Background paper for FACCE-JPI Big Data Workshop

Mario Caccamo[1], Florence Macherez[2], Jan Lassen[3], Gunnar Lischeid[4], Cristina Micheloni[5], Uwe Scholz[6], Paul Wiley[7]

[1] NIAB-EMR, UK; [2] Animal Task Force, Brussels; [3] Aarhus University, Denmark; [4] Leibniz Centre for Agricultural Landscape Research, Germany; [5] Associazione Italiana Agricoltura Biologica (AIAB), Italy; [6] IPK-Gatersleben, Germany; [7] FACCE-JPI Secretariat/BBSRC, UK

1. Introduction to FACCE-JPI

The Joint Programming Initiative on Agriculture, Food Security and Climate Change (FACCE-JPI) was launched in 2010, bringing together 22 member countries[1]. Its aim is to build the European Research Area tackling the challenges at the intersection of agriculture, food security and climate change that cannot be addressed solely at the national level. This is being realised through the alignment and integration of national and European research programmes, the funding of new research programmes, and through exploring innovative approaches for the member countries to work together to address the challenge of ensuring a secure food supply to an ever-increasing global population in the context of climate change.

In 2016, FACCE-JPI updated its Strategic Research Agenda, which outlines the research challenges in FACCE-JPI's remit across five Core Themes, and details impact-driven research priorities in each of these themes which will be the focus of FACCE-JPI's efforts over the coming years. In addition, a number of priorities were identified that cut across the five Core Themes, including one concerning big data in agriculture:

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)

This priority was identified by the FACCE-JPI Governing Board for inclusion in the 2016-2018 Implementation Plan as a subject for an exploratory workshop to identify the key big data issues in FACCE-JPI's remit (the intersection of agriculture, food security and climate change) and to explore what role FACCE-JPI can take in realising the potential of big data both through its existing and future research projects (e.g. maximising the use of data generated; providing guidelines on data management) and through taking action to address some of the challenges raised in this paper. The paper serves to set the landscape and identify the key questions that could be considered at the workshop.

2. Current landscape of "Big Data" in Agriculture, Food Security and Climate Change

In general, "big data" is quite an issue in various sectors, but it has not received the same level of attention within agriculture as some other areas in life sciences (e.g. genomics). On the other hand, increasingly more data are now available in the agricultural sector that could be used in a much more systematic way then has been done hitherto. However, different than for other sectors, there are many different data holders like remote sensing data, digital soil maps, data of various authorities in the agricultural and environmental field, farmers' data (for example, sensors on modern tractors and harvesters) etc. that hardly communicate with each other. These examples highlight two key aspects of the challenges underlying "big data" solutions: one is the volume of

---

[1] Austria, Belgium, Czech Republic, Cyprus, Denmark, Estonia, Finland, France, Germany, Ireland, Italy, Latvia, The Netherlands, Norway, Poland, Romania, Spain, Sweden, Switzerland, Turkey and UK. New Zealand joined as an Associate Member in 2016.

the raw data generated by these devices, and the second one is the diversity of the data sources. On one hand, the vast volumes of terabytes require the use of effective and efficient algorithms to translate them into high-quality information, and on the other hand, we also need to focus on integration strategies to combine the information generated by a myriad of different sources. This workshop aims at bringing together these data sources and to pave a way towards a more systematic synergistic use of these data, using modern big data approaches.

Example 1: Plant Phenotyping

In the last years, high-throughput phenotyping has become state-of-the-art in plant research. Examples are automated greenhouse systems for the screening of hundreds of single plants under controlled conditions. Such systems allow the non-destructive screening of plants over a period of time by means of image acquisition techniques, producing Gigabytes of data. In most cases, these are in the form of hundreds of thousands of images files. Analysis tools extract the features and assign values of various traits to each single plant in every development stage. Such datasets require detailed documentation, standardised description of experimental metadata as well as sustainable data storage and publication in order to ensure the reproducibility of experiments, data reuse and comparability among the scientific community. In the field of plant phenotyping, standards to describe the experiments and formats data containers are in development by the community (e.g. www.miappe.org).

