Agriculture, Food Security
and Climate Change

FACCE-JPI Workshop on Big Data

Identifying the potential role of big-data for food security with a focus on collecting data, translating data into information, and promoting and facilitating use of the information by end-users (incl. via open data/knowledge policies)

Copenhagen, 22 November 2017

Aalborg University Copenhagen
A. C. Meyers Vænge 15
2450 København SV
Denmark

Executive Summary and Recommendations

The workshop was set in the context of a background paper which outlined the data challenges in the Agriculture, Food Security and Climate Change intersection and was presented by one of its authors, Mario Caccamo (NIAB-EMR, UK). He highlighted how the increasingly vast quantities of data generated by technologies in modern research require appropriate means to ensure that they can be used to maximum advantage by adhering to FAIR principles (Findable, Accessible, Interoperable and Reusable). FACCE-JPI will have to play a role in promoting these principles to its researchers.

Workshop delegates heard presentations from a wide variety of related initiatives which highlighted the wealth of work, opportunities and facilities that FACCE-JPI could make use of, or learn from when considering its own role in the area of big data. These included the Research Data Alliance (RDA); infrastructures for data management in the life sciences (Elixir) and environmental sciences (ENVRIplus); cloud-based storage and access to data (European Open Science Cloud); an initiative providing active support for data management and analysis at the joint action and project levels which can be costed into grant applications (FAIRDOM); an initiative promoting and facilitating open data within the agriculture and nutrition sectors (GODAN); linking research data through to farm-based tools and designing research together with end-users of data/tools (EIP-Agri); and an innovation hub for delivering data-driven, climate smart agriculture practices to smallholder systems (CGIAR Platform for Big Data in Agriculture). Break-out sessions served to go into more detailed explorations of specific topics.

The combined outputs of the workshop's presentations, break-outs and plenary discussions resulted in the following recommendations to FACCE-JPI:

Recommendations:

1. FACCE-JPI needs to **actively** be an advocate of the FAIR data principles. This can be achieved through:
   - Developing a data sharing policy and making data management plans for future FACCE-JPI calls mandatory (looking to other examples to see which have worked well)
   - Conducting an exercise with past/current FACCE projects to ensure/promote that data generated in the projects is FAIR
   - For specific actions, FACCE-JPI could consider promoting the use of initiatives such as FAIRDOM who can provide cohorts (at the joint action level) of researchers with tailored data management/analysis tools
   - Ensuring the FACCE-JPI research community is connected to the relevant research infrastructures to facilitate the FAIR principles

2. FACCE-JPI should **convene an expert working group** to develop a living roadmap looking at short- medium- and long-term priorities for data in its remit
   - To implement the above data sharing policy and further develop the above FAIR recommendations
   - Identifying how to operationalise interactions with key data initiatives and leverage existing resources for data sharing and appropriate centralisation of data for the benefit of FACCE-JPI
   - To map FACCE-specific data needs and availabilities; should gaps be found where specific facilities/infrastructure is lacking, work with existing infrastructures to address the need
   - Following an analysis of the suggestions in the workshop, to propose actions to be funded by FACCE-JPI (e.g. a Knowledge Hub/Network of Excellence for data in FACCE's remit; exploratory research projects to tackle specific issues of data integration)

3. When designing actions, FACCE-JPI (and applicants to its calls) should consider at the outset the potential translation of research data to usable tools, and **co-design with end-users and industry** to ensure such translation is achievable and to leverage pre-competitive data for mutual benefit.

1. Introduction – Niels Gøtke, Hartmut Stalb and Gabriela Pastori, *FACCE-JPI*

Big Data presents both a challenge and an opportunity for all areas of science, including FACCE-JPI's remit at the intersection of Agriculture, Food Security and Climate Change. This is reflected in the FACCE-JPI Strategic Research Agenda, which includes the research priority:

Identifying the potential role of big-data for food security with a focus on collecting data, translating data into information, and promoting and facilitating use of the information by end-users (incl. via open data/knowledge policies).

FACCE-JPI will use this workshop to begin to address this priority by bringing together relevant experts and initiatives to discuss and advise FACCE-JPI on its role, in particular in the following two issues:

- Exploring how FACCE-JPI can maximise the impact of the data generated in existing and future research projects
- Identifying Big Data challenges in FACCE-JPI's remit, and exploring what FACCE-JPI's role can be in addressing them

The following objectives were identified for the workshop:

- To bring together relevant experts and stakeholders
- To explore research needs and research gaps
- To identify potential application and integration (reassembling) of relevant new and existing data
- To identify how to maximise impact in FACCE-JPI projects through use of existing data
- To identify infrastructures and tools to be used by FACCE-JPI at joint action level
- To provide recommendations to the Governing Board for where FACCE-JPI can contribute to addressing the challenges and opportunities of Big Data in FACCE-JPI's remit

The workshop aims to provide advice to the 22 member countries of FACCE-JPI about what should be done next. Guidance is sought from the workshop delegates as to the next practical steps for FACCE-JPI to take, as well as identifying the challenges in the longer term. This will, of course, have to be done in partnership with others, so learning from other relevant initiatives will be key for FACCE-JPI to develop its own role in the Big Data landscape.

2. Setting the Scene

Two presentations were given to set the scene – firstly, Peter Baumann, co-chair of the Research Data Alliance Big Data Interest Group, gave a view point from outside of the FACCE remit, to give an impression of what the current state of Big Data is from his own expertise. Secondly, in preparation for the workshop, FACCE-JPI developed a background paper identifying key issues within the remit of Agriculture, Food Security and Climate Change. The paper (available here) was presented by one of its authors, Mario Caccamo (EMR-NIAB, UK)

Potential of Big Data for Agriculture, Food Security and Climate Change – *Peter Baumann, Jacobs University Bremen, Germany / rasdaman GmbH; Co-Chair RDA Big Data Interest Group*

Understanding data structure is critical for processing and analysis of big data. The structural variety in big data can be grouped into basic data structures namely arrays, graphs, sets and hierarchies, that typically come in many different variations and combinations. Data relevant to agriculture and climate change are also structured accordingly, for e.g. climate modelling data, satellite imagery, genome data etc. are often structured in arrays, taxonomies as hierarchies, ontologies as graphs and so on. Much of big data comes in the form of raster data (arrays), a data structure consisting of matrix of cells where each cell contains a value representing specific information like temperature, aridity etc. Spatial and temporal sensor data, aerial images, time series simulations and statistical data are some examples of data that are usually stored in raster format.

