**Project Acronym:** DataBio  
**Grant Agreement number:** 732064 (H2020-ICT-2016-1 – Innovation Action)  
**Project Full Title:** Data-Driven Bioeconomy  
**Project Coordinator:** INTRASOFT International

---

**Dissemination level:** PU - Public  
**Type of Document:** Report  
**Contractual date of delivery:** M36 – 31/12/2019  
**Deliverable Leader:** UStG  
**Status - version, date:** Draft – v1.0, 31/1/2020  
**WP / Task responsible:** WP7  
**Keywords:** Business planning, KPIs
Executive Summary

The objective of the deliverable D7.2 is the analysis of the DataBio pilots from a business perspective. Big Data Technology (BDT) is considered to have high potential for creating business value and opportunities for providers of BDT and end-users in the three bio industries considered in the DataBio project: agriculture, forestry and fisheries. Besides creating, implementing and testing the technical performance of concrete BDT applications in these three industries, the goal of the DataBio pilots was also to illustrate the business impact of the technology.

To measure the business impact, for each pilot key performance indicators (KPIs) were developed. Before starting the pilots, baseline values were measured as basis for comparison with achieved results after the execution of the pilots. Even though not all pilots were able to measure the KPIs after the pilot completion, due to problems of getting the right data or unexpected developments during pilot executions, overall it was possible to demonstrate the business potential and impact of BDT in agriculture, forestry and fisheries.

The business impact of BDT is illustrated in three chapters of this deliverable, each dedicated to one of the bioindustries and containing a business analysis of all respective pilots. Wherever possible, each pilot is described in terms of: 1) motivation and objectives, 2) baseline set-up, 3) BDT pipeline as well as data applied including reflection on technology, 4) business impact, 5) how-to guidelines for practice and 6) summary and outlook.

The business impact of agricultural pilots is analysed in Chapter 3. Agricultural pilots were able to illustrate that BDT enables well-informed decision-making for farmers and facilitates more sustainable application of natural resources, namely irrigation water as well as more sustainable farming through lower use of fertilizers and pest disease management. Further business value in terms of higher yields is supported also through yield management, yield prediction and crop improvement with genomic prediction models. Through the targeted decision-making, BDT improves also the overall productivity in agriculture by enabling farmers to better use their working resources. Similar applications provide value also for agricultural authorities by supporting CAP activities and also agricultural insurance.

BDT is introduced to the agricultural market both through commercial companies operating for profit and through state-owned providers of BDT solutions for free. In particular the second type of providers target also societal value creation and preservation by supporting the management of natural resources.

The business value of forestry pilots is analysed in Chapter 4 that contains sections for all forestry pilots. BDT supports value creation in forestry by providing better data about the status and health of the forest. Compared to agriculture, main actors in the forestry pilots are state-organisations responsible for forest management. With BDT technology it is possible to collect and organise data about the forest that can serve as basis for more efficient forest management by private or state owners of forest and also for thriving ecosystems of various companies providing services for the forestry industry. Compared to the agricultural
industry, the creation of the basic data infrastructure in the form of open forest data is driven by state-owned institutions.

The business impact of fisheries is analysed in Chapter 5 containing all fisheries pilots. Fisheries is, compared to agriculture and forestry, the most regulated industry. As the possible catches, i.e. the output of the activities is regulated, BDT is mainly applied to increase productivity of fishing ships and fishing activities by decreasing of costs. The focus is on reduction of fuel used, reduction of time to search for fishes, reduction of by-catches and increase of income through better knowledge of the market.

Despite of the high diversities of the three industries, one common aspect regarding market entrance of BDT is the strong need for cooperation and trust building with the end consumers. In agriculture, BDT application has to be calibrated for pairs of crops and soil (regions), while in forestry data are also bound to specific regions and type of forest trees (also diseases and similar). In fisheries, BDT application have to be calibrated to specific fishing ships. Thus, market entrance requires building of trustful relationships to end customers and exchange and sharing of data particularly historical data.

Overall, the DataBio pilots were able to illustrate the business impact in all three industries and to illustrate market entrance strategies. By focusing on bioindustries BDT also support better preservation and sustainable use of natural resources.
**Deliverable Leader:** Katarina Stanoevska-Slabeva (UStG)

**Contributors:**
- Vera Lenz-Kesekamp, Tabea Vogler (UStG)
- Jukka Miettinen, Renne Tergujeff (VTT)
- Stéphanie Bonnet (SPACEBEL)
- Petr Lukeš (FMI)
- Allan Nielsen (DTU)
- Seppo Huurinainen (Wuudis)
- Virpi Stenman (METSAK)
- Jesus Maria Estrada Villegas (TRAGSA)
- Savvas Rogotis, Costas Mastrogiannis, Veronika Ovcinnikova (NP)
- Nicole Bartelds (NB Advies)
- Ephrem Habyarimana (CREA)
- Karel Charvát Jr. (LESPRO)
- Jaroslav Smejkal (ZETOR)
- Antonella Catucci (e-geos)
- Olimpia Copăcenaru, Adrian Stoica (TerraS)
- Jürgen Decloedt (VITO)
- Karl-Johan Reite (SINTEF Fishery)
- Jose A. Fernandes (AZTI)

**Reviewers:**
- Anna Bozza (CiaoT)
- Renne Tergujeff (VTT)
- Tomáš Mildorf (UWB)
- Karl-Johan Reite (SINTEF Fishery)
- Antonella Catucci (e-geos)
- Caj Södergård (VTT)

**Approved by:** Athanasios Poulakidas (INTRASOFT)

---

**Document History**

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Contributor(s)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>16/12/2019</td>
<td>K. Stanoevska-Slabeva</td>
<td>Initial draft, including early input</td>
</tr>
<tr>
<td>0.2</td>
<td>20/1/2020</td>
<td>K. Stanoevska-Slabeva</td>
<td>Final draft for review</td>
</tr>
<tr>
<td>0.3</td>
<td>21/1/2020</td>
<td>K. Stanoevska-Slabeva</td>
<td>Revised version with review comments incorporated</td>
</tr>
<tr>
<td>0.4</td>
<td>21/1/2020</td>
<td>A. Poulakidas</td>
<td>Formatting and QA</td>
</tr>
<tr>
<td>0.5</td>
<td>29/1/2020</td>
<td>K. Stanoevska-Slabeva</td>
<td>Revision with more emphasis on business</td>
</tr>
<tr>
<td>0.6</td>
<td>30/1/2020</td>
<td>K. Stanoevska-Slabeva, A. Poulakidas</td>
<td>Incorporation of final feedback from pilots</td>
</tr>
<tr>
<td>1.0</td>
<td>31/1/2020</td>
<td>A. Poulakidas</td>
<td>Final version for submission</td>
</tr>
</tbody>
</table>
# Table of Contents

Executive Summary .................................................................................................................. 2
Table of Contents ....................................................................................................................... 5
Table of Figures .......................................................................................................................... 8
List of Tables ............................................................................................................................... 9

1 Introduction ............................................................................................................................ 10
   1.1 PROJECT SUMMARY .................................................................................................. 10
   1.2 DOCUMENT SCOPE .................................................................................................. 12
   1.3 DOCUMENT STRUCTURE .......................................................................................... 12

2 Analysis Approach .................................................................................................................. 14

3 Business Analysis of Agriculture Pilots ................................................................................. 16
   3.1 INTRODUCTION ........................................................................................................... 16
   3.2 PILOT A1.1: PRECISION AGRICULTURE IN OLIVES, FRUITS, GRAPES, AND PILOT B1.2: CEREALS, BIOMASS AND COTTON CROPS_2 ................................................................................................................. 17
      3.2.1 Introduction, motivation and goals of the pilots ...................................................... 17
      3.2.2 Pilots set-up ........................................................................................................... 18
      3.2.3 Technology used ................................................................................................... 19
      3.2.4 Business value and impact .................................................................................... 21
      3.2.5 How-to guidelines for practice ............................................................................. 23
      3.2.6 Summary and outlook .......................................................................................... 24
   3.3 PILOT A1.2: PRECISION AGRICULTURE IN VEGETABLE SEED CROPS .......................................................................................................................... 25
      3.3.1 Introduction, motivation and goals of the pilot ...................................................... 25
      3.3.2 Pilot set-up ........................................................................................................... 25
      3.3.3 Technology used ................................................................................................... 27
      3.3.4 Business value and impact .................................................................................... 29
      3.3.5 How-to guidelines for practice ............................................................................. 30
      3.3.6 Summary and outlook .......................................................................................... 30
   3.4 PILOT A1.3: PRECISION AGRICULTURE IN VEGETABLES_2 (POTATOES) ......................................................................................................................... 31
      3.4.1 Introduction, motivation and goals of the pilot ...................................................... 31
      3.4.2 Pilot set-up ........................................................................................................... 32
      3.4.3 Technology used ................................................................................................... 32
      3.4.4 Business value and impact .................................................................................... 34
      3.4.5 How-to guidelines for practice ............................................................................. 37
      3.4.6 Summary and outlook .......................................................................................... 38
   3.5 PILOT A2.1: BIG DATA MANAGEMENT IN GREENHOUSE ECOSYSTEM ......................................................................................................................... 39
      3.5.1 Introduction, motivation and goals of the pilot ...................................................... 39
      3.5.2 Pilot set-up ........................................................................................................... 40
      3.5.3 Technology used ................................................................................................... 41
      3.5.4 Business value and impact .................................................................................... 44
      3.5.5 How-to guidelines for practice ............................................................................. 45
      3.5.6 Summary and outlook .......................................................................................... 45
   3.6 PILOT B1.1: CEREALS AND BIOMASS CROP .................................................................. 46
      3.6.1 Introduction, motivation and goals of the pilot ...................................................... 46
      3.6.2 Pilot set-up ........................................................................................................... 47
      3.6.3 Technology used ................................................................................................... 47
      3.6.4 Business value and impact .................................................................................... 49
      3.6.5 How-to guidelines for practice ............................................................................. 50
      3.6.6 Summary and outlook .......................................................................................... 52
   3.7 PILOT B1.3: CEREAL AND BIOMASS CROPS_3 ................................................................ 52

Dissemination level: PU - Public
3.7.1 Introduction, motivation and goals of the pilot
3.7.2 Pilot set-up
3.7.3 Technology used
3.7.4 Business value and impact
3.7.5 How-to guidelines for practice
3.7.6 Summary and outlook

3.8 PILOT B1.4: CEREALS AND BIOMASS CROPS_4
3.8.1 Introduction, motivation and goals of the pilot
3.8.2 Pilot set-up
3.8.3 Technology used
3.8.4 Business value and impact
3.8.5 Summary and outlook

3.9 PILOT B2.1: MACHINERY MANAGEMENT
3.9.1 Introduction, motivation and goals of the pilot
3.9.2 Pilot set-up
3.9.3 Technology used
3.9.4 Business value and impact

3.10 PILOT C1.1: INSURANCE (Greece)
3.10.1 Introduction, motivation and goals of the pilot
3.10.2 Pilot set-up
3.10.3 Technology used
3.10.4 Business value and impact
3.10.5 How-to guidelines for practice
3.10.6 Summary and outlook

3.11 PILOT C1.2: FARM WEATHER INSURANCE ASSESSMENT
3.11.1 Introduction, motivation and goals of the pilot
3.11.2 Pilot set-up
3.11.3 Technology used
3.11.4 Business value and impact
3.11.5 How-to guidelines for practice
3.11.6 Summary and outlook

3.12 PILOT C2.1: CAP SUPPORT
3.12.1 Introduction, motivation and goals of the pilot
3.12.2 Pilot set-up
3.12.3 Technology used
3.12.4 Business value and impact
3.12.5 How-to guidelines for practice
3.12.6 Summary and outlook

3.13 PILOT C2.2: CAP SUPPORT (Greece)
3.13.1 Introduction, motivation and goals of the pilot
3.13.2 Pilot set-up
3.13.3 Technology used
3.13.4 Business value and impact
3.13.5 How-to guidelines for practice
3.13.6 Summary and outlook

3.14 SUMMARIZING ANALYSIS OF AGRICULTURAL PILOTS

4 Business Analysis of Forestry Pilots
4.1 INTRODUCTION
4.2 PILOT 2.2.1: EASY DATA SHARING AND NETWORKING
4.2.1 Introduction, motivation and goals of the pilot
4.2.2 Pilot set-up
4.2.3 Technology used .............................................................................................................. 90
4.2.4 Business value and impact ............................................................................................... 91
4.2.5 Summary and outlook ...................................................................................................... 91
4.3 PILOT 2.2.2: MONITORING AND CONTROL TOOLS FOR FOREST OWNERS .............. 91
4.3.1 Introduction, motivation and goals of the pilot ................................................................. 91
4.3.2 Pilot set-up ....................................................................................................................... 92
4.3.3 Technology used .............................................................................................................. 92
4.3.4 Business value and impact ............................................................................................... 93
4.3.5 How-to guidelines for practice ......................................................................................... 94
4.4 PILOT 2.3.1: FOREST DAMAGE REMOTE SENSING ......................................................... 94
4.4.1 Introduction, motivation and goals of the pilot ................................................................. 94
4.4.2 Pilot set-up ....................................................................................................................... 94
4.4.3 Technology used .............................................................................................................. 95
4.4.4 Business value and impact ............................................................................................... 97
4.4.5 How-to guidelines for practice ......................................................................................... 99
4.4.6 Summary and outlook ..................................................................................................... 99
4.5 PILOT 2.3.2-FH: MONITORING OF FOREST HEALTH .................................................. 100
4.5.1 Introduction, motivation and goals of the pilot ................................................................. 100
4.5.2 Pilot set-up ....................................................................................................................... 101
4.5.3 Technology Used ............................................................................................................ 101
4.5.4 Business value and impact ............................................................................................... 102
4.5.5 How-To guidelines for Practice ...................................................................................... 103
4.5.6 Summary and outlook ..................................................................................................... 104
4.6 PILOT 2.3.2-IAS: INVASIVE ALIEN SPECIES CONTROL – PLAGUES ....................... 104
4.6.1 Introduction, motivation and goals of the pilot ................................................................. 104
4.6.2 Pilot set-up ....................................................................................................................... 105
4.6.3 Technology used .............................................................................................................. 106
4.6.4 Business value and impact ............................................................................................... 107
4.6.5 How-to guidelines for practice ......................................................................................... 107
4.6.6 Summary and outlook ..................................................................................................... 107
4.7 PILOT 2.4.1: WEB-MAPPING SERVICE FOR THE GOVERNMENT DECISION MAKING ......................... 108
4.7.1 Introduction, motivation and goals of the pilot ................................................................. 108
4.7.2 Pilot set-up ....................................................................................................................... 109
4.7.3 Technology used .............................................................................................................. 109
4.7.4 Business value and impact ............................................................................................... 109
4.7.5 How-to guidelines for Practice ...................................................................................... 110
4.7.6 Summary and outlook ..................................................................................................... 110
4.8 PILOT 2.4.2: FINNISH FOREST DATA BASED METSÄÄN.FI SERVICES ...................... 111
4.8.1 Introduction, motivation and goals of the pilot ................................................................. 111
4.8.2 Technology used .............................................................................................................. 113
4.8.3 Business value and impact ............................................................................................... 114
4.8.4 Summary and outlook ..................................................................................................... 116
5 Business Analysis of Fishery Pilots ....................................................................................... 118
5.1 INTRODUCTION .................................................................................................................... 118
5.2 THE COMMERCIAL FISHING BUSINESS ....................................................................... 118
5.3 SUMMARY OF THE FISHERIES PILOTS ........................................................................ 119
5.3.1 Datasets and challenges for fisheries pilots ................................................................. 120
5.3.2 Overall progress and benefits from the pilots .............................................................. 121
5.3.3 The Tuna Fisheries Pilots ............................................................................................... 123
5.3.4 The Small Pelagic Fisheries Pilots ............................................................................... 125
5.4 CLASSIFICATION OF KPIs IN FISHERY TRIALS ........................................................... 127
5.5 Pilots A1 and B1: Oceanic Tuna Fisheries Immediate Operational Choices and Planning ........................................128
  5.5.1 Pilot A1: Oceanic tuna fisheries immediate operational choices .................................................................129
  5.5.2 Pilot B1: Oceanic Tuna Fisheries Planning ..................................................................................................130
  5.5.3 Data processing flow ....................................................................................................................................131
  5.5.4 Data and results visualization ........................................................................................................................132
  5.5.5 Business value and impact for oceanic tuna pilots .......................................................................................135
  5.5.6 Summary and outlook ................................................................................................................................139

5.6 Pilot A2: Small Pelagic Fisheries Immediate Operational Choices .................................................................139
  5.6.1 Introduction, motivation and goals of the pilot ..............................................................................................139
  5.6.2 Pilot set-up .....................................................................................................................................................140
  5.6.3 Technology used ............................................................................................................................................140
  5.6.4 Business value and impact ..........................................................................................................................142
  5.6.5 How-to guidelines for practice .....................................................................................................................142
  5.6.6 Summary and outlook ................................................................................................................................142

5.7 Pilot B2: Small Pelagic Fisheries Planning .......................................................................................................143
  5.7.1 Introduction, motivation and goals ...............................................................................................................143
  5.7.2 Pilot set-up .....................................................................................................................................................143
  5.7.3 Technology used ............................................................................................................................................144
  5.7.4 Business value and impact ..........................................................................................................................146
  5.7.5 How-to guidelines for practice .....................................................................................................................146

5.8 Pilot C1: Small Pelagic Fish Stock Assessment .................................................................................................146
  5.8.1 Introduction, motivation and goals ...............................................................................................................146
  5.8.2 Pilot set-up .....................................................................................................................................................147
  5.8.3 Technology used ............................................................................................................................................148
  5.8.4 Business value and impact ..........................................................................................................................150
  5.8.5 How-to guidelines for practice .....................................................................................................................150
  5.8.6 Summary and outlook ................................................................................................................................151

5.9 Pilot C2: Small Pelagic Market Predictions and Traceability ..............................................................................151
  5.9.1 Introduction, motivation and goals ...............................................................................................................151
  5.9.2 Pilot set-up .....................................................................................................................................................152
  5.9.3 Technology used ............................................................................................................................................153
  5.9.4 Business value and impact ..........................................................................................................................155
  5.9.5 How-to guidelines for practice .....................................................................................................................156
  5.9.6 Summary and outlook ................................................................................................................................157

6 References .........................................................................................................................................................158

Table of Figures

Figure 1: Pilot A1.1 Aggregated Findings ..................................................................................................................21
Figure 2: Aggregated results of Pilot B1.2 in comparison with the target values ......................................................21
Figure 3: GS advantage: higher response to selection harnessing quantitative and population genetics with GEBV-driven intercrosses shortening generation intervals ..................................................................................43
Figure 4: Overall genomic prediction and selection roadmap ................................................................................43
Figure 5 - Left to right: NDVI image from multispectral RPAS data; RGB mosaic; thermal image over RGB mosaic; DSM ........................................................................................................................................49
Figure 6: Crops classification and irrigation needs ................................................................................................52
Figure 7: High-level overview of the affected area (flood incident), colour coded with the output of the followed damage assessment procedures .........................................................................................66
Figure 8: Technology pipeline for pilot C1.2 ...........................................................................................................70
Figure 9: Premium volume distribution for crop insurance in Europe .....................................................................72
Figure 10: Romania - total declared area and number of plots registered for CAP support (2019). Data Source: Agency for Payments and Intervention in Agriculture (APIA), Romania ................................................. 75
Figure 11: Number of farm holdings in Europe by economic size ........................................................................ 88
Figure 12: Forestry TEP .................................................................................................................................. 96
Figure 13: An example of Metsä.fi map layer consisting of multiple datasets ................................................... 112
Figure 14: Example of pilot data processing pipeline on a high abstraction level ............................................. 114
Figure 15: Example of Laatumetsä mobile solution with related map service .................................................. 115
Figure 16: Overview of fisheries pilots ............................................................................................................ 120
Figure 17: Plot of the active buoys deployed by the whole Basque fleet for one month of 2009, each black dot is the position sent from the buoy to the vessel (n=1,250,000) .................................................................. 124
Figure 18: Overview of data sources, stake holders and components in the pelagic fisheries ............................ 126
Figure 19: Scheme of the data processing flow ................................................................................................ 131
Figure 20: Example of web visualization of WMS-T services provided by CMEMS in the Indian Ocean .......... 132
Figure 21: Conceptual diagram showing a Bayesian network to forecast tuna biomass based on satellite data and models combined with fisheries data ................................................................................ 132
Figure 22: Bayesian network visualizing tipping points in the relationship between environmental conditions and the different levels of tuna captures .................................................................... 133
Figure 23: Bayesian networks showing environmental condition when the captures where high .................... 133
Figure 24: Bayesian network showing the higher probability of high captures given several favourable environmental conditions ........................................................................................................ 134
Figure 25: Scale show areas of higher probability of finding high biomass of tuna. Green circles show successful fishing attempts and in red circles failed fishing attends. Thin lines at sea show Economic Exclusive Zones (EEZs) showing territorial waters where only the country fleets and authorized fleets can fish. ................................................. 134
Figure 26: Total sailed nautical miles and fishing days (3 ships) ........................................................................ 136
Figure 27: Total consumed fuel oil and fuel oil consumed per kg of catch (3 vessels) ........................................ 137
Figure 28: Average fuel oil consumption and average vessel speed during sailing (3 vessels) ......................... 138
Figure 29: Variation in daily mackerel price ..................................................................................................... 155
Figure 30: Screenshot from the web portal, where filtering of historical data is facilitated ................................. 156

List of Tables

Table 1: The DataBio consortium partners ...................................................................................................... 11
Table 2: Overview of Analysis Aspects ........................................................................................................... 14
Table 3: Overview of Agricultural pilots .......................................................................................................... 17
Table 4: Overview of activities across region in pilot A1.1 and Pilot B1.2 ........................................................ 18
Table 5: Quantification of business gains (baseline – pilot value) in pilot A1.1 and Pilot B1.2 ......................... 22
Table 6: KPIs of the A2.1 pilot ......................................................................................................................... 44
Table 7: Pilot B1.3 KPIs .................................................................................................................................. 55
Table 8: Pilot B1.4 KPIs .................................................................................................................................. 58
Table 9: Pilot C1.1 KPIs .................................................................................................................................. 66
Table 10: Pilot C2.1-Romania KPIs ................................................................................................................ 80
Table 11: KPIs of the pilot C2.2 ....................................................................................................................... 85
Table 12: Overview of Forestry pilots ............................................................................................................. 89
Table 13: Pilot 2.3.1 KPIs ............................................................................................................................... 98
Table 14: Data production by DataBio fisheries pilots ....................................................................................... 120
Table 15: Development of TRL levels in the fisheries pilots ............................................................................ 122
Table 16: Fishery Pilot A1 assessment criteria. ................................................................................................. 135
1 Introduction

1.1 Project Summary

DataBio (Data-driven Bioeconomy) is a H2020 lighthouse project focusing on utilizing Big Data to contribute to the production of the best possible raw materials from agriculture, forestry, and fishery/aquaculture for the bioeconomy industry in order to produce food, energy and biomaterials, also taking into account responsibility and sustainability issues.