By looking separately into the different subject areas of FACCE-JPI, similar activities regarding the Big Data challenges can be identified. In all fields of application, the use of ontologies and controlled vocabularies to describe metadata is state of the art. Standardised data formats are used or are under development. These formats are applied to stored and published data sets. The best choices to publish the structured data are international repositories. In case of less structured data sets the assignment of Digital Object Identifiers (DOIs) is one of the standard procedures. Often, systems to publish DOI data sets are operated institutionally. The current situation shows that a lot of activities are running but they are mostly not aligned between FACCE-JPI's constituent areas. The compilation of used ontologies, controlled vocabularies, running repositories, provided access interfaces and data publication policies is the first step to overcome this issue. As the next task, the harmonisation of data formats and access interfaces should be started. Finally, the definition of possibilities to integrate data sets from different applications should be evaluated, in order to enable a domain-wide use of information.

Changes in climate patterns have the potential to affect agriculture in a number of ways. Indeed, intensive farming and climate change are closely interlinked contributing directly to global processes that are intimately interrelated. In recent years, we have increased our knowledge on how these processes relate and interfere with each other but much remains to be done to establish how a sustainable balance can be achieved. As an initial step, a better understanding on how large-scale climate impact studies can be used to input to growers' and agronomists' decision making will help to exploit new opportunities and advance sustainable intensification in agriculture under a changing climate.

Example 2: Field trials

The ability to validate crops in field trials under agronomy regimes and in a diversity of environments is a key step towards understanding the performance of different varieties. Field trials also play an important role in the monitoring of pathogens and early identification of potential threats to yields. As with clinical trials, it is also very important to be able to record negative results and the underlying metadata that could help supporting other studies and experiments potentially years after the trials have been conducted. An increasing number of field trials are being implemented with public-sector funding with requirements for the data to be collected using adopted standards and made available to the scientific community. As an example, we can consider grant-funded projects that include the implementation of field trials as part of the research project. Typically, a medium-scale field trial will collect data over a period of several months for tens of parameters amounting for several gigabytes of digital data. There is no available repository to collect and distribute this kind of data. Most of the publicly available field trials data is maintained in local databases by the same people that generate them, thus failing one of the key principles of FAIR data (see Section 2.2). Traditionally data collection for field trials has been both time consuming and labour intensive. With the use of UAVs and other high-throughput devices we can now collect vast amounts of measurements in only few hours. Therefore, another important factor to consider is the speed at which data can be gathered, as this imposes another strain in data management approaches. Field trials are also playing a key role in supporting experiments related to climate change with several projects currently setting up long-term studies in a diversity of locations covering a number of environments with the purpose of monitoring and assessing the effects of changes in weather patterns on plants, animals and resources.

Example 3: Livestock

Several tools have been developed and implemented for use in daily management on farms. On farms with many animals, the amount of time available for each animal gets more and more limited and therefore objective tools can be used to predict the biological state of an animal. This could include activity data, positioning tools, on line measurement of milk but also sequence data, and imaging tools that create vast amount of data when recording continuously. Also, the development of drones and sensors can be used to optimise management. These data can be used to make decisions on health, reproduction, slaughter and production efficiency. As an example, currently many projects are studying the usefulness of using sequence data from samples from the gastrointestinal tract (GIT). These projects are performed in several species such as dairy, beef, pig and poultry. Outcomes of such studies might be better grouping of animals according to their GIT composition and/or probiotics, or defaunation of specific GIT microbes to improve a given phenotype. These studies could be of great importance as a mitigation strategy for methane from ruminants as this is produced by microbes in the rumen. Also various infrared spectral data of milk, feed and faeces are being used to predict a number of phenotypes.

