



# D1.4 DATA

# MANAGEMENT PLAN

## WP 1

21 Aug. 2017

Document describing how data will be managed in the IoF2020 Large-Scale Pilot, including guidelines for data management in IoT in Agriculture



IoF2020 has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 731884. Visit [iof2020.eu](http://iof2020.eu) for more information about the project.

## DOCUMENT IDENTIFICATION

<b>Project Acronym</b>	IoF2020
<b>Project Full Title</b>	Internet of Food and Farm 2020
<b>Project Number</b>	731884
<b>Starting Date</b>	January 1st, 2017
<b>Duration</b>	4 years
<b>H2020 Call ID &amp; Topic</b>	IOT-01-2016
<b>Date of the DoA</b>	2017-2021
<b>Website</b>	<a href="http://www.iof2020.eu">www.iof2020.eu</a>
<b>File Name</b>	D1.4 Data Management Plan 1.0.doc
<b>Date</b>	21 August 2017
<b>Version</b>	1.0
<b>Status</b>	Final
<b>Dissemination level</b>	PU: Public
<b>Author</b>	Sjaak Wolfert
<b>Contact details of the coordinator</b>	George Beers george.beers@wur.nl



## PROJECT SUMMARY

**The internet of things (IoT) has a revolutionary potential. A smart web of sensors, actuators, cameras, robots, drones and other connected devices allows for an unprecedented level of control and automated decision-making. The project Internet of Food & Farm 2020 (IoF2020) explores the potential of IoT-technologies for the European food and farming industry.**

The goal is ambitious: to make precision farming a reality and to take a vital step towards a more sustainable food value chain. With the help of IoT technologies higher yields and better-quality produce are within reach. Pesticide and fertilizer use will drop and overall efficiency is optimized. IoT technologies also enable better traceability of food, leading to increased food safety.

Nineteen use-cases organised around five trials (arable, dairy, fruits, meat and vegetables) develop, test and demonstrate IoT technologies in an operational farm environment all over Europe, with the first results expected in the first quarter of 2018.

IoF2020 uses a lean multi-actor approach focusing on user acceptability, stakeholder engagement and the development of sustainable business models. IoF2020 aims to increase the economic viability and market share of developed technologies, while bringing end-users' and farmers' adoption of these technological solutions to the next stage. The aim of IoF2020 is to build a lasting innovation ecosystem that fosters the uptake of IoT technologies. Therefore, key stakeholders along the food value chain are involved in IoF2020, together with technology service providers, software companies and academic research institutions.

Led by the Wageningen University and Research (WUR), the 70+ members consortium includes partners from agriculture and ICT sectors, and uses open source technology provided by other initiatives (e.g. FIWARE). IoF2020 is part of Horizon2020 Industrial Leadership and is supported by the European Commission with a budget of €30 million.

## EXECUTIVE SUMMARY

Data or Big Data rapidly has become a new resource or asset in the current economy, also in the agricultural sector. This development leads to several issues that have to be addressed such as data availability, quality, access, security, responsibility, liability, ownership, privacy, costs and business models. In Europe several initiatives have already been established to work on this in a general context (e.g. the EU open data policy) and more specifically in agriculture (e.g. GODAN, COPA-COGECA).

IoF2020 has to address these issues by developing a Data Management Plan (DMP). This document (D1.4) provides a first version of a DMP containing a general overview of relevant developments and a first inventory of needs and issues that play a role in the use cases. This has resulted in a number of general guidelines for data management in IoF2020:

- Research papers that are derived from the project should be published according to the open access policy
- Research data should be stored in a central repository according to the FAIR principles: findable, accessible, interoperable and reliable
- The use cases in IoF2020 should be clearly aligned to the European Data Economy policy and more specifically in line with the principles and guidelines provided by the stakeholder community *i.e.* COPA-COGECA.

A number of follow-up actions were identified to establish open, transparent data management in IoF2020:

- Investigate what research data is involved in the use cases and other project activities and define how they should be treated according to the open data policy.
- Participate actively in the debate and developments of the European Data Economy and data sharing in agriculture.
- Analyze and explore the use cases in a deeper way in order to identify which data management issues potentially play a role and define plans how to deal with them.
- Concerning collaboration with other projects, a more systematic and structural approach should be explored in order to maximize the benefits and impact of the mutual activities on data management.

As a result the DMP will be continuously adapted and updated during the project's period.

# TABLE OF CONTENTS

<b>EXECUTIVE SUMMARY</b>	<b>4</b>
<b>ABBREVIATIONS</b>	<b>7</b>
<b>1. INTRODUCTION</b>	<b>8</b>
<hr/>	
1.1. CONTEXT AND BACKGROUND	8
1.2. OBJECTIVE OF THIS DOCUMENT	10
1.3. OUTLINE	10
<b>2. APPROACH</b>	<b>11</b>
<b>3. EXTERNAL DEVELOPMENTS</b>	<b>12</b>
<hr/>	
3.1. DATA MANAGEMENT DEVELOPMENTS	12
H2020 Open Access Policy	12
Open Research Data pilot	13
European Data Economy	17
Agricultural Data developments	19
3.2. CONSEQUENCES FOR IOF2020	24
<b>4. RELEVANT INITIATIVES AND PROJECTS</b>	<b>25</b>
<hr/>	
4.1. PROJECTS AND INITIATIVES	25
GODAN	25
EIP-Agri	27
AgInfra+	28
BigDataEurope	31



DATABIO	33
4.2. RELATIONSHIP WITH IOF2020	35
<b>5. INVENTORY OF USE CASES</b>	<b>35</b>
<b>6. DATA MANAGEMENT PLAN</b>	<b>43</b>
<hr/>	
6.1. GENERAL GUIDELINES FOR DATA MANAGEMENT IN IOF2020	43
6.2. FURTHER ACTIONS AND RECOMMENDATIONS	43
<b>7. CONCLUSIONS</b>	<b>44</b>
<hr/>	



## ABBREVIATIONS

CIARD	Coherence in Information for Agricultural Research for Development
DMP	Data Management Plan
EIP	European Innovation Partnership
GODAN	Global Open Data for Agriculture and Nutrition
H2020	Horizon 2020 the 8 <sup>th</sup> Research Framework Programme of the European Union
I(C)T	Information (and Communication Technology
IPR	Intellectual Property Rights
OA	Open Access
OADA	Open Agriculture Data Alliance
ODI	Open Data Institute
ORD	Open Research Data
UC	Use Case

# 1. INTRODUCTION

## 1.1. CONTEXT AND BACKGROUND

Big Data is becoming a new resource, a new asset, also in the agricultural sector. Big Data in the agricultural sector includes enterprise data from operational systems, farm field sensor data (e.g. temperature, rainfall, sunlight), farm equipment sensor data (from tractors, harvesters, milking robots, feeding robots), data from wearable animal sensors (neck tag, leg tag), harvested goods and livestock delivery vehicles sensor data (from farms to processing facilities) etc. Increasingly, Big Data applications, Big Data initiatives and Big Data projects are implemented and carried out, aiming for improving the farm and chain performance (e.g., profitability and sustainability) and support associated farm management decision making.

In the IoF2020 different use cases are taking place in which data also plays a key role involving farm companies that share their (big) farm data with enterprises and organisations who strive to add value to that data. This implies that the data from one party is combined with data from other parties in the chain, and then analysed and translated into advices, knowledge, or information for farmers. In this way Big Data becomes an asset in supporting farmers to further improve their business performance (e.g., higher yield, better quality, higher efficiency). These data-driven developments often involve collaborations between agri-IT companies, farmers' cooperatives and other companies in the food supply chain. These business-to-business initiatives and interactions are increasingly conducted through inter-organisational coordination hubs, in which standardised IT-based platforms provide data and business process interoperability for interactions among the organisations.

In Figure 1 an example of such a network of collaborating organisations is provided. It involves four stakeholders around the farmer (i) sperm supplier, (ii) milk processor, (iii) feed supplier and (iv) milking robot supplier. Multiple farms can be involved, and for each stakeholder relation data are collected from the farm and collected in a data platform. All stakeholders can then receive multiple datasets back from this platform, depending on authorization settings. This has to be arranged and governed by some form of network administrative organization. It can be expected that the IoF2020 use cases can be modelled into a similar picture.