There has been significant increase in the amount, volume and complexity of data collected from different sources like sensor feeds and through modelling and simulations. This poses a need to enhance the ability to collect, process and analyse these massive data sets to underpin the study of Earth systems that enables application of knowledge in sectors like agriculture. In this context, United States Geological Survey (USGS) and specifically the Landsat 8 working group coined the term 'analysis-ready data' to indicate data that can be easily exploited by both (non-expert) humans and tools. The concept is to process and aggregate raw data close to the source of collection, where expertise, facilities and infrastructure are available. This principle of 'analysis-ready data' was extended towards the concept of a 'data cube', a multi-dimensional data aggregation to include spatio-temporal data like time series, and properties such as elevation, bathymetry etc.

One of the challenges for analysis-ready data is the lack of homogeneity to allow linking related data sets and concepts for complex data analysis. The Open Geospatial Consortium (OGC), International Organization for Standardization (ISO), and OASIS (Advancing Open Standards for Information Society) provide a suite of standards for accurate and efficient processing and analysis of geodata for agriculture and climate change. The OGC standards include different standards to allow data manipulation for a variety of geographic features from particular geographic data servers e.g. WFS (Web Feature Service) for sharing data on geographic features from specific sources, WMS (Web Map Service) for visualising geographic data on world wide web, WCS (Web Coverage Service) for provision and visualising of geospatial data from a web server etc. Some examples where such standards are employed include projects like EOfarm Startup where raster data is managed through OGC standards framework. A similar framework is deployed for water quality monitoring using data from sources like Landsat8, Sentinels, and RapidEye. Some of the features that can be analysed include land surface temperature, canopy greenness, vegetation etc. Another example is the BigPicture Project where site specific variations are monitored and analysed using data from satellite images. It allows for recommendations to be derived for targeted measures like application of plant protection products, fertiliser placement etc.

Big data processing platforms and data management systems are required to mine large volumes of complex data. Rasdaman or raster data management is an array database management system that provides a flexible, scalable, standards based service of storing and retrieving multi-dimensional spatio-temporal data, images, and statistical data such as river discharge, soil moisture, precipitation etc. It is a blueprint for big datacube standardisation. Given datacube advances to support earth data systems, the European Space Agency (ESA) is currently preparing for a European Earth Data Observation (EO) Datacube Initiative with an aim of advancing data analytics, enabling a flexible framework for services through open standards, maintaining a European technology lead, and enlarging EarthServer as a reference example.

For the future, it is important to consider the emergence of artificial intelligence (AI) and machine learning (ML) which will have an impact on processing and analysis of complex data. There will be challenges associated with implementation and decision making through AI and ML that will need to be addressed.

Big Data in the FACCE context: overview of FACCE-JPI Background Paper on Big Data; *Mario Caccamo - NIAB-EMR, UK*

There are many challenges and opportunities relating to data in the remit of FACCE-JPI, i.e. Agriculture, Food Security and Climate Change. We live in a unique time, where technologies, science and connectivity are all closely interrelated, and all have a wealth of data associated with them. Typically, technologies are developed bottom-up, and then find applications in science. It is only when the impact of the technology is realised in science that data standards are considered, however this is often too late.
There is a generic pipeline for data in the biosciences – a large volume of data generated through technologies at source, translated to information, generating knowledge, which is then interpreted to form discipline-specific biological answers.

An example of this pipeline is plant phenotyping. Phenotyping platforms gather large volumes of heterogeneous data in a very efficient way from many different sources and using different technologies, e.g. one facility in the USA generates 5TB of data per day. However, it is not possible to keep all of this data, and so it must be translated to useful information. An additional challenge is to then make the step from using the data-driven technologies in science to making them relevant and accessible to end users.

There are other examples of data pipelines throughout the FACCE-JPI remit, such as in field trials or in livestock monitoring.

Regardless of the particular area of science or the specific technology used to generate data, one thing is critical to ensure that the use of these data can be maximised is that they must be FAIR. That is they should adhere to the following principles:

- Findable (e.g. indexed and searchable, and machine accessible)
- Accessible (in repositories and using open-access policies or well-defined licences)
- Interoperable (e.g. using Application Programming Interfaces [APIs] and appropriate ontologies)
- Reusable (e.g. data can be integrated)

Application Programming Interfaces allow different applications to use a common language, facilitating integrated data sharing and data use.

There are also challenges that are not technical in nature. These include building trust in new technologies, perceptions on the use and sharing of data, and having appropriate regulatory frameworks in place to facilitate this. An example where this is being addressed, from outside of FACCE's remit, is the Pistoria Alliance in the life sciences. This brings together life science companies, vendors, publishers and academic groups in pre-competitive collaboration to lower barriers to innovation.

A number of different stakeholders need to be considered for FACCE-JPI in the context of data:

- **Policy Makers and Funders** – promoting and encouraging FAIR principals; data management plans
- **Researchers** – is data in the right place and accessible, are models functional at different scales, how can they adopt FAIR principles
- **Trainers/Education** – ensure there is not a skills shortage in a new generation of (data) scientists; promoting FAIR prinicples
- **Service Providers** – opportunities for new business (with a need for sustainable business models); vehicles for innovation. Incentivising FAIR principles
- **End users –** affordable; easy-to-interpret; trust, adoption and sharing. How can growers benefit from FAIR principles?

