Shared protocols and data template in agronomic trials

D. Cammarano[1], P. Martre[2], D. Drexler[3], X. Draye[4], A. Sessitsch[5], N. Pecchioni[6], J. Cooper[7], H. Willer[8], A. Voicu[2], P. Hinsinger[2]

[1]James Hutton Institute, Errol Road, Invergowrie, U.K.; [2]UMR LEPSE, INRA, Montpellier SupAgro, 2 place Viala, 34 060, Montpellier, France; [3]ÖMKi, Hungarian Research Institute of Organic Agriculture, H-1033 Budapest; [4]Universite Catholique de Louvain, Croix du Sud 2, Louvain, Belgium; [5]Austrian Institute of Technology, Konrand-Lorenz-Strasse, Tulln, Austria;[6]Research Centre for Cereal and Industrial Crops, Foggia, Italy; [7]University of Newcastle, Newcastle, U.K.; [8]FiBL - Research Institute of Organic Agriculture, Switzerland.

A paper from the Proceedings of the
14th International Conference on Precision Agriculture
June 24 – June 27, 2018
Montreal, Quebec, Canada

Abstract. *Due to the overlap of many disciplines and the availability of novel technologies, modern agriculture has become a wide, interdisciplinary endeavor, especially in Precision Agriculture. The adoption of a standard format for reporting field experiments can help researchers to focus on the data rather than on re-formatting and understanding the structure of the data. This paper describes how a European consortium plans to: i) create a "handbook" of protocols for reporting definitions, methodologies and parameters measured/calculated; and ii) how a data-template for field data was created and will be linked to the "handbook". The overall goal of the EU-funded project Solutions for Solutions for improving Agroecosystem and Crop Efficiency for water and nutrient use (SolACE) is to help European agriculture face major challenges, such as increased rainfall variability and reduced use of N and P fertilizers in order to satisfy both economic and ecological goals. The "Handbook of Protocols" and the "Data Template" have been created to achieve a flexible, standard, and clear documentation linked with the data itself to facilitate interchange of data among project's partners and any statistical analysis and modelling of different datasets.*

Keywords. *Data template, Protocols, Data standard, agronomic data, field experiment.*

Introduction

Agricultural research has been always an interdisciplinary field. For example, agronomists and soil scientists work together to improve soil health and crop production. Due to the overlap of many disciplines and the availability of novel technologies this interdisciplinarity has widened. The use of drone technologies, soil and plant sensing, daily (or hourly) weather data, GPS and GIS, resulted in the generation of big datasets that require certain computational skills to interpret. In addition, the adoption of Decision Support Systems (DSS) in agricultural decision-making means that more data need to be analyzed and correctly interpreted to guide agronomic in-field decision. Therefore, an increased exchange of data among researchers from different scientific fields means that a common vocabulary and organization level should be adopted. This is particularly true when data will be shared and used in Precision Agriculture (PA) because of the abovementioned reasons.

Furthermore, agricultural data can be analyzed and utilized to understand global impact on food safety and security with the aims of improving human livelihood (UN 2015). For this reason, public-funded research has generally an Open Access (OA) policy to comply with. The European Union (EU) provides clear guidelines for the OA data and data management plans (EU 2017). White and van Evert (2008) discussed the importance of promoting data sharing in agricultural research.

In several research fields, data are available through accessible databases, such as genomics data (Mewes et al. 2002; Arend et al. 2016), global daily weather data from the NASA-AgMERRA (Ruane et al. 2016), and NASA Prediction of World Energy Resource (NASA-POWER) (Stackhouse 2012). Similarly, soil scientists have created global soil datasets like the S-World (Stoorvogel et al. 2016), SoilGrids (Hengl et al. 2017), the Global Soil Organic Carbon Map (GSOCmap) (Yigini et al 2017). On the other hand, agronomic data coming from field research have seldom approached a global level of organization. Although individual initiatives and systems have been proposed, there is no commonly accepted framework to report and store field data (van Evert et al 1999; Bostick et al 2004; White et al 2013).

