**Scope Guidelines**

**Data Descriptors** submitted to *Scientific Data* should provide detailed descriptions of valuable research datasets, including the methods used to collect the data and technical analyses supporting the quality of the measurements. Data Descriptors focus on helping others reuse data, rather than testing hypotheses, or presenting new interpretations, methods or in-depth analyses. Relevant datasets must be deposited in an appropriate public repository prior to Data Descriptor submission, and their completeness will be considered during editorial evaluation and peer review. The data must be made publicly available without restriction in the event that the Data Descriptor is accepted for publication (excepting reasonable controls related to human privacy issues or public safety).

### A global soil health database for soil health assessment

Jinshi Jian1, Ryan D. Stewart1\*, Xuan DU2,

1. School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, VA, USA

2. Department of Hydraulic Engineering, Yangling Vocational & Technical College, Yang Ling, Shaanxi, China

Corresponding author: Ryan D. Stewart (ryan.stewart@vt.edu)

### Abstract

*170 words maximum*

The Abstract should succinctly describe the study, the assay(s) performed, the resulting data, and their reuse potential, but should not make any claims regarding new scientific findings. No references are allowed in this section.

Field studies have been performed for decades on cropland conservation management, but the data from historical papers are not integrated to serve future cropland soil quality management. Reliable soil health database (SoilHealthDB) is essential to investigate cash crop yield, soil physical, chemical, and biological properties when cropland alter from traditional management to conservation management. Here, we collect, digitized, and integrated a global soil health dataset based on published ﬁeld measurements. We extracted 5,241 entries of data from 281 published studies for four main cropland conservation management methods: cover crop, no-tillage, agro-forestry systems, and organic farm. Data are from 324 geographic sites (location with different latitude and longitude) in 45 countries around the world. The SoilHealthDB including 42 soil health indicators and 46 background information. Our database provides a common data-frame-work for sharing soil health measurements in the future. For some soil health indicators with abundance of comparisons, meta-analysis can be applied to analyze its response under conservation management. SoilHealthDB could also support a comprehensive analysis of soil health changes related with cropland conservation management, therefore support a soil health calculator developing.

|  |  |
| --- | --- |
| Design Types | data integration objective |
| Measurement Type(s) | 42 soil health indicators (details please see table 2) |
| Technology Type(s) | data extracted from published papers |
| Factor Type(s) |  |
| Sample Characteristic(s) | Cover crop, no-tillage, organic farm, soil health, soil quality, |

### Background & Summary

*700 words maximum*

The Background & Summary should provide an overview of the study design, the assay(s) performed, and the data generated, including any background information needed to put this study in the context of previous work and the literature, and should reference literature as needed. The section should also briefly outline the broader goals that motivated collection of the data, as well as their potential reuse value. We also encourage authors to include a figure that provides a schematic overview of the study and assay(s) design.

Cropland soil degradation due to natural vegetation remove, unsuitable agricultural activities, and erosion is among the main factors causing soil quality and crop yield decreases. According to a recent report from the Food and Agriculture Organization of the United Nations (FAO), one-third of soils in the world are infertile due to unsustainable land-use management practices1. Cropland conservation management practices, including the use of cover crops within rotations, change from traditional tillage to no-tillage, have been proposed as a way to increase soil carbon and soil health, and many on-site experiments have been conducted to identify the effects of including cover crops and no-tillage on soil properties. Despite the concept of conservation management has been proposed for decades, there is little consistency in what indicators should be measured and how to actually quantify the soil quality improvement after conservation management. In addition, contradictions exist among studies, for example, introducing cover crop as rotation during fallow season usually enhances soil organic carbon2, yet sometimes does not3–5.

In order to averaging means across different individual study with heterogeneous variation, ssystematic review and meta-analysis were conducted in the past to evaluate the effects of cover crops6, no-tillage7,8, organic farm9, and agroforestry systems10 on crop yield and soil properties. However, there are still many challenges in understanding the factors impacting soil health at global scale owing largely to lack of global observation-based dataset of soil health with corresponding information on soil health indicators, background information, and management practices. Therefore, it is essential to collect the historical published data and integrate them together to support a comprehensive quantitative assessment on soil quality improvement by conservation management.

We collected studies which compared agricultural production with traditional management and conservational management, publication meet the criteria were digitized and the data were integrated into a global soil health database (SoilHealthDB). This dataset is freely available and can be continuously update by including newly publication. Such a global soil health dataset based on ﬁeld measurements can be used for providing statistical analysis (e.g., meta-analyses) on a specific soil health indicator, to understand factors impacting soil health and cash crop yield. In addition, our soil health data can be potentially upscaled to a global observation-based soil health calculator that would provide valuable information for scientists in agriculture and soil health improvement and for farmers to determine future cropland management planning. Lastly, the SoilHealthDB including soil erosion, runoff, and nutrient leaching indicators, provide an opportunity to evaluate the non-point pollution under traditional and conservational management, the principal sources of which are conversion of natural vegetation covered land to intensive agriculture, where animal manures, atmospheric deposition, and commercial fertilizers contribute a major water contaminant in the globe.

