the reporting format itself, including its guidance for data and metadata fields, vocabularies, units, definitions, and templates; and 3) discusses potential applications, limitations and complicating factors, the potential to include additional measurements, and how this reporting format can support future data re-analyses, management, and archiving efforts.

2. Methodology

2.1. Metadata specification and level of focus

The goal of this effort was to define both the *data* as well as accompanying *metadata* requirements and formats that would balance parsimony with interpretability and data reuse. In general, metadata document the content, format, and context of a data product (Michener and Jones, 2012), most critically describing who created, collected and managed the data, the data content and format, and when and where it was collected. Additional metadata information can include information on storage, generation, processing, quality control, and the study context (Fegraus et al., 2005).

We focused on a relatively small, core set of metadata aimed at documenting soil respiration flux measurements (and/or, as noted below, CO_2 concentration data). In general, we assume that this format is not responsible for site-level documentation and metadata nor for standardizing lower-level conventions on file formats, character encoding, or numeric representations (Fig. 1). Our goal was a reporting format that would be most relevant and useful for synthesis activities, and thus one distinguishing between natural and experimental measurement conditions. Controlled vocabularies (Soranno et al., 2015) were used when necessary for consistent metadata reporting. Finally, we sought to balance between optimizing data discoverability and usability, while not placing undue burden on data contributors (Fegraus et al., 2005); in our experience, the more onerous the data archival process is, the fewer the datasets that will be archived, leading to their near-inevitable loss over time (Vines et al., 2014).

2.2. Review of existing standards and database efforts

A critical first step was reviewing and learning from previous work in this area. Early R_S databases mainly consisted of syntheses of knowledge and previously published studies: Schlesinger (1977) summarized knowledge about the annual carbon balance of detritus and soil, for example, while Hibbard et al. (2005) synthesized annual R_S estimates pulled from larger flux networks. These early data collections were typically organized in unstructured tables in the publication itself, making subsequent reuse difficult. The Soil Respiration Database (SRDB; Bond-Lamberty and Thomson, 2010; Jian et al., 2020) offered a more usable structure, as it was (and remains) a synthesis of published hand-measured, chamber-level annual R_S data available in four standardized tables in machine-readable form: two data and two metadata, with loosely defined controlled vocabularies for many fields. The SRDB consequently has become a widely used resource, with the original paper cited over 300 times to date. The recently published COSORE (Bond-Lamberty et al., 2020) offered a philosophy and structure for continuous RS data that we drew from in our initial work designing this standard.

We also studied and leveraged lessons from the design and format

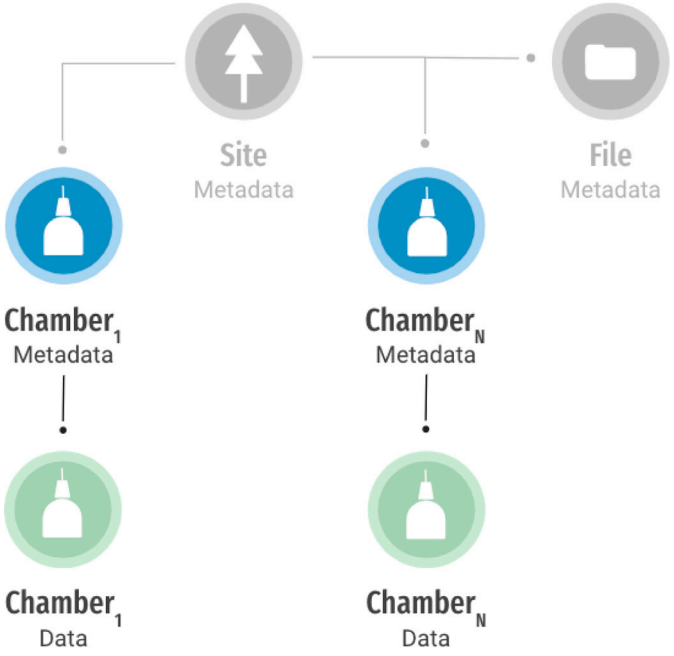


Fig. 1. This reporting format focuses on chamber-level *metadata* and *data* for soil respiration, R_s, the soil-to-atmosphere CO_2 flux. Metadata about the larger research site, data creators and contributors, and file encodings and standards are critical but assumed to be specified elsewhere.

decisions made by related ecological and earth sciences databases, in particular those with hierarchical tables linking metadata, site, and observational data together. These included the soil radiocarbon database ISRaD (Lawrence et al., 2020) and the soil incubation data database SIDb (Schädel et al., 2020). We examined their choices of ancillary data, handling of arbitrary temporal averaging periods, and choices made with respect to complexity versus completeness. As noted above, we also benefited from the concurrent development of COSORE (Bond-Lamberty et al., 2020), an effort to assemble an open community database of continuous and long-term R_S datasets. The simultaneous development produced interactions between COSORE and this reporting format that benefited both efforts. In particular, it meant that the nascent format was repeatedly confronted with real-world datasets, forcing us to consider carefully the tradeoffs of various choices.