One global effort to create a standardized and harmonized dataset for agronomic data were the International Benchmark Sites Network for Agrotechnology Transfer (IBSNAT) and the International Consortium for Agricultural Systems Applications (ICASA) projects. Since 1983 the IBSNAT first, then the ICASA tried to create standards for agricultural data, reported in a simple way that would be easy to share among researchers (White et al 2013). Their approach was adopted by the Agricultural Model Intercomparison and Improvement Project (AgMIP; www.agmip.org) which used the ICASA data standard as a tool for harmonizing and sharing data to quantify the impacts of climate change on food security and to improve crop models (Porter et al., 2014). One aspect that was evident from such effort was the possibility of harmonizing and exchanging data between field experimentalists, crop and economic modelers, climate scientists and stakeholders. This lead to a series of results on the impact of climate change on smallholder farmers in Sub-Saharan Africa and South-East Asia (Rosenzweig et al. 2015).

However, in many projects it is hard to harmonize field data, especially when multiple research groups are scattered in different locations. Researchers usually tend to report data in the format and units with which they are most familiar. For example, grain yield can be reported at a given percentage of humidity or at dry weight and if this information is not reported it might generate some error in interpreting the data. The soil organic carbon can be measured in several ways, but if it is determined by combustion method it might overestimate its content in Calcareous soils. This means that a different user who knows neither the method nor the soil type might draw some erroneous conclusion. Furthermore, definitions can mean different things to different scientists as highlighted by Passioura (2002) for Water Use Efficiency (WUE). If the origin, definition, and methodology of data collection are not clear any information extracted from them might be misinterpreted. Also, if the field data are not consistently reported in a way that is common to many it might lead to errors in using and interpreting the results.

Therefore, the aims of this paper were to demonstrate how an European consortium is going to: i) create an "handbook" of protocols for reporting definitions, methodologies and parameters measured/calculated; and ii) how a data-template for field data was created and will be linked to the "handbook".

Project overview

The overall goal of the EU-funded project *Solutions for Solutions for improving Agroecosystem and Crop Efficiency for water and nutrient use* (SolACE) is to help European agriculture respond to major challenges, such as increased rainfall variability and reduced use of N and P fertilizers in order to satisfy both economic and ecological goals. The project will do so by designing solutions (strategies and tools) that combine novel crop genotypes and agroecosystem management innovations to improve water and nutrient use efficiency over a range of agricultural contexts across pedoclimatic regions of Europe. It is clear from this short description that SolACE involves a significant amount of trans-disciplinary work. In addition, 25 partners from 14 Countries are taking part of the project and many different experimental fields, in different agro-environmental conditions, will be set up (Figure 1). The project began in May 2017 and is scheduled to end in May 2022.

Fig 1. Geographical distribution of the project's partners.

There are two major innovation loops in the project, one dealing with the breeding/molecular aspects and another with the agronomic/agroecosystem management. In both cases, field or glasshouse experiments will be established where wheat or potato will be grown over several years under different stress levels. In addition, farmers' field tests will be conducted, further increasing the amount of field data generated. In this context SolACE aimed at drafting a Data Management Plan (DMP) that complied with the EU regulation; but at the same time, the project seeks to promote data sharing at the various steps of SolACE implementation, and to ensure data quality and comparison through the definition of standardized protocols for data collection and processing. The subsequent sections provide more details about these two issues.