### Methods

The Methods should include detailed text describing any steps or procedures used in producing the data, including full descriptions of the experimental design, data acquisition assays, and any computational processing (e.g. normalization, image feature extraction). See the [detailed section in our submission guidelines](https://www.nature.com/sdata/publish/submission-guidelines#sec-5) for advice on writing a transparent and reproducible methods section. Related methods should be grouped under corresponding subheadings where possible, and methods should be described in enough detail to allow other researchers to interpret and repeat, if required, the full study. Specific data outputs should be explicitly referenced via data citation (see Data Records and Data Citations, below).

Authors should cite previous descriptions of the methods under use, but ideally the method descriptions should be complete enough for others to understand and reproduce the methods and processing steps without referring to associated publications. There is no limit to the length of the Methods section.

**Data collection**

We designed a soil health data framework, hereafter named SoilHealthDB, which include 46 background information (Table 1), and 42 soil health indicators (Table 2). We conducted a systematic literature search of studies on field comparisons between traditional and conservational management. Currently, we targeted four major conservational management methods: cover crop (CC), no-tillage (NT), organic farm (OG), and agro-forestry systems (AFS).

Publication was searched and collected from three sources: (1) previous soil health related tools such as the “Research Landscape Tool” that compiles soil health results into a searchable database and includes publication and research projects; (2) citied papers from previous meta-analysis or review papers such as Poeplau & Don (2015), Alvarez *et al.* (2017), Sileshi (2009), and Gattinger *et al.* (2012); and (3) additional publications were gathered via a literature search using ISI Web of Science, Google scholar, and the China National Knowledge Infrastructure (CNKI) for papers. We used the key words “Soil health” or “soil quality” and “conservational management” or “cover crop” or “green manure” or “organic farm” or “no-till” or “agroforestry systems” in ISI Web of Science, Google scholar, and the CNKI to search published papers. Papers from peer reviewed journals, conference collections, theses, and dissertations are included. No other restriction was considered on the language, publication data, and other filtering criteria. We collected totally more than 500 papers, we then read through articles and used the following criteria to determine whether the publication would be included in this study: (1) experiments were conducted in the field or research station; (2) the publications reported comparison between control (traditional management) and treatment (conservational management); (3) provide at least one comparison between control and treatment of soil health indicators (Table 2). Within these constraints, 281 papers.

Data are digitized from tables and figures, the software Data Thief was used to read the data from figures (version III, <http://datathief.org/>). Background information was extracted from the publication and fit into 46 background indicators (Table 1). When latitude and longitude were not reported in the literature, they were collected from <https://www.findlatitudeandlongitude.com> based on the site’s name, elevation was identified by latitude and longitude in <https://www.freemaptools.com/elevation-finder.htm> when it was not reported in the original paper. The extracted data from 281 papers resulted in 5241 comparisons across the globe (Figure 1), approximately 20 comparisons in one study, and some of the comparisons are not independent. We allocate a unique experiment ID to a comparison if the CC group, cash crop group, site, tillage, fertilization, soil depth, termination, or rotation is different within a comparison (Figure 2), which resulted in a total of 1177 experiments. We assume each experiment is independent of each other.

After the location information is carefully checked, sites’ climate region can be identified according to climate Koppen classification (<http://koeppen-geiger.vu-wien.ac.at/>) and sites’ latitude and longitude (please see the supplemental R code). The missing MAT and MAP values were filled by a global air temperature and precipitation data from the Center for Climate Research at the University of Delaware (Willmott *et al.*, 2001), the MAP and MAT were calculated based on the monthly precipitation and temperature between 1961 and 2015. Soil texture was grouped into coarse (sand, loamy sand, and sandy loam), medium (sandy clay loam, loam, silt loam, and silt), and fine (clay, sandy clay, clay loam, silty clay, and silty clay loam) based on The Cornell Framework of Soil Health Manual 16.

Finally, the cash crops were grouped into corn, soybean, wheat, other monoculture, corn-soybean rotation (CS), corn-soybean-wheat (CSW), and other rotation of more than two cash crops. The cover crops were grouped into broadleaf, grass, legume, mixed of more than two legumes (LL), mixed of legume and grass (LG), and other mixed of more than two cover crops (MTT). Soil sampling depths were grouped into 0-10cm, 0-20cm, 0-30cm, and 30-100cm (Figure 3). It should be note that we group the cash crop, cover crop, and soil sampling depth in our own way, the user can regroup them according to the detailed cash crop, cover crop, and soil sampling depth information according to their research objects.

**Code availability**

For all studies using custom code in the generation or processing of datasets, a statement must be included in the Methods section, under the subheading "Code availability", indicating whether and how the code can be accessed, including any restrictions to access. This section should also include information on the versions of any software used, if relevant, and any specific variables or parameters used to generate, test, or process the current dataset.

All the data processing and data view were conducted under R (version 3.5.1)14. Source code is available on the Open Science Framework. The code is detailed commended with instructions for use. The intent of the code is to illustrate how some missing background information were fill back other data sources, and to explain how some soil health indicators are grouped based on the basic information. The code and the SoilHealthDB are freely for individual, academic, research, and commercial usage, but it could not be repacked or sold without written permission.