Finally, we surveyed large networks like FLUXNET, AmeriFlux, and the Integrated Carbon Observation System (ICOS), which are both diverse and complex, and require extensive standardization for functionality. These networks focus on flux data from eddy covariance towers, meaning there are numerous, continuous, data streams flowing into these databases, requiring standards on all levels of data, as well as provenance (traceability) throughout. For example, FLUXNET and AmeriFlux use common unit names and timestamps for fluxes, which allows for compatibility between databases. ICOS (https://www.icos-cp.eu/) adopts EU data standards for spatial data to its data products at all levels (https://inspire.ec.europa.eu/) and emphasizes an end-to-end computational chain that can regenerate flux datasets from raw observations if needed. The reporting format described here was designed with an eye towards future interoperability with these efforts.

2.3. Community consultation

Community engagement provides critical feedback necessary for a usable and broadly accepted data format (Sansone et al. 2019). We used a survey aimed at users, managers, curators, and data advisors of earth science data (specifically chamber-level gas flux data) to gain consensus on the structure, reproducibility, and usability of soil respiration data. We collected feedback from 17 respondents on both the goals and structural details of the proposed format. The survey was sent to the community in three phases and designed to engage the broader community on the importance of specific variables, data types, and general structure. It included questions on the most important variables to include, eddy-covariance tower compatibility, and how to ensure proper data provenance. Respondents prioritized the inclusion of ancillary measurements (88% ranked high importance), transparent data provenance (64%), and site / chamber-level metadata (47%) (Fig. 2).

The reporting format was presented at a public webinar hosted by ESS-DIVE in July 2020 and opened to community feedback and discussion, and then in November 2020 posted to a public repository on GitHub (https://github.com/ess-dive-community/essdive-soil-respiration).

3. Results

Based on the research of existing standards and community engagement described above, this Rs reporting format includes chamber-level metadata (Table 1) and data (Table 2). Six required and nine optional metadata fields specify the measurement conditions: location, instrument, any experimental manipulations, etc. Three required and 20 optional data fields provide, for a given measurement period that is defined by beginning and ending timestamps, a mean R_S flux (or optionally CO_2 concentrations, from which fluxes could be recomputed; see discussion below) and associated ancillary measurements.

We did not attempt to define site-level metadata. Site-level descriptions and metadata are a common problem and need throughout field ecology and earth sciences (Fegraus et al., 2005; Reichman et al., 2011) and, with two exceptions (Table 3), we saw no point in reinventing these metadata here. The exceptions are 1) defining an offset from Coordinated Universal Time, needed for unambiguous timestamp interpretation, and 2) providing a mechanism for attaching raw data (as downloaded from measurement instruments) to a dataset. Inclusion of raw data constitutes an attempt to 'future-proof' datasets (Ely et al., 2021; Rogers et al., 2017), allowing for future re-computation using new methods, and is discussed further below.

All data reporting format documentation is publicly available in a GitHub repository (https://github.com/ess-dive-community/essdive-soil-respiration) and also as a user-friendly GitBook website (https://ess-di