Handbook of Protocols and Methodology

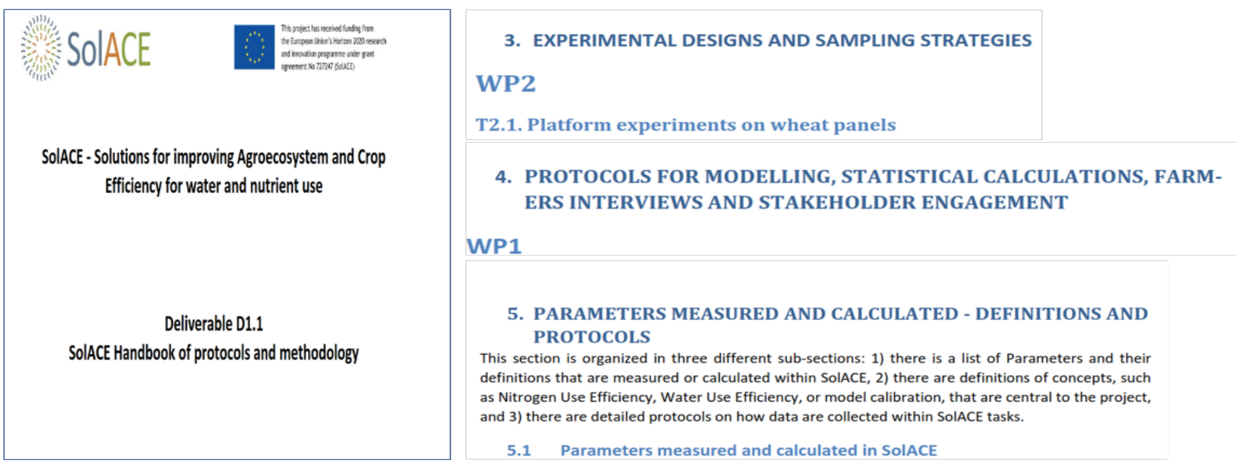
The SolACE Handbook of Protocols and Methodology is intended for all SolACE project partners. It is envisioned to be a living handbook describing the standards to use for data collection. This document, generated at the beginning of the project, will be constantly reviewed, revised, and updated throughout the project. After the first year this document already contains most of the information regarding the field experiments that are about to start.

The Handbook of Protocols and Methodology describes the parameters that are to be measured and/or calculated in the SolACE project as well as the protocols describing the methodology for measurements or calculations. Moreover, the Handbook includes the definitions of central concepts that are of key importance to the project, which shall be interpreted the same way throughout SolACE. The Handbook focuses primarily on the parameters and definitions that are used by more than one project partner, and it sets – wherever possible – common guidelines for conducting measurements and calculations. The purpose of the Handbook is to guarantee methodological consistency within the project, and, where possible, the comparability of results – regardless of whether it is greenhouse, field, or laboratory investigations. Because of the above-mentioned reasons, the Handbook is a "*living document*", ready for development throughout the project. Its content evolves together with the project, considering the possible changes and adaptations that inherent to scientific work.

The Handbook does not intend to repeat detailed protocols that can be cited from available literature, but it aims to document which protocols have been chosen to ensure that SolACE partners have a common understanding and approach for specific tasks. If it proves impossible to harmonize methodologies and protocols for a topic, the Handbook will indicate all applied methods, and will include the reasoning behind this decision. This way, we aim to keep track of the logical development of the project and ensure that methods and protocols are discussed and thought through within the consortium, in order to achieve as much coherence and transparency as possible.

A further aim of the Handbook is to design a system for collecting and managing the data generated from the literature review, data mining, and new investigations during SolACE. For this purpose, guidelines for the collection, storage, and quality control of data are included in the Handbook. The Handbook describes the standards to use for the data collection, and it is used by all work packages.

The first version of this living handbook has been generated and it is available internally. Because it is a "*living"* document it will be further expanded and updated as the project goes. An extract of several section headings is showed in Figure 2. In the future, it will be made available as a format like Google Doc, or as a Wiki, according to what will be decided in the subsequent years.


SolACE

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 727247 (SolACE)

SolACE - Solutions for improving Agroecosystem and Crop Efficiency for water and nutrient use

Deliverable D1.1
SolACE Handbook of protocols and methodology

3. EXPERIMENTAL DESIGNS AND SAMPLING STRATEGIES

WP2

T2.1. Platform experiments on wheat panels

4. PROTOCOLS FOR MODELLING, STATISTICAL CALCULATIONS, FARMERS INTERVIEWS AND STAKEHOLDER ENGAGEMENT

WP1

5. PARAMETERS MEASURED AND CALCULATED - DEFINITIONS AND PROTOCOLS

This section is organized in three different sub-sections: 1) there is a list of Parameters and their definitions that are measured or calculated within SolACE, 2) there are definitions of concepts, such as Nitrogen Use Efficiency, Water Use Efficiency, or model calibration, that are central to the project, and 3) there are detailed protocols on how data are collected within SolACE tasks.