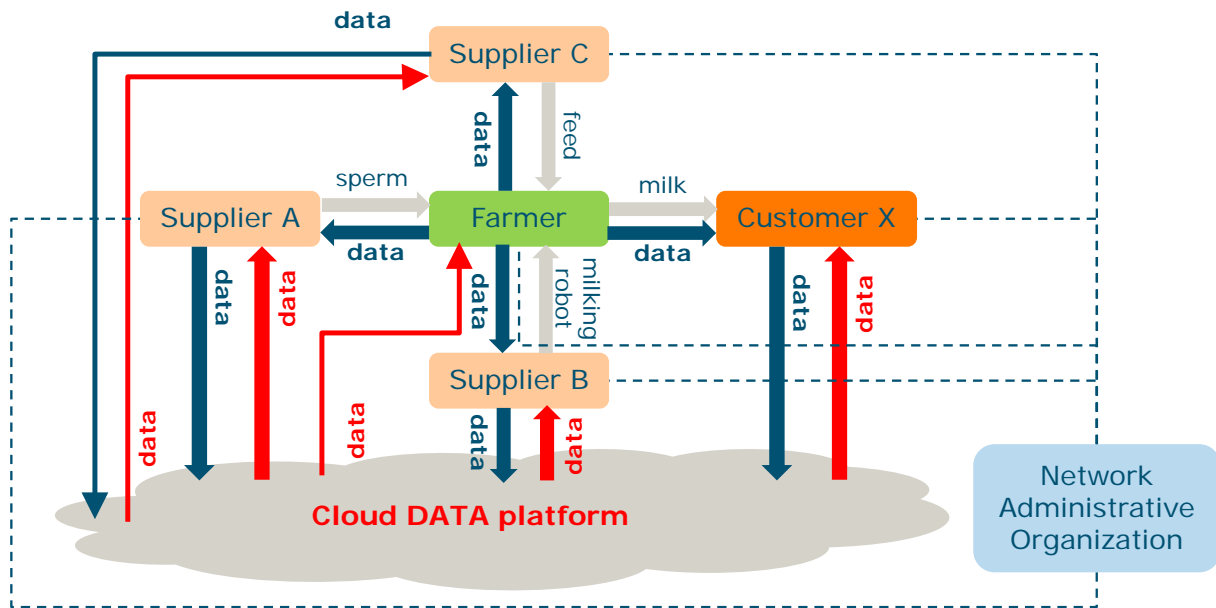


Figure 1 Example of a data sharing network in dairy farming in the agri-food sector

There are multiple ways of expanding and intensifying such a Big Data network. For instance, at the end of the market chain, consumers may play a role in future. They could be interested in Big Data as a quality check for products, and they could in principle provide consumer based data to the platform. Moreover, the role of government can be relevant. While governments seem to be interested in open data possibilities, this may cause uncomfortable privacy concerns for suppliers in the network, who prefer to keep business models away from the public.

The management of Big Data initiatives comprises challenges of privacy and security, which impact discouragements and distrust among farmers. Trust is considered to be a starting point in increasing Big Data applications. Many involved companies are refraining from sharing data because of the fear of issues such as data security, privacy and liability. Corresponding to the different stages of the data value chain (data capture, data storage, data transfer, data transformation, data analytics and data marketing), several key issues of Big Data applications have been identified and can be summarised as data availability, data quality, access to data, security, responsibility, liability, data ownership, privacy, costs, and business models. Ownership and value of data appear to be important issues in discussions on the governance of Big Data driven inter-organisational applications. A growing number of Big Data initiatives address privacy and security concerns. Also Big Data applications raise power-related issues that sometimes can lead to potential abuses of data. And many companies expect to develop new business models with data but are in a kind of deadlock and afraid of taking the first step. So the challenge is how companies could or should deal with these inter-organisational governance issues as they are considered as (the most) hindering factors for fulfilling the promises and opportunities of Big Data in agri-food sector.



Against this background, the network structure raises multiple questions in the scope of data management. For instance, how is the communication among the different actors? On what conditions do farmers take part, and how easy is it to enter or leave the network? What is the role of a network administrative organization and how openly do they perform together with partners of the network?

## 1.2. OBJECTIVE OF THIS DOCUMENT

Task 1.4 'Development Data Management Plan' is meant to address these questions by developing a plan and specific guidelines for the use cases and the project as a whole concerning data management. The Data Management Plan (DMP) will be developed, outlining:

- how (research) data will be collected, processed or generated within the project;
- what methodology and standards will be adopted;
- whether and how this data will be shared and/or made open;
- and how this data will be curated and preserved during and after the project.

The DMP aims to ensure that IoF2020 activities are compliant with the H2020 Open Access policy and the recommendations of the Open Research Data pilot. The DMP will furthermore explain how the project will be connected with other past and on-going initiatives such as EIP-Agri, agINFRA and global channels, such as OpenAIRE, CIARD and GODAN.

Under this task an Open Access Support Pack will be developed translating the generic H2020 requirements and recommendations into specific guidelines and advice that can be applied in the project.

This document (D1.4) will provide first version of Data Management Plan containing a general overview of relevant developments and a first inventory of needs and issues that play a role in the use cases. The application of the DMP by all IoF2020 partners will be continuously monitored under this task and an updated version of the DMP including more detailed specific support packs for the use cases (D1.5) will be delivered in Month 36.

## 1.3. OUTLINE

The remainder of this document is organized as follows. Chapter 2 will describe the approach how the results were found. This results in an overview of general developments on data management, described in Chapter 3. Chapter 4 will provide a short overview of relevant past and on-going initiatives and projects on data management in agriculture. Then a first inventory of the needs and potential issues of the use cases will be described. Based on the findings in these Chapters, a

concrete Data Management Plan will be provided in Chapter 6, followed by some general conclusions in Chapter 7.

## 2. APPROACH

The approach that was followed to generate this report consists of the following steps (see Figure 2):

1. Identification and description of relevant external developments in the field of data management in general and more specific for agriculture (see Chapter 3).
2. Identification and description of relevant initiatives and/or projects (in the recent past of on-going) in the field of (agricultural) data management (see Chapter 4).
3. Based on the results of the previous two steps a quick scan of the needs and issues in the IoF2020 use cases is made (see Chapter 5).
4. Based on the results of all previous steps a preliminary version of the IoF2020 Data Management is defined that provides general guidelines for the use cases (see Chapter 6).

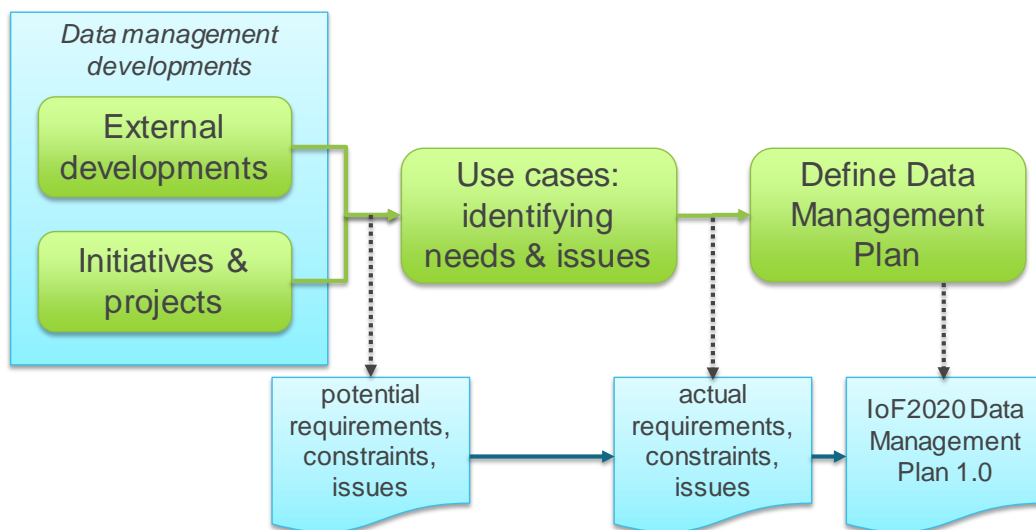


Figure 2 Steps that were taken to achieve a first version of the Data Management Plan for IoF2020

After these steps the data management plan will be further refined and tailored into a specific support pack for the use cases, but this is outside the scope of this deliverable. The results of this deliverable will be iteratively updated and refined as a living document but consolidated in a next deliverable in Month 36 (D1.5).

## 3. EXTERNAL DEVELOPMENTS

In this chapter we will highlight and briefly describe some data management developments that are most relevant to a H2020 innovation action project such as IoF2020 in Section 3.1. In Section 3.2 we will then describe the possible consequences for IoF2020 of these developments.

### 3.1. DATA MANAGEMENT DEVELOPMENTS

#### H2020 Open Access Policy<sup>1</sup>

Open access (OA) can be defined as the practice of providing on-line access to scientific information that is free of charge to the user and that is re-usable. In the context of R&D, open access to 'scientific information' refers to two main categories:

- Peer-reviewed scientific publications (primarily research articles published in academic journals)
- Scientific research data: data underlying publications and/or other data (such as curated but unpublished datasets or raw data)

It is now widely recognised that making research results more accessible to all societal actors contributes to better and more efficient science, and to innovation in the public and private sectors. The Commission therefore supports open access at the European level (in its framework programmes), at the Member States level and internationally.

#### *Peer-reviewed scientific publications*

All projects receiving Horizon 2020 funding are **required** to make sure that any peer-reviewed journal article they publish is openly accessible, free of charge (article 29.2. Model Grant Agreement).