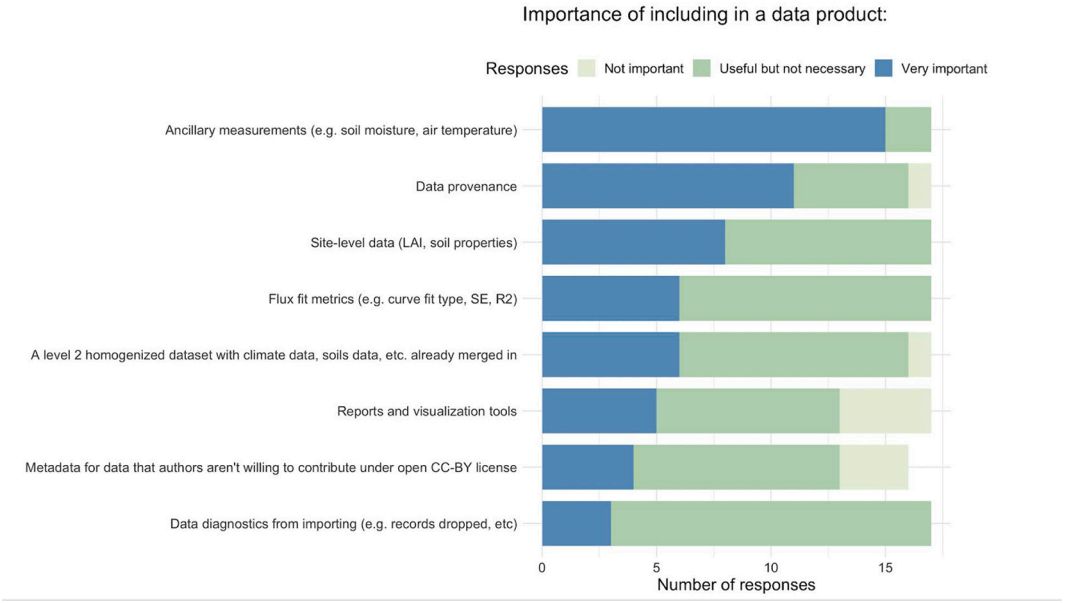


Fig. 2. Survey responses ranking the importance of various attributes for a soil respiration data product and reporting format. Blue represents "very important to include", green is "useful but not necessary", and light green is "not important". Data based on repeated surveys of researchers and data specialists (total $N = 17$). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Chamber-level metadata, including field name, description, format, unit, and requirement level. Each row in the table provides metadata for a given measurement location, typically a chamber and/or collar at the soil surface used to measure soil respiration; there will thus be as many rows as there are chambers for a given dataset. The longitude and latitude fields are intended to be variable-resolution, depending on available data, and may be the same across chambers, i.e. providing only a site-level location. "Logical" fields are intended to be given as "T/F" or "TRUE/FALSE". A unit entry of "CL" means that entry is a controlled list, and accepts only certain predefined values.

Field Name	Description	Format	Unit	Required?
CHAMBER_ID	Unique (within dataset) chamber identifier; must appear in flux data table as well	character		Y
MSMT_VAR	Type of flux measured. Options include: "Rs" (soil respiration), "Rh" (heterotrophic respiration only), "Reco" (ecosystem respiration), or "NEE" (net ecosystem exchange)	character	CL	Y
TREATMENT	Soil or ecosystem treatment applied at measurement location; default is "None"	character	CL	Y
LONGITUDE	Decimal longitude of measurement location, positive = east	numeric	°	
LATITUDE	Decimal latitude of measurement location, positive = north	numeric	°	
ELEVATION	Elevation of measurement location	numeric	m	
AREA	Soil surface measurement area	numeric	cm^2	
VOLUME	Volume of measurement chamber	numeric	cm^3	
COLLAR_DEPTH	Depth of collar insertion	numeric	cm	
OPAQUE	Opaque chamber?	logical		Y
PLANTS_REMOVED	Plants removed from inside the collar?	logical		Y
FAN	Mixing fan in chamber?	logical		
SPECIES	Comma-separated list of dominant species (genus + species) at measurement location, following http://www.theplantlist.org/	character		
SENSOR_DEPTHS	Comma-separated depths of solid-state sensors (applicable only for gradient-based measurement methods)	character	cm	
INSTRUMENT	Measurement instrument (i.e. model)	character	CL	Y
SOFTWARE_VERS	Instrument software version	character		
NOTES	Additional information about this measurement location	character		

ve.gitbook.io/continuous-soil-respiration-reporting-format/).
Community-developed data reporting formats are rarely static documents, and we have established these resources so that users of the data reporting format can submit GitHub issues that will help to prioritize any updates to the format. The GitHub repository will always contain the most-up-to-date version of the data reporting format and documentation, and major releases will be publicly archived in ESS-DIVE, a permanent data archive. The repository also provides templates to guide researchers putting their data into the reporting format (Fig. 3).

We use semantic versioning for this reporting format to track and indicate changes. Semantic versioning (https://semver.org/) follows an

Table 2
Fluxdata and associated ancillary data, including field name, description, format, unit, and requirement level. Each row in the table gives the average soil respiration flux, R_s, for a measurement period, along with diagnostic information and ancillary measurements such as air temperature, soil moisture, etc. The two timestamp fields TIMESTAMP_BEGIN and TIMESTAMP_END define the beginning and end, respectively, of the averaging period in local standard time. They are up to 14-digit integers depending on the data's temporal resolution, and may be given as YYYY (for annual data), YYYYMM (for monthly data), etc., to a maximum resolution of YYYYMMDDHHMMSS. A unit entry of "CL" means that entry is a controlled list, and accepts only certain predefined values.

Field Name	Description	Format	Unit	Required?
CHAMBER_ID	Unique (within dataset) chamber label; must appear in chamber data table as well	character		Y
TIMESTAMP_BEGIN	Timestamp in local standard time: beginning of averaging period	integer	Variable; see table caption	Y
TIMESTAMP_END	Timestamp in local standard time: end of averaging period	integer	Variable; see table caption	Y
DRY_CO2	Mean chamber CO_2 concentration during measurement period	numeric	ppmv	
FLUX_CO2	Soil respiration CO_2 flux (positive = to atmosphere)	numeric	μmol CO_2 m^{-2} s^{-1}	
FLUX_SE_CO2	Standard error of flux computation	numeric	μmol CO_2 m^{-2} s^{-1}	
CRVFIT_CO2	Flux computation method ("Lin" or "Exp" for linear and exponential)	character	CL	
R2_CO2	R^2 of flux computation	numeric	fraction	
ERROR_CODE	Error code generated by analyzer or during import; empty or zero indicates no error	numeric		
CO2_AMB	Ambient CO_2 concentration at measurement location	numeric	ppmv	
TAIR_AMB	Ambient air temperature at measurement location	numeric	°C	
TAIR	Air temperature inside measurement chamber	numeric	°C	
PAR	Photosynthetically active radiation inside measurement chamber	numeric	μmol m^{-2} s^{-1}	
RH	Relative humidity inside measurement chamber	numeric	%	
PAR_AMB	Photosynthetically active radiation outside measurement chamber	numeric	μmol m^{-2} s^{-1}	
VPD_AMB	Vapor pressure deficit at measurement location	numeric	Pa	