### Data Records

The Data Records section should be used to explain each data record associated with this work, including the repository where this information is stored, and to provide an overview of the data files and their formats. Each external data record should be cited using the data citation format presented at the end of this template (e.g. "Data resulting from Method X can be found in xxxxx.txt (Data Citation 1)"). A data citation should also be placed in the subsection of the Methods containing the data-collection or analytical procedure(s) used to derive the corresponding record.

Tables should be used to support the data records, and should clearly indicate the samples and subjects (study inputs), their provenance, and the experimental manipulations performed on each (please see Tables and Submitting Experimental Metadata, below). They should also specify the data output resulting from each data-collection or analytical step, should these form part of the archived record.

Data records are reported in a single table including 5241 rows and 268 columns. Each row corresponds to as many as 42 comparisons (if all soil health indicators have data), and each column corresponds to one background information or soil health indicator. The names, attributes, and description of the background information and soil health indicators are shown in Table 1 and Table 2, respectively. Background information columns are grouped by categories further denoted as “attributes” (Table 1). Detailed description on each background information and soil health indicator can be found in Table 1 and Table 2.

### Technical Validation

The Technical Validation section should present any experiments or analyses that are needed to support the technical quality of the dataset. This section may be supported by figures and tables, as needed. *This is a required section*; authors must provide information to justify the reliability of their data.

Possible content **may include:**

* experiments that support or validate the data-collection procedure(s) (e.g. negative controls, or an analysis of standards to confirm measurement linearity)
* statistical analyses of experimental error and variation
* phenotypic or genotypic assessments of biological samples (e.g. confirming disease status, cell line identity, or the success of perturbations)
* general discussions of any procedures used to ensure reliable and unbiased data production, such as blinding and randomization, sample tracking systems, etc.
* any other information needed for assessment of technical rigour by the referees

Generally, this **should not include:**

* follow-up experiments aimed at testing or supporting an interpretation of the data
* statistical hypothesis testing (e.g. tests of statistical significance, identifying differentially expressed genes, trend analysis, etc.)
* exploratory computational analyses like clustering and annotation enrichment (e.g. GO analysis).

Quality control was performed to guarantee the quality of the data. Each paper was carefully read at least twice, and special attentions were paid to the tables, figures, and method parts where most of the soil health indicator comparisons and background information are from. It should note that different matrix may involve in a same soil health indicator (e.g., soil total nitrogen, soil organic nitrogen, or soil inorganic nitrogen maybe reported in different papers to represent the soil nitrogen indicator, ID 5 in Table 2), therefore, it is important the measure matrix, unit, and other detailed descriptions should be recorded in the comments columns. After the data extraction, we compared the digitized data against the tables or figures from the original paper again to make sure the data are loaded correctly. Whenever before a new paper was extracted, we first used the bibliography database manager Mendeley to check whether it is a duplicate of previous papers (details please see the supplemental reference document).

After the data extraction, we performed an elaborate examination of the data quality using R (version 3.5.1). The formats of each column (numerical or string) were checked to correct the mistyping in the numerical columns (e.g., all soil health indicator, and some background information columns like latitude and longitude). For each soil health indicator, we calculated the response ratio, which is the value of treatment divide by the value of control, and plotted the frequency distribution of response ratio for each soil health indicator, and returned to the original articles for checking extreme values. We also made visualization of data distribution for the numeric background columns that contain numeric values (e.g. latitude, elevation) and manually checked the outliers by validating them in the original papers. For the location of each site, we plotted the latitude and longitude by country and check whether there are sites from a specific country fall outside the borderline. For those sites, we checked the extracted latitude and longitude information with location information from the original paper (latitude, longitude, site name, and country name). For some sites located nearby the coastal area, few sites fall in the sea probably due to the precision of values. For these sites, we corrected slightly the longitude and latitude to the near land if the reported coordinates are extracted correct.

### Usage Notes

*This section is optional*

The Usage Notes should contain brief instructions to assist other researchers with reuse of the data. This may include discussion of software packages that are suitable for analysing the assay data files, suggested downstream processing steps (e.g. normalization, etc.), or tips for integrating or comparing the data records with other datasets. Authors are encouraged to provide code, programs or data-processing workflows if they may help others understand or use the data. Please see our [code availability policy](http://www.nature.com/sdata/policies/editorial-and-publishing-policies#code-avail) for advice on supplying custom code alongside Data Descriptor manuscripts.

For studies involving privacy or safety controls on public access to the data, this section should describe in detail these controls, including how authors can apply to access the data, what criteria will be used to determine who may access the data, and any limitations on data use.

In SoilHelthDB, each soil health indicator may have multiple units, but the unit between each comparison will always the same. The information about unit will be recorded under comments column. The user should check the unit before any data processing and data use. We suggest the user play with the data using the code we provided, the code already include elaborate explanation and should be easy to follow, the user can contact the correspond authors when they have questions on running the code and using the data. It should also note that for some soil health indicators, we do not extracted enough comparisons from the 281 papers, we those encourage the user expand the data based on the SoilHelthDB data framework. Another important notice is for all the soil health indicators (except soil carbon), multiple measure matrix may involve, and with different units. Without data filtration and converting, only response ratio can be analysed. But more analysis can be applied with data filtration and converting.