#### *Research data*

The Commission is running a **pilot on open access** to research data in Horizon 2020: the Open Research Data (ORD) pilot. This pilot takes into account the need to balance openness with the protection of scientific information, commercialisation and Intellectual Property Rights (IPR), privacy concerns, and security, as well as questions of data management and preservation. The pilot applies to research data underlying publications but beneficiaries can also voluntarily make other datasets

---

<sup>1</sup> The information in this section was taken from:

<http://ec.europa.eu/research/openscience/index.cfm?pg=openaccess> accessed 22 Aug. 2017



open. Participating projects are required to develop a Data Management Plan, in which they will specify what data will be open. In previous work programmes, the ORD Pilot was limited to some specific areas of Horizon 2020. Starting with the 2017 work programme, however, the ORD pilot was extended to cover **all thematic areas** of Horizon 2020, thus realising the Commission's ambition of "open research data per default" (but allowing for opt-outs).

*More information*

For details of how open access applies to beneficiaries in projects funded under Horizon 2020, please see the [Guidelines on Open Access to Scientific Publications and Research Data](#) and/or the [Guidelines on data management](#).

Also:

- [Participants Portal](#)
- [OpenAIRE](#)
- [Open access in FP7](#)

## **Open Research Data pilot<sup>2</sup>**

*What is the open research data pilot?*

Open data is data that is free to access, reuse, repurpose, and redistribute. The Open Research Data Pilot aims to make the research data generated by selected Horizon 2020 projects accessible with as few restrictions as possible, while at the same time protecting sensitive data from inappropriate access.

If your Horizon 2020 project is part of the pilot, and your data meets certain conditions, you must deposit your data in a research data repository where they will be findable and accessible for others. Don't panic - you are not expected to share sensitive data or breach any IPR agreements with industrial partners. You do not need to deposit all the data you generate during the project either – only that which underpins published research findings and/or has longer-term value. In addition to supporting your research's integrity, openness has many other benefits. Improved visibility means your research will reach more people and have a greater impact – for science, society and your own career. Recent studies have shown that citations increase when data is made available alongside the publication; these papers also have a longer shelf-life.

---

<sup>2</sup> The information in this section was taken from <https://www.openaire.eu/opendatapilot> accessed at 22 Aug. 2017



*Which H2020 strands are required to participate?*

Projects starting from January 2017 are by default part of the Open Data Pilot. If your project started before earlier and stems from one of these Horizon 2020 areas, you are automatically part of the pilot as well::

- Future and Emerging Technologies
- Research infrastructures (including e-Infrastructures)
- Leadership in enabling and industrial technologies – Information and Communication Technologies
- Nanotechnologies, Advanced Materials, Advanced Manufacturing and Processing, and Biotechnology: ‘nanosafety’ and ‘modelling’ topics
- Societal Challenge: Food security, sustainable agriculture and forestry, marine and maritime and inland water research and the bioeconomy - selected topics in the calls H2020-SFS-2016/2017, H2020-BG-2016/2017, H2020-RUR-2016/2017 and H2020-BB-2016/2017, as specified in the work programme
- Societal Challenge: Climate Action, Environment, Resource Efficiency and Raw materials – except raw materials
- Societal Challenge: Europe in a changing world – inclusive, innovative and reflective Societies
- Science with and for Society
- Cross-cutting activities - focus areas – part Smart and Sustainable Cities.

Maybe data sharing is not appropriate for your project; the [EC's Guide on Open Access Scientific Publications and Research Data](#) lists conditions that would allow or require you to opt out of the pilot. In that case please consider if a partial opt-out is possible.

*What is a data management plan (DMP)?*

To help you optimise the potential for future sharing and reuse, a Data Management Plan (DMP) can help you to consider any problems or challenges that may be encountered and helps you to identify ways to overcome these. A DMP should be thought of as a “living” document outlining how the research data collected or generated will be handled during and after a research project. Remember, the plan should be realistic and based around the resources available to you and your project partners. There is no point in writing a gold plated plan if it cannot be implemented!

It should describe:

- The data set: What kind of data will the project collect or generate, and to whom might they be useful later on? The pilot applies to (1) the data and metadata needed to validate results in scientific publications and (2) other curated and/or raw data and metadata that may be required for validation purposes or with reuse value.
- Standards and metadata: What disciplinary norms will you adopt in the project? What is the data about? Who created it and why? In what forms it is available? Metadata answers such questions to enable data to be found and understood, ideally according to the particular standards of your scientific discipline. Metadata, documentation and standards help to make your data Findable, Accessible, Interoperable and Re-usable or FAIR for short.
- Data sharing: By default as much of the resulting data as possible should be archived as Open Access. Therefore legitimate reasons for not sharing resulting data should be explained in the DMP. Remember, no one expects you to compromise data protection or breach any IPR agreements. Data sharing should be done responsibly. The DMP Guidelines therefore ask you to describe any ethical or legal issues that can have an impact on data sharing. Furthermore,
- Archiving and preservation: Funding bodies are keen to ensure that publicly funded research outputs can have a positive impact on future research, for policy development, and for societal change. They recognise that impact can take quite a long time to be realised and, accordingly, expect the data to be available for a suitable period beyond the life of the project. Remember, it is not simply enough to ensure that the bits are stored in a research data repository, but also consider the usability of your data. In this respect, you should consider preserving software or any code produced to perform specific analyses or to render the data as well as being clear about any proprietary or open source tools that will be needed to validate and use the preserved data.

The DMP is not a fixed document. The first version of the DMP is expected to be delivered within the first 6 months of your project, but you don't have to provide detailed answers to all the questions yet. The DMP needs to be updated over the course of the project whenever significant changes arise, such as new data or changes in the consortium policies or consortium composition. The DMP should be updated at least in time with the periodic evaluation or assessment of the project as well as in time for the final review. Consider reviewing your DMP at regular intervals in the project and consider how you might make use of scheduled WP and/or project staff meetings to facilitate this review.

*What practical steps should you take?*

1. When your project is part of the pilot, you should [create a Data Management Plan](#). Your institution may offer Research Data Management support to help you planning.

2. Also, you should [select a data repository](#) that will preserve your data, metadata and possibly tools in the long term. It is advisable to contact the repository of your choice when writing the first version of your DMP. Repositories may offer guidelines for sustainable data formats and metadata standards, as well as support for dealing with sensitive data and licensing.
3. As noted earlier, you do not need to keep everything. Curating data requires time and effort so you want to make sure that you are putting your effort into the outputs that really matter. Select what data you'll need to retain to support validation of your finding but also consider any data outputs that may have longer term value as well – for you and for others.

### Links

- EC's Guide on Open Access to Scientific Publications and Research Data in Horizon 2020 (updated August 25, 2016) [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)
- EC's Guidelines on Data Management in Horizon 2020 (updated July 26, 2016): [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)
- EC's Agenda on Open Science: <https://ec.europa.eu/digital-agenda/en/open-science>
- DMPonline tool: <https://dmponline.dcc.ac.uk/>
- DCC How to Write a DMP guide: <http://www.dcc.ac.uk/resources/how-guides/develop-data-plan>
- DCC How to Select What Data to Keep guide: <http://www.dcc.ac.uk/resources/how-guides/five-steps-decide-what-data-keep>
- DCC How to Licence Research Data guide: <http://www.dcc.ac.uk/resources/how-guides/license-research-data>
- RDNL video The what, why and how of data management planning: <http://datasupport.researchdata.nl/en/start-de-cursus/ii-planfase/datamanagementplanning/>
- Software Sustainability Institute's Software Management Plan: [https://www.software.ac.uk/sites/default/files/images/content/SMP\\_Checklist\\_2016\\_v0.1.pdf](https://www.software.ac.uk/sites/default/files/images/content/SMP_Checklist_2016_v0.1.pdf)



## European Data Economy

Building a European data economy is part of the Digital Single Market strategy. The initiative aims at fostering the best possible use of the potential of digital data to benefit the economy and society. It addresses the barriers that impede the free flow of data to achieve a European single market.

### *General need for action*

[Digital data](#) is an essential resource for economic growth, competitiveness, innovation, job creation and societal progress in general. The EU needs to ensure that data flows across borders and sectors. This data should be accessible and reusable by most stakeholders in an optimal way. A coordinated European approach is essential for the development of the data economy, as part of the [Digital Single Market strategy](#). The European Commission adopted a [Communication](#) on "Building a European Data Economy", accompanied by a [Staff Working Document](#) on January 2017, where it:

- looks at the rules and regulations impeding the free flow of data and present options to remove unjustified or disproportionate data location restrictions, and
- outlines legal issues regarding access to and transfer of data, data portability and liability of non-personal, machine-generated digital data.

The European Commission has launched a [public consultation](#) and [dialogue with stakeholders](#) on these topics to gather further evidence. This process will help identify future policy or legislative measures that will unleash Europe's data economy. The development of the European Data economy is one of the three emerging challenges identified in the [mid-term review](#). The actions to be implemented are:

- to prepare a legislative proposal on the EU free flow of data cooperation framework (autumn 2017)
- to prepare an initiative on accessibility and re-use of public and publicly funded (spring 2018)

In addition, the Commission will continue its work on liability and other emerging data issues. For more details, read the [Communication](#).

### *Facing the challenge - removing data localisation restrictions: the free flow of data*

Free flow of data means the freedom to process and store data in electronic format anywhere within the EU. It is necessary for the development and use of innovative data technologies and services. In order to achieve the free flow of data, the European Commission will collect more evidence on data location restrictions and assess their impacts on businesses, especially SMEs and start-ups, and public sector organisations. The Commission will also discuss the justifications for and proportionality of those data location restrictions with Member States and other stakeholders. It will then take justified and appropriate follow-up actions, in line with [better regulation principles](#), to address the issue.