(continued on next page)

Table 2 (*continued*)

Field Name	Description	Format	Unit	Required?
PRECIP_AMB	Precipitation at measurement location	numeric	mm	
Tx	Soil temperature at *x* cm (for additional "Tx" fields, follow same format, e.g. T5, T15)	numeric	°C	
SMx	Volumetric soil moisture at *x* cm depth (for additional "SMx" fields, follow the same format, e.g. SM5, SM15)	numeric	$m^3\ m^{-3}$	
WTD	Water table depth at measurement location, positive numbers are depth	numeric	cm	
NOTES	Additional information about this measurement	character		

Table 3
Addition metadata relating to the entire dataset, including field name, description, format, unit, and requirement level. This table will have only one row per dataset. Note that in general, this reporting format assumes (cf. Fig. 1) that site-level metadata (soil information, ecological classification, etc.) are provided elsewhere.

Field Name	Description	Format	Unit	Required?
RAW_DATA	Link to raw data files	character		
UTC_OFFSET	Site offset from Coordinated Universal Time (UTC)	integer	hours	Y

x.y.z format, where *x* is the major version number (changing only when there are major changes to the format that provide fundamental new capabilities and/or may break existing scripts); *y* is the minor version number (signifying smaller but significant changes); and *z* the patch number (documentation typo fixes, or other changes that are completely backwards compatible). Following each official (major) release, a DOI will be issued and the reporting format permanently archived by Zenodo (https://zenodo.org/) and/or ESS-DIVE (https://ess-dive.lbl.gov/).

4. Discussion

4.1. Towards FAIR data

Reporting formats and, ultimately, data standards provide consistency and interpretability, making data more findable (by providing a pathway to data archiving), accessible (through free and open data repositories), and usable (Stall et al., 2019). More specifically, they are necessary (but not sufficient) for data being findable, accessible, interoperable, and reusable (FAIR, Wilkinson et al., 2016). The reporting format proposed here supports these goals by clearly specifying the location, time domain, instrumentation, and errors of R_S measurements. It is also a format developed with considerable community input, following recent research (Sansone et al. 2019) suggesting that community approaches lead to greater buy-in, even as they come with their own challenges.

It does not, however, aim to be sufficient for FAIR by itself; that is, this is not intended to be a stand-alone format. In particular, we have not attempted to define important metadata components such as data contributor metadata, site-level metadata (e.g. ecosystem type), or information about file format conventions or encoding, despite their undoubted importance (Vandenbussche and Vatant, 2011). The assumption is that these format specifications will be provided via common conventions adopted by networks and repositories such as NEON (Schimel et al., 2007), ESS-DIVE (Varadharajan et al., 2019) and AmeriFlux (Novick et al., 2018), LTER (Moore, 2016), or ICOS (Op de Beeck et al., 2018).

As discussed by Ely et al. (2021), funder mandates for data archival can be a burden for data contributors. We hope, however, that reporting formats such as this one relieve some of this burden by providing clear, community-agreed upon specifications that are straightforward and align with common data collection practices. Moreover, we provide user-friendly documentation and data reporting format templates that are intended to help users adopt and adhere to the formats. A well-designed reporting format also enables better data quality control, accelerates re-use and thus impact of shared data (Piwowar and Vision, 2013), and provides collaborative opportunities across research groups. An important challenge across much of science remains, however: ensuring that formal recognition of dataset re-use occurs, whether through data citations or another mechanism (Agarwal et al., 2021; Groth et al., 2020; Reichman et al., 2011). This is not straightforward if, for example, the original dataset is combined with other data into a larger database such as COSORE (Bond-Lamberty et al., 2020). Until citation-tracking systems enable adequate attribution of dataset

Data File

* Required fields are shaded blue. Any desired parameters would be additional columns. *

* For additional variables not listed here, please see ESS-DIVE data dictionary format guidelines (link in instructions.md) *

To use this template, please remove Rows 1-4 and save in Comma Separated Values format.

