# Artificial Intelligence and Augmented Intelligence for Automated Investigations for Scientific Discovery

PhenoHarmonIS

14-18/05/2018

Agropolis Scientific Park, Montpellier, France

Dr Samantha Kanza

AI3SD Network+

18/05/2018

AI3SD-Event-Series:Report-2

**Network: Artificial Intelligence and Augmented Intelligence for Automated Investigations for Scientific Discovery**

Principal Investigator: *Professor Jeremy Frey*
Co-Investigator: *Professor Mahesan Niranjan*
Network+ Coordinator: *Dr Samantha Kanza*

# Contents

# 1 Event Details

| | |
|---|---|
| Title | PhenoHarmonIS 2018 |
| Organisers | Elizabeth Arnaud from Bioversity International, France Sandrine Auzoux from CIRAD, Centre de coopération internationale en recherche agronomique pour le développement (Agricultural Research and International Cooperation Organisation) |
| Dates | 14-18/05/2018 |
| Programme | Programme |
| No. Participants | 110 Participants - representing 40 institutions (see Section 5) |
| Location | Agropolis Scientific Park, Montpellier, France |
| Sponsors | Bioversity International, CGIAR, Cirad, Inra, Phenome, Planteome, AGAP, Exelerate, #DigitAg, Aida |

# 2 Event Summary and Format

PhenoHarmonIS is an invite only week long workshop run every two years that is focused on harmonizing agronomic data using semantic web technologies. The scientific domains that were represented in this workshop are conservation, breeding, crop traits, agronomy and agro-ecology. The invited participants ranged from agronomists in many different areas of agriculture to data scientists and ontologists working in the agricultural domain. The participants were encouraged to provide feedback on the standards and tools that they use in their data management and analysis to identify which tools are useful, and where the issues and gaps currently lie. The workshop was also intended to assess progress made since the last PhenoHarmonIS conference in 2016, to see what tools and standards have been adopted and developed by the community.

The workshop was made up of presentations on a range of topics in the semantic agro space, interactive breakout sessions and some visits to other locations[1] to demonstrate the use of technology in agriculture. The main presentations were all run sequentially such that it was possible to attend each one. A number of breakout sessions were run in parallel to one another across several days, and some ran across multiple sessions. The breakout sessions also included flash talks and presentations that linked specifically to that section (either a presentation of the tool to be discussed and evaluated, or projects that were related to the breakout topic). There were also two social events and plenty of time for networking and discussions outside of the presentations and breakout sessions. This was the format monday through thursday and the friday was used for specific project group meetings.

# 3 Event Background

This workshop is organised and run within the Crop Ontology Community , a community that is aiming to improve data quality and representation within the agricultural domain such that it can aid the farming community. Before the PhenoHarmonIS workshop was established the

---

[1]These locations were DiaScope, an experimental unit conducting research on how plants adapt to diverse environments, and three different areas in the Cirad Campus: Rhizoscope platform (phenotyping of root system architecture of rice plants grown on glass beads); PHIV Cell imaging platform (medium throughput phenotyping and image analysis of stem anatomy for biomass quality in Sorghum); and the Abiophen greenhouse (exploration of rice phenotypes under higher $CO_2$.

crop ontology community ran several other related workshops. The Crop Ontology Community of Practice Workshop was the original precursor to PhenoHarmonIS and was run in 2011 and 2012. The initial workshops were designed to attract the interests of scientists in the ontologies that this community had been creating and to encourage their participation within this community. These tools are still being developed and some of the workshops at PhenoHarmonIS involved interactive walkthroughs and trials of these tools, with opportunities to provide feedback. The PhenoHarmonIS workshops have expanded from this to provide a place to facilitate collaboration and connections within this community and to give others in this community an opportunity to present their tools and current projects both for exposure and feedback.

# 4    Talks, Topics and Breakout Sessions

A range of topics were covered in the 32 talks that were given at this workshop (full list here). The workshop was started off with a 'Beyond the Census of Things' talk by Brian King which outlined the big picture of the agricultural ontological domain, providing a high level overview of the current status of this area, and how the current tools and technologies can be progressed in the future. The following talks ranged from a wide variety of technical presentations detailing specific ontologies or platforms that have been created, to a few very agronomic based talks that focused on specific approaches to crop breeding and the current challenges in using technologies in farming in developing countries. There were also some talks and discussions centered around the issues of data sharing and applying FAIR principles to these processes.