3. Working with Others

The European landscape of researchers, initiatives and organisations working towards addressing challenges in big data and data management is vast and complex. In order to understand FACCE-JPI's position in this landscape, and where links and synergies can be made to advance the topic in our own remit, a number of relevant actors presented their own viewpoint in this landscape. Full presentations are available here, and are summarised in brief as follows:

European Open Science Cloud - *Wim Haentjens, DG-Research and Innovation, European Commission*

The Food 2030 policy framework, which aims to future proof food systems by 2030, has set the context for the European Open Science Cloud (EOSC). The overall context is that the challenges facing food systems are such that ambitious research and innovation programmes are required that comply with the principles/criteria of sustainability, resilience, responsibility, diversity, competitiveness and inclusivity. The proposed Food 2030 framework outlines the priorities for Europe. A food systems approach must be applied in order to address the below mentioned priorities and systemic issues in these areas:

- Nutrition for sustainable and healthy diets
- Climate smart and environmentally sustainable food systems
- Circularity and resource efficiency of food systems
- Innovation and empowerment of communities

Research breakthroughs, innovation and investment, open science and international collaboration are drivers of these priorities.

In this context, the EC is seeking to bring stakeholders together to improve the innovation ecosystem, to work with international partners and to further open science to fields relevant to food systems. The Food

2030 expert group was launched in September 2017 and the Food 2030 World Food Day conference took place in October 2017. Following these events, FIT4FOOD 2030 was launched in November 2017. Member state mapping of food systems research and innovation took place in December 2017.

The first report of the High Level Expert Group on EOSC highlighted the transition from data-sparse to data-rich research science and pointed out that whilst data loss is happening and is significant, data growth is staggeringly rapidly. The expert group concluded that it is necessary to support researchers in a transition towards data-rich research and offered recommendations that led to the realisation of European Open Science Cloud (EOSC). Some past and ongoing activities in EOSC include EC communications on 'European Cloud Initiative', H2020 open data policy, and the 1st EOSC Summit of June 2017. The EOSC Declaration of October 2017 outlines the principles of implementation of EOSC and actions list of specific actions to be implemented in the context of EOSC.

To drive EOSC forward, investments will be made in H2020 calls between 2018-2020 (INFRAEOSC, Food Cloud) to respond to researchers' need, an EOSC Stakeholders Forum has been held in November 2017 and an EOSC Roadmap is to be established to define the governance structure, broad architecture and financing of EOSC.

GODAN (Global Open Data for Agriculture and Nutrition) – *André Laperrière, GODAN Executive Director*

The idea for GODAN (Global Open Data for Agriculture and Nutrition) was conceived by the G8 in 2013 at the International Conference on Open Data for Agriculture. The initiative is a knowledge sharing platform that supports the proactive sharing of open data to make information about agriculture and nutrition available, accessible and usable to deal with the urgent challenge of ensuring world food security. This is in the context of a data-intense landscape, e.g. in 2025 there is predicted to be 1000 satellites orbiting the Earth generating data. In the same year, the world will generate 180 zettabytes (180 followed by 21 zeros) of data. GODAN brings together over 600 member organisations from all sectors who have not previously interacted, and helps them to navigate the data and put them in touch with data relevant to them. Their resources adhere to the FAIR principles described above, but extending the "accessible" principle to mean that the data is accessible in a way that enables users to understand and make sense of the data.

Open data is needed to make better choices in agriculture, and this is recognised in the private sector. Agriculture accounts for one third of GHG emissions, and is central to many of the Sustainable Development Goals, however there are few collaborative platforms to exploit and integrate datasets to enable best practices that will help to adapt to future weather variability.

Case studies were presented which exemplified the importance of open data and the translation of these data into tools that can be directly used by farmers, emphasising the need for intermediaries who are able to interpret and shape the data and make them usable:

- The "Climate FieldView" system is a precision agriculture tool that can be used in a tractor and allows for live simulations (using open data), such as implications of using or not using fertiliser.
- A simple low-tech input station to gather on-farm weather data that can be shared and used in predictive tools that integrate weather data with data on soil conditions to inform farmers when it is necessary to irrigate

- Artificial Intelligence for data integration can be used in apps at the users' fingertips, such as photographing an insect, and instantly receiving tailored advice on pest management.

Further information and brochures are available on the GODAN website: www.godan.info

ENVRIPlus: Reusable solutions to common challenges in Environmental Research Infrastructures – *Zhiming Zhao, University of Amsterdam*

Addressing global challenges such as climate change and food security requires a systems approach towards understanding the Earth's subsystems and their interaction with the environment. Research Infrastructures (RIs) in the environment thematic area provide key tools for addressing these challenges. ENVRIplus is a H2020 project that brings together Research Infrastructures (RIs), projects and networks in environmental and earth systems science in order to create a coherent, interdisciplinary and interoperable cluster of RIs across Europe. The project has a duration of four years with a budget of 15M€. It brings together 37 partners and more than 20 RIs across Europe under six common themes namely technical innovation, data for science, access, social relevance, knowledge transfer, and communication and dissemination.

The data-related activities of ENVRIPlus RIs are well integrated under the 'data for science' theme of the project. The theme aims to address similar challenges faced by many RIs and share solutions to common challenges and issues. These include supporting systems level science, interfacing with virtual research environments (VREs), and re-using technologies e.g. from e-infrastructures. Some of the activities in this theme have aimed at understanding common challenges and requirements in data from a multi viewpoint modelling approach, and to provide a common ontological framework based on the ENVRI Reference Model (RM). The 'data for science' theme relates to the design, implementation and operation of environmental RIs and focusses on key issues of data identification and citation, processing, optimisation, curation, cataloguing, and provenance, supported by a common architecture design, meta information linking and a reference model (RM). ENVRI RM illustrates common characteristics of environmental RIs in Europe from various perspectives.

Some of the highlighted achievements in ENVRIPlus includes an improved ENVRI Reference Model that offers a framework to promote interoperability between infrastructures, development of Open Information Linking for Environmental RIs (OIL-E) to provide a framework for linking knowledge resources (standards, models, vocabularies etc.) used by different environmental RIs, a Knowledge base with semantic description and reasoning tools derived from various RIs, and a service portfolio containing information about services that ENVRIPlus offers to the ENVRI community and beyond.

ELIXIR: From Research Infrastructure to Application - *Paul Kersey, EMBL-EBI*

We live in an age when there are massive quantities of data being generated (faster than we can deposit it) and this is increasing exponentially. Particularly in the case of the food sector, data production is very distributed which can cause bottlenecks and a challenge to integrate these different data. There is a large number of different databases available (around 1,800 for molecular biology alone), but the challenge is how we can be more strategic in investment, how we can ensure that resources are interoperable, and how we can harvest national investments to ensure a common European data landscape.