Columns	Chamber_ID	Timestamp_Begin	Timestamp_End	Flux_CO2	Dry_CO2	Flux_SE_CO2	CrvFit_CO2	R2_CO2	Error	C
Contains	*Unique chamber label; must appear in chamber data table as well*	*ISO timestamp begin of averaging period (up to a 12-digit integer as specified by the data's temporal resolution)*	*ISO timestamp END of averaging period (up to a 12-digit integer as specified by the data's temporal resolution)*	*CO2 flux (positive = to atmosphere)*	*Mean chamber CO2 concentration during measurement period*	*Standard error of CO2 flux computation*	*CO2 flux computation method ("Lin" or "Exp" for linear and exponential)*	*R2 of CO2 flux computation*	*Error generated by analyzer or during import*	*A c a n k*
Units		YYYYMMDDHHMMSS	YYYYMMDDHHMMSS	µmol CO2 m-2 s-1	ppmv	µmol CO2 m-2 s-1		fraction		p
Example	Chamber_1	20201106115505	20201106115605	2.15	435.22	0.008	Lin	0.88	Chamber never closed.	
Example	Control	2020110611	2020110611	0.58	451.89	0.002	Exp	0.9		

Fig. 3. Example of the template provided in the Github documentation at https://github.com/ess-dive-community/essdive-soil-respiration, corresponding to Table 2 (not all fields are shown here). Required fields are indicated in blue. Users can download and populate this template to easily follow the reporting standard. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

collections, the association with and thus credit to the original data contributor is lost if subsequent users cite only the primary data collection.

4.2. Enabling future reproducibility and re-analyses

This reporting format focuses on R_S *fluxes,* computed from the fundamental observation of CO_2 concentration (whether made through mass spectrometry, via infrared gas analyzer, or solid-state in-ground sensor) measured over a period of time (Tang et al., 2003; Xu et al., 2006). Of course, we cannot foresee all possible data uses in the future, or anticipate novel processing methods that may be developed. Thus, future-proofing, i.e., preserving unprocessed instrument output (Rogers et al., 2017) is highly desirable. Archiving raw instrument data in anticipation of future advances in algorithms, science questions, etc. is desirable across many domains of science (Sandve et al., 2013). Centralized earth science programs with end-to-end data systems, like ICOS (Integrated Carbon Observation System Research Infrastructure, 2020), have largely accomplished this goal.

This reporting format supports future-proofing in three ways. First, along with (or instead of) the computed R_s fluxes, it allows for the optional reporting of CO_2 concentrations at individual timepoints, which enables re-computation of the fluxes, for example with a different algorithm. This is useful but limited, as any custom instrument settings and/or researcher protocols are not preserved. For a more powerful capability and following Ely et al. (2021), this format provides for the archival of complete (raw) instrument output. *Re*-processing these raw data would be a complex step, but instrumentation outputs typically record significantly more information about the measurements, e.g. the analyzer state and settings as well as precise start and end times of all analyzer measurements and actions. Finally, the format is designed to be responsive to the current but also future needs of data contributors and user communities (see "Future Developments" below) and thus not marginalize or exclude any groups in the research community (Bezuidenhout 2020).

4.3. Limitations and compromises

As described above, any reporting format or data standard must decide the level of metadata detail that balances depth and breadth while maintaining the format's practicality and usability for data contributors. There are many possible metadata additions one could imagine that might increase the utility or benefit of R_S deposited data, such as ecosystem disturbance history, instrument dead band and repetition settings, etc. However, it remains unclear that the theoretical future gain would be worth the very tangible current burden; few of the many original metadata fields in the widely-used SRDB (Bond-Lamberty and Thomson, 2010) have ever been used in any analysis, for example. In contrast, many experimental and sampling details may be included in protocol descriptions or supplementary tables in publications. Importantly, nothing in this proposed format restricts data contributors from including more metadata detail or data types in addition to those listed in Tables 1–3, although metadata included in these non-standardized formats are difficult to use when data from many sources are combined in analyses.

4.4. Future developments

This proposed reporting format is a dynamic document, available online at https://github.com/ess-dive-community/essdive-soil-respiration and subject to further feedback and/or change as needed; it is emphatically not intended to be a finalized standard imposed on the community. Providing feedback (via GitHub issues or via email; this is documented on the webpage above at https://github.com/ess-dive-community/essdive-soil-respiration#how-to-contribute) allows users–data contributors, data consumers, and other interested parties–to raise issues, provide feedback, and prioritize changes and growth of the format in a public space with full records (version control) of changes. The published reporting format can be revised with minor edits, ensuring users can easily access the latest update. One obvious extension would be to add other greenhouse gases such as CH_4 into this framework, paralleling e.g. the capabilities of COSORE (Bond-Lamberty et al., 2020). Regardless of the exact future direction taken, or changes made, we hope it will contribute to enabling broader and FAIRer use of 'long tail' scientific data by researchers worldwide.

Declaration of Competing Interest

None.

Acknowledgements

B.B.-L., S.C.P., and R.C.-O. 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. D.S.C. was supported by the AmeriFlux Management Project funded by the DOE's Office of Science under Contract No. DE-AC02-05CH11231. We thank Kim Ely for her feedback and constructive criticism.