### Acknowledgements

The Acknowledgements should contain text acknowledging non-author contributors. Acknowledgements should be brief, and should not include thanks to anonymous referees and editors or effusive comments. Grant or contribution numbers may be acknowledged.

This work is supported by the Natural Resources Conservation Service, U.S. Department of Agriculture, under NRCS Conservation Innovation Grant 69-3A75-14-260.

### Author contributions

Each author’s contribution to the work should be described briefly, on a separate line, in the Author Contributions section.

Jinshi Jian and Ryan D. Stewart conceived the design of the data framework. Jinshi Jian and Xuan Du extracted and integrated the data from papers to the SoilHealthDB. Jinshi Jian drafted the manuscript. K.P. and all authors revised and approved the manuscript.

### Competing interests

A competing interests statement is required for all papers accepted by and published in *Scientific Data*. If there is no conflict of interest, a statement declaring this must still be included in the manuscript.

The authors declare no competing interests.

### Figures

Figure images should be provided as separate files and should be referred to using a consistent numbering scheme through the entire Data Descriptor. In most cases, a Data Descriptor should not contain more than three figures, but more may be allowed when needed. We discourage the inclusion of figures in the Supplementary Information – all key figures should be included here in the main Figure section.

For initial submissions, authors may choose to supply a single PDF with embedded figures.

Authors are encouraged to consider creating a figure that outlines the experimental workflow(s) used to generate and analyse the data output(s).

G:\My Drive\MyResearch\29. PstDoc\NRCS\Manuscript_ScientificData\DataAndR\1 site_plot.tiff

**Figure 1.** **The spatial distribution of sites from cover crop, no-tillage, organic farm, and agro-forestry system across the globe.** The numbers in the parenthesis are number of sites. The size of sign in the map represents the number of comparisons within each site.



**Figure 2.** **Flow chart shows the procedure of meta-analysis.** Historical cover crops paper across global were collected and these meet the criteria were digitized and integrated into the dataset, the missing bulk density and standard deviation were estimated based on the regression modeling; soil carbon stocks and sequestration were further calculated. To identify the experiment ID of pair comparisons, if the cash crop, site, tillage, fertilizer level, cover crop, soil sampling depth, cover crop termination, and cash crop rotation are same, we assign them a same experiment ID, otherwise, a different experiment ID was assigned.

G:\My Drive\MyResearch\29. PstDoc\NRCS\Manuscript_ScientificData\DataAndR\SoilSamplingDepth.tif

**Figure 3. Diagram showed how soil sampling depths were separated into 0-10cm, 0-20cm, 0-30cm, and 30-90cm layers.**

### Figure Legends

### Figure legends begin with a brief title sentence summarizing the purpose of the figure as a whole, and continue with a short description of what is shown in each panel and an explanation of any symbols used. Legends must total no more than 350 words, and may contain literature references. The first sentence of the legend will be used as the title for the figure. It should contain no references of any kind, including to specific figure panels, data citations, bibliographic citations or references to other figures or panels.

### Tables

Authors are encouraged to provide one or more tables that provide basic information on the main ‘inputs’ to the study (e.g. samples, participants, or information sources) and the main data outputs of the study; also see the additional information on providing metadata on page 6. Tables in the manuscript should generally not be used to present primary data (i.e. measurements). Tables containing primary data should be submitted to an appropriate data repository.

Authors may provide tables within the Word document or as separate files (tab-delimited text or Excel files). Legends, where needed, should be included in the Word document. Generally, a Data Descriptor should have fewer than ten tables, but more may be allowed when needed. Tables may be of any size, but only tables that fit onto a single printed page will be included in the PDF version of the article (up to a maximum of three).