2.1. Technologies are developed bottom-up

As with other areas in life sciences, the recent advent of inexpensive high-throughput technologies promises to revolutionise agriculture. It is now possible to measure almost anything imaginable and in only few hours we can collect vast amounts of data using all kinds of different approaches. We can fly drones over large crop fields loaded with a multitude of cameras, install sophisticated sensors to monitor livestock behaviour, access publicly available satellite images and monitor current crop prices; all from the comfort of a desk in an office that can and probably will be remotely located. The challenge, however, remains: how can we make sense of the data? Realising the value of the data collected and how this can be translated into applicable knowledge is at the heart of the Big Data revolution.

Most of the devices that are now available to growers, agronomists and scientists have been developed in a bottom-up fashion focused on specific issues first and only looking into the data they generate after the initial prototypes are already available to the users. In many ways, data in new technologies comes as an afterthought. The added value of the data, however, can only be realised if it is available in a format and following release policies that not only facilitate but also encourage the integration with other datasets. Although this point is probably valid in other areas of science, it is particularly important for the needs of modern agriculture which combines data from so many disciplines and is extremely dynamic in nature.

2.2. FAIR guiding principles for data management

Undoubtedly there is a sense of urgency around improving the infrastructure underlying data management in agriculture. This impacts a diverse set of stakeholders ranging from scientists to farmers. A number of guiding concepts known as the FAIR data principles have been proposed by the Big Data scientific community to establish approaches for data stewardship that maximise the opportunity for integration and added value. FAIR data refers to making information resources

- Findable
- Available
- Interoperable
- Reusable

The data should be findable not just by humans but also indexed to be searchable by machines which in the times of the Internet of Things (IoT) is particularly relevant. By the same token, the data should be available in repositories that ideally are governed by open-access policies or otherwise have well-defined licences. It is, however, all about making the data interoperable to be combined with other datasets; specifically, this refers to the availability of robust and well-documented application programming interfaces (APIs) and the use of ontologies. As a conclusion from the above, the data should be reusable and ready to be integrated with other datasets. The availability of up to date and sound metadata describing datasets is a key step towards delivering FAIR data.

2.3. Trust and adoption

As discussed above, to generate value out of agriculture data, they often need to be aggregated and combined from different sources. We not only rely on information generated by tractors, drones, sensors and other farm devices but also need input from external provides such as meteorological data, financial advisers and chemical and feed compounds databases. Equally, growers and producers are required to provide information about their operations to comply with regulations such as relating to the use of fertilisers and pesticides, or required for applying for subsidies. Much of this information is often perceived as sensitive or providing a competitive advantage, resulting in barriers for data sharing and the adoption of open standards.

The inability to clearly defined data provenance and provide mechanisms for crediting the generators of data has also become a hurdle for data sharing. Data sharing agreements based in the Toronto Agreement (doi:10.1038/461168a) could serve as a guide for similar approaches in agriculture. Data sharing with industry can be incentivised focusing on the pre-competitive level, such as the approach taken by the pharmaceutical companies when they set up the Pistoia Alliance (http://www.pistoiaalliance.org).

3. Viewpoints from different stakeholders

3.1. Policy/funder perspective

Data generated through publically funded research should be made available to other researchers and practitioners to exploit and to maximise the impact from the initial research investment. The challenge for funders is how to actively encourage and facilitate the sharing and reuse of data generated through research it has funded for further scientific gains and input to policy. There is a balance needed between bottom-up, community-led approaches (research communities developing and adopting practices and standards specific to their research area) and top-down, funder-led approaches (having policies in place to encourage/mandate open data practices).

At the national level, an example of where a funder has put in place measures to encourage grant holders to manage and share data is the requirement by BBSRC (the UK's main public funder for biological science research) for all applicants to submit a Data Management Plan alongside grant submissions. For multi-national research, such as that funded through FACCE-JPI, there is the additional challenge of multiple national funders having different data sharing guidance. An example of where this has been addressed is the ERA-CAPS network of funders who put in place a common data sharing policy that was agreed by all funders involved in ERA-CAPS calls. The policy provided broad, open guidelines on data management and sharing, and made submitting a Data Management Plan compulsory for applicants. Equally, for Horizon 2020, there are established guidelines on Data Management for projects funded through the Framework Programme, including a template for a Data Management Plan[2].