DataBio has deployed state-of-the-art Big Data technologies taking advantage of existing partners’ infrastructure and solutions. These solutions aggregate Big Data from the three identified sectors (agriculture, forestry, and fishery) and intelligently process, analyse and visualize them. The DataBio software environment allows the three sectors to selectively utilize numerous software components, pipelines and datasets, according to their requirements. The execution has been through continuous cooperation of end-users and technology provider companies, bioeconomy and technology research institutes, and stakeholders from the EU’s Big Data Value PPP programme.

DataBio has been driven by the development, use and evaluation of 27 pilots, where also associated partners and additional stakeholders have been involved. The selected pilot concepts have been transformed into pilot implementations utilizing co-innovative methods and tools. Through intensive matchmaking with the technology partners in DataBio, the pilots have selected and utilized market-ready or near market-ready ICT, Big Data and Earth Observation methods, technologies, tools, datasets and services, mainly provided by the partners within DataBio, in order to offer added-value services in their domain.

Based on the developed technologies and the pilot results, new solutions and new business opportunities are emerging. DataBio has organized a series of stakeholder events, hackathons and trainings to support result take-up and to enable developers outside the consortium to design and develop new tools, services and applications based on the DataBio results.
The DataBio consortium is listed in Table 1. For more information about the project, please visit [www.databio.eu](http://www.databio.eu).

### Table 1: The DataBio consortium partners

<table>
<thead>
<tr>
<th>Number</th>
<th>Name</th>
<th>Short name</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (CO)</td>
<td>INTRASOFT INTERNATIONAL SA</td>
<td>INTRASOFT</td>
<td>Belgium</td>
</tr>
<tr>
<td>2</td>
<td>LESPROJEKT SLUZBY SRO</td>
<td>LESPRO</td>
<td>Czech Republic</td>
</tr>
<tr>
<td>3</td>
<td>ZAPADOCESKA UNIVERZITA V PLZNI</td>
<td>UWB</td>
<td>Czech Republic</td>
</tr>
<tr>
<td>4</td>
<td>FRAUNHOFER GESELLSCHAFT ZUR FOERDERUNG DER ANGEWANDTEN FORSCHUNG E.V.</td>
<td>Fraunhofer</td>
<td>Germany</td>
</tr>
<tr>
<td>5</td>
<td>ATOS SPAIN SA</td>
<td>ATOS</td>
<td>Spain</td>
</tr>
<tr>
<td>6¹</td>
<td>STIFTELSEN SINTEF</td>
<td>SINTEF ICT</td>
<td>Norway</td>
</tr>
<tr>
<td>7</td>
<td>SPACEBEL SA</td>
<td>SPACEBEL</td>
<td>Belgium</td>
</tr>
<tr>
<td>8</td>
<td>VLAAMSE INSTELLING VOOR TECHNOLOGISCH ONDERZOEK N.V.</td>
<td>VITO</td>
<td>Belgium</td>
</tr>
<tr>
<td>9</td>
<td>INSTYTUT CHEMII BIOORGANICZNEJ POLSKIEJ AKADEMII NAUK</td>
<td>PSNC</td>
<td>Poland</td>
</tr>
<tr>
<td>10</td>
<td>CIAOTECH Srl</td>
<td>CiaoT</td>
<td>Italy</td>
</tr>
<tr>
<td>11</td>
<td>EMPRESA DE TRANSFORMACION AGRARIA SA</td>
<td>TRAGSA</td>
<td>Spain</td>
</tr>
<tr>
<td>12</td>
<td>INSTITUT FUR ANGEWANDE INFORMATIK (INFAI) EV</td>
<td>INFAI</td>
<td>Germany</td>
</tr>
<tr>
<td>13</td>
<td>NEUROPUBLIC AE PLIROFORIKIS &amp; EPIKOINIONION</td>
<td>NP</td>
<td>Greece</td>
</tr>
<tr>
<td>14</td>
<td>Ústav pro hospodářskou úpravu lesů Brandýs nad Labem</td>
<td>UHUL FMI</td>
<td>Czech Republic</td>
</tr>
<tr>
<td>15</td>
<td>INNOVATION ENGINEERING SRL</td>
<td>InnoE</td>
<td>Italy</td>
</tr>
<tr>
<td>16</td>
<td>Teknologian tutkimuskeskus VTT Oy</td>
<td>VTT</td>
<td>Finland</td>
</tr>
<tr>
<td>17</td>
<td>SINTEF FISKERI OG HAVBRUK AS</td>
<td>SINTEF Fishery</td>
<td>Norway</td>
</tr>
<tr>
<td>18</td>
<td>SUOMEN METSAKESKUS-FINLANDS SKOGSCENTRAL</td>
<td>METSAK</td>
<td>Finland</td>
</tr>
<tr>
<td>19</td>
<td>IBM ISRAEL - SCIENCE AND TECHNOLOGY LTD</td>
<td>IBM</td>
<td>Israel</td>
</tr>
<tr>
<td>20</td>
<td>WUUDIS SOLUTIONS OY²</td>
<td>MHGS</td>
<td>Finland</td>
</tr>
<tr>
<td>21</td>
<td>NB ADVIES BV</td>
<td>NB Advies</td>
<td>Netherlands</td>
</tr>
<tr>
<td>22</td>
<td>CONSIGLIO PER LA RICERCA IN AGRICOLTURA E L'ANALISI DELL'ECONOMIA AGRARIA</td>
<td>CREA</td>
<td>Italy</td>
</tr>
<tr>
<td>23</td>
<td>FUNDACION AZTI - AZTI FUNDAZIOA</td>
<td>AZTI</td>
<td>Spain</td>
</tr>
<tr>
<td>24</td>
<td>KINGS BAY AS</td>
<td>KingsBay</td>
<td>Norway</td>
</tr>
<tr>
<td>25</td>
<td>EROS AS</td>
<td>Eros</td>
<td>Norway</td>
</tr>
<tr>
<td>26</td>
<td>ERVIK &amp; SAEVIK AS</td>
<td>ESAS</td>
<td>Norway</td>
</tr>
<tr>
<td>27</td>
<td>LIEGRUPPEN FISKERI AS</td>
<td>LiegFi</td>
<td>Norway</td>
</tr>
<tr>
<td>28</td>
<td>E-GEOS SPA</td>
<td>e-geos</td>
<td>Italy</td>
</tr>
<tr>
<td>29</td>
<td>DANMARKS TEKNISKE UNIVERSITET</td>
<td>DTU</td>
<td>Denmark</td>
</tr>
<tr>
<td>30</td>
<td>FEDERUNACOMA SRL UNIPERSONALE</td>
<td>Federu</td>
<td>Italy</td>
</tr>
</tbody>
</table>

¹ Replaced by partner 49 as of 1/1/2018.
² Formerly MHG SYSTEMS OY. Terminated on 27/9/2019.
1.2 Document Scope
This document provides the results of the final business planning of DataBio partners and pilots. It includes business analyses grouped per sector (agriculture, forestry, fishery). This report is intended for the DataBio partners, the European Commission and other parties, including the general public.

1.3 Document Structure
The document is comprised of the following chapters:

**Chapter 1** presents an introduction to the project and the document.

**Chapter 2** provides a short description of the analysis approach applied in the deliverable.

**Chapter 3** contains the business analysis of agricultural pilots.
Chapter 4 contains the business analysis of forestry pilots.

Chapter 5 contains the business analysis of fishery pilots.

Chapter 6 lists the document references.
2 Analysis Approach

The objective of this deliverable is to provide an analysis of the DataBio pilots and results from a business perspective. At the beginning of the analysis the focus was on the following questions:

1. What are the business perspectives of the technology used in the DataBio pilots?
2. Does the DataBio technology and Big Data Technology (BDT) in general provide added value to end users of the three bioindustries considered in the project: agriculture, forestry and fisheries?

The organisations offering BDT or components for BDT for agriculture, forestry and fisheries can be classified as follows:

- Companies that are already on the market and for which offering BDT is either core business or a business part of their portfolio
- State-owned research institutions that are non-profit organisations

Even though both types of organisations are providing BDT to end-users in bio-industries, the organisations in the second category are providing the services without pursuing commercial goals. For these organisations, higher societal goals such as preservation of natural resources (i.e. irrigation water, decrease of use of fertilizers, assessment of forest health, ...) might have priority against commercial goals (e.g. TRAGSA, VITO and others). Commercial companies offering BDT and services (e.g. NP, e-geos, SPACEBEL and others) and/or components are already, even before the project started, on the market either with commercial business models or by providing their components as open source (e.g. IBM). In DataBio rather the combination of these commercial BDT is of interest and not the single business model of each technology provider. The main focus of the business analysis was thus on the question if new pipelines of BDT can provide added value to end-users.

The analysis of each pilot followed, where applicable and possible, the structure described in Table 2.

Table 2: Overview of Analysis Aspects

<table>
<thead>
<tr>
<th>Aspect of Analysis</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Introduction, Motivation, Objectives</td>
<td>First a short introduction regarding the pilot is given together with a description of the motivation for it and its goals</td>
</tr>
<tr>
<td>2. Pilot Baseline</td>
<td>Description of the pilot initial set-up</td>
</tr>
<tr>
<td>3. Technology used, with following sub-topics: “Technology pipeline”, “Data used”</td>
<td>Summary of the technology pipeline on high level describing data collection, processing and visualisation. Furthermore, description of the data that was used and finally reflection, if the data pipeline and</td>
</tr>
<tr>
<td>Section</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>4. Business value and impact</td>
<td>Description and where possible calculation of the business value of the pilot</td>
</tr>
<tr>
<td>5. How to guidelines for practice</td>
<td>Descriptions of guidelines and experiences from the pilot that can be used by interested end-users that want to apply BDT</td>
</tr>
<tr>
<td>6. Summary and Conclusion</td>
<td>Summary and conclusion</td>
</tr>
</tbody>
</table>

Overall, the pilot analysis provides an overview on experiences with the technology and its business impact for end-users.
3 Business Analysis of Agriculture Pilots

3.1 Introduction

The agricultural sector is of strategic importance for the European society and economy. It contains a broad spectrum of industries that at present are facing a series of challenges that affect their production, productivity and profitability. Examples of these challenges on the one hand are crop pests and diseases with increasing resistance, drastic changes due to effects of climate change, and decreasing availability of certain resources such as irrigation water. On the other hand, the fast-growing world population increases the demand for food. To cope with these challenges, new innovative approaches in agriculture are necessary. DataBio pilots aim to provide a contribution in agriculture and focus on the following innovative developments in agriculture:

- Precision agriculture in: a) olives, fruits and grapes; b) vegetable seed crops; c) vegetables (potatoes) – (3 pilots)
- Management in greenhouse eco-system – (1 pilot)
- Cereal and biomass crops – (4 pilots)
- Smart Machinery Management – (1 pilot)
- Insurance in agriculture - (2 pilots)

The objective of the pilots is to illustrate through different usage scenarios that BDT has the potential to result in new business models and/or optimised operational processes when applied in agriculture. The respective KPIs to measure the added value can be classified in the following basic categories:

- **KPIs reflecting the use of resources**: Examples of this type of KPIs are fertilizer consumption, use of irrigation water, working hours spent on paperwork.
- **KPIs reflecting the increase of agriculture outcome**: increase of harvested quantity per field, revenues, market share.
- **Efficiency, productivity and profitability KPIs** calculated by comparing use of resources and resulting outcome.

In total there are 13 agricultural DataBio pilots. In six of them the main BDT providers are companies and in the remaining seven pilots this are independent or state-owned research institutions. The pilots, where companies are the main BDT providers, build upon their existing offerings or research and development activities that are extended and verified in the pilots. Table 3 contains a list of the pilots analysed in this chapter:
Table 3: Overview of Agricultural pilots

<table>
<thead>
<tr>
<th>Task (topic)</th>
<th>Subtask</th>
<th>Pilot group</th>
<th>Pilot</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>T1.2 (A)</strong> Precision Horticulture including vine and olives</td>
<td>T1.2.1</td>
<td>A1: Precision agriculture in olives, fruits, grapes and vegetables</td>
<td>A1.1: Precision agriculture in olives, fruits, grapes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A1.2: Precision agriculture in vegetable seed crops</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A1.3: Precision agriculture in vegetables -2 (Potatoes)</td>
</tr>
<tr>
<td></td>
<td>T1.2.2</td>
<td>A2: Big Data management in greenhouse eco-systems</td>
<td>A2.1: Big Data management in greenhouse eco-systems</td>
</tr>
<tr>
<td><strong>T1.3 (B)</strong> Arable Precision Farming</td>
<td>T1.3.1</td>
<td>B1: Cereals and biomass crops</td>
<td>B1.1: Cereals and biomass crops</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B1.2: Cereals and biomass and cotton crops 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B1.3: Cereals and biomass crops 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B1.4: Cereals and biomass crops 4</td>
</tr>
<tr>
<td></td>
<td>T1.3.2</td>
<td>B2: Machinery management</td>
<td>B2.1: Machinery management</td>
</tr>
<tr>
<td><strong>T1.4 (C)</strong> Subsidies and insurance</td>
<td>T1.4.1</td>
<td>C1: Insurance</td>
<td>C1.1: Insurance (Greece)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>C1.2: Farm Weather Insurance Assessment</td>
</tr>
<tr>
<td></td>
<td>T1.4.2</td>
<td>C2: CAP support</td>
<td>C2.1: CAP Support</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>C2.2: CAP Support (Greece)</td>
</tr>
</tbody>
</table>

3.2 Pilot A1.1: Precision agriculture in olives, fruits, grapes, and Pilot B1.2: Cereals, biomass and cotton crops

This section contains the business analysis of pilots A1.1 and B1.2. Both pilots are considered together as they are provided by the same team and technological partners and are based on the same technological Big Data pipeline that has been adjusted to different application scenarios.

3.2.1 Introduction, motivation and goals of the pilots

Pilots 1 and 6 focus on the development and provision of smart farming services for the production of olives, peaches grapes (pilot A1.1) and arable crops (Pilot B1.2). These services
aim at optimizing agricultural production while at the same time minimizing environmental impact by reducing the use of inputs (natural resources such as water and agrochemicals such as fertilisers). These services provide advice for the fertilization, irrigation and crop protection, adapted to the specific needs of each crop in each area participating in the pilots.

The solutions developed in these pilots are based on a unique combination of technologies such as Earth Observation (EO), Big Data analytics and Internet of Things (IoT) with heterogeneous data including EO data, atmospheric and soil data, facts and scientific knowledge.

In the pilot sites, NP was leading the activities of the pilots, supported by GAIA EPICHEIREIN (business partner), IBM (only in pilot A1.1) and FRAUNHOFER (technology providers) for the execution of their full lifecycle. By the end of the project, a set of validated fully operational smart farming services were developed, adapted for each crop and for the microclimatic and conditions of each area.

3.2.2 Pilots set-up
pilot A1.1 worked with three (3) different crops in three (3) different areas offering a set of services including irrigation, fertilization and crop protection against pests and diseases:

1. **Chalkidiki (Northern Greece)**, where the pilot worked with olive groves of 600 ha for the production of table olives
2. **Stimagka (Southern Greece)**, where the pilot worked with vineyards of 3,000 ha for the production of table grapes
3. **Veria (Northern Greece)**, where the pilot worked with peach orchards covering an area of 10,000 ha.

Pilot B1.2 worked with one (1) crop in one (1) site offering irrigation advisory services:

4. **Kileler (Thessaly)**, where the pilot worked with cotton of 5000 ha

Table 4 provides an overview of the Big Data driven smart services deployed at the four sites:

<table>
<thead>
<tr>
<th>Service</th>
<th>Pilot A1.1 Locations</th>
<th>Pilot B1.2 Location</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Chalkidiki (Olives)</td>
<td>Veria (Peaches)</td>
</tr>
<tr>
<td>Irrigation</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Fertilization</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Crop Protection</td>
<td>+</td>
<td>Exploitation of scientific models for 1 pest and 1 disease</td>
</tr>
</tbody>
</table>
The Big Data based smart advices provided by the technology applied were dedicated to farmers as end users that are supposed to implement the advices. However, the processed decision-relevant data and indices are too complicated for farmers. Thus, and to assure also higher credibility and trust into the results, agronomist and agricultural advisors were involved as intermediaries among technology providers and farmers to interpret the data and to propose concrete decisions for the farmers. The involvement of experts that are trusted by the farmers increases the overall trust in the technology.

The goal in all the cases was to enable better management of production, optimized use of inputs and better yields, through the development of data-driven smart farming services. The specific goals for pilot A1.1 can be summarized as follows:

1. Reduction in the average cost per spraying per hectare for the three crop types by following the advisory services at a given period
2. Reduction in the average number of unnecessary sprays per farm for the three crop types by following the advisory services at a given period
3. Reduction in the average cost of irrigation per hectare for the three crop types following the advisory services at a given period
4. Reduction in the amount of fresh water used per hectare by following the advisory services at a given period
5. Reduction in nitrogen use per hectare following the advisory services at a given period
6. Quantify the % of divergence in the cost of the applied fertilization strategy compared to best practices per hectare (agronomist advice)
7. Increase in production

Pilot B1.2 pursued in a similar way the goals 3 and 4 of pilot A1.1.

3.2.3 Technology used
In this section the necessary investment by the technology provider and the technology end-user is assessed based on the pipeline used by the pilot.

3.2.3.1 Technology pipeline
The technology pipeline of the technical solution applied in pilot A1.1 and Pilot B1.2 consists on a high level of abstraction of data collection, data processing and data visualisation components).