References

Adams, J., 2012. Collaborations: the rise of research networks. Nature 490, 335–336.
Agarwal, D., Damerow, J., Varadharajan, C., Christianson, D., Pastorello, G., Cheah, Y.-W., Ramakrishnan, L., 2021. Balancing the needs of consumers and producers for scientific data collections. Ecol. Inform. 101251.
Baldocchi, D.D., Falge, E., Gu, L., Olson, R.J., Hollinger, D.Y., Running, S.W., Anthoni, P. M., Bernhofer, C., Davis, K.J., Evans, R.D., 2001. FLUXNET: a new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. Bull. Am. Meteorol. Soc. 82 (11), 2415–2434.
Barba, J., Cueva, A., Bahn, M., Barron-Gafford, G.A., Bond-Lamberty, B., Hanson, P.J., Jaimes, A., Kulmala, L., Pumpanen, J., Scott, R.L., Wohlfahrt, G., Vargas, R., 2018. Comparing ecosystem and soil respiration: review and key challenges of tower-based and soil measurements. Agric. For. Meteorol. 249 (Supplement C), 434–443.
Bond-Lamberty, B., Thomson, A., 2010. A global database of soil respiration data. Biogeosciences 7 (6), 1915–1926 [online] Available from: https://www.biogeosciences.net/7/1915/2010/.
Bezuidenhout, L., 2020. Being Fair about the Design of FAIR Data Standards. Digit. Gov.: Res. Pract. 1 (3), 1–7. https://doi.org/10.1145/3399632.
Bond-Lamberty, B., Bailey, V.L., Chen, M., Gough, C.M., Vargas, R., 2018. Globally rising soil heterotrophic respiration over recent decades. Nature 560 (7716), 80–83.
Bond-Lamberty, B., Christianson, D.S., Malhotra, A., Pennington, S.C., Sihi, D., AghaKouchak, A., Anjileli, H., Altaf Arain, M., Armesto, J.J., Ashraf, S., Ataka, M., Baldocchi, D., Andrew Black, T., Buchmann, N., Carbone, M.S., Chang, S.-C., Crill, P., Curtis, P.S., Davidson, E.A., Desai, A.R., Drake, J.E., El-Madany, T.S., Gavazzi, M., Görres, C.-M., Gough, C.M., Goulden, M., Gregg, J., Gutiérrez del Arroyo, O., He, J.-S., Hirano, T., Hopple, A., Hughes, H., Järveoja, J., Jassal, R., Jian, J., Kan, H., Kaye, J., Kominami, Y., Liang, N., Lipson, D., Macdonald, C.A., Maseyk, K., Mathes, K., Mauritz, M., Mayes, M.A., McNulty, S., Miao, G., Migliavacca, M., Miller, S., Miniat, C.F., Nietz, J.G., Nilsson, M.B., Noormets, A., Norouzi, H., O'Connell, C.S., Osborne, B., Oyonarte, C., Pang, Z., Peichl, M., Pendall, E., Perez-Quezada, J.F., Phillips, C.L., Phillips, R.P., Raich, J.W., Renchon, A.A., Ruehr, N.K., Sánchez-Cañete, E.P., Saunders, M., Savage, K.E., Schrumpf, M., Scott, R.L., Seibt, U., Silver, W.L., Sun, W., Szutu, D., Takagi, K., Takagi, M., Teramoto, M., Tjoelker, M.G., Trumbore, S., Ueyama, M., Vargas, R., Varner, R.K., Verfaillie, J., Vogel, C., Wang, J., Winston, G., Wood, T.E., Wu, J., Wutzler, T., Zeng, J., Zha, T., Zhang, Q., Zou, J., 2020. COSORE: A community database for continuous soil respiration and other soil-atmosphere greenhouse gas flux data. Glob. Chang. Biol. https://doi.org/10.1111/gcb.15353 n/a(n/a).
Borgman, C.L., 2012. The conundrum of sharing research data. J. Assoc. Inf. Sci. Technol. 63 (6), 1059–1078.
Dai, S.-Q., Li, H., Xiong, J., Ma, J., Guo, H.-Q., Xiao, X., Zhao, B., 2018. Assessing the extent and impact of online data sharing in eddy covariance flux research. J. Geophys. Res. Biogeosci. 123 (1), 2017JG004277.
Dietze, M.C., 2014. Gaps in knowledge and data driving uncertainty in models of photosynthesis. Photosynth. Res. 119 (1–2), 3–14.
Ely, K., Rogers, A., Agarwal, D., Ainsworth, E., Albert, L., Yang, D., 2021. A reporting format for leaf-level gas exchange data and metadata. Ecol. Inform. 61 (101232), 101232. https://doi.org/10.1016/j.ecoinf.2021.101232.
Fegraus, E.H., Andelman, S., Jones, M.B., Schildhauer, M., 2005. Maximizing the value of ecological data with structured metadata: an introduction to ecological metadata language (EML) and principles for metadata creation. Bull. Ecol. Soc. Am. 86 (3), 158–168.