The main topics covered included:

- Developing Crop Trait Dictionaries
- Curating existing agricultural ontologies
- Use of metadata standards
- Ontology repositories/databases/portals
- Annotation tools
- Data interoperability tools
- Agro-ecology

Many members of this community have worked together on creating tools to aid with the representation and harmonization of agronomic data. The following Ontologies / Standards / Repositories and Tools were discussed:

- Gene Ontology (GO) - Ontology to describe gene functions (molecular function, cellular components, biological processes)
- Crop Ontology (CO) - A set of ontologies to describe crop phenotypes and traits, alongside some additional concepts such as countries, locations, and environmental conditions and methods associated with crop research.
- Agronomy Ontology (AgrO) - An ontology to represent agronomic practices, techniques and variables.
- Environment Ontology (ENVO) - An ontology designed to support the annotation of any organism of biological sample with environment descriptors.
- Planteome Ontologies - A set of reference and species ontologies to describe plants and annotate genomic and phenotypic data.
- Plant Phenotype Experiment Ontology - This ontology is an implementation of the Minimal Information About Plant Phenotyping Experiment (MIAPPE)

- Minimum Information About Plant Phenotyping Experiment (MIAPPE) - This is an open project that aims to harmonize data from plant phenotyping experiments. It contains software to handle plant phenotyping experiment data, and also provides a checklist of the minimum information needed to represent such experiments.
- FAIR (Findable, Accessible, Interoperable and Re-usable) - this is a set of principles for dealing with data.
- COPO Annotation Tool - An annotation tool for plant datasets.
- Breeding API (BrAPI) - A RESTful API to facilitate interoperability between plant breeding databases.
- AgroFIMS - A FieldBook tool to facilitate correct and detailed data entry of agricultural experiments and observations.
- HIDAP (Highly Interactive Data Analysis Platform) - A platform designed to unify best practices across breeders with respect to data collection, data quality and data analysis.

There were also 13 different interactive breakout sessions during this workshop (full list here) that ranged from demonstrating and evaluating existing tools and platforms (such as AgroFIMS), discussing best practices for ontology mapping and many demonstrations of how to annotate agricultural data, and add terms to different ontologies. Other important topics such as AgroEcology were discussed and there were discussion sessions about the current problems facing this community with respect to issues relating to incomplete data, how to (and whether to) unify certain terms across ontologies and how improvements can be made to existing systems.

# 5 Organisations/Participants

The 110 participants represented the following 38 institutions in the following countries:

Commercial Companies

- Agrimetrics - UK
- Agroknow - Greece
- Bayer Crop Science - Belgium
- IBP, Integrated Breeding Platform - Mexico
- Ontocale - Romania
- Syngenta - UK, France & Switzerland

Research Institutes

- Agropolis International - France
- Alfred-Wegener-Institut - Germany
- Bioversity International - France
- Boyce Thompson Institute - USA
- CGIAR - France, Columbia
- CIAT, International Center for Tropical Agriculture - USA
- CIMMYT (International Maize and Wheat Improvement Center) - Mexico
- CIP (International Potato Center) - Peru
- CIRAD - France

- [Earlham Institute](#) - UK
- [EMBL-EBI (European Bioinformatics Institute)](#) - UK
- [Forschungszentrum Jülich](#) - Germany
- [ICRISAT (International Crops Research Institute for the Semi-Arid Tropics)](#) - India
- [IITA (International Institute of Tropical Agriculture)](#) - Africa
- [INRA (French National Institute for Agricultural Research)](#) - France
- [Institute of Plant Genetics, Polish Academy of Sciences](#) - Poland
- [Instituto Gulbenkian de Ciência](#) - Portugal
- [IRD (Institute for Research and Development)](#) - France
- [IRRI (International Rice Research Institute)](#) - Philippines
- [National Institute of Biology](#) - Slovenia
- [National Plant Phenomics Centre](#) - UK
- [Rothamsted Research](#) - UK
- [VIB – UGENT (Center for Plant Systems Biology)](#) - Belgium

Universities

- [Cornell University](#) - USA
- [Oregon State University](#) - USA
- [University of British Columbia](#) - Canada
- [University of Exeter](#) - UK
- [University of Florida](#) - USA
- [University of Minnesota](#) - USA
- [University of Montpellier](#) - France
- [University of Southampton](#) - UK
- [Wageningen University](#) - Netherlands

# 6    Conclusions

This workshop facilitated many useful discussions, both about ontology creation and harmonization, and about agronomic practices in general. It allowed domain experts to connect with one another and share their expertise, and there were three main conclusions.