Data collection: To provide advices related to irrigation, fertilization and crop protection a set of heterogeneous data is required, capturing critical parameters for crop status monitoring in different spatial and temporal resolutions and namely: climate, soil and plant related data, crowdsourced samples, observations and information for the applied farming practices, intra-field – inter-field EO-based vegetation indices. Historical data from at least one cultivating period is required for calibrating/fine-tuning the scientific models that constitute the backbone of the GAIA SmartFarm advisory services for irrigation, fertilization, crop protection. To collect all this data several data collection modules are used:
• In-situ telemetric stations provided by NP, so called GAIAtrons, that collect data related to weather, soil and plant,
• Modules for the collection, pre-processing of earth observation data, the extraction of higher-level data and assignment of vegetation indices at parcel level,
• Android apps for crowdsourcing data from farmers, agricultural advisors and agronomists about the field status and the applied farming practices,
• Specific web-based user interfaces and for collecting and updating the available farm data.

Data processing: The collected data was processed in different combination through several complementary data processing components (both pilot components and DataBio components) provided by different partners. The DataBio components that supported data processing are the following:

• **GAIABus DataSmart Real-time streaming Subcomponent (NP):** This component allows for: the real-time data stream monitoring resulting from NP’s telemetric stations (GAIAtrons) installed in all pilot sites; the real-time validation of data and the real-time parsing and cross-checking.

• **Proton (IBM):** an early warning system for the management of pests and diseases using temporal reasoning for olives, grapes and peaches (used only in pilot A1.1). It exploits the numerical output (risk indicator) of NP’s SmartFarm services that make use of crop and area-tailored scientific models for pest/disease breakouts.

• **Georocket, Geotoolbox, SmartVis3D (FRAUNHOFER):** This component has a dual role: It is a back-end system for Big Data preparation, handling fast querying and spatial aggregations of data, as well as a front-end application for interactive data visualization and analytics.

Data visualisation and presentation: After the data is processed it needs to be provided in an understandable and decision relevant way suitable also for end users. The main component in this category is NeuroCode (NP). Neurocode allows the creation of the main pilot UIs in order to be used by the end-users (farmers, agronomists) and offers smart farming services for optimal decision making. An additional DataBio component providing information visualization functionalities is Georocket (FRAUNHOFER).

The in-situ telemetric stations together with the earth observation and weather information is part of an integrated solution provided by NEUROPUBLIC and GAIA EPICHEIREIN. It provided the starting point for the final solution used by the pilots that was enhanced with complementary components during the project.

### 3.2.3.2 Data used

The specific pilot makes use of four (4) different data types:

• **Sensor measurements (numerical data) and metadata (timestamps, sensor id, etc.):** This dataset is composed of measurements from NP’s telemetric IoT agrometeorological stations (GAIATrons) for the pilot sites. More than 20 GAIATrons were
fully operational at the pilot sites, collecting > 30MBs of data per year each with current configuration (measurements every 10 minutes).

- **EO products in raster format and metadata**: This dataset is comprised of ESA’s remote sensing data from the Sentinel-2 optical products (6 tiles). High volumes of satellite data were processed in order to extract the necessary information about the crops in each parcel participating in the pilot.

### 3.2.4 Business value and impact

Both pilots managed to achieve the expected results for input cost reduction. Aggregated findings can be found at the following figures:

**Figure 1: Pilot A1.1 aggregated findings**

![Figure 1: Pilot A1.1 aggregated findings](image)

**Figure 2: Aggregated results of Pilot B1.2 in comparison with the target values**

![Figure 2: Aggregated results of Pilot B1.2 in comparison with the target values](image)

This was achieved as farmers and agricultural advisors showed a collaborative spirit and followed the advice generated by DataBio’s solutions.

The defined goals were achieved, and also validated in more detail by a set of pilot KPIs which were met in their majority, and in some cases even exceeded the targeted values (as documented in D1.3 [REF-01]). Table 5 summarizes the measured savings of the pilots per hectare:

---

Dissemination level: PU · Public
### Table 5: Quantification of business gains (baseline – pilot value) in pilot A1.1 and Pilot B1.2

<table>
<thead>
<tr>
<th>Saving</th>
<th>Pilot A1.1</th>
<th>Pilot B1.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduction of the average cost of spraying per hectare</td>
<td>250 – 219 = 31 Euro / Hectare</td>
<td>990 – 963 = 27 Euro / Hectare</td>
</tr>
<tr>
<td></td>
<td>250 – 219 = 31 Euro / Hectare</td>
<td>990 – 963 = 27 Euro / Hectare</td>
</tr>
<tr>
<td>Reduction of the average number of unnecessary sprays per farm</td>
<td>5 – 1.4 = 3.6 Number of sprays</td>
<td>4 – 1.8 = 2.2 Number of sprays</td>
</tr>
<tr>
<td></td>
<td>5 – 1.4 = 3.6 Number of sprays</td>
<td>4 – 1.6 = 2.4 Numbers of sprays</td>
</tr>
<tr>
<td>Reduction of the average cost of irrigation per hectare</td>
<td>330 – 198 = 132 Euro / Hectare</td>
<td>3030 – 2007 = 1023 Euro / Hectare</td>
</tr>
<tr>
<td>Reduction of the amount of fresh water used per hectare</td>
<td>817 – 492.4 = 324.6 m³ / Hectare</td>
<td>1868 – 1232 = 636 m³ / Hectare</td>
</tr>
<tr>
<td></td>
<td>817 – 492.4 = 324.6 m³ / Hectare</td>
<td>1868 – 1232 = 636 m³ / Hectare</td>
</tr>
<tr>
<td>Reduction of the nitrogen use per hectare</td>
<td>230 – 161 = 69 Kg / Hectare</td>
<td>220 – 161 = 59 Kg / Hectare</td>
</tr>
<tr>
<td></td>
<td>230 – 161 = 69 Kg / Hectare</td>
<td>220 – 161 = 59 Kg / Hectare</td>
</tr>
<tr>
<td>Quantify % divergence in the cost of the applied fertilization</td>
<td>-40 + (-11.27) = 51.27 % / Hectare</td>
<td>20 – 44 = -24 % / Hectare</td>
</tr>
<tr>
<td></td>
<td>-40 + (-11.27) = 51.27 % / Hectare</td>
<td>20 – 44 = -24 % / Hectare</td>
</tr>
<tr>
<td>Increase in production</td>
<td>10375 – 7010 = 3365 Kg / Hectare</td>
<td>17117 – 18011 = -894 Kg / Hectare</td>
</tr>
<tr>
<td></td>
<td>10375 – 7010 = 3365 Kg / Hectare</td>
<td>17117 – 18011 = -894 Kg / Hectare</td>
</tr>
<tr>
<td>Decrease in inputs focused on irrigation</td>
<td>2670 – 1881 = 789 Kg / Hectare</td>
<td>2670 – 1881 = 789 Kg / Hectare</td>
</tr>
<tr>
<td></td>
<td>2670 – 1881 = 789 Kg / Hectare</td>
<td>2670 – 1881 = 789 Kg / Hectare</td>
</tr>
</tbody>
</table>

It is evident that the business impact of the pilot would be further validated and reach more conclusive insights as KPI measurements from more (and different) cultivating periods get aggregated over the years. More trials would allow to get more business related KPI measurements maximizing the pilot’s impact.

The achieved results allow for the following conclusions regarding the business impact:

- The findings show that technology use results in concrete financial savings per hectare for all considered crop types and regions. As different crop types have different input necessities from an agronomical point of view, the technology used results in different % savings. Scalability and transferability of the technology in different crop types/regions is apparent, as a new set-up would require gathering data for calibration/fine-tuning of the scientific models for irrigation, fertilization and crop
protections of an acceptable amount of time (one cultivating period) prior to producing initial advices to the farmers.

- The findings also show that it was possible to achieve the results because the farmers were cooperative and acted according to the advices proposed by the technology.

Given the above savings a financial gain due to lower costs will be achieved when the price pro service paid by the farmer is lower than the expected savings. For example: in the case of olive trees in Chalcidiki there would be a gain for the single farmer per hectare only if the service costs less than the expected gains. This means that the final gain for the farmer can be calculated based on the following formula:

Financial Gain farmer = Sum of all savings for each combination of crop and location – Sum of cost for the service

From the perspective of the technology providers there will be a financial gain from offering the service only if the income from selling the service covers the cost for providing the service pro hectare.

Financial Gain technology provider = Sum of income pro hectare – Sum of cost for providing the service

Besides these direct costs for the service from the farmers and technology providers point of view resulting from the operation of the technology solution, there are other gains that can be quantified and add to the solution:

- By reducing the number of sprays the farmer increases the productivity of spraying and saves time that he or she can invest in other value creating activities. This also means that the cost for labour decreases as well.
- Further gains can be achieved also by increasing the harvest from the field supported by the technology. Even though this might be difficult to measure, because at the end the quality and quantity of the harvest might depend on many factors than the ones controlled by the technology. However, the more factors influencing the growth and quality of the plants can be controlled by technology, the higher the output in terms of quantity and quality should be.

As multiple parameters (climate and crop type related) are affecting the agricultural production, it became clear that a “one-fits-all” solution is not applicable and several factors need to be taken into consideration in translating the trial results (e.g. biennial bearing phenomenon in olive trees, heavy seasonal/regional rains, multi-year fertilization strategies, etc.).

3.2.5 How-to guidelines for practice

Farmers are constantly struggling to produce more food, in order to meet the increased global demand. At the same time, there is a push towards more sustainable farming practices in order to minimize the environmental impact of agriculture. In this context, the future Common Agricultural Policy (CAP) (which is currently under development) puts its focus on digitization, inviting farmers to produce “more with less”.

Dissemination level: PU -Public
In order to improve farm productivity and increase their profits, farmers were traditionally asked to invest in expensive technological tools and learn how to use them - an offer usually combined with the use of specific brands of agrochemicals. This not only incurred high costs for farmers with a slow depreciation curve (in fact a high percentage of farmers – Greek farmers are in their majority smallholders - did not have the capacity to make such investments), but also required farmers to have digital skills that they lacked.

In order to support the business expansion of BDT that are introduced within the present DataBio pilot, NP and GAIA EPICHEIREIN have already established an innovative business model that allows a swift market uptake - the “Smart-Farming-As-A-Service” model. With no upfront infrastructure investment costs and a subscription fee proportionate to a parcel’s size and crop type, each smallholder farmer can now easily participate and benefit from the provisioned advisory services. The proposed approach takes all the complexity out of the picture, and provides a simple and easy to use advice that both agricultural advisors and farmers can exploit.

Moreover, and as more than 70 agricultural cooperatives are shareholders of GAIA EPICHEIREIN, it is evident that there is a clear face to the market and a great liaison with end-user communities for introducing the pilot innovations and promoting the commercial adoption of the DataBio’s technologies.

Finally, while the proposed data-driven solution of the pilot is appealing to smallholder farmers, it is also applicable to large farms and agricultural cooperatives. Thanks to their increased capacity (e.g. financial and technical), the application of the smart farming services can multiply the benefits for these organizations, as they are applied in larger scale.

3.2.6 Summary and outlook

NP and GAIA EPICHEIREIN have already launched in 2013 their Smart Farming program, called “gaiasense” (http://www.gaiasense.gr/en/gaiasense-smart-farming), which aims to establish a nationwide network of telemetric stations with agri-sensors and use the data to create a wide range of smart farming services for agricultural professionals.

Within the DataBio the quality of the provided services greatly benefited from the collaboration with leading technological partners like IBM and FRAUNHOFER, that specialize in the analysis of Big Data. Moreover, feedback from the end-users and lessons-learnt from the pilot execution significantly fine-tuned and will continue to shape the suite of dedicated tools and services, thus, facilitating the penetration of “gaiasense” in the Greek agri-food sector.

The success of the pilot was established by high profile events [REF-04] and online articles [REF-05] that were promoting the findings of the pilot and consequently the wider adoption of big-data enabled smart farming advisory services in the next years.
3.3 Pilot A1.2: Precision agriculture in vegetable seed crops

3.3.1 Introduction, motivation and goals of the pilot

Eastern Italy is by tradition one of the best areas in the world for seed production. The leader of this pilot, Cooperativa Agricola Cesenate (CAC) is one of the major producers of seeds in the region, joining over 2000 growers from several growing areas along the Adriatic coast of Italy. CAC, through its associated growers, produces “on contract” seed of vegetables, sugar beets, alfa-alfa and many other species for seed companies from all over the world that import the seed and distribute it on the global market.

One of the key factors for obtaining seeds of good quality is the choice of the right harvest time: if harvested too early the vigour of the seeds will be affected; if harvested too late the mature seeds are going to drop to the ground and the best part of the harvest gets lost.

The main focus of this pilot was on monitoring the maturity of seed crops of different species with satellite imagery from ESA’s Sentinel constellation. The objective was to set up a model - combining field observations, satellite and meteorological data - which could predict with reasonable accuracy the optimal date of harvest for each field. This information should allow the CAC fieldsmen who are monitoring the seeds crops to organize their field visits more efficiently, and to plan the seed harvesting operations in an optimized way together with the growers. More precisely, it was expected that the above-described BDT would provide measurable savings in:

- The number of km driven by CAC fieldsmen to the fields
- The number of farms or ha that can be observed at the same time by a single CAC fieldsmen.

Overall, the technology leads to an improvement of the productivity of CAC fieldsmen. Further gains are expected from better estimation of the optimal harvest date and increased efficiency in warehouse and logistics planning.

To this aim, selected seed crops were monitored on-site by the technical personnel of CAC during all stages of growing till maturity and harvest. VITO, an independent research institute and satellite service provider and partner in this pilot, provided Sentinel-1 and -2 satellite images as well as an online platform “WatchITgrow” for crop monitoring and field data registration and developed a model for harvest date estimation.

3.3.2 Pilot set-up

The first trials were initiated in 2017 for sugar beets, with the aim to expand – upon the results achieved – the trials to other seed crops in the next years. In 2017, five sugar beet fields (15 hectares in total) located in the Region Emilia Romagna were selected by CAC. The fields were monitored with WatchITgrow. Sentinel-2 greenness maps and graphs were made available for the fields monitored by CAC.

The “greenness” is actually referring to the fraction of Absorbed Photosynthetically Active Radiation (fAPAR), a measure of the crop’s primary productivity. fAPAR is often used as an
indicator of the state and evolution of crop cover. Low fAPAR values indicate that there is no crop growing on the field (bare soil, fAPAR=0). When the crop emerges, the index will increase until the crop has reached the maximum growing activity (fAPAR=95-100%). Then the fAPAR will decrease again until harvest.

Greenness (or fAPAR) maps can be used to detect variability within a field or between fields. The causes may be diverse and can range from (natural) soil heterogeneity to climate-induced problems such as drought or water logging, or local damages due to pests or diseases, emergence problems, etc. Greenness graphs show the fAPAR evolution throughout the season (field average fAPAR). From this “crop growth curve” information on phenology and crop development is retrieved and a model was designed to decide on the right moment for harvesting.

The results of the first trial were promising:

- Differences in maturity between sugar beet fields and variability within individual fields were well visible from satellite greenness maps.
- Analysis of the growth curve (i.e. greenness index evolution during the season) and discussions with the fieldsmen made CAC seeds and VITO confident that the greenness index can be used to check when the sugar beet seeds are ready to be harvested.

Based on these results for sugar beets it was decided to extend the EO and the in-situ monitoring in 2018 to a larger number of sugar beet seed production fields and to include new seed crops into the trial. Among these some vegetable crops (cabbage and onion) and more field crops (alfa-alfa, sunflower, soybean).

In 2018, in total 61 sugar beet fields, 5 onion, 5 cabbage, 16 sunflower, 3 alfalfa and 2 soybean fields were monitored. In monitoring these fields, especially onion and cabbage, which are early maturing, problems were encountered due to the unstable weather conditions in the production areas in late spring and early summer. The high number of cloudy days prevented the fieldsmen to have access to the Sentinel-2 satellite images during their field checks. So, their reports and checks were not influenced by the satellite data.

The good results for sugar beets that were obtained in 2017 were confirmed in 2018. A stable correlation was found between the satellite derived greenness index and the optimal harvest date. Based on this correlation a «maturity model» was developed and integrated in the WatchITgrow application.

For cabbage, the greenness index appeared to be much more difficult to match with the harvest dates indicated by the fieldsmen. The greenness curve gets its peak during the winter and decreases until starting of blooming in the spring. Besides, the shape of the greenness curve differed between fields and geographical areas. Probably the greenness index is affected by plant density – for which we have different recommendations according to the variety – and by differences in the ratio between female and male lines (note that the male lines are destroyed after the flowering). For the monitored onion fields the greenness curves also showed a high heterogeneity with respect to the harvest dates decided by fieldsmen. For
the above reasons the trials for these two species, cabbage and onion, were not continued in 2019.

The greenness curves of the sunflower fields looked more reliable. The index was closely following the growth of the plants and the tendency was replicated in all the fields monitored. Two fields of soybeans were introduced in the trial, as they were close to the monitored fields of sugar beet and sunflower. The resulting greenness curves appeared to be quite reliable and it was decided to monitor this crop at a larger scale in 2019 and set up a model for estimating the optimal harvest date according to the greenness index.

Finally, in 2019 a large-scale trial was set up which included about 250 hectares of sugar beets (77 fields) and 600 ha of soybeans (41 fields). Greenness maps, growth curves and harvest date estimates were made available in near real time via WatchITgrow. In addition, a few sunflower fields were monitored to be able to finetune the maturity model for this crop.

Optical satellites such as Sentinel-2 do not provide useful observations in cloudy conditions. When cloud free observations are lacking for several weeks interpolation or smoothing techniques cannot bring a solution anymore. The CropSAR technology developed by VITO, which combines Sentinel-2 optical observations with Sentinel-1 radar observations, provides a way to keep on monitoring crop growth and development, independent of weather conditions. This technology was applied at the pilot test site during the 2019 season.

From the trials in 2019 it was found that:

- It is possible to estimate the harvest date from the satellite derived greenness curve with moderate to high accuracy for sugar beets and soybeans.
- The accuracy of the harvest date estimation significantly increases when the CropSAR derived greenness index was used as input, especially in cloudy periods.
- The maturity model that is currently used to «forecast» the harvest date (simple linear approach) could be further improved by using more advanced modelling techniques such as machine learning to predict greenness index values, or by using meteo data (rainfall, temperature) as additional input for maturity modelling.

### 3.3.3 Technology used

#### 3.3.3.1 Technology pipeline

The crop monitoring and maturity assessment process uses Sentinel-2 satellite images as input for generating greenness maps and graphs.

The pilot used the “Proba-V MEP” EO component, managed by VITO, for satellite data processing, analysis and visualization. In this pilot Sentinel-2 (and later also Sentinel-1) satellite data were used for crop monitoring and harvest date estimation.

The frontend of VITO’s WatchITgrow web application was used and adapted following the requirements of this pilot, to display the satellite derived information as processed on the Proba-V MEP. The backend services for maturity assessment that were developed in the frame of this pilot were fully integrated in WatchITgrow.
Finally, a data fusion algorithm (CropSAR) was applied to the trial region to cope with cloudy images inherent to the use of Sentinel-2 satellite data. This data fusion uses both Sentinel-1 and Sentinel-2 data to provide cloud-free time series.

3.3.3.2 Data used
Crop monitoring using Sentinel-2 satellite data is a common practice nowadays. The images provide information with a resolution of 10m, which is in most cases sufficient to detect intra-field variability. Thanks to its wide coverage, Sentinel-2 satellite data also allow comparison of different fields. Although in theory Sentinel-2 images are available every 5 days, in practice observations may be lacking for longer periods due to persistent cloud cover. To overcome this problem VITO applied its CropSAR algorithm which combines Sentinel-2 optical and Sentinel-1 radar data to obtain continuous time series of cloud free observations. The satellite data were made available via the WatchITgrow platform.

In-situ (field) observations were made by CAC fieldsmen. Field location, planting date, development stage, harvest date and germination rates were recorded for the fields and crops included in the trials and were stored in the WatchITgrow database.

Both datasets, the satellite data and the field observations, were then matched and a model was set up for estimating the optimal harvest date of the selected seed crops.

3.3.3.3 Reflection on technology use
As optical satellites are unable to look through clouds, the use of the Sentinel-2 information only has limited accuracy on cloudy days. If clouds persist for several days, the fieldsmen are “blind” and the advantage of the tool fades. The introduction of fused Sentinel-1 and -2 indexes based on optical and radar data can overcome the gap as illustrated for 2019.

In some cases, the 10m resolution of Sentinel-2 turned out to be insufficient. Ditches, side roads or fractions of neighbouring fields disturbed the satellite signal.