Questions for consideration:

- How can FACCE-JPI encourage FAIR principles and maximise data reuse for the research it funds?

---

[2] http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

- Does FACCE-JPI need to implement its own data sharing policy for research it funds (including requiring data management plans for research applications)? Can it adopt existing policies developed for multi-funder research (e.g. ERA-CAPS Data Sharing Policy or Guidelines on FAIR Data Management in Horizon 2020)?
- Are there untapped sources of data that could improve accuracy of models for climate change's impact on agriculture and food security? Accurate models are needed to inform policy at governmental and intergovernmental levels.
- Can publically-funded researchers gain access/make use of privately-generated agricultural data to further the impact of their research?
- Many of the challenges around data sharing are common to all areas of science. However, are there data issues specific to the remit of FACCE-JPI where action can be taken?

3.2. Research perspective

Opportunities and challenges of big data for publically funded research concern agricultural sciences as well as environmental sciences, including biology and geosciences.

So far, crop production research was largely based on lab experiments and field trials. In both cases, only a limited number of influencing factors including site conditions could be considered, and transferability of the results was rather limited. However, making full use of data that are now, in principle at least, available, could in fact turn any arable field or grassland into a field trial. High spatial resolution of the sensor systems could then use spatial heterogeneities and heterogeneities of management as a source of information rather than as a restriction. Correspondingly, transponders, automatic control of forage intake and of milk quality etc. could provide much more information than classical trials with more or less homogeneous cohorts.

At the global scale, about one third of the land mass is currently intensively used by agriculture and forestry, and another third is used extensively. Thus the largest fraction of land is directly affected by human activities with regard to nutrients and pesticides application, greenhouse gas emissions, soil erosion, effects on biodiversity, etc. Thus agriculture is a major issue for natural resources management and environmental sciences.

Some examples are given here. Tractor-mounted N sensors, or harvester-mounted sensors to assess yield provide data at high spatial resolution. GPS data of tractors during soil tillage, combined with data on fuel consumption can be used to assess patterns of wheelspin and thus of soil properties which can then be merged with data of nutrient status and yield. Other sensors can be used to analyse the microbiome of soils and plants, and to investigate the network of plants, bacteria and fungi in the soil. Also, sensors can be used to identify hotspots of plant diseases or weed abundance within single fields. For example, due to increasing herbicide resistance of certain weed species and public concerns over herbicide impacts, new approaches are developed for mechanical weed control based on automatic recognition of single plants. It is expected that in near future rather than a single heavy tractor a multitude of small autonomic robots will perform that task. Those systems could explicitly consider sensitive areas, like nests of breeding birds, or conservation of single endangered weed species that do not harm the crops. Thus a much higher degree of spatial heterogeneity and sensitivity can be accounted for the benefit of both agricultural production and environmental protection.

On the other hand, even large-scale geophysical data could be used. For example, remote sensing gravity data have proven the potential to quantify groundwater extraction for agricultural irrigation, although still at a rather large spatial scale. However, multi-temporal high-resolution laser scanning of the earth surface already today can be used to assess soil erosion or land subsidence due to groundwater extraction at a much smaller scale. Combining thermal and multi-spectrum remote sensing data can be used to assess spatial patterns of actual evapotranspiration even within single fields. Thus not only the need for irrigation can be assessed; the data can equally well be used to assess the crop health status at high spatial resolution, enabling much more focused pest control.

In animal production, to meet policy goals for more efficient use of resources combined with emphasis on better animal health and welfare, practical options should be sought for combining genetic, genomic, metabolomic and phenotypic information to gain a better understanding of biological processes and to improve selection decisions for livestock by exploiting technological gains in these areas. The goal is to provide a toolbox of deliverable products and/or processes to allow these potential advantages to be realised.