|  |  |  |
| --- | --- | --- |
| **Attributes** | **StudyID** | **Description** |
| 1 | ExperimentID | Experiment ID number, please see details in **Figure 2** |
| 2.1 | Author\_F | First author's Family name |
| 2.2 | Author\_G | First author's Given name |
| 3 | YearPublication | Paper publication year |
| 4 | SamplingYear | Sampling year, should be earlier than publication year |
| 5 | Journal | Journal name of paper published |
| 6 | SiteInfor | Site detailed information (and name) |
| 7 | Country | Country name of the site |
| 8 | Latitude | Latitude of the site |
| 9 | Longitude | Longitude of the site |
| 10 | Elevation | Elevation of the site |
| 11 | Tannual | Annual average air temperature |
| 12 | MAT | Mean annual air temperature (Time span may differ from paper to paper) |
| 13 | Pannual | Annual precipitation |
| 14 | MAP | Mean annual precipitation (Time span may differ from paper to paper) |
| 15 | ClimateType | Study site's climate type, obtain from climate koeppon |
| 16 | YearsAfterCoverCrop | How many years cover crop applied before samples taken (e.g., experiment initiated at 1991, sample take at 1995, then fill with 5) |
| 17 | Duration | How many years the whole experiment last for (e.g., experiement started at 1990 and end at 2000, duration is 11 years) |
| 18.1 | Time\_CC | Time of cover crop planted, winter or summer CC |
| 19 | Time\_Comments | Comments about cover crop period |
| 20 | SoilSampling | Soil sampling depth from top to bottom, formated as: top-to-bottom, e.g., 10-to-20. |
| 21.1 | SamplingDepth | Difference between bottom and top, e.g., 10-to-20, fill with 10 |
| 21.2 | SoilDepthGroup | Soil depth grouped based on the sampling depth, see **Figure 2** for more details |
| 21.3 | SurfaceSubsurface | Surface or subsurface indicator, see **Figure 2** for more details |
| 21.4 | SoilBD | Bulk density |
| 22 | SandPerc | Percentage of sand |
| 23 | SiltPerc | Percentage of silt |
| 24 | ClayPerc | Percentage of clay |
| 25.1 | Texture | Soil texture |
| 25.2 | TextureGroup | Soil texture group (coarse, medium, and fine), based on Cornell |
| 26 | SoilpH | Soil pH |
| 27 | SoilTC | Soil total C, this is the background soil carbon information (Unit is %), note that this column is differ from OC\_C\_concentration, OC\_T in the response columns |
| 28 | SOC\_NaturalVeg | Soil organic carbon of nearby natural vegetation land use (Unit is %) |
| 29 | SoilKsat | Soil saturated conductivity |
| 30 | SoilFamily | Soil family or soil group information, should be note that there are many soil classification system in the word |
| 31.1 | CoverCrop | Cover crop type, some literature called catch crop, green manure |
| 31.2 | CoverCropGroup | Cover crop grouped by function or family of cover crop, please see more details in **Table 3** |
| 32.1 | GrainCrop | Grain crop type, also called cash crop |
| 32.2 | GrainCropGroup | Grain crop grouped by function or family of grain crop, please see more details in **Table 4** |
| 33 | Landuse | Landuse type |
| 34.1 | Rotation\_C | Type of rotation/crop sequence for control |
| 34.2 | Rotation\_T | Type of rotation/crop sequence for treatment (cover crop, no-till, organic etc) |
| 34.3 | Rotation\_Diff | Whether type of rotation/crop sequence differ between control and cover crop, yes or no |
| 35.1 | Tillage\_C | Type of tillage for control |
| 35.2 | Tillage\_T | Type of tillage for cover crop |
| 35.3 | Tillage\_Diff | Whether type of tillage differ between control and cover crop, yes or no |
| 35.4 | TillageGroup\_C | Tillage method grouped to CT, RT, or NT of control, details please see **Table 5** |
| 35.5 | TillageGroup\_T | Tillage method grouped to CT, RT, or NT of treatment, details please see **Table 5** |
| 36.1 | Fertilization\_C | Description about fertilization for control |
| 36.2 | Fertilization\_T | Description about fertilization for treatment |
| 36.3 | Fert\_Diff | Whether control and cover crop applied different fertilizer level, yes or no |
| 37 | ControlDescription | Control description, winter fallow, summer fallow, no cover crop, bare soil |
| 38.1 | No\_C | Number of plot in control |
| 38.2 | No\_T | Number of plot in treatment |
| 38.3 | NoSupsample | Number of subsamples |
| 39 | ExperimentDesign | CRD,RCBD, split-design etc. |
| 40 | CCFreshBiomass | Fresh biomass of cover crop (return to soil as green manure) |
| 41 | CCDryBiomass | Dry biomass of cover crop (return to soil as green manure) |
| 42 | CNOfCoverCrop | Carbon to nitrogen ratio of the cover crop dry biomass (determine quality of green manure) |
| 43 | CCTermination | Method of killing cover crop |
| 44 | Conservation\_Type | Including cover crop(CC), No-tillage(NT), organic farm(OGF), Agro-forestry ecosystem (AFS), straw mulching etc. |
| 45 | Conservation\_Description | More details or descriptions on conservation method of cropland |
| 46 | Others | Other meta information about the publication |