Using satellite imagery to estimate the optimal harvest date seems to work fine for sugar beets and soybeans. However, there is a difference between harvest date “estimation”, whereby a full seasonal time series of satellite images is used for a-posteriori determination of the harvest date, and “forecasting”, whereby the harvest date is estimated in near real time from an incomplete time series of satellite images and whereby future weather conditions may influence the actual harvest date. While it was quite straightforward to “estimate” the optimal harvest date for sugar beets and soybeans, “forecasting” this date was found to be much more difficult. More effort needs to be spent on improving the forecasting model. This may be done by using more advanced modelling techniques such as machine learning to predict satellite index values, or by using meteo data (rainfall, temperature) as additional input for maturity modelling. The Proba-V MEP provides access to such data and tools.

Besides continuing with satellite data and adding more data (such as weather data) to the model, it is also important to check with on-site reports the factors which can distort the model parameters.
The usability of the tool also could be further improved in terms of speed and user friendliness. As the fieldsmen are often out of their offices, they need to get the WatchITgrow application adapted to a mobile version with easy access and easy handling.

### 3.3.4 Business value and impact

The goal of this pilot was to provide the fieldsmen with a tool that can predict the harvest date of the major seed crops with sufficient accuracy so they can use this information for organizing more efficiently their visits to the fields and for programming the harvesting operations. Taking into account the maturity stage of each field and the weather forecasts for the days close to harvesting, they can decide when to start the operations and organize the delivery of the harvested seeds to the central warehouse.

Via the online “WatchITgrow” platform provided by VITO, fieldsmen can get more insight in the performance of the crops during the growing period. Some fields may have areas where the greenness parameter shows abnormal values, indicating the presence of limiting factors such as soil characteristics, insufficient irrigation or pests, which require action from the farmer or lead to a separate harvest of the suffering area.

The impact is hard to evaluate. The main advantage which can be expected is an increase of efficiency in the work of fieldsmen. The fieldsmen are visiting periodically the growers and are reporting the status of each field on a database, advising the growers on best practices to get a good yield from their seed crops such as fertilisation, irrigation, pest control. They also organise the harvesting operations and program the combining according to the maturity stage of each field, the weather forecasts and the workflow of the receiving warehouse.

Using a tool which can predict the maturity of each field, they can organise their travels more efficiently and save in the number of drives, the number of kilometres and fuel, and/or they can supervise a wider area. If the use of the tool will be extended to the growers, this efficiency could further increase.

Overall, it was not possible to exactly measure the savings in terms of km driven by the fieldsmen and in terms of number of farms that can be controlled by a fieldsmen. To support the pilot with benchmark data from the fields for comparison with the model data, fieldsmen were actually travelling as usual or even more during the trial. However, the results achieved in the pilot allow for a better estimation of potential improvements of the productivity of fieldsmen: It is expected that the number of km driven by a fieldsmen can be reduced by 15% while at the same time the number of farms that can be observed by a single fieldsmen can increase by 20%.

The platform could be integrated in the database they are currently using to store the data of each field such as planting/sowing dates, time of plant topping, flowering, treatment for preventing diseases, etc. giving further added value to the tool. This will result in very valuable historical data for the fields that can be used for future training of algorithms and as benchmark data.
3.3.5 How-to guidelines for practice

Findings and recommendations from the trials:

- It is possible to estimate the harvest date from the satellite derived greenness curve with moderate to high accuracy for sugar beets and soybeans.
- For other crops (onions, cabbages, etc.) it was not possible to estimate the optimal harvest date in the same way as it is done for sugar beets or soybeans. Different crop characteristics (e.g. low soil cover) or management practices (e.g. male vs female lines and removal of lines during the season) may complicate the analysis of the satellite images.
- For each crop, a dedicated maturity model needs to be set up.
- It generally takes 2 to 3 seasons to set up a reliable model for maturity assessment. Weather conditions may influence the crop’s behaviour and/or the satellite observations.
- Considerable effort needs to be spent on field data collection. Per season data have to be collected for a minimum of 40-50 fields.
- The expertise of the fieldsmen is important when evaluating the satellite-based results e.g. for explaining outliers in the model output.
- As optical satellites are unable to look through clouds, the use of the Sentinel-2 information only has limited accuracy on cloudy days. Algorithms such as CropSAR, which combine Sentinel-2 optical and Sentinel-1 radar data, can be used to overcome this problem and to obtain continuous time series of cloud free observations for crop monitoring and harvest date estimation.
- In some cases, e.g. for small fields, fields surrounded by trees or fields with ditches or roads crossing the field, the 10m resolution of the Sentinel-2 satellite images may not be sufficient and the satellite signal may be disturbed.
- There is a difference between harvest date “estimation” whereby a full seasonal time series of satellite images is used for a-posteriori determination of the harvest date and “forecasting” whereby the harvest date is estimated in near real time from an incomplete time series of satellite images and whereby future weather conditions may influence the actual harvest date. At this stage, we are able to “estimate” the optimal harvest date for sugar beets and soybeans with sufficiently high accuracy, but the model to “forecast” this date needs further development. It could be improved by using more advanced modelling techniques such as machine learning to predict greenness index values, or by using meteo data (rainfall, temperature) as additional input for maturity modelling.

3.3.6 Summary and outlook

The results from the DataBio project have been useful to open the mind of the fieldsmen about the possibility to apply technologies based on satellite imagery to improve the cost effectiveness of the services provided to growers and to customers, gaining a competitive advantage.
The data provided through EO should be integrated with other data such as meteo and sensor data to give farmers and fieldsmen the tools to carry out their activities and to support their decisions in a smarter way.

The involvement in the project was important to get acquainted with the new technologies which are going to enter the market. Growers and fieldsmen are taking knowledge that satellites are not only fit to assist the drive of tractors but that they can also provide useful information for the daily management of the farm. However, technology has a cost and in order to motivate the growers to adopt it the CAP, through Regional Development Plans and incentives, should provide them financial support.

3.4 Pilot A1.3: Precision agriculture in vegetables_2 (Potatoes)

3.4.1 Introduction, motivation and goals of the pilot

Potato has been the major crop in the Netherlands for many years. Due to the reform of the CAP (Europe’s Common Agricultural Policy) the market is changing, and farmers are urged to increase their yields, but in a sustainable way. This means they need to be more conscious of the energy and other resources they use in producing their crops. AVEBE is a cooperative for the potato growing farmers and is supporting their growers in an innovation program called “Towards 20-15-10”, a program which started in 2012. The objectives of this program are to realize in 2020 an average of 15 tons of starch per ha with a variable cost price of €10 per 100 kg of starch. To monitor these objectives farmers are sharing data about their yields and farming practices in study groups.

The goal of this pilot was to support potato farmers with BDT by providing them with information about crop development and expected yields during the growing season. If the farmer knows early that he might not reach the yield goal defined by CAP, he could introduce corrective actions. Thus, to provide the farmers with early insights, NB Advies developed with the help of VITO a system that generates ‘Vigor’ maps for potato growers in the Netherlands. ‘Vigor’ maps are created by using Earth observation and weather data sources combined with field information. These data were collected, stored and processed by an online platform for monitoring and early warning of inhomogeneities in the crop. Yield prediction data was made available in an early stage of the growing season.

The developed online platform provided the farmers more insight by benchmarking their crop during the growth period with crops in the region and/or previous growing seasons. These new insights improve farm management decision making and provide opportunities for timely and more efficient location specific treatment of the crops.

The pilot was able to prove that early insights about the status of potato crops and expected potato yields can be generated based on EO, weather and field data, even though a reliable accuracy of the predictions can only be achieved with sufficient and high-quality historical data.
3.4.2 Pilot set-up

The consortium for the pilot consisted of NB Advies (NL) and VITO (B) in cooperation with AVEBE. The area, where the pilot was executed, is located in the region Veenkoloniën (ca. 51.000 ha) in the North of the Netherlands. This area is characterized by large scale arable farms. In 2007, already 37% of the farms were more than 100 ha in size and this number is growing. In 2019 11 farmers selected one of the fields on their farm to be monitored in the pilot, which resulted in a total of 111 ha that were monitored by the pilot.

For 2 years groups of farmers contributed to the pilot by defining their benchmarking needs. In the first year, a general service based on the ‘WatchItGrow®’ web application was made available to the farmers, providing Fraction of Absorbed Photosynthetically Active Radiation (faPAR) information to the farmers. However, according to the feedback from farmers, the faPAR data was hard to interpret and understand for them. The faPAR maps were useful for showing the inhomogeneity but didn’t provide actionable data. Thus, instead of faPAR in the second year, maps using LAI (Leaf area index) were used. This resulted in a more flexible online platform and more benchmark information for the farmers in the second year. The LAI provided more insight regarding the actual situation of the crops compared to the potential of the field. To inform farmers only when there is a need for their attention, an alert service, i.e. an early warning was established on the platform.

3.4.3 Technology used

3.4.3.1 Technology pipeline

The final result of the pilot is a decision support system (DSS) for potato farmers that is able to provide data about the overall status of the crop and the potential yield based on EO, weather and soil parameters. The DSS involves the following data collection, processing and visualisation technology:

Data Collection: To provide benchmark data for potato crops, five types of data were collected: 1) historical data about crop performance in the past (i.e. emergency date, LAI, greenness, yield development, and actual yield and date of yield; 2) historical data about the field soil (soil texture, soil moisture status and elevation maps); 3) actual data about daily weather (temperature, solar radiation, humidity, precipitation and wind speed); 4) reference values for indexes from literature; and 5) real time EO data (i.e. greenness index) and IOT data (soil moisture status).

Data processing: Data processing involved three steps: 1) calibration and calculation of a crop growth model, 2) real time collection and processing of EO data, 3) benchmarking of the values, i.e. indexes resulting from the growth model and from the analysis of EO data. In the first step, the soil, crop and weather data from field measurements, satellites, weather stations, literature and other sources were collected and, after pre-processing, stored in a database and were used as input in a crop growth model. In order to benchmark crop
performance, the WOFOST\(^3\) crop growth model (FAO) was introduced in the pilot and was calibrated using historical data (2017, 2018) and recent samples. Parallel to the calculation of the growth model, Sentinel-2 data were collected and calculated in real time, providing information about the most recent value of the indexes applied (first faPAR and later LAI). The processing of the EO data involved the following steps (see also D1.3): adjustment of the data with cloud mask and cloud-shadow mask, calculation of a-factor for Weighted Difference Vegetation Index (WDVI), calculation of WDVI from spectral data and calculating LAI for potato fields based on WDVI-LAI correlation data. Finally, in the third step, the model then establishes the benchmark for crop performance: an estimate of the best possible performance under the given set of circumstances.

**Data visualisation:** The DSS is provided through the online platform “WatchItGrow”, i.e. as data as a service for the farmers, in the form of an early warning system that alerts farmers when their attention is needed. The online platform provides crop monitoring and benchmarking services that show the differences, i.e. the variations in the fields. Sentinel-2 satellite images are very helpful for crop monitoring over a large area. But for use in a DSS, it is more useful to show just the field specific information and not the complete images of the whole area.

### 3.4.3.2 Data used

The collected datasets are:

- Sentinel-2 images with an average interval of 5 days.
- Soil characteristics map BOFEK2012\(^4\) spatial dataset for the Netherlands with soil physical units, representing areas of corresponding soil structure and hydrological behaviour.
- Weather data from the Koninklijk Nederlands Meteorologisch Instituut (KNMI) (solar radiation, temperature, precipitation) based on the station closest to the field, example growing season average temperature and daily sum precipitation) measured daily.
- Internet of Things (IoT) soil moisture sensors measured once per hour. In each of the pilot fields soil sensors (IoT) were installed to receive soil moisture data.
- (Historical) yield data related to spatial location of the plot.
- Yield data from experimental station for different varieties.
- Multispectral drone data (for potato-variety specific vegetation index data).
- Field data from farmers (field location, planting data, potato variety, irrigation data).

---


\(^4\) BOFEK 2012: BOFEK is a physical interpretation of the Dutch national Soil Map, scale 1 : 50 000. It is clustering of 315 soil units to 72 soil physical units (based on hydrologic assessments). **BOFEK** provides soil physical characteristics (pF and k(h)) for soil layers per unit.
3.4.3.3 Reflection on the availability and quality of data

The Sentinel-data proved very useful to extract the LAI information. However, during the growing season, there were quite extensive periods (15-20 days), in which no cloud-free images were available. Also, the cloud-shadow gave sometimes disturbing information.

After some difficulties with weather-data services in the first year of the pilot, another KNMI-weather service was implemented, which provided continuous and reliable data.

The IoT soil moisture sensors had some interruption in data communication, especially in the beginning. The signal was improved and proved more reliable after that.

In cooperation with a potato breeding firm also specific crop index data for the varieties of potatoes that are planted by the farmers were collected. The trial fields were monitored by UAV once a month (June, July and August) gathering multi spectral data. By processing the UAV data multiple crop indexes including yield potential were calculated for each plot. The different crop varieties are known to have different phenological development. It was expected that the average crop index values for each variety would show significant differences in crop development between the varieties. There were differences, but they were not significant. This may be due to the weather, which was out of the ordinary in 2019 and might have influenced crop development.

The historical yield data was collected and processed without the spatial location of the sample fields. This made them unusable for correlating it with the historical EO data. Privacy issues raised by the farmers prevented collecting this information afterwards.

The conclusion about availability and quality of data is that there is a lot of data available, but they are not always with a quality suitable for use. When the product is based on third-party service providers, a solid agreement about the availability is necessary. With more demands for service level agreements (SLA) the price of data-services may go up, making it less interesting to use for farmers.

3.4.4 Business value and impact

The goal of this pilot was to establish potato crop monitoring with BDT during the growing season and to provide the potato farmers information about crop development and expected yields. This section summarizes the achieved results.

Crop monitoring: The online platform shows the variability in Leaf Area index (LAI). The LAI represents the area intercepting the solar radiation for crop growth. The pilot showed that maps using LAI are closer to the farmers practice compared to faPAR.

The online platform provided the farmers more insight by benchmarking their crop during the growth period with crops in the region, previous growing seasons, etc. The variability in the field indicated the area which needed attention in the sense of limiting factors, which may be soil characteristics, water, fertilizer or pests. These new insights provide the farmers new opportunities for their farm management decisions for timely and more location specific crop treatment.
Crop monitoring provided valuable information for farmers to inform them about:

- the in-field-variation and the areas requiring inspection as well as site-specific management
- relative performance of their field compared to the surrounding fields
- relative performance of their field compared to the potential
- the need for irrigation (combined with soil moisture data).

The benchmark information was mostly appreciated by the farmers.

The actual added value of the service is hard to tell because there isn’t really a baseline. The farmers were not used to an online crop monitoring system, so the pilot was much about raising awareness about the Big Data approach. The values appreciated most by the farmers were:

- **Field specific information** instead of a general satellite image, which needs to be interpreted by the farmer himself.
- **Alert when new data is available.** Thus, the farmer doesn't have to go and search, even when there is nothing new to find. This gives farmers the opportunity to better allocate and use their time.
- **Crop development benchmark;** farming is a business with a lot of variables, which not all can be controlled by the farmer. Therefore, a well-informed farmer has the advantage to be able to adapt to the circumstances. This benchmark enables farmers to spot problematic fields and areas earlier and to react appropriately to save the crop and yield.
- **Soil Moisture benchmark;** soil moisture is one of the variables which can be monitored in order to plan and manage irrigation facilities accordingly.

Overall, crop monitoring enables farmers to monitor bigger fields and area more efficiently, to allocate their resources in a more productive way to fields and areas that need attention, to identify problematic areas early so that corrective measures can be applied and to apply irrigation in a more efficient way. In sum theses should result in an improvement of the productivity of the farmer and increased harvest. Crop monitoring provides relevant information for farm management and informed decision making.

**Yield improvement:** The crop growth model is a useful tool to show the potential impact of farm management decisions. From literature it is known that the best options for a higher yield in potatoes are created in the spring. In the pilot the effect of later seeding date and subsequently later crop emergence data was tested and simulated. For example, seeding dates varying from April 10th till May 8th resulted in a difference in dry matter in potatoes\(^5\)

\(^5\) Potatoes have a water content that varies from 75 to 85% and the rest is dry matter that contains the nutrients for which potatoes are grown. The feed nutrients are found in the dry matter portion of the potato, not in the moisture portion. Dry matter refers to the material remaining in the potato after removal of water, and the moisture content is the amount of water in the potato.
from 2.9 – 5.3 ton/ha on August 8th. This underlines the known rule that yield improvement is best implemented in spring (please refer for more details to D1.3 [REF-01]).

**Yield prediction:** For the yield prediction the potato growth model needs to be calibrated with historical yield data (for more details see D1.3 [REF-01]). The data for 2017 and 2018 was used to train the system and the data for 2019 was used to test the accuracy of the model.

The findings regarding yield prediction can be summarized as follows:

- In general, the water-limited growth model is under-estimating the yield and the potential compared to the samples for 2019.
- The data available for validation of the WOFOST model proved to be quite limiting the results.
- Only for 2 years data were available for comparison of model data and data from Sentinel-2
- Only 1 year (2018) of field data with location information about the parcel were available
- Weather condition in 2018 and 2019 were quite out of the ordinary
- Yield differences between different varieties influenced the calibration results more than anticipated
- The water limiting effect was quite significant, but soil moisture data about the previous years were not available

Due to limited data availability the algorithm is not sufficiently trained yet for reliable yield predictions.

The potential yield prediction (dry matter) based on the weather data of the last 10 years shows the relative differences between the years, but largely over-estimates the yield at harvest time.

Comparing the model prediction to the actual samples taken in the fields show the same trend for the beginning of July and for harvest-time (mid-September): An over-estimation of the potential and under-estimation of the water limited model calculations for the pilot fields, except for 2018, which was an exceptionally dry year.

The crop growth model proves its benefit for yield prediction purposes, but the accuracy is too limited yet.

**Measurement of goal achievement:** There was also an attempt to measure goal achievement and business impact of the pilot. Major goals of the pilot were the implementation of the 20-15-10⁶ goals that can be summarized as follows:

---

⁶ Avebe project 20-15-10; goals set for 2020.
• **Starch per ha**: From an average of 13.7 tons in 2012 to 15 tons. However, due to unfavourable weather conditions this goal was not achieved and the starch per ha was 5.6 – 11.9 tons per ha.

• **Variable costs per 100 kg starch**: The goal was to decrease costs from 12.5 Euro to 10 Euro per 100 kg starch. It was not possible to measure goal achievement due to missing data.

• **More reliable yield data in % of total realized yield**: The measurement of this KPI was tested but the algorithm was not sufficiently tested

• **Starch content with reference value 20%**: Only this KPI was measured and was up to the reference value.

**Business impact of the technology on general level**: Farming is a business with a lot of variables, which not all can be controlled by the farmer. Therefore, a well-informed farmer has the advantage to be able to adapt to the circumstances. No farming business is the same, no growing season is the same. Therefore, there is a growing need for information generated in a time and location-specific way, which provides a solution in line with the farmer’s needs.

Collection of data from several growing seasons in a row would allow for new, more reliable prediction. Successes of past years will help provide further successes for the future, and any failures will be digitally noted and avoided going forward. A farmer will be able to perform risk assessment based on the data and subsequently change his management accordingly.

Big data sources, like EO and sensor data, provide a continuous flow of data, which will support the development of solutions that best support the farmer in his decision process. Through Big Data sources and IOT devices, the goals around profitability, efficiency and cost management will be achievable.

In this pilot the focus was on yield improvement and prediction. Combined with other cost related KPIs, the yield data can be used for calculation of further productivity KPIs such as the following:

• Higher harvest quantity / Fertilizer consumption
• Higher harvest quantity / Pesticide consumption
• Highest harvest quantity / irrigation water quantity
• Higher harvest quantity / land sq mt
• Higher employee productivity (Revenues / Employee)
• Higher revenues
• Return on Invest (ROI)

**3.4.5 How-to guidelines for practice**

During the pilot several datasets and components were combined in a pipeline and the pipeline was implemented and tested. The combination of data, readily available to the

---

7 Reference: average value 13.7 tons in 2012
8 Reference: average cost 12 – 13 Euro in 2012
farmers were valuable in itself, but the data were more valuable with more context like their performance relative to the surrounding region and to the potential by addressing the so-called yield gap.

For Big Data analysis you actually need Big Data. This was one of the challenges. Sources that give a continuous flow of data, like EO, IOT sensor data and weather data, are very useful. For the analysis they need to be combined with the field data for crop, variety, etc but most important with yield data. Especially the availability of historical yield data with location information during the pilot was too limited to give reliable results. The cooperative has relied for their information on expected yield mostly on the bulk information of the potatoes being delivered to the factory, in which the link to the field, where the potatoes were grown, is lost. In this way a lot of valuable information is lost every year.