There are challenges to make Precision Livestock Farming (PLF) a reliable and validated source of information (from data collected from PLF devices). New techniques for deeper phenotyping, including metabolic profiling, sensor technology, remote sensing, PLF on the one hand, and the genomic revolution (“omics tools and data”) on the other hand, are key elements for successful implementation in the breeding sector. However, these technological developments are so far evolving in two separate worlds. Breaking this fragmentation by bridging knowledge and relevant actors will renew predictive biology approaches, generate new knowledge and is crucial for simultaneously advancing smarter farming and competitive breeding. A main goal is to develop and share innovative pipelines dedicated to high-throughput delivery of big data generation and analysis. Computer databases and data management and analysis facilities are necessary tools for handling the huge amount of data relevant to livestock and for simplifying the localisation, extraction and analyses of relevant information. Such research databases will be shared through common projects and will include original data (genomic data, precision livestock farming data) and classical performance data (dairy production, body weights, health proxies, real health parameters, etc.). When research can demonstrate the added value of such data integration, this will be a powerful impetus to encourage data accessibility.

Exchange and synergistic use of existing data, and new advanced data mining approaches to make more efficient use of large data sets will likely boost our understanding of complex, non-linear relationships and feedback loops between human activity and natural biotic and abiotic processes. This would result in mutual benefit both for agricultural and environmental sciences, and would in the end help to make agricultural management both much more efficient and environmentally friendly. However, that would require not only the ability to merge data from different sources such as satellite and UAV remote sensing, agricultural and environmental authorities and farmers (with widely differing data formats, sources of uncertainties and errors, and monitoring protocols), it would also require sound regulations for data governance, such as access rights, prevention of data misuse, etc.

Questions for consideration:

- How to bring together data and models at different scales? As an example - the alignment of regional or global impact climate models with local crop yield model, or the impact of molecular techniques in animal and crop breeding. How can we foster exchange of data and concepts from different scientific disciplines, e.g., agriculture, environmental sciences, geophysics, etc.?
- How to foster exchange of data and concepts from different scientific disciplines, e.g., agriculture, environmental sciences, geophysics, etc.?
- How can we develop approaches and recommendations for dealing with highly heterogeneous data?
- How can we check data quality using advanced technologies?

3.3. Training/education perspective

As described in many of the examples above, decision-making in agriculture is increasingly becoming more reliant of skills related to data analytics and computational approaches to data manipulation. Nevertheless, still these skills are rarely included in the curriculum of relevant university undergraduate courses resulting in a skills gap that compromises the exploitation of the most advanced technologies. Although this shortage in skills has traditionally impacted scientists and applied researchers as technologies become available, the lack of training is increasingly perceived as one of the main barriers to agronomists and growers for the adoption of new technologies. Thus it is imperative to address this widening gap in skills by providing short-term solutions in the form of training to current professionals as well as investing in the development of a modern curriculum for university courses to support a skills market that is better prepared for a future in agriculture driven by data.

Questions for consideration:

- Can we identify and address current skill gaps in data analytics in agriculture?
- How can we influence the development of curricula for university undergraduate courses (and perhaps earlier) that tackles areas of data analytics?
- Can we prioritise the education of the general public in the new technologies?

3.4. Service provider perspective (commercial and academic)

Data-driven added-value services in the agricultural sectors is emerging as an opportunity for the development of businesses and spin offs. Many of these opportunities are focused on specific areas of expertise such as image capturing services, crop modelling, irrigation scheduling, etc. In most of these examples the focus is on decision-making support and best-practice advice. One of the major challenges for service providers is the access to affordable high-quality reliable data. This point is particularly relevant to the costing of the services and the impact this has on the adoption of novel approaches by the end users (i.e. agronomists and farmers). The development and maturation of new technologies greatly depends on a dynamic and resilient service sector. In many examples the research community will be the early customers of this sector which in turn is driven by access to public funding in the form of research grants. This highlights the importance of the continuous funding of research and the impact of the development of services that will eventually become available to a wider user base.