ELIXIR is a research infrastructure which was formed to address the challenge of biological data management, and has a broad remit covering environment, bioindustries, agriculture and medicine. It is has wide buy-in from across Europe, and the number of countries participating is continuing to grow. ELIXIR is a distributed infrastructure where there is a Hub to coordinate activities between national Nodes based in member countries, with each node acting a hub for national activities. The net result is a hierarchical hub-and-spoke model that aims to coordinate a huge number of resources in a practical manner.

Plants/agriculture is one particular area that ELIXIR is currently interested in. Here there is a mission to facilitate genotype-phenotype analysis for crop and tree species through the following actions:

- Develop standards for representation of genotypic and phenotypic data
- Make data discoverable and interoperable through common APIs
- Annotate and submit key exemplar datasets to relevant public archives
- Develop reusable modules for visualisations
- Disseminate best practices and tools to national projects

Through this work there is the aim to have comprehensive catalogues of genome variation in collected "accessions", comprehensive screening for phenotypic properties, and a Genotype-Phenotype lookup as a powerful tool for breeding and biology. One component of such a system is Ensembl, which is a modular suite of software for genome analysis and visualisation, containing the genomic information for a wide number of organisms (pests, pathogens, crops, animals, pollinators, symbionts etc). This is available through a web-based interface and a series of programmatic interfaces.

However, not all data that would be ideal to have in this and other systems is available. There are a number of reasons for this: sociological barriers (e.g. data ownership and privacy); field phenotypes are environmentally-dependent (there is a need to unpick Genetic x Phenotypic x Environmental interactions); advances are needed in handling phenotypic data (few public archives, weaknesses in standards for data description, hugely variable experimental approaches, a need for standards in metadata and the various levels of metadata). Some of these are being tackled in a coordinated approach, such as developing standards, via MIAPPE (Minimal Information About a Plant Phenotyping Environment) and BrAPI (Breeding API) to provide a means to share standardised, machine-readable, interoperable data and metadata.

Similar activities are also taking place in the livestock sector, and in some ways they are more advanced than in the plant sector. However, overall, agriculture still struggles to get the same levels of publicity and funding when compared to the medical/health sector, and relies more on small-scale collaborative efforts and in-kind contributions. A few flagship projects could have a transformative effect.

ELIXIR is keen to increase its user base and to develop relationships with new user communities, which could include FACCE-JPI funded projects. ELIXIR has also had success in working with pharmaceutical companies in a pre-competitive manner to develop tools for the pharma community. This could be replicated in the food and agriculture sectors. For climate change, there aren't currently strong enough links with climate scientists, and this could be strengthened.

CGIAR Platform for Big Data in Agriculture: Big Data Opportunities in Climate Smart Agriculture –
Andy Jarvis, International Centre for Tropical Agriculture (CIAT)

The new generation of technologies such as survey drones, agribots, GPS-controlled vehicles, and sensors will revolutionise the farms of the future. Using connected digital tools will allow smart use of small farms to better integrate the entire food system. Of 570 million farms around the world, almost 72% are smaller than 1 hectare and there is huge opportunity in digitalisation of agriculture to make data available at small scale to help farmers with the goal of improving production while minimising the cost and burden on resources. Digitising agriculture will help farmers in every aspect of the agricultural value chain from input to on-farm production, harvest and storage, and access to market (e.g. helping farms to plan what and when to plant, share best farming practices, monitor storage conditions, increased ability of smallholder farmers to sell to larger markets by allowing buyers to track crops to sources etc.).

The four major changes that have paved the way for digitalisation in agriculture include the massive surge in access to mobile phones (about 6 billion people have mobile phones, more than people having access to a toilet), an increasing number of satellites in orbit collecting massive amount of earth and geodata (currently 2271 satellite in orbit), smart and cheap sensors becoming standards to agriculture and internet of things for analytics, and improved analytical capacity including scenario modelling.

The CGIAR platform for big data in agriculture aims to increase the impact of agricultural development through big data approaches to solve problems faster, better and at a greater scale. The platform provides leadership in organising open data, convening partners, and demonstrating the power of big data analytics through inspiring projects under the 'organise', 'convene', and 'inspire' principles. The platform convenes current communities of practice that includes crop modelling, data driven agronomy, geo-spatial data, ontologies, and socio-economic data. The Platform also organised the Inspire Challenge 2017 to identify and address agriculture and climate problems using new approaches and innovation processes. The call for proposals were requested in four categories – novel approaches to monitor pests and diseases, data driven farming, disrupting impact assessment, and revealing food systems. A total of 120 proposals were received of which 5 winners were awarded US$ 100,000 innovation grant. Three of them were looking at pest and diseases issues including real time diagnostics for wheat rust (sequencing in the field to understand pathogen in wheat), pest and disease monitoring by using artificial intelligence (smart phone application for automatic sensing of diseases in roots and tubers) and using social media such as Facebook to real-time monitor livestock disease and problems, and provide help-desk and digital extension to farmers through social media. Others include using voice response marketing service (full value chain approach using cell phone to move data from start to finish), and smart phone camera data.

Some examples of big data in climate change and food security include data mining literature and information for climate smart agriculture priority setting to understand the impact of a range of climate smart options and understanding potential impacts. Through digital advisory services, commercial farm data and mobile phone extension systems can be used to identify crop/climate relationships and provide site-specific recommendations to farmers. Another example of using big data is to improve climate services through ICT by using machine learning for more accurate climate forecasting.

The European Innovation Partnership EIP-Agri: Catalysing innovation for productive and sustainable EU agriculture and forestry – *Willemine Brinkman, EIP-Agri Service Point*

EIP-Agri is one of the European Innovation Partnerships, which are collaborations between the Member States and the European Commission. It is an initiative to catalyse innovation in agriculture and forestry in Europe to make it more productive and more sustainable. It consists of Thematic Networks and Multi-actor Projects (funded by H2020) and Operational Groups (funded by the Rural Development Programmes) – there are over 100 Rural Development Programmes which have co-funded between regions, countries and the European Commission. The ultimate goal aim of EIP-Agri is to connect farmers and foresters in the field with specific problems to those elsewhere with solutions as quickly as possible.