For training of the model much more field data is necessary to make the prediction more reliable. Especially the yield data per field is essential field data. For the 2500 farm members and about 44,000 ha (2017), with an average field size of 10 ha, this would mean that there would be 4400 fields for collecting yield data every year. The cooperative can play an important role in collecting this data when they manage the information flow about the yield differently.

In the current situation, the harvest is collected at the head of the field or another location nearby where the harvest can be stored temporarily. When the harvest is collected for transport to the factory just the farmer’s name (or member-id) is registered, but not the field where the product was harvested. With a small adjustment in the registration process this data can be gathered as well. A simple system of locating and registering harvested products per field will add valuable data.

The components for (automated) gathering, processing and analysing the combination of data is in place for future use. It depends on the specific demands form the agribusiness which information services can be implemented.

For research, Big Data is very interesting when the data gathering is automated and the number of data points can be much higher than in experimental fields.

For farmers the analysis provides them insight that would not have been available with only data about their own fields. In this respect is it very important that farmers share their field data with each other, or a trusted party. Privacy issues (and trade secrets) hinder the sharing of the data. A trusted partly, like a cooperative, should provide farmers trust that their data will not be misused and thus facilitate the data sharing which will benefit them all.

3.4.6 Summary and outlook
In the pilot we gained insight about the possibility to apply the technologies provided by Big Data to smart farming services in order to gain a competitive advantage in terms of possible cost-effective services based on satellite imagery to the farmers.
Extensive field trials are expensive and will not predict yield in normal field conditions. The results from the DataBio project have been useful to speed up the process of improving the growth model on the basis of Big Data analysis. The approach contributed to better yield prediction based on the actual growing conditions with a limited number of samples or field trials. Once the model is validated through more empiric tests and observations, the processing industry will be able to enhance their sales process based on the yield prediction.

The availability of historical (field) data is the most crucial factor in order to train Big Data models. For EO data the availability of historical Sentinel-2 will grow the coming years. For the field data the farmers themselves, or the cooperatives, have an important role to play. In order to motivate them to contribute a financial compensation for their data collection would be appropriate.

Big data sources, like EO and sensor data, provide a continuous flow of data, which will certainly support the development of solutions that support the farmer in his decision process. New Business opportunities can be found by:

- Implementing the yield prediction model that was tested in the pilot with AVEBE, but also with other potato processing cooperatives.
- Implementing, with other partners in the Netherlands, the farmer decision support system. This may be the processing cooperatives, but also other stakeholders.
- Elaborating on the potato growth model to create new services like variable rate application and irrigation planning.

3.5 Pilot A2.1: Big data management in greenhouse ecosystem

3.5.1 Introduction, motivation and goals of the pilot

One of the most compelling merits of the genomic selection technology is the possibility to integrate Marker Assisted Selection for yield into practical breeding programmes. This has been a puzzle to breeders, geneticists and other scientists for the last 30 years of Quantitative Traits Loci (QTL) breeding. Genomic selection represents the gold standard approach to expedite cultivar development, and for estimating breeding values upon which superior cultivars are identified and selected. Genomic selection allows superior response to selection, and hence superior breeding progress, due to its intrinsic attributes that expedite breeding works by shortening generation intervals through genomic prediction and selection-driven intercrosses. The Genomic Selection technology is expected to significantly improve genetic gain by unit of time and cost, allowing farmers to grow a better variety sooner relative to conventional approaches, making more income.

The pilot A2.1 was designed to implement Genomics Prediction Models (GS) as a solution to the above-described technological limitations met with current breeding approaches. The pilot was run by a collaborative effort between CREA (Italy) and CERTH (Greece). Genomic data (SNPs) produced in tomato was enough to run Genomic Models but, the size of tomato

Dissemination level: PU -Public
population phenotyped was too low (less than 40) due to the need of the production of new genetic data, in order to assess the genetic variability of the crosses and the collection of environmental and phenotypic data. Therefore, we could not run Genomic Models in Tomatoes. Training Genomic models requires a big size of the training population. As an alternative solution, we ran the genomic models (component C22.03) in sorghum, where a sufficiently bigger population (380) had been genotyped and phenotyped, to improve health-promoting compounds used to manufacture specialty foods.

In the GS approach, different assumptions of the distribution of marker effects were accommodated in order to account for different models of genetic variation including, but not limited to: (1) the infinitesimal model, (2) finite loci model, (3) algorithms extending Fisher’s infinitesimal model of genetic variation to account for non-additive genetic effects. Many problems were modelled including the performance of new and unphenotyped lines, untested environments, single-trait, multi-trait, single-environment, and multi-environment. Models were fed several data types: open-field phenotypic data, biochemical data, phenomic and genomic data. Next, these equations were used to predict the breeding values of genotyped but unphenotyped candidates. In the process, several other Big Data types (e.g., those describing environmental properties) can be used as covariates.

### 3.5.2 Pilot set-up

The first stage of the trials started in 2018. In this year, the CREA’s platform for Genomic prediction and selection was specified to accommodate CERTH’s requirements following a non-conventional breeding approach. For this purpose, CERTH initiated a pilot study for the identification of best tomato crosses bearing desirable traits e.g. organoleptic, nutritional value, tolerance on various environmental conditions. The parental lines were Greek varieties that have been adapted to the local environmental conditions. In order to investigate as many crosses as possible, a holistic approach was applied for the best evaluation of the new genotypes, including: (1) Biochemical characterization and nutritional value assessment; (2) Genomic: next generation sequencing protocols to generate genomic, genotypic dataset; (3) Environmental indoor: air temperature, air relative humidity, solar radiation; (4) environmental outdoor: wind speed and direction, evaporation, rain; (5) farm data: farm logs (work calendar, technical practices at farm level, irrigation information); farm profile (static farm information, such as size, crop type, etc.). Biochemical, genomic and phenomic data were collected in tomato (landraces and several recombinants lines in diverse filial progeny stages) raised in glasshouses.

CERTH also produced the initial molecular data through NGS (Next Generation Sequencing) sequencing technology based on Double Digest RADseq approach and performed the initial analysis and validation based on the STACKS pipeline (available at: [http://catchenlab.life.illinois.edu/stacks/manual-v1/](http://catchenlab.life.illinois.edu/stacks/manual-v1/)). One hundred and thirty-eight samples, originating from 40 tomato lines were included for the study and whole-genome genotyped using the ddRADseq protocol; 10,402 good quality SNPs were produced. The size of the SNP marker matrix was enough to start running the model, but the number of phenotyped individuals was still too low (less than 30) to be usefully used to run the predictive models.
CREA set up and anticipated a GS platform for accommodating the upcoming genomic and phenomic/phenotypic data. In addition, CREA set up a genotyping and phenotyping platform for use as test-bed of the CREA’s Genomic models component.

For the stage two trials, greenhouses for tomato pilot trials were established in Greece, whereas sorghum pilot trials were established in Italy. Tomato lines were genotyped using the double digest restriction-site associated DNA (ddRADseq) approach, while sorghums were genotyped using a genotyping-by-sequencing (GBS) strategy on Illumina next generation sequencing platform. The Biochemical analysis and nutritional value assessment were carried out in the initial parental lines and on the final genotypes as to evaluate the breeding process. For this purpose, a thorough biochemical analysis was carried out implementing both colorimetric and chromatographic methods. Total sugars and soluble solids were measured with a refractometer and expressed as Brix values, total polyphenol content was measured with Folin-Chiocalteu method, total antioxidant activity was assessed with DPPH radical assay, lycopene was measured spectrophotometrically, total flavonoid content was measured with AlCl₃ method, ascorbic acid was assessed with Megazymes ascorbic acid assay kit and amino acids was measured with GC-MS with EZFaastTM Free (Physiological) Amino Acid Analysis kit (Phenomenex). The phenotypic characterization of sorghum and tomato lines was carried out according to international standard operating procedures (IBPGR, UPOV).

### 3.5.3 Technology used

#### 3.5.3.1 Phenomics

In tomatoes, the Biochemical analysis and nutritional value assessment were carried out in the initial parental lines and on the final genotypes as to evaluate the breeding process. For this purpose, a thorough biochemical analysis was carried out implementing both colorimetric and chromatographic methods. Total sugars and soluble solids were measured with a refractometer and expressed as Brix values, total polyphenol content was measured with Folin-Chiocalteu method, total antioxidant activity was assessed with DPPH radical assay, lycopene was measured spectrophotometrically, total flavonoid content was measured with AlCl₃ method, ascorbic acid was assessed with Megazymes ascorbic acid assay kit and amino acids was measured with GC-MS with EZFaastTM Free (Physiological) Amino Acid Analysis kit (Phenomenex). The phenotypic characterization was carried out according to the UPOV guidelines. IoT technology was used to collect environmental indoor data (air temperature, air relative humidity, solar radiation), and environmental outdoor data (wind speed and direction, evaporation, rain).

In sorghums, to analyse total phenols, tannins, flavonoids and antioxidant capacity (TAC), a 10 g sample from each genotype was ground using a Cyclotec Udy Mill (sieve: 0.5mm), the moisture in the sample was determined after they were oven-dried overnight at 105°C, and antioxidants and TAC were analysed in duplicate using 100mg of each sample. For the phenolic compounds the absorbance of samples was measured at 750nm and expressed as gallic acid equivalents (gGAEkg⁻¹ dry mass basis). For condensed tannins and total flavonoids assays, the absorbances were measured at 500nm and 510nm, respectively, and expressed
as μg CE (catechin equivalents) g⁻¹ dry mass basis. TAC was determined using the 2,20-azino-bis/3-ethylbenzthiazo-line-6-sulphonic acid (ABTS) assay and expressed as mmol TE (Trolox equivalents) kg⁻¹ dry basis. IoT technology was implemented to collect and characterize soil, plant, and environmental properties.

3.5.3.2 DNA isolation, next generation sequencing/genotyping, and bioinformatics

In sorghums, DNA was isolated from plantlets using the GeneJET Plant Genomic DNA Purification Kit. The methylation sensitive restriction enzyme ApeKI was used for library preparation, and Genotyping-By-Sequencing (GBS) was carried out on an Illumina HiSeq X Ten platform. The final working matrix consisting of 61,976 high-quality SNPs was used in this work for genomic selection and prediction analytics.

In tomatoes samples, DNA was extracted from young leaves using the NucleoSpin Plant II, Macherey-Nagel kit. Two-hundred and seven NGS libraries were constructed by applying the ddRADseq protocol, using the EcoR1 and MspI restriction enzymes. Next generation sequencing was performed at the Institute of Applied Biosciences of the Centre for Research and Technology Hellas, on an Illumina NextSeq500 platform (Illumina Inc., San. Diego, CA, USA) using the NextSeq™ 500/550 High Output Kit (2 x 150 cycles). The final working matrix consisting of 10,402 high-quality SNPs was obtained.

3.5.3.3 Genomic predictive and selection analytics

Genomic selection represents the gold standard approach to expedite cultivar development, and for estimating breeding values upon which superior cultivars are identified and selected. Genomic selection allows superior response to selection, and hence superior breeding progress, due to its intrinsic attributes that expedite breeding works by shortening generation intervals through genomic prediction and selection-driven intercrosses. The Genomic Selection technology is expected to significantly improve genetic gain by unit of time and cost, allowing farmers to grow a better variety sooner relative to conventional approaches, making more income (Figure 3 and Figure 4).

In the GS approach, different assumptions of the distribution of marker effects were accommodated in order to account for different models of genetic variation including, but not limited to: (1) the infinitesimal model, (2) finite loci model, (3) algorithms extending Fisher’s infinitesimal model of genetic variation to account for non-additive genetic effects. Many problems were modelled including the performance of new and unphenotyped lines, untested environments, single-trait, multi-trait, single-environment, and multi-environment. Models were fed several data types: open-field phenotypic data, biochemical data, phenomic and genomic data. Next, these equations were used to predict the breeding values of genotyped but unphenotyped candidates.
Figure 3: GS advantage: higher response to selection harnessing quantitative and population genetics with GEBV-driven intercrosses shortening generation intervals

Figure 4: Overall genomic prediction and selection roadmap

Several technological scenarios were anticipated and implemented. Cross-validation CV1 reflected prediction of tomato lines that have not been evaluated in any glasshouse trials. Cross-validation CV2 reflected prediction of tomato lines that have been evaluated in some but NOT all target environments (glasshouses). The rationale being that prediction of non-field evaluated lines benefits from borrowing information from lines that were evaluated in other environments (glasshouses). This is critical in cutting costs for varietal adaptability trials of large number of lines in several target environments.
BRR, GBLUP, LASSO, and Bayes B were implemented during the first trial. Under several environments, these algorithms were factorially combined with environments to generate further predictive analytics. For each algorithm, predictive analytics were run on a single environment basis, across environments, marker x environment, and using the approach of reaction norm model.

### 3.5.4 Business value and impact

Genomic predictive and selection (GS) modelling was developed as response to the lengthier and costlier phenotypic selection. In business time to market is important just as the production cost. In addition, specifically for plant breeding, the longer it takes to bring the new cultivar to the market, the shorter will that cultivar stay on the market, in virtue of the naturally-occurring crop degeneration. Some of the most attractive GS attributes are enabling cutting time and cost to cultivar development with high selection accuracy. The high accuracy means that the plant lines selected will breed true to type, implying diminished risks in the breeding processes.

In this pilot, the GS technology showed meaningful results and attractive as reflected by the key performance indices presented in the below table.

**Table 6: KPIs of the A2.1 pilot**

<table>
<thead>
<tr>
<th>KPI short name</th>
<th>KPI description</th>
<th>Goal description</th>
<th>Base value</th>
<th>Target value</th>
<th>Measured value</th>
<th>Unit of value</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>A2.1-KPI-01</td>
<td>Accuracy</td>
<td>Increased accuracy</td>
<td>0.4</td>
<td>0.4-0.7</td>
<td>0.5-0.6</td>
<td>Pearson' r</td>
<td>Pilot was successful</td>
</tr>
<tr>
<td>A2.1-KPI-02</td>
<td>Breeding cycle (years)</td>
<td>Decrease the cycle relative to phenotypic breeding</td>
<td>-</td>
<td>3 times</td>
<td>4 times</td>
<td>Ratio Phenotypic /Genomic selection</td>
<td>Too early to assess</td>
</tr>
<tr>
<td>A2.1-KPI-03</td>
<td>Breeding costs (index)</td>
<td>Decrease costs relative to phenotypic breeding</td>
<td>-</td>
<td>2 times</td>
<td>5 times</td>
<td>Ratio Phenotypic /Genomic selection</td>
<td>Too early to assess</td>
</tr>
</tbody>
</table>

The predictive performance obtained in this pilot were encouraging. Over the two-year trial, with data integration, the four genomic selection models implemented in this pilot performed comparably across traits and are considered suitable to sustain sorghum breeding for antioxidants production and allow important genetic gains per unit of time and cost. In comparison to conventional phenotypic breeding, the genomic predictive and selection modelling allows cutting costs 5 times, and cutting four times the time to cultivar development (Table 6). The results produced in this pilot are expected to contribute to...
genomic selection implementation and genetic improvement of sorghum grain antioxidants for different purposes including the manufacture of health-promoting and specialty foods in Europe in particular, and in the world in general. In addition, the ddRADSeq and NGS GBS genotyping platforms were validated and can be used for sequencing and genotyping (variants calling) services in other plant species and animal husbandry.

Several international seminars and two international webinars (https://vimeo.com/showcase/6623660; https://es.slideshare.net/BDVA/bdv-webinar-series-ephrem-big-data-breakthroughs-for-global-bioeconomy-business) were presented. In addition, four articles were published in specialized international peer-reviewed scientific journals with impact factor [REF-06][REF-07][REF-08]. These scientific papers are expected to exert increasing impact not only in Europe but also across the world.

3.5.5 How-to guidelines for practice

The roadmap for implementing genomic prediction and selection analytics was depicted in the above diagram (Fig. 12). Several scenarios can be modelled including a single trait, multiple traits as index selection, a single environment, and multiple environment. A generic technological flowchart is that, in the genomic predictive and selection modeling, phenotypic and marker data are scored in the training population and fitted into appropriate algorithm to produce individuals’ whole-genome marker effects. Most practically, the training set is the germplasm or a population that best samples the frequency of the genetic information (allele frequency) useful for the breeding program. The marker effects are used in subsequent cycles of selection to compute the genomic estimated breeding values (GEBVs) that are used as predictors of breeding values in testing unphenotyped population. The genomic estimated breeding values are obtained as a product of the estimated marker effects in the training population and the coded marker values obtained in the testing population. To apply genomic selection, GEBVs are obtained in the selection candidates and then used to predict and rank the net genetic merit of the candidates for selection, and superior strains are selected in the process. Genomic predictive and selection modelling is a gold standard for selecting for breeding values, and is well poised to help breeders and seed industries to drastically cut breeding cost and time, and bring new cultivar earlier on the market, generating higher incomes.

3.5.6 Summary and outlook

Current empirical evidence for genomic selection efficiency in plant breeding set to 0.5, the baseline for genomic selection prediction accuracy in plant breeding. Also, recent research works demonstrated that genomic selection accuracy as low as 0.2 can allow substantial within-generation yield improvement. Therefore, the genomic selection model performances obtained in our pilots are high enough to sustain sorghum breeding for antioxidants production and allow important genetic gains per unit of time and cost. In addition to the accuracy, the importance of the genomic selection strategy is also evaluated using other criteria such as the possibility that this technology offers to shorten the breeding cycle with interesting economic returns due to intercrosses driven by genetic predictions, the quick
delivery of novel superior cultivars onto the market and, in the case of antioxidants, genomic selection offers the possibility to select for or against this trait early (e.g., at the seed or seedling stages) without waiting for seed setting or harvest. The genomic selection equations developed in this work can be directly used in sorghum breeding programs and can be adapted to other plant species and animal husbandry. The genomic selection results presented herein and the experimental designs used in this pilot can be implemented in antioxidants genetic investigations and in breeding programs to qualitatively and quantitatively improve the antioxidant production for different purposes including the manufacture of health-promoting and specialty foods.

3.6 Pilot B1.1: Cereals and biomass crop

3.6.1 Introduction, motivation and goals of the pilot

Water scarcity is an increasing common worldwide phenomenon. Hydrologic cycles do not coincide with the annual seasons and there are alternating periods of severe drought with periods of heavy rains. As a general approach, irrigation agriculture is vital to guarantee food security conditions to assure the well-being and progress levels demanded by European Citizenship in the 21st century.

According to FAO estimates, in the first decade of this century, the 17% of irrigated arable crops supplied the 42% of food in the world. By 2020, these irrigated arable crops are expected to provide 50% of food using less water.

Therefore, sustainability of irrigation areas must be promoted, and it is mandatory to solve their specific problems in order to meet their needs. The specific problems are: 1) water scarcity, 2) increase of energy used, 3) absence of tools determining the specifics requirements of each crop at the time, 4) Lack of generalized and interoperable tools, 5) water quality problems, 6) lack of performance of irrigation arable crops, 7) Lack of research in the process of switching to alternative crops: develop pest-resistant local crop varieties, develop crop with low water requiring, etc., 8) no control of needs required to optimize the work. Therefore, the overall challenge is to get a smart agriculture to ensure optimal conditions. It will be necessary to consider social and environmental challenges in order to meet the needs in irrigation areas and turn them into optimized production areas.

Against this background, this DataBio pilot aims to reinforce agribusiness sector adapting the diversification of production to new economic and climate scenarios by use of systems of remote control and remote management in the irrigated areas that are essential to achieve smart agriculture based on IoT (Internet of Things) sensors and Big Data sources.

The technological elements are aimed to get the following objectives: (i) reduction of inputs as water, manure, fertilizers, (ii) reduction in energy consumption, (iii) automation of irrigation systems, (iv) optimization of irrigation areas management, (v) deploy of Big Data in agricultural environment, (vi) modernize agriculture and (vii) traceability control.
Thus, the goal is to design, use and deploy tools and processes to create real Smart Agriculture in irrigation areas and to establish useful processes that can be adjusted to other agri-food chains.