5.1 Parameters measured and calculated in SolACE

Fig 2. The main page of the Handbook of protocols and methodology and chapters dealing with Experimental design and sampling strategies (chapter 3); crop modelling, statistical calculations, farmers interviews and stakeholder engagement (chapter 4); and parameters measured and calculated – definitions and protocols (chapter 5).

Data Template

The data template is based on the ICASA data standard and a detailed description of the dataset exists (White et al 2013). The main properties of this data standard will be briefly described while the main differences between ICASA and SolACE will be highlighted.

The ICASA data standard divides field experiments into three main subsets:

- ***Experiment Metadata***: it includes descriptive information about an experiment, such as the location name, number of replicate, details of the experimentalist, any documentation available, the plot organization;

- ***Management Data***: this section includes information such as the treatments, the soil conditions, weather data, initial conditions of soil prior or at sowing (e.g. soil water content and soil nitrogen), any agronomic management (e.g. sowing date, density, and depth, fertilizer management, and so on), the type of crop and cultivar used, tillage information and so on;

- ***Measured Data***: this section includes the crop and soil data measured either during the growing season at regular times (time-series) or data that have been recorded once during the experiment, like for example grain yield at harvest (summary).

Within each subset there are many ***entities*** like for example the "*soil description*" of the Management Data subset. Within each entity there would be many ***attributes*** like the depth of each layer, the clay content of each soil layer, the soil albedo, and so on. A schematic example of the type of data included in each level of organization can be found in White et al (2013) as well as lines connecting entities that are interrelated. Figure 3 showed a summary of the organization with the three main subsets and a non-exhaustive list of entities and attributes.