### *Exploring the emerging issues relating to the data economy*

The European Commission is currently defining, scoping and articulating the following issues in order to trigger and frame a dialogue with stakeholders:

- Non-personal machine-generated data need to be tradable to allow innovative business models to flourish, new market entrants to propose new ideas and start-ups to have a fair chance to compete.
- Data-driven technologies are transforming our economy and society, resulting in the production of ever-increasing amounts of data. This phenomenon leads to innovative ways of collecting, acquiring, processing and using data which can pose a challenge to the current legal framework.
- Access to and transfer of non-personal data, data liability, as well as portability of non-personal data, interoperability and standards are complex legal issues.

This [consultation](#) process will contribute to the policy choices taken by the European Commission in the future.

### *Useful links*

- Have a look at the [workshops](#) organised on how to build a European data economy.
- [Press release and MEMO - Q&A](#)
- [Communication](#) on Building a European Data Economy
- [Staff Working Document](#) on Building a European Data Economy
- [Factsheet](#) on Building a European Data Economy
- Study on [Measuring the economic impact of cloud computing in Europe](#)
- Study on [Facilitating cross border data flow in the DSM](#)
- Intermediary study on [Cross-border data flow in the Digital Single Market: data location restrictions](#)
- [Speech from Commissioner Oettinger](#) at the Conference "Building European Data Economy" (17 October 2016).
- [Speech from Vice-President Ansip](#) at the Digital Assembly 2016, "Europe should not be afraid of data" (29 September 2016).



### **Agricultural Data developments**

In view of the technical changes brought forth by Big Data and Smart Farming, we seek to understand the consequences for the stakeholder network and governance structure around the farm in this section.

The literature suggests major shifts in roles of and power relations among different players in existing agri-food chains. We observed the changing roles of old and new software suppliers in relation to Big Data and farming and emerging landscape of data-driven initiatives with prominent role of big tech and data companies like Google and IBM. In Figure 3, the current landscape of data-driven initiatives is visualized.

The stakeholder networks exhibits a high degree of dynamics with new players taking over the roles played by other players and the incumbents assuming new roles in relation to agricultural Big Data. As opportunities for Big Data have surfaced in the agribusiness sector, big agriculture companies such as Monsanto and John Deere have spent hundreds of millions of dollars on technologies that use detailed data on soil type, seed variety, and weather to help farmers cut costs and increase yields. Other players include various accelerators, incubators, venture capital firms, and corporate venture funds (Monsanto, DuPont, Syngenta, Bayer, DOW etc.).

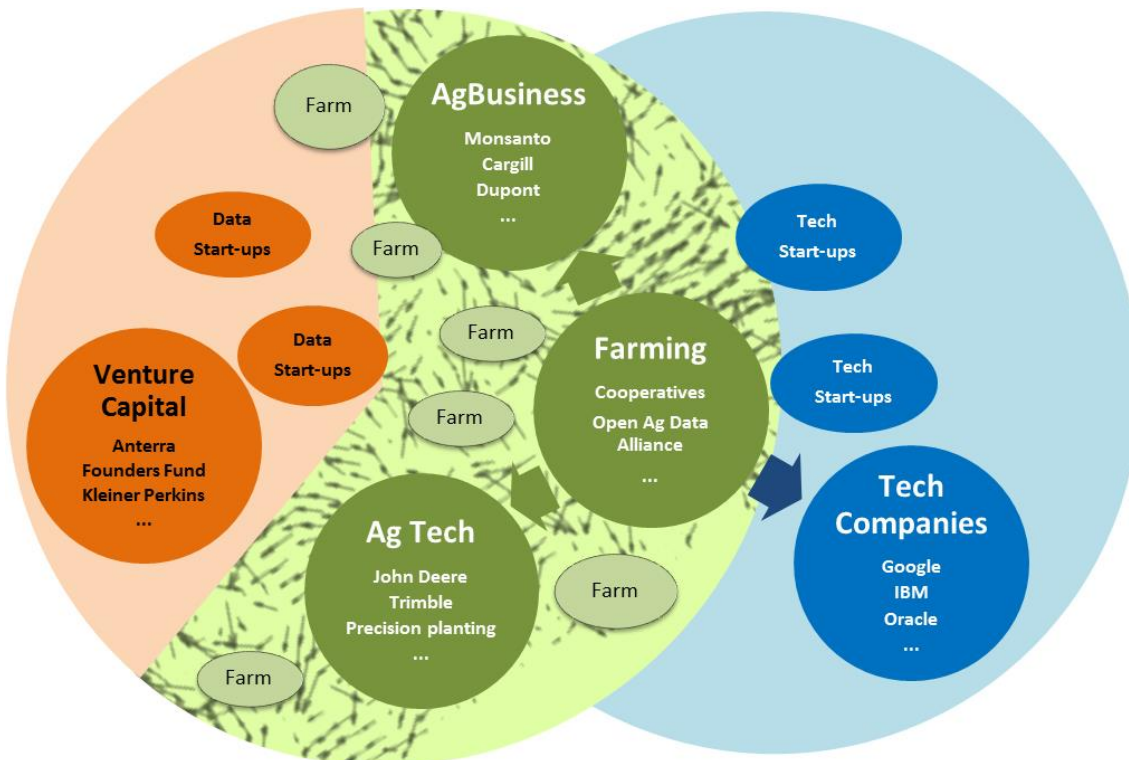


Figure 3 The landscape of the Big Data network with business players.

Monsanto has been pushing big-data analytics across all its business lines, from climate prediction to genetic engineering. It is trying to persuade more farmers to adopt its cloud services. Monsanto says farmers benefit most when they allow the company to analyse their data - along with that of other farmers - to help them find the best solutions for each patch of land.

While corporates are very much engaged with Big Data and agriculture, start-ups are at the heart of action, providing solutions across the value chain, from infrastructure and sensors all the way down to software that manages the many streams of data from across the farm. As the ag-tech space heats up, an increasing number of small tech start-ups are launching products giving their bigger counterparts a run for their money. In the USA, start-ups like FarmLogs, FarmLink and 640 Labs challenge agribusiness giants like Monsanto, Deere, DuPont Pioneer. One observes a swarm of data-service start-ups such as FarmBot (an integrated open-source precision agriculture system) and Climate Corporation. Their products are powered by many of the same data sources, particularly



those that are freely available such as from weather services and Google Maps. They can also access data gathered by farm machines and transferred wirelessly to the cloud. Traditional agri-IT firms such as NEC and Dacom are active with a precision farming trial in Romania using environmental sensors and Big Data analytics software to maximize yields.

Venture capital firms are now keen on investing in agriculture technology companies such as Blue River Technology, a business focusing on the use of computer vision and robotics in agriculture. The new players to Smart Farming are tech companies that were traditionally not active in agriculture. For example, Japanese technology firms such as Fujitsu are helping farmers with their cloud based farming systems. Fujitsu collects data (rainfall, humidity, soil temperatures) from a network of cameras and sensors across the country to help farmers in Japan better manage its crops and expenses. Data processing specialists are likely to become partners of producers as Big Data delivers on its promise to fundamentally change the competitiveness of producers.

Beside business players such as corporates and start-ups, there are many public institutions (e.g., universities, USDA, the American Farm Bureau Federation, GODAN) that are actively influencing Big Data applications in farming through their advocacy on open data and data-driven innovation or their emphasis on governance issues concerning data ownership and privacy issues. Well-known examples are the Big Data Coalition, Open Agriculture Data Alliance (OADA) and AgGateway. Public institutions like the USDA, for example, want to harness the power of agricultural data points created by connected farming equipment, drones, and even satellites to enable precision agriculture for policy objectives like food security and sustainability. Precision farming is considered to be the “holy grail” because it is the means by which the food supply and demand imbalance will be solved. To achieve that precision, farmers need a lot of data to inform their planting strategies. That is why USDA is investing in big, open data projects. It is expected that open data and Big Data will be combined together to provide farmers and consumers just the right kind of information to make the best decisions.

Data ownership is an important issue in discussions on the governance of agricultural Big Data generated by smart machinery such as tractors from John Deere. In particular, value and ownership of precision agricultural data have received much attention in business media. It has become a common



practice to sign Big Data agreements on ownership and control data between farmers and agriculture technology providers. Such agreements address questions such as: How can farmers make use of Big Data? Where does the data come from? How much data can we collect? Where is it stored? How do we make use of it? Who owns this data? Which companies are involved in data processing?