### 3.6.2 Pilot set-up

The product developed by TRAGSA Group with the help of ATOS and IBM Israel has developed a system to generate accurate "irrigation maps" and "vigor maps" of crops, using Big Data Sources as Earth Observation (EO) data and sensors data as inputs. Those maps, from different areas in Spain as Castile and Andalusia, are linked to an informative and management system for early warning of inhomogeneity: therefore, this new service is dedicated to the analytical and accurate finding of crop heterogeneities.

In this pilot BDT has been developed specifically for the Irrigation Community “Let side of Porma River” in Castilla & León Region as final customer. In this pilot area the objective was the reduction of water and energy use in irrigation areas using the following techniques:

- EO Big Data sources and Remote sensing applications for calculation Normalized Difference Vegetation Index (NDVI) and corrected crop factor Kc and balanced against participation, which needs to be measured on site.
- Remotely Piloted Aircraft System (RPAS) for address specific problems: plagues and diseases, irrigation uniformity, soils problems, etc...
- Agroclimatic stations that provide information in situ.
- Telecontrol systems.
- Irrigation equipment.

Due to irregular irrigation, mechanical problems affecting irrigation systems, incorrect distribution of fertilizers or any other sources of inhomogeneity could lead to crops growing differences. This DataBio Service is a very good preventive tool for farmers and landowners in order to avoid production losses and it is a powerful tool for agricultural management of big surfaces.

The Energy - Efficient is oriented to obtain the minimum energy balance in the irrigation areas. In this point the hydrants follow a schedule model designed around a) dynamic irrigation demand, b) contacted power reduction and c) energy dynamic buying in global market.

### 3.6.3 Technology used

#### 3.6.3.1 Technology pipeline

The technological elements are aimed to get the following objectives: (i) reduction of inputs as water, manure, fertilizers, (ii) reduction in energy consumption, (iii) automation of irrigation systems, (iv) optimization of irrigation areas management, (v) deploy Big Data in agricultural environment, (vi) modernize agriculture and (vii) traceability control. The technology applied in this pilot can be summarized as follows:

**Data collection** involved various sensors:
- **Agroclimatic stations** contributing data related to temperature, relative humidity, absolute humidity data and wind parameters
- **Contact sensors**, to determine problems with great precision, speeding up the use of techniques which help to solve these problems
- **Use of humidity sensors** on cropped soil to know its actual conditions in order to determine and control the field capacity
- **Lysimeter** to control the level of nutrients in the field, adjusting the amount of manure and fertilizer needed
- **Control in the parcels** with sprinklers, drippers, metering devices, and similar
- **Elements to control the networks**: pressure switches, pressure reducer, flow meters, manometers
- **Solenoid** valves for an automatic opening and closure
- **Counters**
- **Control in pumping stations**: manometers, flow meters, pumping state, anti-return valves, alarm settings, heating,...
- **Geographic Information Systems**
- **Machinery sensors**
- **Remote sensing and RPAS.** A comprehensive strategy to combine EO and field data was helpful for improved and more efficient agriculture management. Satellite data are suitable for monitoring large areas over time, while RPAS provide specific data for calibration and validation purposes.

**Data processing:** The pilot objective is to integrate EO and RPAS Big Data in decision-making support tools in order to improve water efficiency and agriculture management for irrigated crops. The study area comprehends 100 ha located in Leon province, nevertheless the methodology developed can be extended to the whole users’ community.

Irrigation needed by crops is usually calculated using Kc reference values provided by Food and Agriculture Organization - FAO. The NDVI obtained by means of remote sensing has proven to be useful for calculating the Kc factor adjusted to local conditions. A preliminary test showed promising results for the estimation of Kc using satellite and RPAS in the study area. The Kc values obtained by using NDVI derived from RPAS multispectral images improved those provided by FAO, and showed a very reliable relationship with Kc derived from the satellite SPOT 7. In addition, some products were obtained from RPAS data, including RGB mosaics thermal images and Digital Terrain and Surface Models. These products provide valuable information for different purposes such as the monitoring of plants health or the estimation of growth and biomass.

Linked Open Data Data (LOD) is currently a source of unprecedented visibility for environmental and agricultural data that enables the generation of new communication links as well as a significant advance for research in the agricultural area. LOD allows acquisition, adaptation using and communication from, between and to several data sources.

**Data visualisation:** Data were visualised in several ways as can be seen in Figure 5:
Furthermore, the resulting information was provided in a decision relevant form and was used to drive actuators related to the irrigation system.

### 3.6.3.2 Data used
The pilot used: massive and rapidly updated data, bioagronomic data, sensor data, terrestrial observation data and geographic data from different sources were used, specifically:

- SENTINEL-2: terrestrial observation data owned by the ESA (European Space Agency).
- Orthophotos: terrestrial observation data in image format obtained from the National Geographic Institute of Spain.
- RPAS: terrestrial observation data obtained by thermal and multispectral sensors, owned by TRAGSA.
- SigPAC: spatial data in the ESRI Shapefile format which identify the parcels, owned by Junta de Castilla y León.
- Alphanumeric data from surveys and field visits, owned by TRAGSA.

### 3.6.3.3 Reflection on technology use
The goal accomplished was to design, use and deploy tools and processes to create real Smart Agriculture in irrigation areas and to establish processes which are also useful in other agri-food chains.

Big Data have provided new efficient decision-making tools for helping agricultural development as well as biodiversity protection. New acquired, aggregated and shared data can serve as a breeding ground for extracting and sharing useful information and knowledge among different actors, as well as for combining large data sources (especially regarding weather models and earth observation datasets) with advanced crop and environment models to provide actionable on-farm decisions.

### 3.6.4 Business value and impact
As stated previously, sustainability of irrigation areas must be promoted, and it is mandatory to solve their specific problems in order to meet their needs. Therefore, the following social challenges have been considered by the pilot:
**Sustainable Production:** Selecting better seeds than increase the productivity to attend the increase of demand of food in a limited surface.

**Cost:** Water scarcity and increasing energy costs are the most important threats to irrigated agriculture.

**Risk:** The health security and safety in food is a big preoccupation. It is necessary to guarantee the security and safety in food production.

**Collective decision making:** To support farmers’ decision making in relation to the use of these resources (water, manure and fertilizers) and their management strategy of these resources.

It is important to offer technical solutions to improve the technical-economic and environmental performance of Agriculture. Currently, the irrigated farming in Spain is working on the following lines:

- Implementation of renewable energy in irrigation
- Use of Big Data, Satellite technologies and RPAS for improving water efficiency and early detection of pests.
- Modelling of networks for optimum energy consumption.

Spain has an area of 3.621.722 ha for irrigated agriculture, of which 73% is modernized irrigation pressure and the remaining 27% is irrigated by gravity. Many of them are managed under the control of Irrigation Communities; they would be the addressable market for the BDT developed in this pilot.

During the project an additional irrigation community joined the pilot, so that the covered area increased from 12499.87 ha to 36445.87ha and the number of stakeholders using the tool increased to 300.

The goal is that each Community of Irrigators can choose when to water and irrigate with adequate water volume. This is achieved if irrigation has ICT systems for: a) the organization of irrigation demand; b) procurement of appropriate powers; c) the use of best energy tariffs; d) the use of agro-climatic data to define the accurate irrigation.

The DataBio B1.1 pilot has used different kind of sensors, and actuators distributed in Irrigations Communities in experimental facilities for testing and finally in real scenarios dealing with daily activity and real impact on advances in services and infrastructures that are in place for systemic innovation in Water Communities. The kind of sensors and actuators are very similar in all the modernization irrigation areas and the number varies depending on the considered scenario.

**3.6.5 How-to guidelines for practice**

This DataBio pilot aims to reinforce agribusiness sector adapting the diversification of production to new economic and climate scenarios to systems of remote control and remote
management in irrigated areas where is essential to achieve smart agriculture based Big Data sources.

An algorithm (R language) has been used for the automatic classification of soil cover, which allows the generation of decision trees combining data of different types (cartography, images, BD, etc.). It must be taken into account that these will vary according to the zones, their availability and quality. The classification algorithm was trained using samples of the different uses and coverages to be identified. The sample data are divided into two groups, in this way 80% of the samples are used in the construction of the model. Once the decision tree is generated, it is validated with 20% of remaining samples, not used in its construction.

Using samples from all land uses, a decision tree is generated from which a classification of large LPIS uses is obtained. Using only the agricultural samples, another tree is generated, from which a classification of agricultural crops is obtained. The combination of both classifications will result in the crop layer and soil cover.

After this technical description of the algorithms, it is necessary to emphasize that they had (in the development phase) the following limitations: (i) in LPIS there are no differences between arable crops, so it is not possible to verify if the crop identified by remote sensing coincides with the one existing in the field and (ii) the spatial resolution of Sentinel-2 does not allow the correct identification of woody crops, since it is limited to the response of the crop and / or the plant coverings under it.

The results will collect:

- Null match: agricultural use in LPIS. Classified as non-agricultural.
- Average coincidence: when both in LPIS and in the layer generated, the use is agricultural.
- High coincidence: when both in LPIS and in the layer generated, the crop is of the same type (in both cases, it is a woody crop, or both are herbaceous).
- Perfect match: when the crop is the same in both LPIS and the layer generated by remote sensing.

The cause of the discrepancies should be analysed with the supervision of photo interpreters, with Sentinel-2 images being an important aid for this. Once defined the kind of crop, the developed methodology allows, using temporal series of Sentinel data, the definition of Kc parameter and, using it, the irrigation needs of the specific crop.

Finally, a methodology for the calculation of water needs has been developed and applied to the Genil-Cabra (Andalusia) pilot zone. The farmer association involved in the pilot has provided it with data on irrigated plots and crops from 2017. In addition, the pilot has used Sentinel 2017 images. Using those datasets as initial Big Data sources, a classification process has been developed to obtain the NDVI (Normalized Difference Vegetation Index). This biological index is the basis for the calculation of water needs. In the final cycle of the project, an integration data process has been carried out to harmonize and unify the different
datasets. Figure 6 highlights how using all the previously explained processes is possible to classify the plots accordingly to irrigation needs:

![Figure 6: Crops classification and irrigation needs](image)

3.6.6 Summary and outlook
The final service provides information for precision agriculture, mainly based on time series of high resolution (Sentinel-2 type) satellite images, complemented with IoT sensor data and, in some specific cases defined by profitability, with RPAS data. The final costs saving for farmer communities due to better quality management in agricultural zones, especially focused on irrigated crops, are produced, mainly, by a water and energy better management. Those savings are not easy to calculate at the current stage of products implementation, but a first rough estimation could be close to 5%. Besides this, fertilizers control and monitoring produce, eventually, a prominent economic saving per year and hectare. This better management of hydric and energetic resources is also related to Green-house effect gases reduction, directly linked to better environmental conditions in agriculture.

3.7 Pilot B1.3: Cereal and biomass crops_3

3.7.1 Introduction, motivation and goals of the pilot
This pilot was designed to implement remote sensing, IoT farm telemetry, and proximal sensor network-based Big Data technologies for biomass crop monitoring, predictions, and management in order to sustainably increase farming productivity and quality, while at the same time, minimizing farming and environment associated risks. Biomass crops of interest include biomass sorghum and cardoon which can be used for several purposes including, respectively, biofuel, fiber, and biochemicals, with a high macroeconomic impact. Fiber hemp was anticipated but, due to unexpected farmers aversion, this crop was not included in pilots. The aversion was particularly triggered by a complicated market of the produce.

Crop monitoring enables better use of farmers’ resources as farmers can better decide on which field or part of the farm to focus their attention and activities. It also supports more
accurate and location-specific use of various inputs as irrigation water or fertilizers. Crop yield prediction is important for risk minimization of farm management. Growers and farmers benefit from yield prediction to make informed management and financial decisions. Special cases are seed companies that need to predict the performances of new hybrids in various environments to breed for better varieties.

3.7.2 Pilot set-up
The pilot secured adhesion of private farmers and/or farming cooperatives. During the 2017 and 2018 cropping seasons, 43 sorghum pilots were run covering 240 hectares. The work on this pilot was distributed between CREA, Novamont, and VITO. CREA worked on sorghum, and Novamont on cardoon. VITO provided access to their EO platform “WatchItGrow”, which was also the end-to-end backbone for the technical pipeline used in this pilot.

During 2018, an additional field of cardoon in Umbria region was included in monitoring activity in addition to the fields established in Sardinia and already included in the previous reports. We added Umbria in order to give an example of different cultivation area and cover some of the main areas where cardoon can be cultivated. In 2018, in collaboration with InfAI, CREA was able to extend crop monitoring to foliar diseases in one of the pilot field in Anzola, Italy.

In terms of global sorghum crop disease monitoring, five training and testing fields for crop disease detection had been identified by CREA. Within this diseased field, CREA delimited a most diseased area of about 1000 square meters (~232 m of perimeter) within which leaf disease occurred in about 60 to 70% of the plants. Two foliar diseases were observed, i.e., Anthracnose (most prevalent) and Bacterial stripe. The network worked as it should and detect the fields (sorghum foliar diseases detected with the reliability of 0.925 and sorghum foliar diseases detected with the reliability of 0.861). The network was even able to detect the disease and distinguish it from surrounding areas (for more details see D1.3).

The set was very small. Overall, there were six training sets and two for validation, so the results were limited. The main problem of small datasets is the overfitting – which means that the models are trained too well, precisely to the set of data. In order to overcome overfitting future improvements should focus on:

- Expanding the database
- Augmentation (Expand the database by manipulation)
- Regulation

Up to now, we created 1000 test cases out of our starting point, and the success rate is still high. For the crop monitoring using satellite imageries, forty-three pilot biomass sorghum trials were run by CREA over two cropping seasons in 2017 and 2018. The biomass sorghum pilot trials were mainly established in private farms and co-run by CREA and private farmers and private farming cooperatives operating in the northern Italian communes of Nonantola, Mirandola, and Conselice. Only eight pilots were run in CREA’s experimental station of Cà Rossa (Anzola dell’Emilia) in both 2017 and 2018 cropping seasons. During the 2018 cropping
season, sorghum was monitored for phenology, yields, and foliar diseases. Two cardoon fields were monitored in 2018, one located in the North of Sardinia, as continuation of 2017 work, this cardoon field was established in 2014. The other field is located in Umbria. In the last cultivation period (2017-18) in Umbria the phonological phases were monitored together with the agronomical operations.

The plot sites were geolocated and the coordinates entered into VITO system for monitoring the fAPAR index throughout the cropping season. In addition, Chlorophyl meter and NDVI meters were prepared for respective data collection. Chlorophyl index and NDVI index were collected weekly. Fields were geolocalized, geolocation data saved as kml files before they were integrated into WatchITGrow application. The fAPAR estimates were generated at decametric spatial resolution (10m pixel size), and a temporal resolution of 5 days up to 2-3 days in those areas where the different satellite overpasses overlapped. Spatial resolution refers to the surface area measured on the ground and represented by an individual pixel, while temporal resolution is the amount of time, expressed in days, that elapses before a satellite revisits a particular point on the Earth’s surface. For each experimental field, fAPAR or “greenness” maps were produced, and a growth curve was built, showing the evolution of the fAPAR values throughout the cropping season. To correct for artefacts in the curve (such as abnormally low fAPAR values due to undetected clouds, shadows or haze) and to interpolate fAPAR values between subsequent acquisition dates, a Whittaker smoothing filter was applied on the curve. Finally, the fAPAR values from the curves were used for further analytics.

3.7.3 Technology used

The DataBio technological components implemented in this pilot were developed and deployed by VITO and CREA. Vito provided the platform “WatchITGrow”, while CREA deployed machine learning technology, all of which was the backbone technology and end-to-end solutions of the pilot. The pilot was implemented in a form of advisory services under real-world commercial farms settings. The offered smart farming services include biomass crop monitoring using proximal sensors to derive vegetation indices, and crop growth and yield modelling using fAPAR derived from satellite (Sentinel 2A and 2B) imagery and appropriate machine learning technologies.

3.7.3.1 Reflection on technology use

With respect to the BDT used, the following experiences were made:

- The IoT farm telemetry technology was used in year one for preliminary observation but, this technology revealed itself ill adapted to biomass sorghum as the hardware, particularly the cables, were frequently damaged by rodents. IoT was therefore removed from the trial settings as frequent repairs were becoming a burden.
- Despite the great potential we uncovered in the disease monitoring technology, we nonetheless identified a weakness associated with relying heavily on natural disease inoculum. Indeed, natural inoculum is heterogeneous in the field and diseased areas
can range from a single plant to a few plants, which is greatly challenging in terms of resolution. This investigation was therefore discontinued in 2019.

3.7.4 Business value and impact

The importance of sorghum as food, feed, and biofuel crop cannot be overemphasized. Biomass sorghum demonstrated higher yields with better energy balance relative to major crops of agroindustrial interest. As dedicated biomass sorghum crops are steadily increasing and precision farming is driving agricultural economies worldwide, harnessing satellite technology is well poised to bring about agricultural advantages including cutting farming operational costs. Sentinel-2-derived fraction of absorbed photosynthetically active radiation and the implementation of machine learning technology modelled satisfactorily crop phenology and the aboveground biomass yields up to six months ahead of harvesting, in our sorghum pilots. In addition, attractive key performance indicators were observed as reflected in Table 7.

This outcome from this study is important and can serve several purposes including farmers being able to improve their sorghum biomass business operations through informed decision making in terms of planning ahead field works, logistics, the supply chains, etc. Policymakers and extension services will also benefit from the technologies implemented in this work allowing them early on within season information on potential biomass availability, which is critical to wider energy planning and avoiding energy-related crises.

Table 7: Pilot B1.3 KPIs

<table>
<thead>
<tr>
<th>KPI short name</th>
<th>KPI description</th>
<th>Goal description</th>
<th>Base value</th>
<th>Target value</th>
<th>Measured value</th>
<th>Unit of value</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>CREA-B1.3-KPI-01</td>
<td>Early within season Yields prediction error</td>
<td>Reduce prediction error</td>
<td>5</td>
<td>5</td>
<td>0.16</td>
<td>%</td>
<td>MAPE (% mean absolute percentage error)</td>
</tr>
<tr>
<td>CREA-B1.3-KPI-02</td>
<td>Early yields prediction</td>
<td>Increase the time (months) of prediction before harvest</td>
<td>0</td>
<td>2</td>
<td>6</td>
<td>Numb er of month s before harvest</td>
<td></td>
</tr>
</tbody>
</table>

Several international seminars and two international webinars were presented. In addition, two articles were published in specialized international peer-reviewed scientific journals with impact factor [REF-09][REF-10]. These scientific papers are expected to exert increasing impact not only in Europe but also across the world.
3.7.5  How-to guidelines for practice
Using satellite imageries and supervised machine learning technologies allowed us to model biomass sorghum phenology and carry out early prediction of biomass yields up to six months before harvesting. This pilot combines expertise from Earth observation, ICT, artificial intelligence, and agricultural farming. The Earth observation data were mined to derive the biophysical parameter fAPAR, the agricultural farms provided the information that is critical for modelling farming outcomes, while the artificial intelligence expertise integrated the above information to model the solutions that would later be delivered to stakeholders in the form of advisory services. The equations produced during the data science part of this pilot were made publicly available in two open access publications (https://doi.org/10.3390/agronomy9040203, https://doi.org/10.1007/978-3-030-29852-4_19). These equations can easily be used in sorghum biomass farming businesses. As the data science was done the next big step should be putting the models into production and make them useful for any businesses. This is the beginning of our model operations life cycle including the following (but not necessarily limited to) key focus areas of machine learning engineering: the data pipeline (the data used to make the features used for model training such as fAPAR, phenology, biomass yields records), model training, model deployment, and model monitoring. At this level, the farmer knows how much he will produce early on within season using only satellite imagery-derived fAPAR. In addition, the phenology stages can be monitored handily by the farmers using web capable devices. In the real world, the farmer and other stakeholders will benefit from this technology as an advisory (web) service ether in-house or from third party, depending upon the expertise at the beneficiary level.

3.7.6  Summary and outlook
This pilot was established as a solution to current limitations in crop monitoring in Europe: yields forecasting approaches based mainly on field surveys, censuses, coarser spatial (250-1000m) resolution satellites (e.g., MODIS, SPOT-VEGETATION), all of which are undependable and/or costly. Our pilot was therefore designed to address these shortcomings. The main challenge in this pilot was being able to use high-resolution satellite imageries to predict sorghum biomass yields early within season, and with high precision to avoid Stakeholders’ aversion. The obtained results were highly encouraging. We were able to accurately predict aboveground biomass yields six months before harvesting with the best prediction times identified as days of year 150 and 165 i.e., late May and early June, respectively. These results show that crop monitoring can translate into global business without border. They represent a remarkable opportunity for the farmer and farming cooperatives that can use this information for several business-related purposes. The models developed in this work will also help the extension services and other policy makers for strategic planning purposes including assessing alternative means for energy supply and ways to avoid energy crises.
3.8 Pilot B1.4: Cereals and biomass crops_4

3.8.1 Introduction, motivation and goals of the pilot

The motivation of this pilot is to help farmers increase the efficiency of their business by using methods that do not require a large investment in data and technology data and do not require too much effort from the farmer to acquire and manage the necessary input data.