There are over 3,000 planned Operational Groups, and currently 400 are active, including those that are relevant to Big Data (e.g. Vignoble 2.0 and "Innovative Use of Emerging Technologies to Improve Pig Production Efficiency"). The EIP-AGRI Thematic Networks serve to facilitate the uptake of technologies that are almost ready to go into practice, and to translate them into something usable. There is potential here for follow-up for some of the FACCE-JPI projects.

The EIP-Agri Service Point connects all the different aspects of the network, building a bridge between research and practice, through its website, publications, workshops and seminars. Recently, this has included a workshop on Big Data bringing together a wide range of stakeholders (farmers, advisors, administrators, agroindustry etc.). The workshop identified that a crucial issue around data sharing is the ownership of data. Good practices were recommended to create an enabling environment for data sharing: enhancing transparency and trust when sharing data, accurate terms of use and licencing, and clear incentives. Equally importantly, recommendations were made of what not to do: hiding intentions of data use, over-centralising data, and overregulation.

FACCE-JPI could benefit from EIP-AGRI in the following ways:

- Including the interactive innovation approach in calls by involving end users at the start of projects or in the design of calls
- Sharing calls for EIP-AGRI focus groups
- Explore Operational Groups and other projects for links
- Participate in EIP-AGRI events
- Share FACCE-JPI project results with the EIP-AGRI network (by email: servicepoint@eip-agri.eu)

FAIRDOM – *Andrew Millar, University of Edinburgh*

FAIRDOM (FAIR Data, Operations and Models) originated in Systems Biology, but this area of science shares common challenges with the FACCE-JPI community in that data generated and models used in the respective remits are diverse and heterogeneous. FAIRDOM is a project that merged two data management systems (SEEK and openBIS) to provide consultancy, software and resources for data management.

FAIRDOM operates at the project scale, similar to those seen in FACCE-JPI – e.g. a medium-scale field trial with a few Gigabytes of data, with no available repository to collect or distribute data.
The first part of the resource is the FAIRDOM Hub. This is a public data commons where any user can go to find and/or deposit data. The Hub has been running for two years and has over 90 multi-partner projects using it in self-managed spaces.

The second part of the resource is a highly-customable local installation of the FAIRDOM platform that can connected to in-house storage systems, equipment and data processing routines. This will directly analyse the data that is deposited into the installation. The data deposited can be fed through to be shared on the FAIR Hub catalogue in a systematic way, ensuring (minimum) metadata is captured immediately and associated with data in a consistent way. The system is set up to be able to merge data held in-house, remotely or in FAIRDOM itself in a single logical structure. It can bring together not only datasets, but also models, papers, protocols etc. that can be given a single unique DOI that can be referenced in a publication.

FAIRDOM has a very fine-grained permissions structure which allows users and projects to control how open and accessible they want their files to be, enabling staged sharing of data as a project matures. This is facilitated through permission controls, licences, negotiated access and embargos.

For FACCE-JPI, central coordination of resources for data management is critical. This is not a single central resource, or just a code of practice. There needs to be tools available to fulfil requirements of a code of practice. The model that FAIRDOM has worked on is to have projects factor in 5% of the cost of the research to facilitate data management via FAIRDOM's tools, but to also factor in the time required to learn and engage in data management. Although FAIRDOM comes from a different community, it shares common challenges, and has the resources and knowhow that could be directly useable for FACCE-JPI.

4. Discussion on Quick Wins for FACCE-JPI

Mapping Relationships
The presentations from other initiatives demonstrated that there is a need for **FACCE-JPI to map where the links are to these initiatives, and how we want to work with them**. We can look at other communities more advanced in data sharing and use their solutions. We can look to see how FACCE-funded scientists can get involved with EOSC and the Food Cloud, and how we can link our researchers to the infrastructures that are available.

FAIR Principles
FACCE-JPI has funded many research projects, and will continue to do so in the future. The challenge for FACCE-JPI is to ensure that the data generated through these projects are not lost. **FACCE-JPI has a clear need to ensure that FAIR principles, as described above, are adhered to in FACCE research projects, through developing a Data Sharing Policy, and providing access to the tools necessary to implement the policy** (e.g. data management plans, recommendations for suitable infrastructures/repositories, guidance on data ownership to be considered at the start of projects). Initiatives such as FAIRDOM have shown that trust can be built up in communities, and that this can result in sharing data on terms that are comfortable for those who have generated the data as well as those who are accessing the data. On the technical level, everything is possible to deliver FAIR principles, but we need to ensure that the FACCE community is ready for it, as the type of research funded in the FACCE-JPI remit is a step behind other communities more used to data sharing.

The primary stakeholders will need to be identified to ensure that any data sharing policy developed by FACCE-JPI serves the needs for as many as possible. While it is easy to focus on researchers and academics, ensuring engagement with other stakeholders, including farmers, who would use research data will be important for a successful policy.

Promoting *appropriate* centralisation of data, and leveraging existing tools available through other initiatives (such as ELIXIR or FAIRDOM) will help to drive forwards standardisation of data, and will facilitate interoperability. It is important not to be over-idealistic – set in place some guidelines and put them into practice, this will start the process of building trust and developing good practice. It is an iterative process and will develop over time, but it needs to be started soon for FACCE-JPI.

Existing FACCE-JPI projects
While implementing a Data Sharing Policy will help to promote FAIR principles in FACCE-JPI research funded in the future, there is a need to engage projects which were funded before the policy is put in place. Such projects were not obligated to adhere to open data practices, but FACCE-JPI should be pro-active in contacting the projects not only to promote (rather than mandate) FAIR principles, but to also learn about what datasets they have generated, and what data needs they have. For example, could their research be advanced if they have access to specific datasets that are not currently available? This type of information could feed into other initiatives such as EOSC.

Analysis of data
FACCE-JPI is not in a position to be developing tools for integrating and further analysing its community's data, however, through **building relationships between researchers and research infrastructures**, over time the gaps in analytical capabilities will be identified and can be addressed by those infrastructures.