Groth, P., Cousijn, H., Clark, T., Goble, C., 2020. FAIR data reuse – the path through data citation. Data Intell. 2 (1–2), 78–86. https://doi.org/10.1162/dint_a_00030.
Heidorn, P.B., 2008. Shedding light on the dark data in the long tail of science. Libr. Trends 57 (2), 280–299.
Hibbard, K.A., Law, B.E., Sulzman, J., 2005. An analysis of soil respiration across northern hemisphere temperate ecosystems. Biogeochemistry 73 (1), 29–70.
Integrated Carbon Observation System Research Infrastructure: ICOS Handbook 2020, 2020. ICOS ERIC [online] Available from: https://www.icos-cp.eu/sites/default/files/cmis/ICOS%20Handbook%202020.pdf.
Jian, J., Steele, M.K., Day, S.D., Thomas, R.Q., 2018. Measurement strategies to account for soil respiration temporal heterogeneity across diverse regions. Soil Biol. Biochem. 125, 167–177.
Jian, J., Vargas, R., Anderson-Teixeira, K., Stell, E., Herrmann, V., Horn, M., Kholod, N., Manzon, J., Marchesi, R., Paredes, D., Bond-Lamberty, B., 2020. A restructured and updated global soil respiration database (SRDB-V5). Earth Syst. Sci. Data. https://doi.org/10.5194/essd-2020-136.
Lawrence, C.R., Beem-Miller, J., Hoyt, A.M., Monroe, G., Sierra, C.A., Stoner, S., Heckman, K., Blankinship, J.C., Crow, S.E., McNicol, G., Trumbore, S., Levine, P.A., Vindušková, O., Todd-Brown, K., Rasmussen, C., Hicks Pries, C.E., Schädel, C., McFarlane, K., Doetterl, S., Hatté, C., He, Y., Treat, C., Harden, J.W., Torn, M.S., Estop-Aragonés, C., Asefaw Berhe, A., Keiluweit, M., Della Rosa Kuhnen, Á., Marin-Spiotta, E., Plante, A.F., Thompson, A., Shi, Z., Schimel, J.P., Vaughn, L.J.S., von Fromm, S.F., Wagai, R., 2020. An Open-Source Database for the Synthesis of Soil Radiocarbon Data: International Soil Radiocarbon Database (ISRaD) version 1.0. https://doi.org/10.5194/essd-12-61-2020.
Liu, S., Zheng, Y., Ma, R., Yu, K., Han, Z., Xiao, S., Li, Z., Wu, S., Li, S., Wang, J., Luo, Y., Zou, J., 2020. Increased Soil Release of Greenhouse Gases Shrinks Terrestrial Carbon Uptake Enhancement under Warming. Glob. Chang, Biol. https://doi.org/10.1111/gcb.15156.
Luo, Y., Zhou, X., 2006. Soil Respiration and the Environment. Elsevier/Academic Press, Amsterdam.
McKiernan, E., Bourne, P.E., Brown, C.T., Buck, S., Kenall, A., Lin, J., McDougall, D., Nosek, B., Ram, K., Soderberg, C., Spies, J.R., Thaney, K., Updegrove, A., Woo, K.H., Yarkoni, T., 2016. How Open Science Helps Researchers Succeed, Elife, 2016. https://doi.org/10.7554/eLife.16800.
Michener, W.K., Jones, M.B., 2012. Ecoinformatics: supporting ecology as a data-intensive science. Trends Ecol. Evol. 27 (2), 85–93.
Moore, J.C., 2016. Long-Term Ecological Research and Lessons from Networked Lives, Long-Term Ecological Research. https://doi.org/10.1093/oso/9780199380213.003.0060.
Neylon, C., 2012. Science publishing: open access must enable open use. Nature 492, 348–349.
Nosek, B.A., Alter, G., Banks, G.C., Borsboom, D., Bowman, S.D., Breckler, S.J., Buck, S., Chambers, C.D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D.P., Hesse, B., Humphreys, M., Ishiyama, J., Karlan, D., Kraut, A., Lupia, A., Mabry, P., Madon, T., Malhotra, N., Mayo-Wilson, E., McNutt, M., Miguel, E., Paluck, E.L., Simonsohn, U., Soderberg, C., Spellman, B.A., Turitto, J., VandenBos, G., Vazire, S., Wagenmakers, E.J., Wilson, R., Yarkoni, T., 2015. Promoting an open research culture. Science 348 (6242), 1422–1425.
Novick, K.A., Biederman, J.A., Desai, A.R., Litvak, M.E., Moore, D.J.P., Scott, R.L., Torn, M.S., 2018. The AmeriFlux network: A coalition of the willing. Agric. For. Meteorol. 249 (Supplement C), 444–456.
Op de Beeck, M., Gielen, B., Merbold, L., Ayres, E., Serrano-Ortiz, P., Acosta, M., Pavelka, M., Montagnani, L., Nilsson, M., Klemedtsson, L., Vincke, C., De Ligne, A., Moureaux, C., Marañon-Jimenez, S., Saunders, M., Mereu, S., Hörtnagl, L., 2018. Soil-meteorological measurements at ICOS monitoring stations in terrestrial ecosystems. Int. Agrophys. 32 (4), 619–631.
Phillips, C.L., Bond-Lamberty, B., Desai, A.R., Lavoie, M., Risk, D., Tang, J., Todd-Brown, K., Vargas, R., 2017. The value of soil respiration measurements for interpreting and modeling terrestrial carbon cycling. Plant Soil 413 (1–2), 1–25.