The pilot aims to develop a platform for mapping of crop vigor status by using EO data (Landsat, Sentinel) as the support tool for variable rate application (VRA) of fertilizers and crop protection. This includes identification of crop status, mapping of spatial variability and delineation of management zones.

The main focus of the pilot is on the monitoring of cereal fields by high resolution satellite imaging data (Landsat 8, Sentinel 2) and delineation of management zones within the fields for variable rate application of fertilizers. The main innovation is to offer a solution in form of web GIS portal for farmers, where users could monitor their fields from EO data based on the specified time period, select cloudless scenes and use them for further analysis. This analysis includes unsupervised classification for defined number of classes as identification of main zones and generating prescription maps for variable rate application of fertilizers or crop protection products based on the mean doses defined by farmers in web GIS interface.

Other part was focused on transferring Czech LPIS into FOODIE ontology and to developed effective tools for querying data. This work was done together with PSNC and system is currently supporting open access to anonymous LPIS data through FOODIE ontology and also secure access to farm data.

3.8.2 Pilot set-up

Partners in this pilot who develop and integrate components and services are Lesprojekt, PSNC and UWB. Some steps were consulted with another partner in the DataBio Consortium, NB Advies.

Development of platform is realized on the cooperative 8300 ha farm in Czech Republic, however basic datasets are already prepared for all Czech Republic. So current status of pilot support utilisation of solution on any farm in Czech Republic.

The work was supported by development of platform for automatic downloading of Sentinel 2 data and automatic atmospheric correction. Currently Lesprojekt is ready to offer commercial services with processing satellite data for any farm in Czech Republic.

The pilot farm Rostenice a.s. with 8.300 ha of arable land represents a bigger enterprise established by aggregating several farms in past 20 years. Main production is focused on the cereals (winter wheat, spring barley, grain maize), oilseed rape and silage maize for biogas power station. Crop cultivation is under standard practices, partly conservation practices is treated on the sloped fields threatened by soil erosion. Over 1600 ha is mapped since 2006 by high density soil sampling (1 sample per 3 ha) as the input information for variable application of base fertilizers (P, K, Mg, Ca). Farm machines are equipped by RTK guidance.
with 2-4 cm accuracy. Farm agronomists don’t use any strategy for VRA of nitrogen fertilizers and crop protection because of lack of reliable solutions in CZ.

3.8.3 Technology used

3.8.3.1 Technology pipeline

In this pilot mainly following DataBio components are used and integrated.

- OpenLink Virtuoso: Publishing the Czech farm and open data as Linked Data and allowing querying of the datasets via SPARQL endpoint.
- LESPRO/HSLayers: Visualisation of data and results
- LESPRO/Data model for PA: Integration of various farm data and data from other sources.

3.8.3.2 Data used

- Sentinel-2 L2A, Landsat 5,8 Level 2 Surface Reflectance. These datasets are the main source for estimation of in field variability of yields and definition of management zones. The infield yield variability is calculated from 8-yiers history of EO data.
- Yield maps from grain harvesters - shp point data. Data from harvesters for validation of the trial results
- Farm oriented Linked Data (field and crops, field boundaries in a farm, Yield mass data for some fields) in N-triples format, Linked Open Data (Czech LPIS, Soil maps, erosion zones, water buffers) in N-triples format. Farm data and other datasets from publicly available sources which contains information relevant for farm in way which allows easy analysis and publication of the data in the context with other data.

3.8.4 Business value and impact

To evaluate the results of the pilot, the following KPIs were defined and the values were obtained based on the results of the trials.

<table>
<thead>
<tr>
<th>KPI name</th>
<th>KPI Description</th>
<th>Goal Description</th>
<th>Base Value</th>
<th>Target Value</th>
<th>Measured Value</th>
<th>Unit of value</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>EO processing area</td>
<td>Area of processed EO data</td>
<td>Covering the maximum of pilot farm area</td>
<td>1500</td>
<td>8300</td>
<td>10000</td>
<td>Ha</td>
<td></td>
</tr>
<tr>
<td>Zone delineation accuracy</td>
<td>Accuracy of management zones delineated by field survey and yield maps. Estimated as the deviation to yield zones</td>
<td>Increase the quality of field zoning</td>
<td>50</td>
<td>75</td>
<td>75</td>
<td>%</td>
<td></td>
</tr>
</tbody>
</table>
Fertilizers use efficiency | Increase of fertilizers use efficiency | Increase of fertilizers usage efficiency was influenced by the drought occurrence during the vegetation period 2018 and 2019
--- | --- | ---
5 | 10 | 8 | %

The general way to increase profit of farm is optimization of use of resources to reduce cost of inputs, increase yields, or combination of both of these approaches. The effort to reduce cost of inputs, if handled properly, also has positive effect on environment.

Agriculture is influenced by many external factors and some are difficult to predict. Many farmers are very conservative and reluctant to adopt new technologies unless they are convinced that technology will give them a noticeable economic advantage in order to make investments in this technology profitable. In conditions of uncertainty affecting agriculture, investing in some expensive technologies can be found by many farmers to be risky if, for example, adverse weather conditions can reduce the benefits of a technology or service. One factor that is very important for success is the cost of technology or service. Pilot B1.4 focused on services that make it possible to increase resource efficiency in a very cheap way.

Although some other services and technologies that increase resource efficiency in agriculture can provide better results under favourable conditions, price and uncertainty may be a disincentive for farmers to use these services.

Developing technologies and services that use freely available EO data to calculate long-term in field variability of yields is what makes it possible to offer decision support services dependent on this data at a low cost. Yield maps obtained by monitoring yields on harvesters confirmed the 75% accuracy of results obtained from freely available EO data, which is more than satisfactory given the costs of the service.

Another factor that will support the commercial success of such a service in at least some countries is that the service does not require a high level of adoption of ICT by the farmer. Field boundaries can be obtained from publicly available LPIS data, so cooperation by the farmer in obtaining input data is not required.

These services are particularly suitable for fields with large area, because in the case of small fields, the low spatial resolution of input EO data is an obstacle. The Czech Republic is a country where, for historical reasons, larger fields prevail, compared to most Western European countries, so there are a large number of farms where this service is suitable for a large percentage of their fields.
3.8.5 Summary and outlook

During the pilot, an approach has been tested that allows farmers to increase the productivity of their business with minimal additional investment and minimum service costs. The approach is based on technology lines, using as much as possible data from publicly available sources, thereby reducing data costs.

Due to the spatial resolution of the input earth observation data, this approach is particularly suitable for farms with a large field area. This is the predominant type of farm in the Czech Republic and some other countries.

3.9 Pilot B2.1: Machinery management

3.9.1 Introduction, motivation and goals of the pilot

Management of the work of tractors and other agricultural machines is an important part of planning activities on the farm. Data from these machines is important in the context of other farm data.

Tractor work is an important part of the farm’s costs and its monitoring is needed to calculate the cost of growing crops in each field. This monitoring is also important for quality control of the work performed and for safety reasons. This pilot is focused mainly on collecting telematic data from tractors and other farm machinery and analysis of these data in relation with other farm data. The main goal is to collect and integrate data and receive comparable results. A challenge associated with this pilot is that a farm may have tractors and other machinery from manufacturers that use different telematic solutions and data ownership and sharing policies.

3.9.2 Pilot set-up

The partners participating in this pilot are Zetor, Lesprojekt, FederUnacoma, Este and UWB. Zetor represents tractor manufacturers. Lesprojekt is a partner developing some components and performing data integration. UWB provides and develops other DataBio components used in the pilot. FederunaComa and Este are partners who provide support for the protocols and standards used and Este also incorporates its telematics solution into the pilot.

Machinery management pilot has no spec trial site as the data from tractor during trials are collected wherever the monitored tractors move. Tractors used in both trials are owned by Zetor and collected data are being used mainly by the Testing and Development Department of Zetor company.

Most of the tractors are rented by farmers and Zetor Testing and Development Department monitors them in real-time operation during farm work using Zetor’s telemetry solution. Other tractors are operated in Zetor testing facilities. Farmers use these tractors for their daily activities on the farm, while Zetor uses telemetry to monitor the reliability of tractors and their systems and to identify problems and plan maintenance. Consent to the collection and processing of data is part of the tractor rental contract.
50 tractors owned by Zetor company are involved in trial stage 2. The models of Zetor tractors used in DataBio project involves Crystal 160, Crystal 170 HD, Forterra 140 CL, Forterra 140 HD, Forterra 150 HD, Forterra 140 HSX, Major CL, Major HS, Proxima CL 100, Proxima 110 GP, Proxima 120 HS.

All of the tractors are equipped with monitoring units and telemetry service developed by external supplier and adjusted to Zetor’s need according Zetor’s requests.

The pilot consists of two trial periods as all other agriculture pilots. It was decided that during trial stage 2 additional monitoring units will be deployed on several Zetor tractors in parallel with above mentioned monitoring units to test different ways of data collecting and processing.

The role of Lesprojekt was to use the data gathered in Trial stage 2 in relation with Farm related data from other sources and test if the data can be used for the same purposes as data gathered from other tractors outside DataBio project. As the farmers who uses tractors rented from Zetor company aren’t member of DataBio projects, data about farms and fields will be limited only to those which are publicly available, mainly as part of public LPIS dataset. Lesprojekt also tested data from tractors by other manufacturers gathered outside DataBio project for comparison of information contained in the data and evaluation of their usability farm related analysis.

3.9.3 Technology used

3.9.3.1 Technology pipeline

- UWB/SensLog: Service, for the collection, processing and publication of sensor data. Senslog is required by FarmTelemetry service.
- LESPRO/FarmTelemetry: Extension of SensLog for processing, analysis and publication of data from mobile sensor units. Tractors are considered to be a mobile sensor unit.
- LESPRO/HSLayers: Visualisation of data from tractors and other farm data.
- D2RQ Server: Transformation of the Linked Data from the mapping file of SensLog data and publishing the data on the fly
- LESPRO/Data model for PA: Linking data from tracts with other farm data.

3.9.3.2 Data used

The following datasets were used during the pilot implementation:

- LPIS: Publically available farm data containing information about fields.
- Tractors data in FarmTelemetry: Data gathered before DataBio project used for comparison and Zetor data imported from Zetor telemetry or through Este technology pipeline.
- Original data from Zetor Tractors. Data collected by Zetor in the database of Zetor’s telemetry solution provided by third party developer.
3.9.4 Business value and impact
The trials included tractors owned by Zetor, which are used by farmers in real operation, where the data collected primarily serves the purposes of Testing and Development department of Zetor Company. The data collection process parameters are optimized for the needs of Testing and Development department, and the values of these parameters are different than would be optimal for evaluating the efficiency of agricultural production and the quality of the work.

This applies in particular to the frequency of data collection. This is not a key issue in the case of tractors that are directly owned by the farmer, as the data collection parameters can be adjusted for their needs using the same solution. More details are provided in the deliverable D1.3.

However, while the data collection parameters are not optimal for the purposes of the DataBio project, the technology lines that have been tested during the DataBio project open up new opportunities for Zetor and the partners involved in the pilot.

Currently on the way of Zetor tractors to precision farming, Zetor's management gives priority to further development of the telematics solutions they are currently using. Nevertheless, the pilot explored other options for accessing Zetor data and new ways to use them, and opened the way for these additional access options to be used in parallel with the solution Zetor currently uses, giving the farmer more freedom to use data from their tractors in any way they need.

The decision on further direction of Zetor telemetry and possible opening for third party applications will depend on Zetor management strategy.

3.10 Pilot C1.1: Insurance (Greece)

3.10.1 Introduction, motivation and goals of the pilot
The pilot focused on the evaluation of a set of tools and services for the agricultural insurance market, which aim to eliminate the need for on-the-spot checks for damage assessment and promote rapid payouts. The pilot was based on the fusion of heterogeneous data (EO data, field data) for the assessment of damages at field level.

The methodology of the pilot activities involved the integration of high-power computing and EO-based geospatial data analytics for conducting damage assessment with data from IoT agro-climate stations for field-level condition monitoring. The convergence of the aforementioned technologies in a single dedicated framework is expected to deal effectively with insurance market demands which require a smooth transition from traditional insurance policies (expensive, require human experts for damage assessment) to more flexible index-based insurances. Index-based insurance provides transparency and reduces bureaucracy since it is based on objective predefined thresholds. It has low operational costs requiring minimal human intervention.
On top of that, this new type of insurance can eliminate field loss assessment, adverse selection and moral hazards since the whole process is fully automated, meaning that the point where the pay-out starts (trigger) and the point where the maximum pay-out is reached (exit) are based on a prespecified fixed model per crop. Key stakeholders of the pilot are the farmers, who wish to insure their crops against weather-related systemic perils (e.g. floods, high/low temperatures, and drought), and INTERAMERICAN, a major Greek insurance company with increased interest in agricultural insurance products. The pilot activities were performed mainly at Northern Greece targeting high-impact annual crops (e.g. tomato, maize, cotton, wheat etc.).

3.10.2 Pilot set-up

The pilot worked with annual crops (e.g. tomato, maize, cotton) of high economic interest to the Greek agricultural sector, in several regions of Northern Greece in particular (Evros, Komotini, Thessaly).

The pilot evaluated incidents like floods and heatwaves that fall under the definition of the climate-related systemic perils. The pilot effectively demonstrated how Big Data enabled technologies and services dedicated for the agriculture insurance market can eliminate the need for on-the-spot checks for damage assessment and promote rapid payouts. The role of field-level data has been revealed as their collection and monitoring is important in order to determine if critical/disastrous conditions are present (heat waves, excessive rains and high winds). Field-level data can be seen as the “starting point” of the damage assessment methodology, followed within the pilot. Moreover, regional statistics deriving from this data can serve as a baseline for the agri-climate underwriting processes followed by the insurance companies who design new agricultural insurance products.

NP led the activities for the execution of the full lifecycle of the pilot with technical support from FRAUNHOFER and CSEM. Moreover, INTERAMERICAN was actively engaged in the pilot activities, bringing critical insights and its long-standing expertise into fine-tuning and shaping the technological tools to be offered to the agriculture insurance market.

The goal of the pilot was to enable a better management of the damage assessment process (reduction of the required time) and to support other processes of the insurance companies. Goal achievement was measured by defining specific key performance indicators (KPIs). Baseline KPIs were measured and compared to achievements after the pilot finished. Table 9 provides an overview of baseline and goal achievements.

3.10.3 Technology used

3.10.3.1 Technology pipeline

Data collection: To provide services to the Insurance companies a set of heterogeneous data is required in different spatial and temporal resolutions. Moreover, historical data is critical for shaping insurance products and conducting effective assessments. Data abundance holds the key for creating sound insurance products/tools. More aspects about data requirements
and exploitation can be found in D1.3 [REF-01]. To collect all this data several data collection modules are used:

- In-situ telemetric stations provided by NP, so called GAIAtrons, that collect data related to weather, soil and plant,
- Modules for the collection and pre-processing of earth observation products, the extraction of higher-level products and assignment of vegetation indices at parcel level

**Data processing:** The collected data was processed in different combination through several complementary data processing components (both Pilot components and DataBio components) provided by different partners. The DataBio components that supported data processing are the following:

- **GAIABus DataSmart Machine Learning Subcomponent (NP):** the specific component supports: EO data preparation and handling functionalities. It also supports multi-temporal object-based monitoring and modelling for damage assessment.
- **GAIABus DataSmart Real-time streaming Subcomponent (NP):** This component supports:
  - Real-time data stream monitoring for NP’s GAIAtrons Infrastructure installed in all pilot sites
  - Real-time validation of data
  - Real-time parsing and cross-checking
- **Neural Network Suite (CSEM):** This component was used as a machine learning crop identification system for the detection of crop discrepancies that might derive from reported weather-related catastrophic events.
- **Georocket, Geotoolbox and SmartVis3D (FRAUNHOFER):** This component has a dual role: It is a back-end system for Big Data preparation, handling fast querying and spatial aggregations (data courtesy of NP), as well as a front-end application for interactive data visualization and analytics.

**Data visualisation and presentation:** After the data are processed, they need to be provided in an understandable and decision relevant way suitable also for end users. The main component in this category is NeuroCode (NP). Neurocode allows the creation of the main pilot user interfaces for the end users (insurance companies). An additional DataBio component providing information visualization functionalities is Georocket (FRAUNHOFER).

### 3.10.3.2 Data used

The specific pilot makes use of the following data assets that can be acknowledged for their Big Data aspects (in terms of volume, velocity etc.):

1. **Sensor measurements (numerical data) and metadata (timestamps, sensor id, etc.):**
   This dataset is composed of measurements from NP’s telemetric IoT agro-meteorological stations (GAIAtrons) for the pilot sites. More than 200 GAIAtrons were
fully operational at several agricultural areas of Greece, collecting > 30MB of data per year each with current configuration (measurements every 10 minutes).

2. **EO products in raster format and metadata**: This dataset is comprised of ESA’s remote sensing data from the Sentinel-2 optical products (55 tiles for the whole area of Greece). High volumes of satellite data were processed in order to extract the necessary information about crop health.

3. **Parcel Geometries (WKT), alphanumeric parcel-related data and metadata (e.g. timestamps)**: A dataset comprised of agricultural parcel positions expressed in vectors along with several attributes and extracted multi-temporal vegetation indices associated with them. The volume of this dataset is about 1 GB/year. The update frequency depends on the velocity of the incoming EO data streams and the assignment of vegetation indices statistics to each parcel. Currently, new Sentinel-2 products are available every 5 days approximately and the dataset is updated in regular intervals.

### 3.10.4 Business value and impact

There is a constantly increasing need for agricultural insurance services, due to the adverse effects of climate change and the lack of sufficient compensation frameworks. From their side, insurance companies with offerings for the agricultural sector need to have precise and reliable systems that will facilitate the damage evaluation processes and will ensure swift and fair compensation to those who actually deserve it, thus, allowing follow-up/reactive measures to be undertaken and supporting food security in general.

In the two trial periods of DataBio, tailored agri-insurance tools and services have been developed with and for the agri-insurance companies that perform EO-based damage assessment at parcel level and target towards evolving to next-generation index-based insurance solutions. The pilot results clearly show that data-driven services can facilitate the work of the insurance companies, offering tools that were previously unavailable and were responsible for severe bottlenecks in their day-to-day activities including (see D1.3 for more details):

- long wait for official damage evaluation reports,
- dependence on the human factor,
- difficulties in prioritizing work after receiving several compensation claims.
The defined goals have been achieved and this was validated by a set of KPIs used in the specific pilot that support the exploitation potential of the pilot (see Table 9).

**Table 9: Pilot C1.1 KPIs**

<table>
<thead>
<tr>
<th>KPI description</th>
<th>Base value</th>
<th>Target value</th>
<th>Measured value</th>
<th>Unit of value</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy in damage assessment</td>
<td>No prior information available</td>
<td>&gt;80</td>
<td>95% precision</td>
<td>%</td>
<td>Results are available in real-world data, capturing disasters resulting from extreme weather events (July 2019 - Komotini region - Cotton cultivation affected by floods). As our first priority was to notify and assess the most-affected parcels, validation was focus on positive predictions. Precision reached ~95% effectively showing that data-driven solutions can significantly prioritize and reduce the work required by an expert evaluator.</td>
</tr>
<tr>
<td>Decrease in the required time for conducting an assessment</td>
<td>Several months</td>
<td>Several days</td>
<td>Two weeks approximately</td>
<td>Days, weeks or months</td>
<td>This KPI depends on the availability of reliable EO-data in the post-disaster period. Cloud presence or absence plays a critical role in defining the required time for the assessment. We usually need at least 2 post-disaster EO-based measurements to reach reliable conclusions and based on Sentinel2 measuring resolution,</td>
</tr>
</tbody>
</table>
Number of crop types covered | Initially no crops were being covered by the system | 7 | 7 crop types (based on specific requirements from the insurance company) for all 55 tiles covering Greece. | plain number | 7 major annual crop types were modelled as suggested by the insurance company for the whole Greek arable area (55 tiles x 7 crops = 385 models in total created) and namely: cotton, rice, maize, maize silage, tomato, corn, potato. In addition, continuous NDVI monitoring (measuring NDVI fluctuations before and after a disastrous incident) can be actually applied to any crop type to assess damages at field level.