5. Breakout sessions

Breakout Group A: Demands and availabilities (e.g. for models, new research) of data, and bringing together data from different scales/disciplines

Breakout group A discussed the demands and availabilities (e.g. for models, new research) of data, and bringing together data from different scales/disciplines. The general consensus was that there was some understanding of demands and availabilities of data in the remit of FACCE-JPI at the intersection of agriculture, food security, and climate change, but the overview was patchy. The group suggested that the short term objective for FACCE-JPI could be in carrying out the following streams of activities in parallel:

- **Mapping data generated in the realm of FACCE-JPI** – the intersection of agriculture, food security and climate change. This includes mapping data generated by projects that have been funded by FACCE-JPI. In doing this, FACCE projects could also identify gaps in data.
- **Mapping needs and demands of data required based on specific research topics:** The FACCE-JPI Strategic Research Agenda (SRA) could serve as the source for the research topics. It will be essential to identify research communities (farmers, policy makers etc.) in carrying out the mapping exercise.

The group also discussed some requirements for mapping data availabilities and demands in order to help understand and fulfil critical gaps in data that are needed for undertaking analysis related to agriculture, food systems and climate change effects. Research topics and research communities should be clearly identified and a good representation of countries is required to understand the status of available regional and national data. In addition, researchers from other disciplines with big data experiences will also help identify gaps and inter-disciplinary opportunities.

Several instruments were suggested to help undertake mapping exercises. This includes short targeted visits and workshops to bring together experts and stakeholders from specified and from different disciplines at a pan-European level and focussing on 'use-cases' to demonstrate use of data. Multi-stakeholder interactions is required to increase visibility and accessibility of data, data platforms and frameworks. **There was a general agreement around the need for a network of excellence and knowledge hub.** Through this instrument, regular meetings would be organised to build a comprehensive, dynamic mapping. Learning and exchange activities between and within disciplines and at European level could also be organised within the framework of the network / knowledge hub. These activities will help address the goals of mapping demands and needs as well as building best user cases linking to current status of availability of data. The group noted the need for infrastructure and resources for long term continuous surveillance of data and for an integrated approaches to address the dynamic developments of data needs and demands.

In the plenary discussion, there were differing opinions on what the purpose of the Knowledge Hub or Network of Excellence proposed by the group would be. It could be used for mapping datasets and gaps, and developing best-use cases that have come from FACCE-JPI projects, but it would need sufficient resources to be more than just a network. There are other initiatives in the landscape that FACCE could encourage the use of, or help to shape, such as EOSC Science Demonstrators. FACCE-JPI has a strategy to help with the valorisation the research projects that it funds, and so case studies relating to datasets and best practice could also form part of this on-going work. **A small focus group could be set up to determine the best route for FACCE-JPI to take in this space, and what FACCE-JPI would want to achieve both in the short and long term for such a network.**

Breakout Group B: Academic demands for private-sector data

Group B's topic for discussion was identifying academic needs for private-sector data. The group considered the private sector to be both industry AND farmers, as they are both part of the private sector and each generate data that could be of interest to academia. The group first identified issues and proposed solutions to overcome these issues:

- It is not a one-way street: knowing what private sector needs are from academic data is just as important.
- The value of data being shared needs to demonstrated, and incentives for sharing data need to be provided to those who hold the data. This is not just monetary, it is about demonstrating the benefit that can be jointly realised by sharing data between private sector and academia
- An understanding of what specific datasets are needed for FACCE-JPI research from the private sector will enable a more targeted dialogue with the private sector

- The effort of providing data should be minimal, particularly for farmers. E.g. ensuring that the method for depositing data for their own use can also be used to share it more widely.
- There may be data held in the private sector that is crucial to climate change research – this must be a priority to try and access for public research. Using publically-held data to leverage private sector data would be a possibility, but it was unclear how to achieve this.

Potential solutions that FACCE-JPI could implement were identified:

- **When designing FACCE-JPI actions, use a multi-actor approach that includes the private sector at the outset to create a win-win situation.** There is a need to identify pre-competitive data in our remit where there would be added value to both academia and the private sector to combine datasets
- To try and publicise good practice, there could be a FACCE-JPI award or specific case studies for research that has successfully leveraged private sector data. This would serve to demonstrate the value of data.
- A workshop could be held to specifically identify the sources of data. This could be done in conjunction with national ministries to leverage national experiences of private sector collaboration, and bring in farmer cooperatives and advisors.

The mapping of sources of private sector data suggested in this group could be combined with that suggested by Group A. Having a clear idea of what relevant public datasets are available will also be of use to the private sector, and could help leverage access to pre-competitive, privately-held data. Concerns were raised that this may be beyond the scope of what FACCE-JPI could/should be doing, rather it could have a role in working with others who may be doing something similar, and to make smaller steps towards industry engagement.

FACCE-JPI could consider coordinating some funding across ministries to facilitate some pre-competitive industrial/academic collaboration. There are others working in this space (e.g. EARSC, Catapults etc.), so tasking a small group in FACCE-JPI to clearly define FACCE's role in this would be important before embarking on any activities.

Breakout Group C: Making big data relevant for end-users; translating data into tools

The third group considered how to make big data relevant to end-users, and how to translate data into useful tools for these end-users. They considered end-users to be a broad group of stakeholders including farmers, farmer groups, cooperatives, agricultural service provision companies. Other researchers were also considered an important end user of big data, as they are able to pick up FACCE-JPI results and take the research further down the application chain.

As mentioned in the other groups, there needs to be security factored into any data provided for use by end users (e.g. farm data, which may contain health or business data specific to particular farms). This is particularly true for those end-users who are also generators of data, such as farmers. There is a circularity whereby data generated on farm can be delivered into a "backend" repository of data, consolidated with other farm and public research data, and fed through (via an open API) to SMEs to develop on-farm tools (the "front end"). Clarity that such tools are powered by data provided by farmers also helps to build trust in and commitment to data sharing, as the benefits of doing so can be demonstrated. A simple method of providing on-farm data would be essential for this, e.g. by combining it with reporting required through the Common Agricultural Policy.