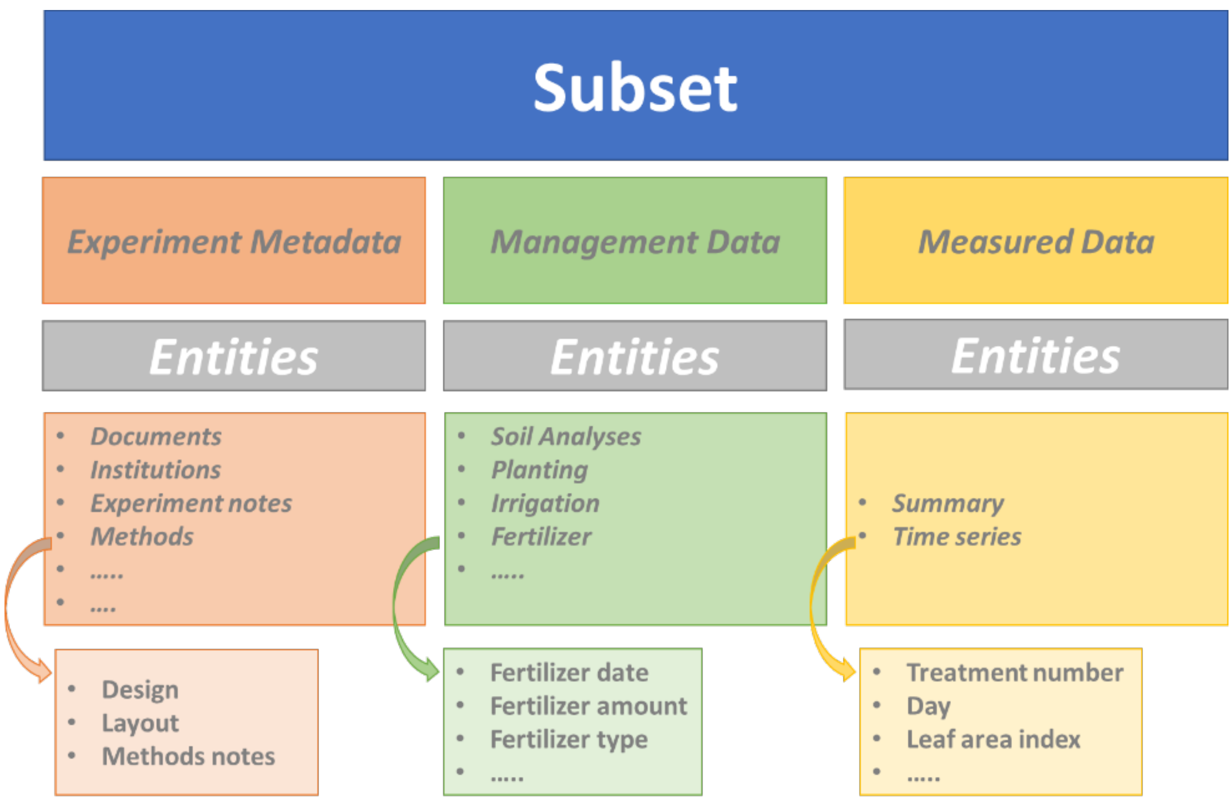


Fig 3. Summary of the levels of organization within the ICASA data standard (adapted from White et al 2013). The blue box represents the subset. There are three subsets in the data standard: Experiment Metadata, Management Data, Measured Data. Within each subset there are several entities and within each entity (e.g. fertilizer) there would several attributes describing it (e.g. fertilizer type, amount, application date, application depth, and so on). There would some connections and relationships among entities but for clarity they are not shown here.

The SolACE field data will be entered in a spreadsheet that has been created on purpose for this experiment and will be trialed during the growing season 2018 on potato from a field experiment in Hungary. The main feature of this spreadsheet is that it contains 5 Sheets:

- General instructions: which contains the instructions on how to fill the spreadsheet and how to name it. The file will be named following the 2-digits Code for the location (e.g. Foggia = FG) + 2-digits code for the crop type (e.g. Wheat = WH) + 2-digits code experiment type (e.g. N levels = NL);

- Weather Data: this sheet contains information of the daily weather data (WTH) recorded by the local weather station;

- Soil Data: all the information regarding the soil chemical and physical properties will be input in the SOL sheet. The SOL sheet contains both the data collected once during the length of the experiment such as Clay content, and the soil water and nitrogen content that can be collected several times during the growing season;

- Management Data: All the agronomic and management data (MGT) will be placed in this spreadsheet. Information like planting time, planting density, planting depth, fertilization time, depth and type are some of the information that will populate this MGT sheet;

- Plant Data: The time-series data and the summary data (Fig. 3) are reported in this PLT sheet.

In SolACE the name of the variables is standardized following the ICASA Standard for the naming convention; in addition, these variable names are linked to the Handbook of Protocols and Methodology where each Work Package (WP) will report what are the measuring in their experiments. An example of the data template is shown in Figure 4 where the MGT sheet is shown for a hypothetical experiment. The first four columns are the treatment number (TRTNO), the replicate number (RP), the Fertilizer amount (FE) and the irrigation type (IR: e.g. a full irrigated experiment will be labelled as "100"; a rainfed will be labelled as "0", and a deficit-irrigation will be reported as "x"). Then, subsequent columns will describe the type of cultivar (CRID; in this case durum wheat "WHD"), the ID of the cultivar, the cultivar name (CUL_NAME), the code for the type of crop residues left (if any) (ICPCR), the percentage of residues incorporation (ICRIP), the aboveground dry weight of the residues (ICRAG, kg dm ha^{-1}), the planting year (PLYR), the planting day of year (PLDOY), the planting population (PLOP), the planting distribution (PLDS), the row spacing (PLRS, cm), the planting depth (PLDP, cm), the total irrigation applied (IR_TOT, mm), number of irrigation (IR_#), the year of the irrigation (IRYR), the day of year of the irrigation (IRDOY), the Irrigation operation (IROP). IROP is a variable that is linked to another code, for example sprinkler, furrow, drip will have different codes associated to it.