Questions for consideration:

- How can the public sector prioritise funding in areas of research relevant to the development of new data-driven technologies? (pump-priming grants, proof-of-concepts, etc)
- How can we promote an integration with industry that will also promote a commitment of resources from the private sector?
- How can we make more relevant the results of large-scale climate-change impact studies to the day-to-day decision-making process of growers and agronomists?

3.5. End user perspective

End users in relation to big data form a very diverse set. Two that are considered here are companies (developing equipment and tools for use on farms) and the farmers themselves who could make use of data-driven tools to optimise on-farm solutions.

Communication between and within layers of industry, from data generators to end users, is extremely important and necessary to maintain a focus on big data in order to keep industry involved through funding and interest in future developments. Industry can be seen in several layers: on the one hand, companies (in collaboration with the scientific community) need to develop new equipment. On the other hand, farmers need easy-to-interpret daily management tools which are consistent, reliable and which are a good investment. Farmers don't need background information on how the data are collected and analysed, but rather tools that provide solutions which can be used in their everyday routine.

Also, in order to have an end user perspective, it is necessary that there is cooperation and interdisciplinary projects so that improvement in one area doesn't lead to disadvantages in other areas – e.g. if reduced methane production directly from ruminants leads to less digested feed which, in turn, leads to higher production of nitrous oxide in soil, nothing is achieved and the end user has invested without any gain.

Questions for consideration:

- How to ensure that available data is translated into really useful and affordable tools to be used on a day-to-day basis for farmers.
- How to ensure that farmers are able/willing to provide data generated on the farm (e.g. considering legal issues, competitive advantage etc.)
- How to integrate independent datasets across different scales (especially climate data) e.g. to really inform what crops are grown at farm level.
- How to ensure that data-driven tools that seek to optimise farm management in the context of climate change do not have contrary trade-off effects elsewhere.

4. Objectives of the workshop

In light of the background covered in this paper, the Working Group of the FACCE-JPI workshop has identified the following objectives for the workshop:

- The purpose of the workshop is:
    - To explore research needs and research gaps
    - To identify potential application and integration (reassembling) of relevant new and existing data
    - 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
- The workshop aims to:
    - To bring together relevant experts and stakeholders
    - 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

4.1. Potential discussion points for workshop

In order to achieve the objectives above, several questions and potential discussions points have been identified in this paper that could be taken forward in the workshop, however, this will need to be considered and prioritised by the FACCE-JPI bodies (the workshop Working Group, the Scientific and Stakeholder Advisory Boards, and the Governing Board) ahead of the workshop to ensure that the workshop outcomes are tailored to the needs of the JPI. These questions are summarised below:

From the introduction:

- DOIs are typically used for public scientific data - can private/industrial data be identified and accessed in an equivalent manner?

From the policy section:

- How can FACCE-JPI encourage FAIR principles and maximise data reuse for the research it funds?
- Does FACCE-JPI need to implement its own data sharing policy for research it funds (including requiring data management plans for research applications)? Can it adopt existing policies developed for multi-funder research (e.g. ERA-CAPS Data Sharing Policy or Guidelines on FAIR Data Management in Horizon 2020)?
- Are there untapped sources of data that could improve accuracy of models for climate change's impact on agriculture and food security? Accurate models are needed to inform policy at governmental and intergovernmental levels.
- Can publically-funded researchers gain access/make use of privately-generated agricultural data to further the impact of their research?
- Many of the challenges around data sharing are common to all areas of science. However, are there data issues specific to the remit of FACCE-JPI where action can be taken?

From the research section:

- How to bring together data and models at different scales? As an example - the alignment of regional or global impact climate models with local crop yield model, or the impact of molecular techniques in animal and crop breeding. How can we foster exchange of data and concepts from different scientific disciplines, e.g., agriculture, environmental sciences, geophysics, etc.?
- How to foster exchange of data and concepts from different scientific disciplines, e.g., agriculture, environmental sciences, geophysics, etc.?
- How can we develop approaches and recommendations for dealing with highly heterogeneous data?
- How can we check data quality using advanced technologies?