There was recognition in the group that this is not a specific issue for FACCE-JPI to resolve (what FACCE-JPI funds is quite a way from the daily business of the farmer), but it could have a role in identifying research datasets to feed into the production line for such tools, and working with other initiatives, such as eROSA who are developing an e-infrastructure roadmap for open science in agriculture. This could be achieved through an inventory of data and data categories that are in the intersection of the three areas of Food Security, Agriculture and Climate Change (i.e. to have a role in cataloguing data, rather than generating or hosting datasets, and supporting an environment where stable data can be collected and standardised). **Small exploratory research projects funded by FACCE-JPI to specifically look at combining datasets in the three sectors of FACCE-JPI** would not cost much, but could have impact in contributing towards the production line of generating end-user tools.

6. Final Discussion and Recommendations

Many of the discussions throughout the workshop have identified areas where FACCE-JPI can play a role, but there is further work that is needed to work out precisely what that role will be within what is a very complex and interlinked landscape of big data.

There are some short-term, easily-implemented actions for FACCE-JPI to take. The first of these is for **FACCE-JPI to pro-actively be an advocate of FAIR principles,** and for these to become embedded in FACCE-JPI actions in the future. Further work can be done to **engage with recently finished and existing FACCE-JPI projects** to ensure that data generated in these projects is made as open as possible to maximise the wider use and impact of the data generated through FACCE-JPI funding.

Several other ideas from the workshop could be further developed by **setting up an Expert Working Group** made up of members of the Governing Board and data experts from the community and related initiatives. **This group could then work on a roadmap for FACCE-JPI to identify the actions to be taken in the medium and longer terms.** Building on the background paper, this group would take up the suggestions made in the workshop and develop them into either concrete actions that can be taken forwards by FACCE-JPI or identifying where FACCE-JPI needs to work with others to ensure that our specific remit is included in developments of other, broader data initiatives. This group would also be charged with **implementing a Data Sharing Policy for FACCE-JPI** that operationalises our need to be an advocate of FAIR data. This was considered to be an urgent and immediate action coming out of the workshop that should not wait until a FACCE roadmap for big data is finalised.

The FACCE-JPI roadmap for big data would need to a living document that is regularly reviewed in light of developments in the area, and outcomes of any pilot actions that are launched as early outputs of the roadmap. The research community should be consulted as the roadmap develops. This will help to identify gaps (e.g. in resource, computational capability, standards etc) that FACCE-JPI could then work with research infrastructures to address.

Recommendations:

In light of all of the discussions and suggestions from the workshop, the following recommendations are made to the FACCE-JPI Governing Board for consideration:

1. FACCE-JPI needs to **actively** be an advocate of the FAIR data principles. This can be achieved through:
   - Developing a data sharing policy and making data management plans for future FACCE-JPI calls mandatory (looking to other examples to see which have worked well)
   - Conducting an exercise with past/current FACCE projects to ensure/promote that data generated in the projects is FAIR
   - For specific actions, FACCE-JPI could consider promoting the use of initiatives such as FAIRDOM who can provide cohorts (at the joint action level) of researchers with tailored data management/analysis tools
   - Ensuring the FACCE-JPI research community is connected to the relevant research infrastructures to facilitate the FAIR principles

2. FACCE-JPI should **convene an expert working group** to develop a living roadmap looking at short- medium- and long-term priorities for data in its remit
   - To implement the above data sharing policy and further develop the above FAIR recommendations
   - Identifying how to operationalise interactions with key data initiatives and leverage existing resources for data sharing and appropriate centralisation of data for the benefit of FACCE-JPI
   - To map FACCE-specific data needs and availabilities; should gaps be found where specific facilities/infrastructure is lacking, work with existing infrastructures to address the need
   - Following an analysis of the suggestions in the workshop, to propose actions to be funded by FACCE-JPI (e.g. a Knowledge Hub/Network of Excellence for data in FACCE's remit; exploratory research projects to tackle specific issues of data integration)

3. When designing actions, FACCE-JPI (and applicants to its calls) should consider at the outset the potential translation of research data to usable tools, and **co-design with end-users and industry** to ensure such translation is achievable and to leverage pre-competitive data for mutual benefit.

7. Acknowledgements

FACCE-JPI wishes to thank all of the delegates of the workshop for their time and active engagement with the discussions in the workshop. Thanks also go to the speakers, moderators and rapporteurs, to the FACCE-JPI Working Group (with members of the Governing Board, Stakeholder Advisory Board, Scientific Advisory Board and Secretariat) who organised the workshop, and to the authors of the background paper. Special thanks to Aalborg University Copenhagen for hosting the workshop, and to Per Mogensen for local logistical support.

Annex 1: Agenda

FACCE-JPI Workshop on Big Data

22nd November 2017

Aalborg University Copenhagen, A. C. Meyers Vænge 15, 2450 Copenhagen, Denmark

AGENDA *(vers 15 November 2017)*

Chair: Gabriela Pastori, BBSRC, UK

08:30 – 09:00	***Registration - Coffee***
09:00 – 09:15	**Introduction**
09:00 – 09:15	**Welcome, Introduction to FACCE-JPI and aims of the workshop** *Niels Gøtke, FACCE-JPI Governing Board, Hartmut Stalb, FACCE-JPI Chair and Gabriela Pastori, BBSRC*
09:15 – 09:55	**Setting the scene**
09:15 – 09:35	**Plenary: Broad view of the overarching potential of Big Data** *Peter Baumann - Jacobs University Bremen, Germany; Co-Chair RDA Big Data Interest Group*
09:35 – 09:55	**Big Data in the FACCE context: overview of FACCE-JPI Background Paper on Big Data** *Mario Caccamo - NIAB-EMR, UK*
09:55 – 11:00	**Working with Others - I**
09:55 – 10:15	**European Open Science Cloud -** *Wim Haentjens, DG-Research and Innovation, European Commission*
10:15 – 11:00	**Short presentations from other initiatives/projects:** what they are doing, how can FACCE work with them, synergies • GODAN (Global Open Data for Agriculture and Nutrition) – *André Laperrière, GODAN Executive Director* • FAIRDOM – *Andrew Millar, University of Edinburgh, UK* • ENVRIplus – *Zhiming Zhao, University of Amsterdam, Netherlands*
11:00 – 11:30	***Coffee + Networking***
11:30 – 12:15	**Working with Others - II**
11:30 – 12:15	**Short presentations from other initiatives/projects:** what they are doing, how can FACCE work with them, synergies • Elixir – *Paul Kersey, European Bioinformatics Institute (EMBL-EBI), Cambridge, UK* • EIP-Agri (European Innovation Partnership for Agricultural Productivity and Sustainability) – *Willemine Brinkman, EIP-Agri Service Point* • CGIAR Platform for Big Data in Agriculture – *Andy Jarvis, International Center for Tropical Agriculture (CIAT)*
12:15 – 13:00	**Plenary Discussion – I**
12:15 – 13:00	**Quick wins for FACCE-JPI:** • Working with others – reflections on morning's presentations, identifying links between FACCE-JPI and others • Getting the most from FACCE-JPI research portfolio – implementing FAIR policies, promoting access to infrastructures, encouraging networking of FACCE projects' data managers… **Wrap up of morning** – capturing recommendations