**Table 1. Description and attributes of background information in the SoilHealthDB.**

|  |  |  |  |
| --- | --- | --- | --- |
| **ID** | **Indicator** | **Description** | **Comments** |
| 1 | Biomass | Biomass of cash crop except for yield (e.g., stem, leave, root) | Unit varied, unit is kg/hm2 if without explanation |
| 2 | Yield | Yield of grain(cash) crop | Unit: kg/hm2 if without explanation |
| 3 | BD | Soil Bulk density | Always g/cm3 |
| 4.1 | OC\_Conc | Soil organic carbon concentration (Unit is %) | SOC = SOM/1.72 |
| 4.2 | OC\_Stock | Soil organic carbon stock (Unit is Mg/ha) | Require BD, depth to calculate |
| 4.3 | OC\_SEQ | Soil organic carbon sequestration rate (Mg/ha/cm/yr) | There are two methods to calculate |
| 5 | N | Soil Nitrogen | Soil total nitrogen (mg kg-1) if without explanation |
| 6 | P | Soil Phosphorus | Unit is mg kg-1 if without explanation |
| 7 | K | Soil Potassium | Unit is mg kg-1 if without explanation |
| 8 | pH | Soil pH | No unit |
| 9 | CEC | Soil cation exchange capability | Unit is cmolc kg-1 if without explanation |
| 10 | EC | Soil electric conductivity | Unit is dS m–1 if without explanation |
| 11 | BS | Soil base Saturation | Unit is mg kg-1 soil if without explanation |
| 12 | Aggregate | Soil aggregation | There are multiple ways to measure soil aggregation, the Unit also varied, it is g/kg if without explanation |
| 13 | Porosity | Soil porosity | % if without explanation |
| 14 | Penetration | Soil penetration resistance | Unit is Mpa if without explanation |
| 15 | Infiltration | Soil infiltration rate | Unit is mm/h if without explanation |
| 16 | Ksat | Field saturated hydraulic conductivity | Unit is cm/h if without explanation |
| 17 | Erosion | Soil erosion or wind erosion | Unit is kg/ha if without explanation |
| 18 | Runoff | Runoff | Unit is Mg/ha if without explanation |
| 19 | Leaching | Soil nutrient leaching | Unit varied |
| 20 | ST | Soil temperature | ℃ |
| 21 | SWC | Soil water content | Gravimetric water content Unit is kg/kg if without explanation |
| 22 | AWHC | Available water hold capacity | Unit is g/g if without explanation |
| 23 | Weed | Weed of the cropland | Unit is kg/ha if without explanation |
| 24 | Diseases | Diseases of the cropland | Unit varied |
| 25 | Pests | Pests of cropland | Unit is µg/g if without explanation |
| 26 | SoilFauna | Earthworm, authors, nematode (soil should benefit from their exists) living in the soil | Unit varied |
| 27 | Fungal | Bacterial, fungal, mycorrhizal in the soil | Unit varied |
| 28 | O-Microbial | Other microbial indicators | Unit varied |
| 29 | Enzyme | Enzyme activity specifically beta-glucosidase activity and phenol oxidase | Unit varied |
| 30 | Cmin | Soil mineralizable carbon | Unit is mg/kg if without explanation |
| 31 | Nmin | Soil mineralizable nitrogen | Unit is mg NO3-N / kg soil if without explanation |
| 32 | N2O | Soil N2O efflux | Unit is kg/ha if not noted |
| 33 | SIR | Soil substrate-induced respiration | Unit is µg/g if without explanation |
| 34 | CO2BTest | Soil CO2 burst test respiration | Unit is mg CO2 /kg soil /d if without explanation |
| 35 | CO2 | Soil respiration (some literature called CO2 efflux, CO2 flux) | kg/ha if without explanation |
| 36 | CH4 | Soil CH4 emission | kg/ha if not noted |
| 37 | MBC | Microbe Biomass Carbon | In unit of: mg/kg if without explanation |
| 38 | MBN | Microbe Biomass Nitrogen | Unit is mg/kg if without explanation |
| 39 | Microelement | Mn, Zn, Cu etc. | Fill with 9999 if a paper reported Microelement, we do not record the actual comparison yet |
| 40 | SQI | Soil quality indicator, soil health indicator | Fill with 9999 if a paper reported Microelement, we do not record the actual comparison yet |
| 41 | ESS | Ecosystem services indicator | Fill with 9999 if a paper reported Microelement, we do not record the actual comparison yet |
| 42 | Texture | Cover crop effect on soil texture compare with control | Fill with 9999 if a paper reported Microelement, we do not record the actual comparison yet |
|  | Notes | Other notes, comments |  |

**Table 2. Description and attributes of soil health indicators in the SoilHealthDB.** Note that in the data sheet, each indicator has 5 columns, recording information for control, treatment, standard deviation (SD) for control, SD for treatment, and comments, respectively.

### References

Bibliographic information for any works cited in the above sections, using the standard Nature referencing style.

1. FAO. *The state of the world’s land and water resources for food and agriculture (SOLAW): managing systems at risk*. (The Food and Agriculture Organization of the United Nations and Earthscan, 2011).

2. Kaye, J. P. & Quemada, M. Using cover crops to mitigate and adapt to climate change. A review. *Agron. Sustain. Dev.* **37,** (2017).

3. Bandick, A. K. & Dick, R. P. Field management e  ects on soil enzyme activities. *Soil Biol. Biochem.* **31,** 1471–1479 (1999).

4. Idowu, O. J. *et al.* Use of an integrative soil health test for evaluation of soil management impacts. *Renew. Agric. Food Syst.* **24,** 214–224 (2009).

5. Ndiaye, E. L., Sandeno, J. M., McGrath, D. & Dick, R. P. Integrative biological indicators for detecting change in soil quality. *Am. J. Altern. Agric.* **15,** 26 (2000).

6. Poeplau, C. & Don, A. Carbon sequestration in agricultural soils via cultivation of cover crops - A meta-analysis. *Agric. Ecosyst. Environ.* **200,** 33–41 (2015).

7. Cooper, J. *et al.* Shallow non-inversion tillage in organic farming maintains crop yields and increases soil C stocks: a meta-analysis. *Agron. Sustain. Dev.* **36,** (2016).

8. Luo, Z., Wang, E. & Sun, O. J. Can no-tillage stimulate carbon sequestration in agricultural soils? A meta-analysis of paired experiments. *Agric. Ecosyst. Environ.* **139,** 224–231 (2010).

