CONSORTIUM FOR PRECISION CROP NUTRITION

DATA SHARING AND MANAGEMENT POLICY

This document describes the general principles for sharing and management of data, algorithms and other information adopted by the Consortium for Precision Crop Nutrition (CPCN). The policy will be implemented through a Terms & Conditions (T&C) agreement signed by all members of the consortium. The consortium may also adopt a generic end-user license agreement (EULA) entered into between a software developer and users of the software.

Introduction

The CPCN was established to bring together public and private partners to co-innovate on crop nutrition principles, standards, and algorithms. It aims to provide a platform that can facilitate and stimulate collaboration through open exchange of information, data, algorithms and ideas to accelerate the development and rapid dissemination of digital crop nutrition advisory tools. It provides an environment for the partners to build trust, share data and other critical resources, and learn from each other. It is essential to create a data-exchange framework that supports the strategic objectives of CPCN, and to stimulate partnerships outside the confines of the consortium.

The rapid advancements in data science have revolutionized agricultural research in recent years. The development of powerful data analytics has allowed researchers to develop algorithms to extract knowledge using large datasets collected across geographies. This has necessitated data sharing among researchers for different disciplines, creating a currency that can readily be exchanged among partners. Creating a transparent and clear data-exchange framework will accelerate research and permit members to quickly develop algorithms for decision support systems that benefit farmers and businesses supporting them. Advances in web infrastructure and analytics enable remote sharing and access of data across geographies. Aggregation of historically collected data can provide valuable information to enable understanding complex environments and development of decision tools.

For transparency and to reduce the sensitivity challenges around data sharing, this data policy provides general guidelines for CPCN members in accordance with established or community data standards (e.g. industry standard metadata schemas, controlled vocabularies, ontologies etc.) and guidelines for collecting, storing, and utilizing data and other information. This policy is defines the principles and core operational practices for the CPCN. It is likely to evolve further as the consortium advances and learns from its own activities. Through this document, CPCN promotes the transparent collection, processing, analysis, and sharing of crop nutrition data. It also promotes privacy protection and safeguarding personal identifiable information of farmers or others from whom data may get sourced. The Consortium will identify suitable platforms that can be used for data sharing and management in compliance with this policy and its specific T&Cs. The Consortium will not generate or negotiate individual data or software licenses, but instead will utilize appropriate standard licenses (where possible they should be machine-readable licenses, e.g. Creative Commons, GPL, MIT, GNU etc.) that are accepted by all its members.

Principles

The following will serve as the guiding principles for data sharing and management for the Consortium.

1. Maximize collaboration through sharing, access and use of data, methods, literature, algorithms and other information needed for making better crop nutrition decisions.

2. Remove obstacles to data sharing. Utilize and contribute to improving existing standards for data generation, normalization and integration, including metadata. Promote new data generation.
3. Honor and acknowledge data ownership in all activities, through the use of DOI so that data can be formally cited and also allow tracking impact of datasets.
4. To the extent possible, follow open access and open-source approaches so that innovation in a pre-competitive space can be stimulated.
5. Adhere to the FAIR Data Principles[1], i.e. data must be Findable, Accessible, Interoperable and Reusable.
6. Develop minimum data requirements for agronomic research, which will be validated and published for wider use. Comply with data privacy and security guidelines and consider other ethical issues, including, as applicable, specific regulations such as the GDPR (General Data Protection Regulation)[2], the HITRUST CSF framework[3], or CARE principles for indigenous data governance.[4]
7. Implement good data governance and risk management.
8. All members of the community follow a code of conduct comprised of:
    Be considerate
    Be respectful
    Be collaborative
    Be pragmatic
    Support others in the community
    Get support from others in the community

IMPLEMENTATION GUIDELINES

Data Governance Committee

Implementation of the data policy and specific T&Cs will be overseen by the Data Governance Committee, which serves as a subcommittee of the Steering Committee. The composition of the Data Governance Committee may change based on need and as decided by the Steering Committee. The Data Governance Committee will report to the Steering Committee and through that to the membership to seek feedback and approval of the policy and its implementation. The Data Governance Committee will also be the primary body for resolving any potential conflicts that may arise.

What type of data?

The consortium may develop and use a simple, generic data submission and sharing agreement that aligns with the principles stated above. The consortium will also make use of Digital Object Identifier (DOI) services and registration[5] to facilitate better data tracking and acknowledgement of data use.

To enable the full analytical potential, and acknowledging the principles stated above, the consortium will strive to collect and manage most of its data as geo-referenced data. Data can be shared using existing spreadsheets then ingested into suitable platforms, which can come in different formats. Once it is placed in the different platforms data will be stored in a format that allows long-term storage and access.

[1] https://www.go-fair.org/fair-principles/
[2] https://gdpr-info.eu/
[3] https://hitrustalliance.net/product-tool/hitrust-csf/
[4] https://www.gida-global.org/care
[5] https://www.doi.org/

Data categories and levels of restrictions

Level	Category	Description	Access restriction
1	Farm data including household surveys	• These datasets potentially include personal or field identifiable information that may lead back to the farmer. Data will include yield estimates and farmer management practices. The information may include demographic information and socio-economic data by location, season/year	• Restrictions/prohibitions on re-identification of individual names or fields. (It is advised to get explicit disclosure permission from the data subject) • Data access and sharing to comply with licenses that cover particular datasets. (It is encouraged that data be covered by standard licenses e.g. Creative Commons) • Requirements to acknowledge the providers of data • Requirements to share any publications developed using the data
2	Field research/agronomic data	• This is replicated (by field and site) raw data that is collected from on-station and on-farm field research or agronomic trials. • Farmers can also serve as replicates in some cases. • Emphasis will be placed on georeferenced data to allow better analytical approach • Data can include soil and plant spectral libraries where available. • The data will include site description. • These datasets should not have personal identifiable socio-economic data of farmers or any individuals. (While it is acknowledged that in some cases data analyses are weakened or rendered useless with anonymized data, it is encouraged to utilize platforms that allow the user to perform data analytics without downloading original data).	• Restrictions on further sharing of data with third parties • Requirements to acknowledge the providers of data • Requirements to publish or post findings from any analyses done with the data • Requirements to share any publications developed using the data • Requirements to share any agronomy advisory decision tools developed using the data and include proper citation on any software licenses
3	Aggregated/summarized data	• These datasets include highly summarized	• Open access

		data across treatments, sites, seasons. These datasets can be aggregated across main treatments, different sites or years, depending on data collectors' preferences.	• Appropriate acknowledgement of data providers
4	Geospatial data	• This includes remote sensing data, images, digital soil maps • Consortium members produce this data which may be useful for developing certain applications • Members may want to and will be encouraged to contribute to different layers of this data.	• Access depends on restrictions that come with the data
5	Scientific publications	• This includes peer-reviewed journal articles, books, book chapters, reports and any other papers	• Access is guided by copyright and license conditions depending on the publication • Appropriate acknowledgement of data providers
6	Methods and protocols	• Includes plant and soil sampling protocols, laboratory, field, calibration or validation protocols • These will serve in the development of standard methods	• Open access, including the general public
7	Algorithms and software	• Open source , or proprietary source code	• Access through a generic EULA adopted by the consortium (Open Source), or access through specific EULA's issued by the software developer (Proprietary IP) •

Period of agreement

- This policy is a living document and will be reviewed regularly by the Consortium and adjusted as needed.