13:00 – 14:00	***Lunch*** ***At 13:30, during the lunch break, there will be the option of participating in a short tour of one of two labs on the campus. Delegates can choose either:*** ***1. Smart Digital Devices for Food – Bent Mikkelsen, or*** ***2. Robot Lab – Lazaros Nalpantidis***
14:00 - 15:30	**Break-out groups: Specific data issues - I**
14:00 – 14:10 14:10 – 15:30	**Introduction to break-out groups:** instructions on outcomes needed from groups **Parallel break-out groups** tackling a key topic each (World Café format – delegates divided into three groups which rotate around the three topics): • **Group A:** Mapping data demands and availabilities (e.g. for models, new research); Bringing together data from different scales/disciplines • **Group B:** Academic demands for private-sector data • **Group C:** Making big-data relevant for end-users; translating data into tools
15:30 – 15:45	***Coffee***
15:45 – 16:00	**Break-out groups: Specific data issues - II**
15:45 – 16:00	**Reporting back from break-out groups (3 x 5') –** can continue into common discussion
16:00 - 16:50	**Common discussion**
16:00 – 16:50	Translating break-out discussions into agreed recommendations: • What can FACCE-JPI do in each of the three breakout topics? Prioritisation of issues • Who does FACCE-JPI need to work with to achieve results/success? • Are there other Big Data issues that haven't been covered in the workshop that FACCE-JPI is in a position to address? • What are the agreed recommendations to be made to the FACCE-JPI Governing Board o Ideas for joint working, research, networking…
16:50 – 17:00	**Wrap up + Conclusions**

Annex 2: Delegate list

Name	Affiliation	Country
André Laperrière	GODAN	United Kingdom
Andrew Millar	University of Edinburgh / FAIRDOM	United Kingdom
Andy Jarvis (via Skype)	CIAT / CGIAR Platform for Big Data in Agriculture	Colombia
Antonio Sánchez-Padial	INIA / FACCE-JPI	Spain
Bent Egberg Mikkelsen	Aalborg University / RICHFIELDS	Denmark
Bettina Heimann	EURAGRI	Denmark
Björn Usadel	Forschungszentrum Jülich / EMPHASIS	Germany
Bruce McCallum	Ministry of Business, Innovation and Employment	New Zealand
Davide Guariento	BioSense Institute / IoF2020	Serbia
Florence Macherez	Animal Task Force / FACCE-JPI StAB	France
Frank Ewert	ZALF / FACCE-JPI SAB	Germany
Gabriela Pastori	BBSRC / FACCE-JPI	United Kingdom
Gudrun Langthaler	Research Council of Norway / FACCE-JPI GB	Norway
Gunnar Lischeid	ZALF / FACCE-JPI	Germany
Hartmut Stalb	Federal Ministry of Food and Agriculture / FACCE GB Chair	Germany
Jonas Kathage	European Commission, Joint Research Centre	Germany
Jürgen Vangeyte	ILVO / ICT-Agri	Belgium
Karin Andeweg	Wageningen UR / Global Research Alliance	Netherlands
Knud Erik Skouby	Aalborg University, Copenhagen	Denmark
Manju Bura	BBSRC / FACCE-JPI Secretariat	United Kingdom
Manuel Lainez	INIA / FACCE-JPI GB	Spain
Marijn van der Velde	European Commission, Joint Research Centre	Netherlands
Mario Caccamo	NIAB-EMR / FACCE-JPI	United Kingdom
Maris Kruuse	Estonian Agricultural Research Centre	Estonia
Matti Pastell	Natural Resources Institute Finland (LUKE) / FACCE-JPI	Finland

Name	Affiliation	Country
Nicolas Gengler	ULg-GxABT / FABRE-TP	Belgium
Niels Gøtke	DASHE / FACCE-JPI GB	Denmark
Paul Kersey	EMBL-EBI / ELIXIR	United Kingdom
Paul Wiley	BBSRC / FACCE-JPI Secretariat	United Kingdom
Peer Berg	Norwegian University of Life Sciences / Animal Task Force	Norway
Per Mogensen	Danish Agency for Science and Higher Education / ICT-Agri	Denmark
Peter Baumann	Jacobs University / RDA, OGC, ISO, INSPIRE	Germany
Richard Lloyd	Innovation for Agriculture / 4D4F Data Driven Dairy Decisions for Farmers	United Kingdom
Sirli Pehme	Ministry of Rural Affairs	Estonia
Stefan Lampel	Project Management Jülich / FACCE-JPI GB	Germany
Stefanie Margraf	Project Management Jülich / FACCE-JPI Secretariat	Germany
Valeria Forlin	European Commission, DG-CLIMA	Belgium
Willemine Brinkman	EIP-Agri Service Point	Netherlands
Wim Haentjens	European Commission DG-RTD / FOOD 2030	Belgium
Yurus Emre Aydin	Bilimtek Teknoloji AS	Turkey
Zhiming Zhao	University of Amsterdam / ENVRIPLUS, VRE4EIC and SWITCH	Netherlands