From the training/education section:

- Can we identify and address current skill gaps in data analytics in agriculture?
- How can we influence the development of curricula for university undergraduate courses (and perhaps earlier) that tackles areas of data analytics?
- Can we prioritise the education of the general public in the new technologies?

From service providers' perspective:

- How can the public sector prioritise funding in areas of research relevant to the development of new data-driven technologies? (pump-priming grants, proof-of-concept, etc)
- How can we promote an integration with industry that will also promote a commitment of resources from the private sector?
- How can we make more relevant the results of large-scale climate-change impact studies to the day-to-day decision-making process of growers and agronomists?

From the end users' perspective:

- How to ensure that available data is translated into really useful and affordable tools to be used on a day-to-day basis for farmers.
- How to ensure that farmers are able/willing to provide data generated on the farm (e.g. considering legal issues, competitive advantage etc.)
- How to integrate independent datasets across different scales (especially climate data) e.g. to really inform what crops are grown at farm level.
- How to ensure that data-driven tools that seek to optimise farm management in the context of climate change do not have contrary trade-off effects elsewhere.

Annex 1: Non-exhaustive list of relevant initiatives that FACCE-JPI could work with or should be aware of:

Agrimetrics is the UK big data agritech centre (http://www.agrimetrics.co.uk) founded as a partnership between four founding members: University of Reading, NIAB, SRUC and Rothamsted Research. Agrimetrics is one of the four new centres for agricultural innovation funded by the UK government following the strategy for agricultural technologies published in July 2013. Agrimetrics is specifically focused on data-driven technologies and the development of expertise and capabilities in analytics, bioinformatics, translational research and knowledge exchange in crops, livestock and food, and sustainability.

BrAPI - The Breeding API specifies a standard interface for plant phenotype/genotype databases to serve their data to crop breeding applications. It is a shared, open API, to be used by all data providers and data consumers who wish to participate. http://docs.brapi.apiary.io

CGIAR Platform for Big Data in Agriculture - The goal the platform is to harness the capabilities of big data to accelerate and enhance the impact of international agricultural research. This 6-year platform (2017-2022) will provide global leadership in organizing open data, convening partners to develop innovative ideas, and demonstrating the power of big data analytics through inspiring projects. http://blog.ciat.cgiar.org/cgiar-platform-for-big-data-in-agriculture/

Data Driven Dairy Decision For Farmers (4D4F) aims at developing a network for dairy farmers, dairy technology suppliers, data companies, dairy advisors, veterinarians and researchers to improve the decision making on dairy farms based on data generated by sensors**.** http://www.4d4f.eu/

EIP-AGRI – The European Innovation Partnership for Agricultural Productivity and Sustainability (EIP-AGRI) works to foster competitive and sustainable farming and forestry that 'achieves more and better from less'. It has held a number of workshops and semimars addressing big data and data sharing https://ec.europa.eu/eip/agriculture/en

ELIXIR - ELIXIR is an intergovernmental organisation that brings together life science resources from across Europe. These resources include databases, software tools, training materials, cloud storage and supercomputers. The goal of ELIXIR is to coordinate these resources so that they form a single infrastructure. This infrastructure makes it easier for scientists to find and share data, exchange expertise, and agree on best practices. https://www.elixir-europe.org/

EMPHASIS - ESFRI has identified “Plant Phenotyping” as a priority for the European research area and EMPHASIS has been listed on the ESFRI Roadmap as an infrastructure project to develop and implement a pan-European plant phenotyping infrastructure. Within EMPHASIS European partners from academia and industry are working together. Existing national and European phenotyping networks are integrated in this program. EMPHASIS activities are structured in five categories of phenotyping infrastructure: controlled conditions, intensive field, lean field, modelling and e-infrastructure.
http://emphasis.plant-phenotyping.eu/