There is also a growing number of initiatives to address or ease privacy and security concerns. For example, the Big Data Coalition, a coalition of major farm organizations and agricultural technology providers in the USA, has set principles on data ownership, data collection, notice, third-party access and use, transparency and consistency, choice, portability, data availability, market speculation, liability and security safeguards<sup>3</sup>. And AgGateway, a non-profit organization with more than 200 member companies in the USA, have drawn a white paper that presents ways to incorporate data privacy and standards<sup>4</sup>. It provides users of farm data and their customers with issues to consider when establishing policies, procedures, and agreements on using that data instead of setting principles and privacy norms. The European farmers and agri-cooperatives association COPA-COGECA has recently also published their 'Main principles underpinning the collection, use and exchange of agricultural data'<sup>5</sup>. The principles concern:

- Ownership of farm data – data produced on the farm or during farming operations should be owned by the farmers themselves. If this data is used, also indirectly through combined services, the farmer should be somehow compensated for this.
- Ownership of the underlying rights to derived data - In any case, it should be clear when farm data is used and for what purpose; the farmer should be in full control of his/her data. When farm data is used by third parties anonymization and security are of utmost importance.

---

<sup>3</sup> <https://www.fb.org/issues/technology/data-privacy/privacy-and-security-principles-for-farm-data>

<sup>4</sup> [https://s3.amazonaws.com/aggateway\\_public/AgGatewayWeb/WorkingGroups/Committees/DataPrivacySecurityCommittee/2017-03-31%20Data%20Privacy%20and%20Use%20White%20Paper%20-%201.2.pdf](https://s3.amazonaws.com/aggateway_public/AgGatewayWeb/WorkingGroups/Committees/DataPrivacySecurityCommittee/2017-03-31%20Data%20Privacy%20and%20Use%20White%20Paper%20-%201.2.pdf)

<sup>5</sup> <http://www.copa-cogeca.be/Download.ashx?ID=1566736&fmt=pdf>

- Duration, suspension and termination of supply – farmers must be provided with the possibility to opt out of a contract on data use with the right to delete all historical data.
- Guarantee of compliance with laws and regulation – data collection should not be in conflict with general laws (e.g. on privacy) and data should not be used for unlawful purposes
- Liability – in contracts on data use intellectual property rights of farmers and agri-cooperatives must be protected and liabilities must be clearly described. It is not always possible to go into all possible details so there should be a good balance between what is written in the contract and trust between partners.

Based on these type of principles several codes of conduct/practice with concrete models for contracts have been developed such as the New Zealand Farm Data Code of Practice<sup>6</sup>, the Dutch BO-Akkerbouw Code of Conduct<sup>7</sup>, and Ag Data Transparency has established a certifying procedure<sup>8</sup>. More similar codes and model contracts will pop-up in the near future.

The ‘Ownership Principle’ of the Big Data Coalition states that “We believe farmers own information generated on their farming operations. However, it is the responsibility of the farmer to agree upon data use and sharing with the other stakeholders (...).” While having concerns about data ownership, farmers also see how much companies are investing in Big Data. In 2013, Monsanto paid nearly 1 billion US dollars to acquire The Climate Corporation, and more industry consolidation is expected. Farmers want to make sure they reap the profits from Big Data, too. Such change of thinking may lead to new business models that allow shared harvesting of value from data.

In conclusion, Big data applications in Smart Farming will potentially raise many power-related issues. There might be companies emerging that gain much power because they get all the data. In the agri-food chain these could be input suppliers or commodity traders, leading to a further power shift in

---

<sup>6</sup> <http://www.farmdatacode.org.nz/>

<sup>7</sup> <https://www.bo-akkerbouw.nl/wp-content/uploads/2016/12/GEDRAGSCODE-datagebruik-akkerbouw-v2-170329.pdf>

<sup>8</sup> <http://www.agdatatransparent.com/>

market positions. This power shift can also lead to potential abuses of data e.g. by the GMO lobby or agricultural commodity markets or manipulation of companies. Initially, these threats might not be obvious because for many applications small start-up companies with hardly any power are involved. However, it is a common business practice that these are acquired by bigger companies if they are successful and in this way the data still gets concentrated in the hands of one big player. It can be concluded that Big Data is both a huge opportunity as a potential threat for farmers.

### 3.2. CONSEQUENCES FOR IOF2020

The following consequences for IoF2020 could be derived from the developments that were described in the previous section:

- 1) Any peer-reviewed journal article should be published openly accessible
- 2) Research data should be stored in a central repository where it is findable and accessible by others
  - a) Since IoF2020 is an Innovation Action type of project, the first question is whether research data is actually involved; this has to be identified first
  - b) It should be further explored whether data sharing is appropriate in IoF2020 and if a partial opt-out for the Open Research Data pilot is appropriate
- 3) A data management plan needs to be written for IoF2020 according to the guidelines provided by the ORD Pilot
- 4) It should be identified if and to what extent the data-driven technologies in the IoF2020 use cases are aligned with the European Data Economy policy.
  - a) Since this policy is also under debate, IoF2020 should actively participate in this debate and development of this Data Economy.
- 5) It should be explored which issues around agricultural data sharing (e.g. data ownership/control, rights to use data, etc.) are (potentially) playing a role and if principles and guidelines, especially provided by IoF2020 partner COPA-COGECA, are applicable. And if yes, what type of contracts should be set-up?





## 4. RELEVANT INITIATIVES AND PROJECTS

In this chapter several projects and initiatives are described that are dealing with (Agricultural) Data and are potentially relevant for IoF2020.

### 4.1. PROJECTS AND INITIATIVES

#### GODAN<sup>9</sup>

GODAN 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. It is a rapidly growing group, currently with over 584 partners from national governments, non-governmental, international and private sector organisations that have committed to a joint [Statement of Purpose](#).

The initiative focuses on building high-level support among governments, policymakers, international organizations and business. GODAN promotes collaboration to harness the growing volume of data generated by new technologies to solve long-standing problems and to benefit farmers and the health of consumers. GODAN encourages collaboration and cooperation between stakeholders in the sector.

The GODAN initiative was launched At the 2012 G-8 Summit, G-8 leaders committed to the New Alliance for Food Security and Nutrition, the next phase of a shared commitment to achieving global food security.

As part of this commitment, they agreed to “share relevant agricultural data available from G-8 countries with African partners and convene an international conference on Open Data for Agriculture, to develop options for the establishment of a global platform to make reliable agricultural and related information available to African farmers, researchers and policymakers, taking into account existing agricultural data systems.”

In April 2013, the commitment to convene an international conference on Open Data for Agriculture was fulfilled when the G8 International Conference on Open Data for Agriculture took place.

This conference worked to ‘obtain commitment and action from nations and relevant stakeholders to promote policies and invest in projects that open access to publicly funded global agriculturally relevant data streams, making such data readily accessible to users in Africa and world-wide, and ultimately supporting a sustainable increase in food security in developed and developing countries.

---

<sup>9</sup> The information in this section is based on <http://www.godan.info/about>, accessed 30 Aug. 2017



The GODAN initiative was a by-product of this conference and was announced at the Open Government Partnership Conference in October 2013.

Any organization that supports open access to agriculture and nutrition data can become member of GODAN. Partners include government, donors, international and not-for-profit organizations and businesses. GODAN partners support the shared principles based on GODAN's Statement of Purpose:

- Agricultural and nutritional data to be available, accessible, usable and unrestricted
- Partners aim to build high level policy and private sector support for open data
- Encourage collaboration and cooperation across existing agriculture, nutrition and open data activities and stakeholders to solve long-standing global problems

GODAN partners commit to:

- Host regular conversations with our peer multilateral and local organisations to identify and share best practices and determine how to more effectively share data and provide useable analysis for local application
- Recruit new partners to GODAN

GODAN activities and its Secretariat are financially supported by the US Government, the UK Department for International Development (DFID), the Government of the Netherlands, FAO, Technical Centre for Agricultural and Rural Cooperation (CTA), GFAR, The Open Data Institute (ODI), the CGIAR and CABI.

*More information:*

- [Statement of Purpose](#)
- [Theory of Change](#)
- [Ownership of Open Data: Governance Options for Agriculture and Nutrition](#)
- [A Global Data Ecosystem for Agriculture and Food](#)



## EIP-Agri<sup>10</sup>

The agricultural European Innovation Partnership (EIP-AGRI) works to foster competitive and sustainable farming and forestry that 'achieves more and better from less'. It contributes to ensuring a steady supply of food, feed and biomaterials, developing its work in harmony with the essential natural resources on which farming depends.

The European Innovation Partnership for Agricultural productivity and Sustainability (EIP-AGRI) has been launched in 2012 to contribute to the European Union's strategy 'Europe 2020' for smart, sustainable and inclusive growth. This strategy sets the strengthening of research and innovation as one of its five main objectives and supports a new interactive approach to innovation: *European Innovation Partnerships*.

The EIP-AGRI pools funding streams to boost interactive innovation. Having an idea is one thing, turning it into an innovation action is another. Different types of available funding sources can help get an agricultural innovation project started, such as the **European Rural Development policy** or the EU's research and innovation programme **Horizon 2020**. The EIP-AGRI contributes to integrating different funding streams so that they contribute together to a same goal and duplicate results.

Rural Development will in particular support **Operational Groups** and **Innovation Support Services** within a country or region. **Horizon 2020** will fund multi-actor projects and thematic networks involving partners from at least three EU countries. Other policies may offer additional opportunities.