9. Gattinger, A. *et al.* Enhanced top soil carbon stocks under organic farming. *Proc. Natl. Acad. Sci.* **109,** 18226–18231 (2012).

10. Shi, L., Feng, W., Xu, J. & Kuzyakov, Y. Agroforestry systems: Meta-analysis of soil carbon stocks, sequestration processes, and future potentials. *L. Degrad. Dev.* (2018). doi:10.1002/ldr.3136

11. Stewart, R. D. *et al.* What we talk about when we talk about soil health. *Agric. Environ. Lett.* (2018).

12. Alvarez, R., Steinbach, H. S. & De Paepe, J. L. Cover crop effects on soils and subsequent crops in the pampas: A meta-analysis. *Soil Tillage Res.* **170,** 53–65 (2017).

13. Sileshi, G. *Evidence for impact of green fertilizers on maize production in sub-Saharan Africa*. (2009).

14. R, C. T. *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013*. (2014).

15. Nelson, D. W. & Sommers, L. E. Total carbon, organic carbon, and organic matter. *Methods soil Anal. part 3—chemical methods* 961–1010 (1996).

16. Moebius-Clune, B. N., D.J. Moebius-Clune, B.K. Gugino, O. J. I., R.R. Schindelbeck, A.J. Ristow, H.M. van Es, J.E. Thies, H. A. S. & M.B. McBride, K.S.M. Kurtz, D.W. Wolfe, and G. S. A. *Comprehensive Assessment of Soil Health: The Cornell Framework Manual*. (Cornell University, 2016).

### Data Citations

Data citations provide bibliographic information for any data records described or used in the manuscript. See further details below.

### Additional Formatting Information

**Referencing Figures, Tables, and other content**

**The Word document may reference Figures (e.g. Fig. 1), Tables (e.g. Table 1), and Supplementary Information (e.g. Supplementary Table 1, or Supplementary File 2, etc.). When information from metadata documents must be referred to, it should also be included in the main manuscript as Tables, and formatted in a way that suits human readability. To refer to the ISA-Tab metadata records within the manuscript, use the phrase “see associated Metadata Record”.**

**Citation format**

References should be numbered sequentially, first throughout the text, then in tables, followed by figures and, finally, boxes; that is, references that only appear in tables, figures or boxes should be last in the reference list. Only one publication is given for each number. Only papers that have been published or accepted by a named publication or recognized preprint server should be in the numbered list; preprints of accepted papers in the reference list should be submitted with the manuscript. Published conference abstracts, numbered patents, and archived code with an assigned DOI may be included in the reference list. Grant details and acknowledgments are not permitted as numbered references. Footnotes are not used.

Scientific Data uses standard Nature referencing style. All authors should be included in reference lists unless there are six or more, in which case only the first author should be given, followed by ‘et al.’. Authors should be listed last name first, followed by a comma and initials (followed by full stops, '.') of given names. Article titles should be in Roman text; only the first word of the title should have an initial capital and the title should be written exactly as it appears in the work cited, ending with a full stop. Book titles should be given in italics and all words in the title should have initial capitals. Journal names are italicized and abbreviated (with full stops) according to common usage. Volume numbers and the subsequent comma appear in bold. The full page range should be given where appropriate. See the examples below:

**Journal Article**:

1. Schott, D. H., Collins, R. N. & Bretscher, A. Secretory vesicle transport velocity in living cells depends on the myosin V lever arm length. *J. Cell Biol*. **156**, 35‐39 (2002).

**Book** ‐ Book titles should be given in italics and all words in the title should have initial capitals:

1. Hogan, B. *Manipulating The Mouse Embryo: A Laboratory Manual* 2nd edn (Cold Spring Harbor Laboratory Press, 1994)

**Publicly available preprint:**

1. Babichev, S. A., Ries, J. & Lvovsky, A. I. Quantum scissors: teleportation of single-mode optical states by means of nonlocal single photon. Preprint at http://arXiv.org/quant-ph/0208066 (2002).

**Code:**

1. Gallotti, R. & Barthélemy, M. Source code for: The multilayer temporal network of public transport in Great Britain. *Figshare* https://dx.doi.org/10.6084/m9.figshare.1249862.v1 (2014).

**Online material** ‐ Stable documents hosted on the web may be cited in the main reference list, using the format below. Websites or dynamic web resources should be cited by embedding the URL in the main article text:

1. Manaster, J. Sloth squeak. *Scientific American Blog Network* http://blogs.scientificamerican.com/psi-vid/2014/04/09/sloth-squeak (2014).

**Technical or government report:**

1. Akutsu, T. *Total Heart Replacement Device.* Report No. NIH-NHLI-69 2185-4 (National Institutes of Health, 1974).

**Data Citations**

In-text data citations should be of the form (Data Citation 1), referring by number to a data record listed at the end of the document. Data citations may occur anywhere in the manuscript (except the Abstract), but in general each dataset central to the Data Descriptor should be cited at least once in the Data Records, and also indicated in the associated portion of the Methods section.

Data citations should be listed in the order they are cited in the manuscript. All authors or dataset submitters should be listed for each record in the “Data Citations” section. Please do not use “*et al.*” to shorten author lists, and please do not include dataset titles in this section.