Piwowar, H.A., Vision, T.J., 2013. Data reuse and the open data citation advantage. PeerJ 1, e175.
Reichman, O.J., Jones, M.B., Schildhauer, M.P., 2011. Challenges and opportunities of open data in ecology. Science 331 (6018), 703–705.
Rogers, A., Medlyn, B.E., Dukes, J.S., Bonan, G., von Caemmerer, S., Dietze, M.C., Kattge, J., Leakey, A.D.B., Mercado, L.M., Niinemets, Ü., Prentice, I.C., Serbin, S.P., Sitch, S., Way, D.A., Zaehle, S., 2017. A roadmap for improving the representation of photosynthesis in earth system models. New Phytol. 213 (1), 22–42.
Rosenthal, R., 1979. The file drawer problem and tolerance for null results. Psychol. Bull. 86 (3), 638–641.
Sandve, G.K., Nekrutenko, A., Taylor, J., Hovig, E., 2013. Ten simple rules for reproducible computational research. PLoS Comput. Biol. 9 (10), e1003285.
Sansone, S.-A., McQuilton, P., Rocca-Serra, P., Gonzalez-Beltran, A., Izzo, M., Lister, A.L., Thurston, M., 2019. FAIRsharing Community: 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.
Schädel, C., Beem-Miller, J., Aziz Rad, M., Crow, S.E., Hicks Pries, C.E., Ernakovich, J., Hoyt, A.M., Plante, A., Stoner, S., Treat, C.C., Sierra, C.A., 2020. Decomposability of Soil Organic Matter Over Time: The Soil Incubation Database (SIDb, Version 1.0) and Guidance for Incubation Procedures. https://doi.org/10.5194/essd-12-1511-2020.
Schimel, D., Hargrove, W., Hoffman, F., MacMahon, J., 2007. NEON: a hierarchically designed national ecological network. Front. Ecol. Environ. 5 (2), 59.
Schlesinger, W.H., 1977. Carbon balance in terrestrial detritus. Annu. Rev. Ecol. Syst. 8, 51–81.
Soranno, P.A., Bissell, E.G., Cheruvelil, K.S., Christel, S.T., Collins, S.M., Fergus, C.E., Filstrup, C.T., Lapierre, J.-F., Lottig, N.R., Oliver, S.K., Scott, C.E., Smith, N.J., Stopyak, S., Yuan, S., Bremigan, M.T., Downing, J.A., Gries, C., Henry, E.N., Skaff, N. K., Stanley, E.H., Stow, C.A., Tan, P.-N., Wagner, T., Webster, K.E., 2015. Building a multi-scaled geospatial temporal ecology database from disparate data sources: fostering open science and data reuse. Gigascience 4, 28.
Stall, S., Yarmey, L., Cutcher-Gershenfeld, J., Hanson, B., Lehnert, K., Nosek, B., Parsons, M., Robinson, E., Wyborn, L., 2019. Make scientific data FAIR. Nature 570 (7759), 27–29.
Tang, J., Baldocchi, D.D., Qi, Y., Xu, L., 2003. Assessing soil CO2 efflux using continuous measurements of CO2 profiles in soils with small solid-state sensors. Agric. For. Meteorol. 118, 207–220.
Vandenbussche, P.-Y., Vatant, B., 2011. Metadata recommendations for linked open data vocabularies. Version 1, 2011–2012.
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, 182796–182813.
Vines, T.H., Albert, A.Y.K., Andrew, R.L., Débarre, F., Bock, D.G., Franklin, M.T., Gilbert, K.J., Moore, J.-S., Renaut, S., Rennison, D.J., 2014. The availability of research data declines rapidly with article age. Curr. Biol. 24 (1), 94–97.
Wallis, J.C., Rolando, E., Borgman, C.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.
Wang, X., Wang, C., Bond-Lamberty, B., 2017. Quantifying and reducing the differences in forest CO2-fluxes estimated by eddy covariance, biometric and chamber methods: a global synthesis. Agric. For. Meteorol. 247, 93–103.
Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3, 160018.
Wolkovich, E.M., Regetz, J., O'Connor, M.I., 2012. Advances in global change research require open science by individual researchers. Glob. Chang. Biol. 18 (7), 2102–2110.
Wu, L., Wang, D., Evans, J.A., 2019. Large teams develop and small teams disrupt science and technology. Nature 566 (7744), 378–382.
Xu, L., Furtaw, M.D., Madsen, R.A., Garcia, R.L., Anderson, D.J., McDermitt, D.K., 2006. On maintaining pressure equilibrium between a soil CO 2 flux chamber and the ambient air. J. Geophys. Res. 111 (D8), 225.
Xu, Z., Tang, S., Xiong, L., Yang, W., Yin, H., Tu, L., Wu, F., Chen, L., Tan, B., 2015. Temperature sensitivity of soil respiration in China's forest ecosystems: patterns and controls. Appl. Soil Ecol. 93, 105–110.