The EIP-AGRI brings together innovation actors (farmers, advisers, researchers, businesses, NGOs and others) at EU level and within the rural development programmes (RDPs). Together they form an EU-wide EIP network. EIP Operational Groups can be funded under the RDPs, are project-based and tackle a certain (practical) problem or opportunity which may lead to an innovation. The Operational Group approach makes the best use of different types of knowledge (practical, scientific, technical, organisational, etc.) in an interactive way. An Operational Group is composed of those key actors that are in the best position to realise the project's goals, to share implementation experiences and to disseminate the outcomes broadly. The first Operational Groups are currently being set up in several EU countries and regions.

The Rural Networks' Assembly, which was launched in January 2015, coordinates two networks - the EIP-AGRI Network and the European Network for Rural Development (ENRD). The Assembly includes several subgroups, one of them being the permanent Subgroup on Innovation for agricultural productivity and sustainability. This Subgroup on Innovation will support the EIP-AGRI Network.

---

<sup>10</sup> The information in this section is based on <https://ec.europa.eu/eip/agriculture/en/about>, accessed 30 Aug. 2017



The EIP-AGRI website has exciting and interactive features. All visitors can voice their research needs, discover funding opportunities for innovation projects and look for partners to connect with. Through the website's interactive functions, users can share **innovative project ideas** and practices, information about **research and innovation projects**, including projects' results, by filling in the available easy-to-use e-forms. Various EIP-AGRI-related **publications** are available for download on the website, providing visitors with information on a wide range of interesting topics. Future functionalities will be developed for **Operational Groups** and European funds managing authorities once the programmes start. Through this collaborative effort, the EIP-AGRI website will become a one-stop-shop for agricultural innovation in Europe.

The **EIP-AGRI network** is run by the European Commission (DG Agriculture and Rural Development) with the help of the **EIP-AGRI Service Point**. The Service Point offers a wide range of tools and services which can help you further your ideas and projects. It also facilitates networking activities; enhancing communication, knowledge sharing and exchange through conferences, **Focus Groups**, workshops, seminars and publications.

A recent series of workshops on Agricultural Data is particularly relevant:

- 'Data revolution: emerging new data-driven business models in the agri-food sector', Sophia, Bulgaria, 22-23 June 2016. <https://ec.europa.eu/eip/agriculture/en/event/eip-agri-seminar-%E2%80%98data-revolution-emerging-new>
- 'Data Sharing: ensuring fair sharing of digitisation benefits in agriculture' Tuesday 4-5 April 2017, Bratislava, Slovakia. <https://ec.europa.eu/eip/agriculture/en/event/eip-agri-workshop-data-sharing>
- 'Digital Innovation Hubs: mainstreaming digital agriculture', 1-2 June 2017, Kilkenny, Ireland. <https://ec.europa.eu/eip/agriculture/en/event/eip-agri-seminar-digital-innovation-hubs>

*More about EIP-AGRI:*

- [Brochure: EIP-AGRI network](#)
- [Brochure: EIP-AGRI Service Point](#)

#### **AgInfra<sup>11</sup>**

**AGINFRA+** aims to exploit core e-infrastructures such as [EGI.eu](#), [OpenAIRE](#), [EUDAT](#) and [D4Science](#), towards the evolution of the AGINFRA data infrastructure, so as to provide a sustainable channel addressing adjacent but not fully connected user communities around Agriculture and Food.

---

<sup>11</sup> The information in this section is based on <http://www.plus.aginfra.eu/>, accessed 28-Sep-2017.



To this end, the project will develop and provide the necessary specifications and components for allowing the rapid and intuitive development of variegating data analysis workflows, where the functionalities for data storage and indexing, algorithm execution, results visualization and deployment are provided by specialized services utilizing cloud based infrastructure(s).

Furthermore, **AGINFRA+** aspires to establish a framework facilitating the transparent documentation and exploitation and publication of research assets (datasets, mathematical models, software components results and publications) within AGINFRA, in order to enable their reuse and repurposing from the wider research community.

**AGINFRA** is the European research hub and thematic aggregator that catalogues and makes discoverable publications, data sets and software services developed by Horizon 2020 research projects on topics related to agriculture, food and the environment.

It is part of the broader vision of the European research e-infrastructure “European Open Science Cloud”, a synergy between [OpenAIRE](#), [EUDAT](#), [GEANT](#), [EGI](#), [LIBER](#).

With the integration of big data processing components from projects like the Horizon 2020 BigDataEurope and the FP7 [SemaGrow](#), [AGINFRA](#) evolves into a big data analytics capable e-infrastructure for agri-food, to respond to the needs of three (3) adjacent yet not fully connected user communities:

- **H2020 SC1** Health
- **H2020 SC2** Food security and sustainable agriculture
- **H2020 SC5** Climate action and environment

**AGINFRA+** addresses the challenge of supporting user-driven design and prototyping of innovative e-infrastructure services and applications. It particularly tries to meet the needs of the scientific and technological communities that work on the multi-disciplinary and multi-domain problems related to agriculture and food. It will use, adapt and evolve existing open e-infrastructure resources and services, in order to demonstrate how fast prototyping and development of innovative data- and computing-intensive applications can take place.

This project builds upon the extensive experience and work of its partners, who are key stakeholders in the e-infrastructures ecosystem. It also implements part of a strategic vision shared between **Agroknow**, the National Agronomic Research Institute of France (**INRA**), the Alterra Institute of the Wageningen University & Research Center (**ALTERRA**), the National Institute for Risk Assessment of Germany (**BfR**), and the Food and Agriculture Organization (**FAO**) of the United Nations - the latter one, not participating as a funded beneficiary, but supporting the project and its activities. These stakeholders are part of a core group of internationally recognised players (including the Chinese Academy of Agricultural Sciences) aiming to put in place a free global data infrastructure

for research and innovation in agriculture, food and environmental science. This data infrastructure will become an incubator of the large infrastructure investments that global donors (including the European Commission) make in the field of agricultural research around the world.

**AGINFRA+** will evolve and develop further the resources and services of the AGINFRA data infrastructure, which has been developed in the context of the FP7 [agINFRA](#) project. The new project will build upon core components of AGINFRA, such as:

- the federated data and software registry of [CIARD RING](#),
- the [AGINFRA API](#) gateway for indexing and hosting executable software components for advanced data processing & analysis,
- the open source software stack for data analysis, indexing, publication and querying developed by projects such as FP7 [SemaGrow](#) and H2020 [Big Data Europe](#),
- the semantic backbone of the Global Agricultural Concept Scheme (GACS1) that has been based upon the alignment of FAO's AGROVOC with the USDA's National Agricultural Library Thesaurus and CABI's Thesaurus,
- the advanced research data set processing & indexing demonstrators developed within FP7 SemaGrow for specific scientific communities such as [Trees4Futures](#) and [gMIP](#).

The envisaged pilots will focus on three societal challenges that are of primary importance for our planet and for humanity:

- Food safety risk assessment and risk monitoring, addressing H2020 SC1 Health, demographic change and well-being.
- Plant phenotyping for food security, addressing H2020 SC2 Food security, sustainable agriculture and forestry, marine and maritime and inland water research, and the Bioeconomy.
- Agro-climatic and Economic Modelling, addressing H2020 SC5 Climate action, environment, resource efficiency and raw materials.

In order to realize its vision, AGINFRA+ will achieve the following objectives:

- identify the requirements of the specific scientific and technical communities working in the targeted areas, abstracting (wherever possible) to new AGINFRA services that can serve all users;
- design and implement components that serve such requirements, by exploiting, adapting and extending existing open e-infrastructures (namely, OpenAIRE, EUDAT, EGI, and D4Science), where required;
- define or extend standards facilitating interoperability, reuse, and repurposing of components in the wider context of AGINFRA;



- establish mechanisms for documenting and sharing data, mathematical models, methods and components for the selected application areas, in ways that allow their discovery and reuse within and across AGINFRA and served software applications;
- increase the number of stakeholders, innovators and SMEs aware of AGINFRA services through domain specific demonstration and dissemination activities.

The development of fully defined demonstrator applications in each of the three application areas will allow to showcase and evaluate the AGINFRA components in the context of specific end-user requirements from different scientific areas.

### **BigDataEurope<sup>12</sup>**

Big Data Europe will undertake the foundational work for enabling European companies to build innovative multilingual products and services based on semantically interoperable, large-scale, multi-lingual data assets and knowledge, available under a variety of licenses and business models.

Big Data Europe aims to:

- Collect requirements for the ICT infrastructure needed by data-intensive science practitioners tackling a wide range of societal challenges; covering all aspects of publishing and consuming semantically interoperable, large-scale, multi-lingual data assets and knowledge.
- Design and implement an architecture for an infrastructure that meets requirements, minimizes the disruption to current workflows, and maximizes the opportunities to take advantage of the latest European RTD developments, including multilingual data harvesting, data analytics, and data visualization.

Societal challenges and their Big Data focus areas are:

- Health - heterogeneous data linking and integration, biomedical semantic indexing
- Food & Agriculture - large-scale distributed data integration
- Energy - real-time monitoring, stream processing, data analytics, decision support
- Transport - streaming sensor network and geospatial data integration
- Climate - real-time monitoring, stream processing and data analytics

---

<sup>12</sup> The information in this section was taken from <https://www.big-data-europe.eu/>, accessed 28-Sep-2017.