TRTNO	RP	FE	IR	CRID	CUL_ID	CUL_NAME	ICPCR	ICRIP	ICRAG	PLYR	PLDOY	PLPOP	PLDS	PLRS	PLDP	IR_TOT	IR_#	IRYR	DOY	IROP
1	1	0	100	WHD	01	Simeto	WHD	0	500	2006	298	350	Row	17	4	150	6	2007	100	IR004
1	1	0	100	WHD	01	Simeto	WHD	0	500	2006	298	350	Row	17	4	150	6	2007	120	IR004
1	1	0	100	WHD	01	Simeto	WHD	0	500	2006	298	350	Row	17	4	150	6	2007	150	IR004
1	2	0	100	WHD	01	Simeto	WHD	0	500	2006	298	350	Row	17	4	150	6	2007	100	IR004
1	2	0	100	WHD	01	Simeto	WHD	0	500	2006	298	350	Row	17	4	150	6	2007	120	IR004
1	2	0	100	WHD	01	Simeto	WHD	0	500	2006	298	350	Row	17	4	150	6	2007	150	IR004
1	3	0	100	WHD	01	Simeto	WHD	0	500	2006	298	350	Row	17	4	150	6	2007	100	IR004
1	3	0	100	WHD	01	Simeto	WHD	0	500	2006	298	350	Row	17	4	150	6	2007	120	IR004
1	3	0	100	WHD	01	Simeto	WHD	0	500	2006	298	350	Row	17	4	150	6	2007	150	IR004
2	1	100	100	WHD	01	Simeto	WHD	0	500	2006	298	350	Row	17	4	150	6	2007	100	IR004
2	1	100	100	WHD	01	Simeto	WHD	0	500	2006	298	350	Row	17	4	150	6	2007	120	IR004
2	1	100	100	WHD	01	Simeto	WHD	0	500	2006	298	350	Row	17	4	150	6	2007	150	IR004
2	2	100	100	WHD	01	Simeto	WHD	0	500	2006	298	350	Row	17	4	150	6	2007	100	IR004
2	2	100	100	WHD	01	Simeto	WHD	0	500	2006	298	350	Row	17	4	150	6	2007	120	IR004
2	2	100	100	WHD	01	Simeto	WHD	0	500	2006	298	350	Row	17	4	150	6	2007	150	IR004
2	3	100	100	WHD	01	Simeto	WHD	0	500	2006	298	350	Row	17	4	150	6	2007	100	IR004
2	3	100	100	WHD	01	Simeto	WHD	0	500	2006	298	350	Row	17	4	150	6	2007	120	IR004
2	3	100	100	WHD	01	Simeto	WHD	0	500	2006	298	350	Row	17	4	150	6	2007	150	IR004

Fig 4. Example of the Management Data sheet (MGT) from the data template. The first column reports the Treatment number (TRTNO), the second one the Replicate number (RP), the third the fertilizer level of the experiment (FE), the IR column reports if the experiment was fully watered (100) or drought (0) or irrigated at deficit-irrigation (x). a complete list of the variable name can be found on the DSSAT Foundation website (dssat.net/data/standards)

When it comes to plant and soil data, not all the research groups will measure the same thing or all the same variables. For example, one research group will measure soil bulk density, soil moisture, texture, nitrogen, while another will only have texture and water. Therefore, one location will have a "SOL" sheet with more variables than another. This will not be a problem as groups have the freedom to include whatever columns they need, so long as they report the column header using the ICASA data standard naming convention. This will facilitate, on top of the standardization, any statistical analysis of the dataset any time during the project. It will help to input the data into several crop simulation models, and functional and structural models. It will also help to get all the data into a Life Cycle Analysis assessment. All this without any additional data processing/cleaning.

If variables are not already in the ICASA dictionary, they will be added in a way that follows the ICASA rationale. For example, the Life Cycle Analysis requires additional variables that are not listed in ICASA, and they will be created and added to our extended version of the ICASA dictionary. Then, additional talks will be held with the scientists that curate the ICASA dictionary to understand if these additional variables could potentially become officially an ICASA data standard. In SolACE a well-structured stand-alone dictionary mirroring the ICASA one has been created for the length of the project. At the end of the project, a meeting with the scientists keeping the formal ICASA data dictionary will be held to decide about the added value of a formal merge.

The rationale for creating a dictionary rather than an ontology is because with quantitative traits the important point is to relate to units of measurements, validation or quality control criteria and measurement protocols. Ontologies are needed most for qualitative traits, reflected in their widespread use in fields like genomics and biomedical research.