ENVRIplus - ENVRIplus is a Horizon 2020 project bringing together Environmental and Earth System Research Infrastructures, projects and networks together with technical specialist partners to create a more coherent, interdisciplinary and interoperable cluster of Environmental Research Infrastructures across Europe. http://www.envriplus.eu/

ERA-CAPS is the ERA-Net for Coordinating Action in Plant Sciences. The ERA-CAPS network has a Data Sharing Policy for the multinational research it funds. This could be a model for FACCE-JPI funded research. http://www.eracaps.org/sites/default/files/content/joint-calls/3rd-call/ERA-CAPS%20Data%20Sharing%20Policy_2016.pdf

EU-PLF - Precision Livestock Farming (PLF) could assist livestock producers through automated, continuous monitoring of the animals. The observation data can be translated into key indicators on animal welfare, animal health, productivity and environmental impact. A number of PLF tools have been developed at laboratory levels and as prototypes. The project has allowed the development of new business models while linking high tech SME's to European industry players to create new PLF-products with global impact. http://www.eu-plf.eu/

FAIRDOM –FAIRDOM's aim is to establish a data and model management service facility for Systems Biology. Its prime mission is to support researchers, students, trainers, funders and publishers by enabling Systems Biology projects to make their Data, Operating procedures and Models, Findable, Accessible, Interoperable and Reusable https://fair-dom.org/

Internet of Food and Farm 2020: Large-scale project granted €30 million from Horizon 2020 to foster a large-scale take-up of Internet of Things (IoT) technologies in the European farming and food value chain in the next 4 years. It will consolidate Europe's leading position in the global IoT industry by fostering a symbiotic ecosystem of farmers, food industry, technology providers and research institutes. The project brings together 73 partners from 16 countries under the coordination of Wageningen University & Research. https://iof2020.eu/iof/iof2020

RDA-IGAD - The Research Data Alliance (RDA) Interest group on Agricultural Data (IGAD) is a discipline-specific group that promotes good practices in the research domain, including data sharing policies, data management plans, and data interoperability and provides a forum for sharing experiences and providing visibility to research and work in agricultural data. Furthermore IGAD has served as a platform and and is structured in focused domain-specific Working Groups. https://www.rd-alliance.org/rda-disciplines/rda-and-agriculture

Science Europe is currently developing standard protocols for Data Management Plans that can be used by any funder for any type of research (with some tailoring to specific research communities) - once finished this could be a resource used by FACCE-JPI for research calls. http://www.scienceeurope.org/policy/working-groups/research-data/

The **Wheat Initiative** (http://www.wheatinitiative.org) was launched in 2011 following the endorsement from the G20 Agriculture Ministries with the objective to provide a framework to establish and develop strategic research in wheat. The initiative promotes the creation of Expert Working Groups in areas of priority. One of these groups is focused on the development of a Wheat Information System (WheatIS - http://wheatis.org) which has prioritised areas of big data. The main objective of the WheatIS is to provide a single-access web-based system to data resources and bioinformatics tools for the wheat community.

Annex 2: FACCE-JPI Big Data Workshop working group and acknowledgements

The FACCE-JPI Big Data Workshop was organised by a working group comprising:

Governing Board	Niels Gøtke	Danish Agency for Science and Higher Education
Scientific Advisory Board	Frank Ewert	Leibniz Centre for Agricultural Landscape Research (ZALF)
Stakeholder Advisory Board	Florence Macherez	Animal Task Force
Stakeholder Advisory Board	Bram Moeskops	TP Organics
Secretariat Lead	Paul Wiley	BBSRC
Secretariat Support	Stefanie Margraf	Project Management Jülich
Secretariat Support	Manju Bura	BBSRC
Local Logistics Support	Per Mogensen	Danish Agency for Science and Higher Education

The Working Group would like to express their warm gratitude to the authors of this background paper, which has been invaluable in setting the agenda for the workshop, and to Aalborg University for providing meeting facilities for the workshop on their Copenhagen campus.