Modus: A Standard for Big Data

Dave Nerpel, Jason Ellsworth, Aaron Hunt

Insero Solutions, Ellensburg WA; Wilbur-Ellis, Kennewick WA; Agrian, State College PA

A paper from the Proceedings of the
13th International Conference on Precision Agriculture
July 31 – August 4, 2016
St. Louis, Missouri, USA

Abstract. Modus Standard is a system of defined terminology, agreed metadata and file transfer format that has grown from a need to exchange, merge and trend agricultural testing data. The three presenters will discuss steps taken to develop the system, benefits to data exchange, current user base and additions being made to the standard.

Keywords. *Big Data, Soil Data, Agricultural Data, Data Management, Soils Labs, Farm Management Information Systems*

Issues Addressed by the Modus Standard:

Over 200 testing facilities, private and public, in North America generate agricultural test results for a large number of users. Some of these have regional and international reach. Analysis results are used for crop decisions, sometimes saved for historical trending. Some data is from field composite sampling, some from multi-sample point data set used for understanding and treating within-the-field variations. The data is ingested into an array of FMISs (Farm Management Information Systems) or kept for historical trending on particular fields.

Over time an array of data transfer formats have developed, many FMISs have evolved unique requirements forcing testing labs to report in a large number if file formats, managing a difficult process of getting the right data in the right format to the right user. Long names for analytes, extraction methods, analysis and observation procedures and units of

measurement were often labeled with differing abbreviations from one information system to another.

Industry and academic attempts to amass, trend and analyze data have been frustrated by:

- Lack of common terminology of reporting field testing results
- Lack of procedure metadata
- Inadequate structures for data communication

Increasing focus on farm data has brought to light several issues. There are a number of enterprise attempts underway to store and trend large amounts of data over time and geographies. Additionally, agricultural industry groups such as the International Plant Nutrition Institute are attempting to understand long term patterns to plant food use and cropping practices. At a time when soil and tissue testing is growing rapidly there has become a clear need to deal with ag laboratory analysis results in a more transparent way. Industry members have also come to the understanding that data transparency is key in identifying solutions to environmental challenges.

M2M (machine to machine) data movement, e.g., moving data from one computer or data system to another, requires commonality of terms and transfer format.

Developing the Solution

Discussion of a standardized approach to data handling of lab testing data was initiated with an in-concept sketch of the problem at InfoAg 2013 by the presenters of the current session. There was large agreement to the identification of the need from the 100 or so participants at the session. This included FMIS developers, equipment companies, soils labs and an array of other stockholders in the agricultural data world.

A number of interested parties initiated the development of a solution to this problem over the next year. The originator group was comprised of individuals from high volume agricultural analysis labs, academia and information management systems. The name "Modus Standard" was adopted. A full description of the effort and the Project Charter project can be found at: *www.modusstandard.org*. The project is being moved forward through scheduled and impromptu webinars and there have to date been two face-to-face meetings of the group.

Current Status

Current components of the standard cover (1) nomenclature (2) sample metadata and (3) file transfer format for test results for: soil, plant tissue, feed, nematode and water for agronomic purposes.

2015 has been the first year of significant system use for Modus. Some entities have adopted use of Modus with out any direct interaction with the presenters here or the originator group. They have just adopted since the standard is free-to-use and a good practice for their business and customer base. Since the 'membership' to the group is pretty amorphous, getting an exact quantification of use has been difficult but one active FMIS has brought in in excess of 300,000 samples using Modus in their first 6 months of adoption.

Terminilogy Standardization

Dr. Jason Ellsworth heads the effort to standardize terminology for lab tests and determine a "Modus ID" for each unique analyte, lab procedure and unit of measure combination. The goal here is not to in any way limit innovative testing procedures but to denominate the same procedures with the same terms where possible.

Below is an example found at the Modus website:

S-P-AA-NH4AC.04	Phosphorus
S-P-B1-1:10.01.03	Phosphorus
S-P-B1-1:7.01.03	Phosphorus
S-P-B2-1:10.01.03	Phosphorus
S-P-B2-1:7.01.03	Phosphorus
S-P-CALD.01.03	Phosphorus
S-P-H3A1.01.04	Phosphorus
S-P-KEL.01.03	Phosphorus
S-P-LANC.04	Phosphorus
S-P-M1.04	Phosphorus
S-P-M2.04	Phosphorus
S-P-M3.01.03	Phosphorus
S-P-M3.04	Phosphorus
S-P-MKEL.01.03	Phosphorus
S-P-MMOR.01.03	Phosphorus
S-P-MOR.01.03	Phosphorus
S-P-BIC.01.03	Phosphorus
S-P-PRS.04	Phosphorus
S-P-RES.04	Phosphorus
S-P-SP.04	Phosphorus
S-P-TRG.01.03	Phosphorus
S-P-W1:1.01.03	Phosphorus
S-P-W1:5.01.03	Phosphorus
S-P-W1:1.01	Phosphorus
S-P-W10:1.04	Phosphorus
S-P-W20:1.04	Phosphorus
S-P-EPA3050.04	Phosphorus
S-PFF-M3.04	Phosphorus Fixation Factor

From this you can see there are 27 unique phosphorus tests for soil being accounted for in the standard. Each has a unique combination of lab procedures including extraction method, measurement method, primary and secondary common names, reporting units, etc. In the Analysis Nomenclature currently being accounted for there are in excess of 360 unique tests accounted for. There is similar details in Plant Tissue and the other data types.

Here the goal is to call exactly the same tests by the same Modus ID from data originator to user. This allows data handlers with data from different sources to trend data as well as making data delivery considerably easier for data

originators. They simply deliver the users Modus data in the same format regardless of end use. If there are tests that a particular originator wants to keep confidential it can be listed as "unknown" this is rarely done.

XML Data Transfer

Aaron Hunt heads the effort to standardize the data transfer format. All the above listed data types are accounted for with XML transfer. XML allows for great data detail in a human and machine readable format that makes transfer from a data originators Lab Information System to a Farm Management Information System software with an extremely high level of data integrity, accuracy and speed.

It is best for interested parties to find an example of the Submit Schema at www.modusstandard.org > Modus BitBucket Repo > Modus Schema > Source. The XML format accounts for all the complexity of the Analysis nomenclature, all the data types in the same standard format. If some lab choses to report just a few of the hundreds of potential analyses the XML is just populated with the relevant data. The unpopulated data fields are understood by the systems to have no reported values.

Adoption and User Comments

The use of Modus Standard has been rapid. Reception has been very positive because for the ease it has brought to data transfer. Based on a 80:20 factor many very large data originating labs and users of data have been the initial adopters. Others are following along.

Here are a few comments:

- Chris Langford, VP of Ag Services, Waypoint Analytical: *Modus has just makes getting complete and good information out a lot easier*
- Dr. Terry Griffin, Assistant Professor, Kansas State University: *Modus is an enabling technology for Big Data*
- Jeff Dearborn, Chief Agronomy Officer, Agrian: *Modus allows us to ingest data from numerous labs and have it complete and usable*

Summary

Based on common need for simplicity and transparency the speakers and the Modus Standard user group have enacted a solution to complex data management for agricultural testing. Others are welcome and encouraged to adopt the standard. We welcome reactions and input from all facets of the ag community.