Please use full repository names and check for consistency with the names in our list of [recommended data repositories](http://www.nature.com/sdata/policies/repositories), FAIRsharing.org or re3data.org.

Please write DOIs using full URL notation.

For data with a digital object identifier (DOI) this should be in the format:

Lastname1, Initial1A. Initial1B., Lastname2, Initial2A. Initial2B., … & LastnameN, InitialNA. Initial NB. *Repository\_name* DOI (YYYY).

Example citation for data with a DOI:

1. Perkins, A. D., Lee, M., & Tanentzapf, G. *Figshare* http://dx.doi.org/10.6084/m9.figshare.806269 (2014).

For data identified by accession ID, repositories may not provide the data creator (author). In these cases the data citation format should be *Repository\_name* Accession\_ID (YYYY)

Example citation for data with an accession identifier:

1. *GenBank* PRJNA244495 (2014).

[Further guidance on data citation](http://blogs.nature.com/scientificdata/2016/07/14/data-citations-at-scientific-data/).

**Manuscript metadata tables**

Authors are are asked to submit one or more tables presenting, at a minimum, the samples and subjects employed in the study, the experimental, observational and analytical manipulations performed on each, and the data outputs resulting from these manipulations with their manuscript. Please also see Data Records and Tables, above; please note that, should you wish to create more complex metadata records, template Study and Assay files are available either in a ZIP archive with this document or for separate download from the “[Submission guidelines](http://www.nature.com/sdata/publish/submission-guidelines)” page of the *Scientific Data* website.

Here, we provide four generic ‘Table 1’ examples, including two experimental study examples, one observational study example, and an example for an aggregated dataset of the type that may result from a meta-analysis.

**Experimental study Table 1 example**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Subjects** | **Protocol 1** | **Protocol 2** | **Protocol 3** | **Protocol 4** | **Data** |
| Mouse1 | Drug treatment | Liver dissection | RNA extraction | RNA-Seq | GEOXXXXX |
| Mouse2 | Drug treatment | Liver dissection | RNA extraction | RNA-Seq | GEOXXXXX |
| Mousen | Drug treatment | Liver dissection | RNA extraction | RNA-Seq | GEOXXXXX |

**Experimental study with replicates Table 1 example**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **Protocol 1** | **Protocol 2** | **Samples** | **Protocol 3** | **Data** |
| CellCulture1 | Drug treatment | RNA extraction | TechnicalRep1a | Microarray hybridization | GEOXXXXX |
| CellCulture1 | Drug treatment | RNA extraction | TechnicalRep2a | Microarray hybridization | GEOXXXXX |
| CellCulture1 | Drug treatment | RNA extraction | TechnicalRep3a | Microarray hybridization | GEOXXXXX |
| CellCulture2 | Drug treatment | RNA extraction | TechnicalRep1b | Microarray hybridization | GEOXXXXX |
| CellCulture2 | Drug treatment | RNA extraction | TechnicalRep2b | Microarray hybridization | GEOXXXXX |
| CellCulture2 | Drug treatment | RNA extraction | TechnicalRep3b | Microarray hybridization | GEOXXXXX |

**Observational study Table 1 example**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sample** | **Geographical location** | **Geoposition** | **Protocol** | **Data** |
| Body of water 1 | location name | latitude, longitude, altitude | Measurement of surface temperature | dataFile1 |
| Body of water 2 | location name | latitude, longitude, altitude | Measurement of surface temperature | dataFile2 |
| Body of water n | location name | latitude, longitude, altitude | Measurement of surface temperature | dataFile3 |

**Data aggregation study Table 1 example**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Sample** | **Sample number** | **Temporal range** | **Protocol 1** | **Protocol 2** | **Data** |
| Database URL 1 | Dataset 1 | Number of samples in the dataset | Range of measurements reported in the dataset | Data assimilation procedure | Method to generate output data | dataFile1 |
| Database URL 1 | Dataset 2 | Number of samples in the dataset | Range of measurements reported in the dataset | Data assimilation procedure | Method to generate output data | dataFile1 |
| Database URL 2 | Dataset n | Number of samples in the dataset | Range of measurements reported in the dataset | Data assimilation procedure | Method to generate output data | dataFile2 |

**Machine accessible metadata records**

All Data Descriptor manuscripts published in *Scientific Data* are accompanied by detailed machine-accessible metadata files in ISA-Tab format depicting the workflow used to generate the accompanying datasets. In order to facilitate the creation of these records, authors may be asked to complete metadata templates for their data by the editorial office, if the Data Descriptor is of interest to the Editorial Board.

**Depositing your data to an appropriate repository**

Your *Scientific Data* manuscript will not be sent to review unless the dataset(s) described therein have been deposited in an appropriate public repository ([please see our list of recommended repositories](http://www.nature.com/sdata/policies/repositories)). Should a specific repository not be available for your field or data-type, or should the repository of your choice not permit confidential peer-review, you may upload your data to one of our [recommended generalist repositories](https://www.nature.com/sdata/policies/repositories#general). Integrated submission systems are available for both figshare and Dryad.