Some WP do not have a need for a data template. WP3 has experiments on microbiome, they report their protocols into the "*Handbook*" but there is no practical implementation of linking their protocols with a data template because they use bioinformatic tools to extract the data to produce directly the information they need. In practice, they already have a standardized pipeline that is described in the protocols.

Summary

In conclusion, the SolACE experiment aims at harmonizing, standardizing and documenting field and glasshouse experiments to allow and facilitate the exchange of the data among project partners. At the end of the project, a clearly defined data set will be made available following the European Union regulations on data sharing and open access. But, if the data are well formatted it will help other researchers to focus on the data rather than on re-formatting and understanding the structure of the data.

The goal of the Handbook of Protocols and Methodology and the Data Template is to have a flexible, standard, and clear documentation linked with the data itself to facilitate interchange of data among project's partners and any statistical analysis and modelling of different datasets.

It is unlikely that agricultural data will reach a common unique data standard, but the attempts from this project, along with the ones from other international projects like AgMIP, ICASA, and DSSAT Foundation might help a wider adoption of a similar concept following a shared data dictionary to advance a better and more rationale usage of field data from agricultural experiment.

Acknowledgements

This project has received funding from the European Union's H2020 research and Innovation Programme under grant agreement No 727247 (SolACE). Jeff White and Cheryl Porter for their useful feedback on the issues of data standardization.

References

Arend, D., Junker, A., Scholz, U., et al. (2016). PGP repository: a plant phenomics and genomics data publication infrastructure. *Database,* 1–10.

Bostick, W.M., Koo, J., Walen, V.K., Jones, J.W., Hoogenboom, G. (2004). A web-based data exchange system for crop model applications. *Agronomy Journal*, 96, 853–856.

European Commission (2017). Open Access. https://ec.europa.eu/research/openscience/index.cfm?pg=openaccess. Accessed 29 April 2018.

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B.M., Ruiperez Gonzalez, M., Kilibarda, M. et al. (2017). SoilGrids250m: global gridded soil information based on Machine Learning. *PLoS ONE*, doi:10.1371/journal.pone.0169748.

Mewes HW, Frishman D, Güldener U, et al. (2002). MIPS: a database for genomes and protein sequences. *Nucleic Acids Research,* 30, 31-4.

Porter C.H., Villalobos C., Holzworth D., Nelson R., et al. (2014). Harmonization and translation of crop modeling data to ensure interoperability. *Environmental Modelling & Software*, 62, 495–508.

Rosenzweig, C., Elliot, J., Deryng, D., Ruane, A.C. (2014). Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. *PNAS.* 111, 3268-3273.

Ruane, A.C., Goldberg, R., Chryssanthacopoulos, J. (2015). AgMIP climate forcing datasets for agricultural modeling: Merged products for gap-filling and historical climate series estimation, *Agr.icultural and Forest Meteorology.* 200, 233-248.

Stackhouse, P.W. (2012). NASA Climatology Resource for Agroclimatology. http://power.larc.nasa.gov. Accessed 10 April 2018.

Stoorvogel, J.J., Bakkenes, M., et al. (2016). S-World: A Global Soil Map for Environmental Modelling. *Land Degradation & Development.* 28, doi.org/10.1002/ldr.2656.

UN General Assembly, (2015). Transforming our world: the 2030 Agenda for Sustainable Development (A/RES/70/1). http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E. Accessed 29 April 2018.

van Evert, F.K., Spaans, E.J.A., Krieger, S.D., Carlis, J.V., Baker, J.M. (1999). A database for agroecological research data. 1: data model. *Agronomy Journal*, 91, 54–62.

White, J.W., Hunt, L.A., Boote, K.J., Jones, J.W., et al. (2013). Integrated description of agricultural field experiments and production: The ICASA Version 2.0 data standards. *Computers and Electronics in Agriculture*, 96, 1-12.

White, J.W., van Evert, F.K. (2008). Publishing agronomic data. *Agronomy Journal,* 100, 1396–1400.

Yigini, Y., Baritz, R., Vargas, R.R. (2017). Soil Organic Carbon Mapping; Cookbook Manual. Rome, FAO.