There were many discussions about the re-use of ontologies, with many participants showing an obvious keenness to ensure that this practice is upheld. However, despite this attitude, it has become obvious that this still often doesn't happen. Papers are often published detailing the creation of new ontologies, and many ontologies are created in the agricultural domain; however these are often not shared. Furthermore, there still remain many issues around the area of harmonizing / aligning ontologies. The discussions during the workshop highlighted this lack of harmonization and the fact that there are many similar but slightly different ways of doing things in agriculture (for example, a discussion about measuring crop height covered four different ways of doing this almost immediately). It is clear that the necessary 'best practices' for harmonizing these terms and creating extensible re-usable ontologies have not fully been developed yet, and that there is still a lot more work to be done in this area.

The second conclusion that was formed was that despite today's technological eras, we are still lacking in best practices to record the necessary day to day farming data that is needed to continue to make advancements in this field. There are many factors that contribute to this. Many farmers in developing countries don't necessarily have access to the resources they need to formally and electronically record this data; and even for developed farmers, there is a combination of a resistance to changing their existing practices, and a lacking in electronic systems that would facilitate this improvement. Essentially, in order to move forward in this domain, either the people collecting the data need to change their practices, or new ways of handling incomplete and inconsistent data need to be formed in order to make the best out of the data that we have.

Finally, the use of semantic web technologies remains an underappreciated endeavour thus far. The ever present 'chicken and egg' issues still remain with respect to ontologies. Many need to be persuaded of their value before they will be willing to use them or expend effort developing them, but this effort is required in order to generate that value. Additionally, funding bodies aren't always eager to fund purely ontological / data management areas of research. However, some recommendations were made as to how to mitigate this and to obtain funding to further work on these areas. We need to work towards pinpointing where one can get value from them and to create projects that aim to provide this value, and ensure that a solid data management plan is included in the project design, thus enabling further research and development to continue in this area and hopefully show why it's so important.

Feedback from the conference participants also suggested a desire to have more representatives from developing countries at the next event, to enable a better understanding of the challenges of being close to the field.

# 7    Related Events

For those who are interested in getting involved with this community and attending related events there are some additional opportunities that will occur before the PhenoHarmonIS 2020 conference.

CGIAR Platform for Big Data in Agriculture Convention in Kenya (3-5th October 2018) - this event will include some follow ups from PhenoHarmonIS 2018.

International Data Week in Botswana (5-8th November 2018) - this event, run by co-data is for data scientists and researchers of all disciplines. It will cover some of the issues with african data, which is highly pertinent to the PhenoHarmonIS community.

Upcoming events of interest can be found on the AI3SD website events page.
http://www.ai3sd.org/events/ai3sd-events
http://www.ai3sd.org/events/events-of-interest

# 8    References / Related Reading

1. Arnaud, E., Matteis, L., Laporte, M.A., Duruflé, H., Espinosa, H., Hyman, G., Shrestha, R., Portugal, A., Yves, P., Chibon, M.D. and Akintunde, A., The Crop Ontology. Buttigieg, P.L., Pafilis, E., Lewis, S.E., Schildhauer, M.P., Walls, R.L. and Mungall, C.J., 2016. The environment ontology in 2016: bridging domains with increased scope, semantic density, and interoperation. *Journal of biomedical semantics*, *7*(1), p.57.

2. Cooper, L., Meier, A., Laporte, M.A., Elser, J.L., Mungall, C., Sinn, B.T., Cavaliere, D., Carbon, S., Dunn, N.A., Smith, B. and Qu, B., 2017. The Planteome database: an integrated resource for reference ontologies, plant genomics and phenomics. *Nucleic acids research*, *46*(D1), pp.D1168-D1180

3. Gene Ontology Consortium, 2004. The Gene Ontology (GO) database and informatics resource. *Nucleic acids research*, *32*(suppl_1), pp.D258-D261.

4. Pommier, C., Cornut, G., Letellier, T., Michotey, C., Neveu, P., Ruiz, M., Larmande, P., Kersey, P.J., Cwiek, H., Krajewski, P. and Coppens, F., 2018. Data standards for plant phenotyping: MIAPPE and its implementations [W785].