Session 2.8 Information Technologies and Data

Oral Presentation

Title: **Enhancing Discoverability and Re-use of CGIAR's Agricultural Data: Challenges and Progress**

Authors: Medha Devare and the Open Access and Data Management Communities of Practice
CGIAR

Abstract: CGIAR's 15 Centers and other entities involved in agricultural research and development are charged with tackling complex challenges at a variety of scales, but research outputs are too often not easily discoverable or reusable. CGIAR is attempting to enhance discovery and reuse of its data for models and other tools through the Open Access and Open Data initiative and the development of a platform to harness the power of big data and ICTs. The OA/OD initiative focuses on creating a culture of data sharing, and provides support for harmonized data through the development of common metadata, ontologies, and strengthened collaboration and coordination around tools and approaches. This foundational work will render CGIAR outputs interoperable, ensuring they are discoverable via integrated and contextualized views across Centers and programs, type (e.g., publications, data, etc.), and discipline (e.g., genetic/genomic; agronomy; breeding; socioeconomic, and other sectors). The contents of most repositories at CGIAR Centers are not generally easily discoverable or inter-linked (e.g., agronomic trial data with socioeconomic or adoption data in the same geography). In the absence of such interoperability-mediated integration, "open" is of limited utility. The overall objective, then, is to make CGIAR's trove of research data and associated information accessible for indexing and interlinking by a robust, demand-driven cyberinfrastructure for agriculture, ensuring that research outputs are Findable, Accessible, Interoperable and Re-usable (FAIR) for simulation, analytics, and visualization tools to enhance innovation and impact. This presentation will review the challenges to sharing and mining CGIAR data effectively, and the progress towards addressing these.

Oral Presentation

Title: **DataMill: a new application to interface researchers' database with crop models**

Authors: Myriam Adam[1], S. Auzoux[2], R. Loison[2], and F. Affholder[2]
[1] CIRAD, UMR AGAP, ICRISAT West Central Africa, Burkina Faso, [2] CIRAD, UR AIDA, France

Abstract: DataMill is an application built to improve access and re-use of agronomic data for crop modelling. DataMill extracts data from a standardized database and then translates these data to compatible model-ready formats for multiple crop models. DataMill is made of: (i) layouts of each input/output file from selected crop models, (ii) a Microsoft Access database (DataMill_DB) that structures all the variables of input/output files, (iii) and a Visual Basic executable file that converts the data from DataMill_DB into model input/output files in native format. The next step is to input researchers' data into DataMill_DB through queries. This step will be crucial and depends on how each researcher is structuring and storing its data.

DataMill facilitates the creation of model input files, one of the major bottlenecks in the use of crop models. Currently, DataMill is working for SARRA-H and DSSAT formats input files and is in development for APSIM and STICS. The concept of DataMill and an example of its use will be presented at the conference.