- Social Sciences - statistical and research data linking and integration
- Security - real-time monitoring, stream processing and data analytics, image data analysis

The **Food and Agriculture pilot (SC2)** within BDE is focusing on viticulture. The problem of discovery and linking of information is present in every major area of agricultural research and agriculture in general. This is especially true in viticulture where different research methodologies produce a great amount of heterogeneous data from diverse sources; scientists need to be able to find all this information so as to analyse and correlate it to provide integrated solutions to the emerging problems in the European and global vineyard. These problems arise largely because of the impact of climate change and therefore the exploitation of the appropriate grapevine varieties is very important. Factors to bear in mind include the intensity of diseases, the intensification of the cultivation, the proper implementation of precision viticulture systems that affect the quality of viticultural products and their role in human health.

The overall goal of the SC2 Pilot is to demonstrate the ability of Big Data technologies to complement existing community-driven systems (e.g. [VITIS](#) for the Viticulture Research Community) with efficient large-scale back-end processing workflows. The pilot deployment is organised in three Cycles with different targeted objectives:

- **Pilot Cycle 1 (SC2 Pilot Pitch-Deck)** - The goal of this Pilot Cycle is to showcase a large-scale processing workflow that automatically annotates scientific publications relevant to Viticulture. The focus of the first demonstrator cycle is on the Big Data aspects of such a workflow (i.e. storage, messaging and failure management) and not on the specificities of the NLP modules/tools used in this demonstrator.
- **Pilot Cycle 2 (SC2 Pilot Maturity / Functionality Expansion)** - The goal of this Pilot Cycle is to showcase the ability of scalable processing workflows to handle a variety of data types (beyond bibliographic data) relevant to Viticulture.
- **Pilot Cycle 3 (Lowering SC2 Community Boundaries)** - The goal of this Pilot Cycle is to provide an engaging, intuitive graphical web interface addressing key data-oriented questions relevant to the Viticulture Research Community, and if possible, intuitive interfaces for end-users for sharing and linking their on-the-field generated data.

In SC2 Pilot Cycle 1, content mainly refers to open scientific publications relevant to Viticulture, available at FAO/AGRIS and NCBI/PubMed in PDF format (about 26K and 7K publications respectively). In Cycle 2, the content pool has been extended to include:

- Weather Data, available via publicly available APIs (e.g. OpenWeatherMap, Weather Underground, AccuWeather etc.)





- User-generated data, e.g. geotagged photos from leaves, young shoots and grape clusters, ampelographic data, SSR-marker data etc.

Additional data sources include:

- Sensor Data, measuring temperature, humidity and luminosity retrieved from sensors installed in selected experimental vineyards,
- ESA Copernicus Sentinel 2 Data, for selected experimental vineyards.

The goal of the inclusion of these data is to complement the existing SC2 Pilot Demonstrator Knowledge Base so as to support complex real-life research questions, based on the correlation of environmental conditions with real observations on crop production and quality.

### **DATABIO<sup>13</sup>**

The Data-Driven Bioeconomy project (DataBio) focuses on the production of best possible raw materials from agriculture, forestry and fishery for the bioeconomy industry to produce food, energy and biomaterials taking into account responsibility and sustainability

In order to meet the above objectives, DataBio is controlling and putting to use the innovative ICTs and information flows centered mostly around the use of proximal and remote sensors, in order to provide a streamlined Big Data Infrastructure for data discovery, retrieval, processing and visualizing, in support to decisions in bioeconomy business operations.

The main goal of the **DataBio project** is to show the benefits of Big Data technologies in the raw material production from **agriculture, forestry and fishery/aquaculture** for the bioeconomy industry **to produce food, energy and biomaterials** responsibly and sustainably.

**DataBio** proposes to deploy a state of the art, big data platform on top of the existing partners' infrastructure and solutions – the **Big DATABIO Platform**.

The work will be continuous cooperation of experts from end user and technology provider companies, from bioeconomy and technology research institutes, and of other partners.

DataBio is working on pilots in three areas: Agriculture, Fishery and Forestry. The Agricultural pilots are divided into:

- **Precision Horticulture including vine and olives**

---

<sup>13</sup> The information in this section is based on <https://www.databio.eu/en/>, accessed 28-Sep-2017



- A1. Precision agriculture in olives, fruits, grapes and vegetables
- A2. Big Data management in greenhouse eco-systems
- **B. Arable Precision Farming**
  - B1. Cereals and biomass crops
  - B2. Machinery management and environmental issue
- **C. Subsidies and insurance**
  - C1. Insurance
  - C2. CAP support

In the pilots also associated partners and other stakeholders will be actively involved.

The selected pilots and concepts will be transformed into pilot implementations using co-innovative approaches and tools where the bioeconomy sector end users, experts and other stakeholders will give input to the user and sector domain understanding for the requirement specifications for ICT, Big Data and Earth Observation experts and for other solution providers in the consortium.

Based on the preparation and requirement specifications work, the pilots are implemented using and selecting the best suitable market ready or almost market ready Big Data and Earth Observation methods, technologies, tools and services to be integrated to the common **Big DATABIO Platform**.

During the pilots the close cooperation continues and feedback from the bioeconomy sector user companies will be harnessed in the technical and methodological upgrades for pilot implementations.

Based on the pilot results and the new solutions also new business opportunities are expected.

In addition during the pilots the end users are participating in trainings to learn how to use the solutions and developers also outside the consortium will be active in the Hackathons to design and develop new tools, services and application for the platform.

**Databio's expected achievements include but are not limited to:**

- Demonstrate increase of productivity in bioeconomy
- Increase of market share of Big Data technology providers in the bioeconomy sector
- More than double the use of Big Data technology in bioeconomy



- Leveraging additional target sector investments by a factor of >5
- More than 100 organizations in demonstrations
- Liaison with other Big Data actions
- Closely working with BDVA

## 4.2. RELATIONSHIP WITH IOF2020

Through overlapping partners between the mentioned initiatives and projects and IoF2020, there is already quite a natural ground for collaboration and knowledge exchange on data management. Especially IoF2020's coordinator Wageningen University & Research is involved in most relevant initiatives and projects. However since we are dealing with relatively large initiatives, projects and organizations, knowledge exchange and collaboration it is not automatically guaranteed. Specific attention needs to be paid to this. Work on very similar specific pilot areas (e.g. viticulture, agricultural machinery data, etc.) would be a good starting point for this.

Since the beginning of the IoF2020 project there were already several occasions (conferences, workshops, etc.) where explicit connections were made with these projects and initiatives. It was decided to setup an explicit collaboration with the DataBio project.

## 5. INVENTORY OF USE CASES

Based on the issues that are identified and described in the previous chapters, a first scan of the IoF2020 use cases was made.

A quick scan for the constraints to data privacy and security was conducted by a questionnaire at use case level. The results are listed in **Table 1**.

**Table 1** Potential constraints to data privacy and security for each use case.

Trial and use cases	Data privacy & security constraints
<b>Arable Trial</b>	
UC 1.1. Within-field management zoning	Farmer's data can only be used outside his company only if he agrees. Other partners provide proprietary software tools and information.
UC 1.2 Precision Crop Management	The data collected by the Bosch systems are intended to remain confidential and will be the property of the

	<p>farmer who has acquired a system. The Bosch cloud must be able to guarantee this confidentiality and the security of its data.</p> <p>The API-AGRO interface, which is an API management platform operated by a consortium represented in the Use Case by Arvalis, allows to add a layer of management of the confidentiality and possible opening to other actors of those data. While remaining under the control of the owner of the systems, it will be possible to communicate the acquired data according to conditions which still have to be defined while guaranteeing a control by the owner of the data of the destination that he wants to give to his data.</p> <p>The Arvalis models and agro-climatic references that will be used for DSS are the property of Arvalis.</p>
UC1.3 Soya Protein Management	No constraints so far about data privacy and security in UC 1.3.
UC1.4 Farm Machine Interoperability	The constraints will be dependent on the agreements achieved with the partners and farmers involved in the UCs that we collaborate with, i.e. UC 1.1 and UC 1.3. UC 1.4 members will have to sign agreements with the farmers for the access to their data, but also NDAs for specific software elements and equipment components with the other use case participants.
<b>Dairy Trial</b>	
UC2.1 Grazing Cow Monitor	No specific data protection issues arise
UC2.2 Happy Cow	The platform is secured to prevent access to cow/farm data by default. Login opens the data associated to the user. Communication is encrypted.
UC2.3 Silent Herdsman	An agreement on the use of the data acquired on each trial site has to be established. The principle is that the data will be owned by the farmer and will be made available for the purposes of the project. Any public reporting of the key findings regarding the trial farms must be anonymised and cleared by the data owners.
UC2.4 Remote Milk Quality	Up to our present knowledge, there are no constraints.

	There will be no data exchanged. Data of users will be monitored and handled but not used. Only alerts and checks are exchanged
<b>Fruit Trial</b>	
UC3.1 Fresh table grapes chain	No issues mentioned.
UC3.2 Big wine optimization	<p>None till the edition of this document. However, the following considerations will be taken into account during the implementation phase, others could come according to the discover needs:</p> <ul style="list-style-type: none"> <li>• All the communications are expected to be carried out over DTLS (Datagram Transport Layer Security) using DTLS 1.2 version, it means ECC asymmetric cryptography for the key negotiation and authentication, in conjunction with AES symmetric cryptography for the efficient and optimal data exchange.</li> <li>• Since the protocol, expected to be used, is OMA LwM2M, it includes all the security from this protocol for defining an access control for the list of servers that can interact with the devices and also the definition of a bootstrap server and commissioning service to guarantee the secure provisioning of all the credentials and configuration details.</li> </ul>
UC3.3 Automated olive chain	Farmers involved in the action would like to remain privacy their personal data.
UC3.4 Intelligent fruit logistics	<p>Collection of data on customer sites using our IoT-RTI:</p> <ul style="list-style-type: none"> <li>• Who owns the data collected on premise by the IoT-enabled RTI?</li> <li>• Who can do what with the data?</li> <li>• Which data can be shown to whom (data-driven business models!)?</li> <li>• Encryption of data</li> </ul> <p>Questions regarding security of IoT-Technology used in the pilot:</p> <ul style="list-style-type: none"> <li>• Is there a danger of hacking, data capturing by</li> </ul>

	<p>others, interference with communication?</p> <ul style="list-style-type: none"> <li>• Would it be possible to manipulate collected data?</li> <li>• Can the central database be accessed by external parties?</li> <li>• To which security level is it possible to encrypt the data?</li> </ul>
<b>Vegetable Trial</b>	
UC4.1 City farming leafy vegetables	<p>The data generated in a city farm, in general, are related to plant growth and the operation of a city farm. Data related to persons are out of scope. The data are owned by the owner of the city farm. The data collected in a city farm represent a value from an economic point of view. Access to the data in general only takes place with the consent of the owner. Data access should be protected. Also, any data link should be secured.</p> <p>Measures need to be in place against unauthorized or unlawful access or processing of data as well as against accidental loss or damage of data. Data access rights management should be in place.</p>
UC4.2 Chain-integrated greenhouse production	<p>Data should be owned by farmers and agri-business, and privacy issues should be respected according to European regulation. As for Intellectual Property, a consortium Agreement (CA) will be negotiated between all partners, settling among other things the internal organization of the consortium, reflecting what has been described about the project management structure. With respect to this Use Case 4.3, Background IP remains with the partner(s) who created it and Foreground IP (if any) goes to the partner(s) that developed it.</p>
UC4.3 Added value weeding data	<p>Yes, farmers need to provide access to their machine data and crop data.</p>
UC4.4 Enhanced quality certification system	<p>No issues mentioned</p>

Meat Trial	
UC5.1 Pig farm management	Raw pig farm data is confidential and owned by the farmer – and must only be accessed in aggregated fashion. Aggregated feed data must be kept in the farm and only accessed by Cloud Service component when actually needed. All accesses should be logged so that farmer can be informed about what is being read by whom. The same is true for slaughterhouse data.
UC5.2 Poultry chain management	<p>Privacy:</p> <ul style="list-style-type: none"> <li>• Delivering data to platforms (who owns data...)</li> <li>• The data remain the ownership of the provider – since SADA is an integrated company, all raw data is property of SADA</li> <li>• Support on legal and ethical issues on workers manipulation model</li> </ul> <p>Security:</p> <ul style="list-style-type: none"> <li>• Authentication of the data delivered to the platform</li> </ul>
UC5.3 Meat Transparency and Traceability	Data privacy and security must be handled within this project. Farms may provide information which can result in detailed insights on their internal processes. Data access restriction is provided by the access layer. This layer implements a set of static and dynamic access rules for EPCIS events, supporting different roles and actors. Data providers (farmer) stay owner of their data and decide which supply chain partner can access what kind of information.

From this table it can be generally concluded that there is quite some difference between the several use cases in thinking about data management issues. Some use cases have thought about this in much detail and are very aware of the various issues that might arise. Other use cases say that they don't see any issue or only have thought about in a general way.

The results from this scan will be shared and discussed between the use cases and it is expected that they can learn from each other. This will result in a more generic list of data management issues that in the end again can be made specific for each use case.

For all trials and use cases together an inventory was made on what type of data is expected as is presented in Table 2.

*Table 2 Data that is expected to be generated by all use case and trials*

Type	Origin	Format	Estimated size
Dissemination material: <ul style="list-style-type: none"> <li>• press releases,</li> <li>• leaflets,</li> <li>• audio-visual material,</li> <li>• posters,</li> <li>• images/photos</li> </ul>	IoF2020 consortia generated	Adobe Photoshop (.psd) MS PowerPoint (.ppt, pptx) JPEG (.jpg, .jpeg) MS Word (.doc, .docx) Adobe Acrobat Reader (.pdf) Adobe Illustrator (.ai) Hard copy	0.5 Gb
Demographic and personal data of third parties interviewed	Interviews, questionnaires, cooperation agreements, invitation letters to participate in the pilots, application forms, informed consent forms	Google forms MS Excel (.xls, .xlsx) MS Word (.doc, .docx) Adobe Acrobat Reader (.pdf) Comma Separated Values (.csv)	0.5 Gb
Demographic and personal data of partners within Use Cases	Survey, questionnaires	Google forms MS Excel (.xls, .xlsx) MS Word (.doc, .docx) Adobe Acrobat Reader	0.5 Gb



Type	Origin	Format	Estimated size
		(.pdf) Comma Separated Values (.csv)	
Project reports/deliverables with internal reviewing process	WP2 team generated	MS Excel (.xls, .xlsx) MS Word (.doc, .docx)	0.5 Gb
Project reports/deliverables with external reviewing process	WP2 team generated	MS Excel (.xls, .xlsx) MS Word (.doc, .docx)	0.5 Gb
Contact details of project partners and advisory/scientific board(s) (Name, Email, Phone, Skype ID...)	Survey, questionnaires	MS Excel (.xls, .xlsx) MS Word (.doc, .docx) Adobe Acrobat Reader (.pdf)	0.5 Gb
Guidelines for consortium members	WP2 team generated	MS Word (.doc, .docx) Adobe Acrobat Reader (.pdf) MS PowerPoint (.ppt, pptx)	0.5 Gb
Outputs generated at project events	Agendas and meeting minutes, Attendance sheets	MS Word (.doc, .docx) Adobe Acrobat Reader (.pdf) MS PowerPoint (.ppt, pptx)	0.2 Gb
Dataset produced by aggregating data during project implementation	Interview Focus group discussion	MS Word (.doc, .docx) Adobe Acrobat Reader (.pdf)	1 Gb

Type		Origin	Format	Estimated size
		Observation Survey	MS PowerPoint (.ppt, pptx)	
A dataset documenting and providing evidence for either a report or a publication produced in the context of project activities	Peer-reviewed scientific publications	Interview Focus group discussion	MS Word (.doc, .docx)	0.5 Gb
	Scientific publications with internal reviewing process	Observation Survey Desk research etc.	Adobe Acrobat Reader (.pdf)	

## 6. DATA MANAGEMENT PLAN

### 6.1. GENERAL GUIDELINES FOR DATA MANAGEMENT IN IOF2020

From the previous Chapters in this document the following general guidelines for data management in IoF2020 can be derived:

- Research papers that are derived from the project should be published according to the open access policy
- Research data should be stored in a central repository according to the FAIR principles: findable, accessible, interoperable and reliable
- The use cases in IoF2020 should be clearly aligned to the European Data Economy policy and more specifically in line with the principles and guidelines provided by the stakeholder community *i.e.* COPA-COGECA.

### 6.2. FURTHER ACTIONS AND RECOMMENDATIONS

The following steps should be taken to further develop data management in IoF2020:

- Investigate what research data is involved in the use cases and other project activities and define how they should be treated according to the open data policy.
- Participate actively in the debate and developments of the European Data Economy and data sharing in agriculture.
- Analyze and explore the use cases in a deeper way in order to identify which data management issues potentially play a role and define plans how to deal with them.
- Although many collaborative actions with other relevant projects and initiatives are already taking place and there are many natural connections through the project partners, a more systematic and structural approach should be explored in order to maximize the benefits and impact of the mutual activities on data management.

This Data Management Plan and its future updates will be actively communicated with all partners in IoF2020 and the use cases in particular. It is planned to have a workshop on this issue, possibly combined with other similar issues such as 'Ethics' during the annual project meeting in Spring 2018 organized by WP4.



## 7. CONCLUSIONS

Data or Big Data rapidly has become a new resource or asset in the current economy, also in the agricultural sector. This development leads to several issues that have to be addressed such as data availability, quality, access, security, responsibility, liability, ownership, privacy, costs and business models. In Europe several initiatives and projects have already been established to work on this in a general context (e.g. the EU open data policy) and more specifically in agriculture (e.g. GODAN, COPA-COGECA).

The consequences for IoF2020 are partly mandatory (open access publishing and open research data) and for the other part guiding principles that also have to be further explored. In this document we have presented a first version of a Data Management Plan with concrete actions to be taken in order to establish open, transparent data management in IoF2020.