3.10.5 How-to guidelines for practice
The proposed solution is based on mature technologies and high-quality data, in order to ensure high accuracy and quality for the designed tools and services. EO-based methodologies were used in order to extract useful information from EO products for:

- damage assessment targeting towards a faster and more objective claims monitoring approach just after the disaster,
- the adverse selection problem. Through the use of high-quality data, it will be possible to identify the underlying risks associated with a given agricultural parcel, thus, supporting the everyday work of an underwriter,
- large scale insurance product/risk monitoring, that will allow the insurer to assess/monitor the risk at which the insurance company is exposed to from a higher level

More and more insurance companies are interested in entering the agricultural market, which exhibits high value, due to its vulnerability to extreme weather phenomena. However, before they integrate such technology- and data-driven tools, they need to be persuaded that these tools will help them reduce operational costs by minimizing the human intervention and ensuring high quality of services. The involvement of one of the largest insurance companies in Greece in this pilot (INTERAMERICAN) helps in bringing the proposed solution closer to the market and with their precious feedback, it will be more easily available for commercial exploitation.

3.10.6 Summary and outlook
The proposed pilot worked on the provision of a set of data-powered tools supporting agric- insurance use case scenarios (damage assessment, underwriting, continuous monitoring). It makes use of various heterogeneous data, advanced data analytics algorithms and modern technologies while at the same time it takes any complexity out of the picture. What became
evident is the fact that data abundancy holds the key in delivering even more precise solutions. Data abundancy can address issues relevant to the multi-parametric nature of the problem as different climate-related perils affect dissimilarly different crop types within their various phenological stages.

In the context of DataBio, NP has initiated a close collaboration with INTERAMERICAN (https://www.interamerican.gr/), a major Greek insurance company with a clear target in co-designing and jointly launching next-generation agri-insurance products in the near future (including index-based products). This collaboration among NP and INTERAMERICAN is expected to continue (contributing to the sustainability of the projects outcomes) in the next years, also as part of another high-profile research project, H2020 e-shape (https://e-shape.eu/), where NP is a key partner in S6P4 “Resilient and Sustainable ecosystems including Agriculture and food” and INTERAMERICAN the pilot’s co-designer. Also investigated is the possibility to offer the agri-insurance services of INTERAMERICAN alongside with the Smart farming ones of NP as part of a joint exploitation plan (and vice – versa, i.e. Smart farming services alongside agri-Insurance ones). This will allow both companies to widen their market share.

3.11 Pilot C1.2: Farm Weather Insurance Assessment

3.11.1 Introduction, motivation and goals of the pilot

Agricultural insurance protects against loss of or damage to crops or livestock. It has great potential to provide value to farmers and their communities, both by protecting farmers when shocks occur and by encouraging greater investment in crops. This concept is particularly evident if considering current challenges related to climate change effects and increasing of world population. However, in practice insurance effectiveness has often been constrained by the difficulty of designing good products and by demand constraints. The objective of proposed pilot is the provision and assessment on a test area of services for agriculture insurance market, based on the usage of Copernicus satellite data series, also integrated with meteorological data, and other ground available data by using bid data methods and AI tools.

Among the needs of the insurances operating in agriculture besides the more consolidated procedures of damage assessment by means of Earth Observation techniques, one of the most promising is the evaluation of risk parameters estimation down to parcel level.

For the risk assessment phase, the integrated usage of historical meteorological series and satellite derived indices, supported by proper modelling, allow to tune EO based parameters in support to the risk estimation phase. The availability of this information allows a better estimation of potential risky areas and then a more accurate pricing and designing of insurance products. These advantages could positively impact the increase of insurance penetration. Moreover, the definition of key parameters related to the field lost by using machine learning based approaches will support the design of innovative insurance products as parametric insurance that are currently very promising for farmer protection.
3.11.2 Pilot set-up

The pilot has been realized considering potato crop in Netherland. In particular the following products have been generated (for more details see D1.3 [REF-01]):

**Weather risk map:** weather-based risk map is intended to show the occurrence of extreme weather events, heavy rains in particular, in order to identify areas with possible high damage frequency.

**Intra-field analysis:** single parcel analysis to detect the growth homogeneity and evidencing irregular areas in the parcel, providing an indicator of field anomalies.

Detection of parcels with anomalous behaviours and identification of more influencing parameters: identify the parameters (weather or soil related) with the dominant impact on the crop yield.

Trying to identify the parameters (weather or soil related) with the dominant impact on the crop yield such as Normalized Difference Vegetation Index (NDVI) measurements the following approach was first considered:

For the 2017 dataset we went through the following steps for each one of the crop types (potato):

1) Split parcels into two datasets.
2) Use the first part of the dataset for the clustering and create groups using satellite, meteorological measurements and soil characteristics aggregated on the level of one or two months considering a full growing season from March to October.
3) Characterize / label each group based on the NDVI values of their parcels.
4) For each parcel try to identify in which cluster / group belongs considering its measurements from March to October.
5) After selecting the group, it belongs, use the prediction model that have been trained in the measurements of the parcels that belong in the same cluster and predict NDVI values.

Due to the limited number of usable measurements for the different parcels for the half of the dataset, we could not apply the prediction and feature selection per cluster. For that reason, we used the full dataset of 2017 considering SAR and meteorological measurements (such as precipitation, cumulative precipitation, temperature and cumulative temperature) and soil characteristics for the prediction of NDVI values after 14 days or any other preferable time window, e.g.: use the SAR and meteorological measurements for the 30/06/2017 and predict NDVI value for 14/07/2017. The goal was to try to identify which are the dominant parameters that affect the growing of the parcels for each crop type. For the prediction and feature importance we used random forests. The higher the value of the importance for a
feature the stronger the correlation with the NDVI value. The performed activity reveals that temperature is a factor with high impact on NDVI of potatoes.

Partners involved in the pilot and their role have been:

- Satellite Service Providers and Research and technology Organization (e-GEOS, VITO, Exus): Added value maps and products providing information for risk and damage assessment to be used by insurances in the agriculture domain;
- ICT Expert (Exus): provision of machine learning technology
- Meteorological and Environmental EO service provider (MEEO): Providing weather data and value-added products
- End Users and local agronomic expert (NBAdvice): definition of requirements/provision of input crop data/ validation of the service

3.11.3 Technology used

3.11.3.1 Technology pipeline

The following figure summarizes the technology pipeline used in the pilot:

![Technology Pipeline for pilot C1.2](image)

3.11.3.2 Data used

The objective of proposed pilot has been the provision and assessment on a test area of services for agriculture insurance market, based on the usage of Copernicus satellite data series, also integrated with meteorological data and other ground available data. For the risk assessment phase, the integrated usage of historical meteorological series and satellite-derived indices, supported by proper modelling, demonstrated the potential of EO based products in support to the risk estimation and parametric insurance design phase. Nevertheless, the lack of data about losses from the Insurance (due to privacy issues) did not
allow to perform the initially planned activities. As said the developed method needs an additional set up phase together with the final user (availability of losses dataset is essential for the analysis) to reach an adequate maturity level but preliminary results are very promising for supporting risk assessment for potatoes.

3.11.4 Business value and impact

Results are promising in terms of general procedures and methods. These need to be tested over larger areas and compared with validation data provided by the final users (insurance). The data availability is a crucial challenge for this market considering the very restricted dissemination level of the information and the high competitive level. In fact, the insurance companies are not interested in supporting the development of products that can be available also for their competitors. To overcome these potential limitations, a set up phase of the service in operative environment is necessary in close cooperation with the insurance company involved. This collaboration has the potential to transform the tested methods into operative services filling the existing gap between prototype development and final product.

In order to analyse the benefit of the tested technology for the insurance industry (risk estimation also by means of machine learning), it is important to define the three levers of value in insurance market:

1. Sell More
2. Manage Risk Better
3. Cost Less to Operate

The activity performed in the pilot impacts essentially the point “Cost Less to Operate”. One clear way to reduce operating costs in insurance is to add information and increase automation to complex decision-making processes, such as underwriting. To keep processing costs in check, many insurance carriers have a goal to increase the data available in support to a more precise and automatic risk evaluation in support of the underwriting. In fact, the use of decision management technologies like risk maps, machine learning, and artificial intelligence the insurance can reduce the time spent to analyse each contract and focus team members on higher value activities. Moreover, the identification of parameters that most affect the crop yield performed in the pilot, can support an innovative insurance typology called “parametric insurance”. This particular insurance typology is revolutionizing the insurance industry allowing to dramatically cut operative costs removing the in-field direct controls.

The first step in building a parametric product is determining the correlation between the crop losses and a particular index representative of the climate event associated to the loss. The activity performed in the pilot by using machine learning approach is to identify the most important parameter affecting the crop yield that can be the basis for a parametric or index-based insurance.

In terms of business impact, quantify the potential effect of the proposed solution for the Insurance Industry is a complex issue considering the work necessary to transform the
methodology in an operative service. Just to provide some business projection it can be considered that direct European agricultural insurance premiums in 2016 were 2.15 million euros (estimated by Munich RE).

![Crop Insurance in the EU - Premium Volume m€³](image)

**Figure 9: Premium volume distribution for crop insurance in Europe**

It can be considered that around 70% of this amount is spent by Insurances to reimburse damages and the remaining 30% is used to pay internal costs and re-insurances. Considering this dimension and considering the row and very preliminary estimation obtained by the pilot, it is possible to assume that the cost that can be saved by using EO based services in support of risk assessment is around 2% of the total cost used by the insurance to pay internal costs.

### 3.11.5 How-to guidelines for practice

The remote sensing literature offers numerous examples proposing Earth Observation techniques to support insurance, for example in the assessment of damage from fire and hail. To date, however, few operational applications of remote sensing for insurance exist and are operative. Many scientific papers claiming potential applications of remote sensing, typically stress the technical possibilities, but without considering and prove its contribution in terms of “value” for the insurer. The discrepancy between the perceived potential and the actual uptake by the industry is probably the result of two main reasons:

- technological solutions not adequate and too expensive, in relation to the valued added
- over-optimistic assumptions by the remote sensing community, regarding the industry’s readiness to adopt the information by remote sensing.

Despite this situation, EO can still play a central role in supporting the insurance market in agriculture trying to design services that can really bring value to the users. This is the case of supporting in field verification and parametric insurance products (innovative insurance products). The present pilot investigates these services demonstrating the potentiality and opening up the route for new collaboration with users.
As said, the methodology needs to be a pre-operational set up phase in close collaboration with the insurance company. In fact, the developed method can be applied to different areas and crops.

### 3.11.6 Summary and outlook

The objective for the pilot was to find useful services for the insurance to gain more insight about the risk and the impact of heavy rain events for crops in the Netherlands. Potato crops are very sensitive to heavy rain, which may cause flooding of the field (due to lack of runoff) and saturation of the soil. This may cause the loss of the potato yield in just a few days. Areas of greater risk can be charged with higher costs for the farmer. The investigated correlation among precipitation and losses can also support the identification of index for parametric insurance products.

Moreover, instead of just raising the premium, the intention of the pilot was to be able to create awareness and incentives for farmers to prevent losses. Therefore, the services served multiple purposes.

Weather is an important factor in crop insurance, because it represents a critical aspect influencing yield. The analysis of the long-term precipitation, categorized in threshold values, for intense rain events, gave insight in the areas with higher risk. During climate change, these numbers may change. Therefore, a service about the changing patterns is an interesting service.

In the pilot, the relation between one single event and the potential yield loss have been analysed. For this purpose, an annotated set of data, where actual losses were determined, was necessary. Because of the privacy issues related to sharing the damage data, the location of damages fields could not be pinpointed precisely enough for correlation to the EO data. Without the details about historical events this relationship could not be determined. Based on the information available to the team, it has been possible to determine the events and give information about damage risk for other areas. A service, based on the alert that a heavy rain event took place, would be useful for gaining insight about the impact on other locations.

In order to find the most limiting aspect in the crop development we created a dataset based on the Sentinel-2 raster size to combine NDVI with SAR, precipitation (cumulative), temperature and soil type. Splitting up the dataset in subsets per potato type the precipitation was the most determining factor. Unfortunately, we could not find the connection with the heavy rain, because the training set was not sufficient for that analysis. The developed methodology, however, is valuable for further analysis, not limited to insurance topics and can be extended to other crops in support to risk assessment and also for design new insurance products such as parametric insurance.
3.12 Pilot C2.1: CAP Support

3.12.1 Introduction, motivation and goals of the pilot

In the framework of EU Common Agricultural Policy (CAP), farmers can have access to subsidies from the European Union, that are provided through Paying Agencies operating at national or regional level. For the provision of the subsidies, Paying Agencies must operate several controls in order to verify the compliance of the cultivation with EU regulations. At present, the majority of the compliance controls are limited to a sample of the whole amount of farmers’ declarations due to the increased costs of acquiring high and very-high resolution satellite imagery. Moreover, they are often focused on a specific time window, not covering the whole lifecycle of the agriculture land plots during the year.

The free and open availability of Earth Observation data is bringing land monitoring to a completely new level, offering a wide range of opportunities, particularly suited for agricultural purposes, from local to regional and global scale, in order to enhance the implementation of Common Agricultural Policy (CAP). Nowadays, satellite image time series are increasingly used to characterize the status and dynamics of crops cultivated in different agricultural regions across the globe.

Pilot C2.1 CAP Support provides products and services, based on specialized highly automated techniques for processing Big Data, in support to the CAP and relying on multi-temporal series of free and open EO data, with focus on Copernicus Sentinel-2 data. The main goal of the approach is to provide services in support to the National and Local Paying Agencies and the authorized collection offices for a more accurate and complete farm compliance evaluation - control of the farmers’ declarations related to the obligation introduced by the current CAP. The pilot services demonstrate the implementation of functionalities used for supporting the subsidy process in verifying specific requests set by the EU CAP.

3.12.2 Pilot set-up

Pilot C2.1 CAP Support addresses two different situations, materialized through the two areas of interest: one located in Italy, managed by e-Geos and one in Romania, managed by Terrasigna.

Under the framework of the DataBio project, Terrasigna ran CAP support monitoring service trials during 2017, 2018 & 2019 for 10 000 sqkm AOI in Southeastern Romania with the following features:

- large plots (the land parcels have been chosen in order to be as large as possible with a minimum degree of land fragmentation);
- diversity of crops (the selected area contains as many different types of crop types as possible);
- accessibility (any point or parcel within the area could be easily accessed during field campaigns / field observations and was situated relatively close to Bucharest).
Moreover, during the second phase of the pilot activities, the service was extended at national-scale level, providing results for the whole territory of Romania, for both 2018 and 2019. The total surveyed area exceeded 9 million ha, corresponding to more than 6 million plots of various sizes and shapes. 21% of the total number of plots within the test areas had surfaces below 1 ha. Several challenges had to be addressed: the large area to be surveyed, characterized by geographical variability and the presence of small/narrow plots, crop diversity and high cloud coverage. Therefore, the service tailored for agriculture monitoring in Romania had to:

- be able to address small/narrow plots distributed over diverse location;
- provide results for a broad variety of crops;
- make use of the Copernicus Sentinel temporal resolution;
- provide early warnings to the decision makers.

Figure 10: Romania - total declared area and number of plots registered for CAP support (2019). Data Source: Agency for Payments and Intervention in Agriculture (APIA), Romania
The analysis for the Romanian area of interest was entirely based on Terrasigna's toolbox for crop determination, consisting of a set of in-house developed algorithms for calculating CAP support-related products.

In relation to the Italian case, the objective of the trial has been to set up a quick methodology, based on the computation of markers, in relation to predefined scenarios in terms of crop type and reference periods, during which agricultural practices must take place, to detect LPIS\GSAA parcel anomalies in terms of crop type or crop family, with respect the last update (LPIS) or farmer’s declaration (GSAA) and to re-classify the parcel itself. The methodology works at parcel level, therefore several markers as ploughing, presence, harvesting, are computed for each parcel depending on the specific crop type. The workflow is based on the following steps:

- Download of Sentinel-1 and Sentinel-2 satellite data from repositories.
- Preprocessing of Sentinel-2 data in order to mask clouds and related shadows
- Generation of spectral indices from pre-processed Sentinel-2 satellite data, also by composing data from different images, to be used for markers computation
- Intersection of Sentinel-2 spectral indices and pre-processed Sentinel-1 data with parcels to be monitored
- Computation of markers at parcel level

For the definition of markers, it must be considered that each of them must be defined according to the geographic location, and specific algorithm and related parameters must be identified, therefore requiring a proper tuning by leveraging on time series analysis. This operation is supported by the analysis, for each crop, of its spectral behaviour along time, in order to identify from a mathematical point of view, markers related to specific activities.

The methodology has been applied on the AOI of the project in Veneto (Varese Province) where the LPIS 2016 was available.

3.12.3 Technology used

3.12.3.1 Technology pipeline

All the components used within the pilot are EO-related. The execution for the Romanian area of interest was entirely based on Terrasigna's toolbox for crop determination, consisting of a set of in-house developed algorithms for calculating CAP support-related products. Following an automatic learning process, the system becomes capable of recognizing several types of crops, of the order of several tens. The processing chain used during Trial 2 included the following activities:

A. Data ingestion
   - Earth Observation data used within the framework of the CAP Support Pilot is derived from two different sensors (Sentinel-2 and Landsat-8), which requires an effort to harmonize the spatial resolution and the footprint of the native pixel grids.

B. Scene classification
• Use of statistical parameters for the crop classification (obtaining the native structure of semantic clusters and applying them at tile level);
• Granting of a semantic profile for the individual classified scenes (the pixels get the fuzzy labels belonging to the crop class).

C. Time series analysis

D. Construction of graphical products and analytical data
• Concatenation of tile-level results;
• Delivery of single channel or RGB maps illustrating crop types, crop compliance, classification confidence etc.;
• Extraction of numerical, quantitative syntheses based on the delivered results.

The following products have been obtained:

• Maps of the main types of crops, for an annual agricultural cycle completed;
• Intermediate maps with the main types of crops, during an annual agricultural cycle (they may serve as early alarms for non-observance of the declared crop type);
• Early discrimination maps between winter and summer crops;
• Layers of additional information, with the degree of confidence for the crop type maps delivered;
• Maps of the mismatches between the crop type declared by the farmer and the one observed by the application;
• NDVI maps nationwide for a period of time, uncontaminated by clouds and cloud shadows;
• Lists of parcels with problems, in order of the surfaces affected by inconsistencies;
• National maps with RGB aspect mediated for a period of time, uncontaminated by clouds and shadows.

The validation of the results against independent sources revealed promising results, with an accuracy higher than 90% for more than 10 crop types.

Anomalies detection: The markers computed in relation to predefined scenarios in terms of crop type, reference periods and specific thresholds, during which agricultural practices must take place, have been implemented in a decision model to verify parcel’s correct classification. The model has been run for each parcel of the macro-classes considered as suitable for the automatic detection of anomalies.

The specific pilot made use of the following technological DataBio components:

• **FedEO Gateway** – Main purpose: Data Management (Collection, Curation, Access) – EO Collection Discovery, EO Product Discovery, Catalog, Metadata.
• **FedEO Catalog** – Main purpose: The component was used in combination with the FedEO Gateway and Data Manager to setup a complete chain to retrieve and index Sentinel-1, Sentinel-2 or Landsat data and other data available through FedEO on a local processing platform.
**Data Manager** – Main purpose: The component was used in combination with the FedEO Gateway and FedEO Catalog to setup a complete chain to retrieve and catalog Sentinel-1, Sentinel-2 or Landsat data (SciHub and CMR/USGS) and other data available through FedEO on a local processing platform.

**Mosaic Cloud Free Background Service** – Main purpose: Data management and Data curation - keeping an up to date collage (mosaic) of Sentinel-2 and Landsat-8 images, covering the area of interest (AOI) with the latest, cloud free satellite scenes; the fusion and harmonization between images are made only at RGB level, mainly for eye inspection, but also for other possible advanced processing; the whole process chain is independent and self-content, based on cloud and shadows mask extraction, histogram matching procedures and, finally, a pixel based analysis. Backgrounds are updated automatically, soon after a new raw scene is available during the whole trial stage 2 period.

**EO Crop Monitoring Service** – Main purpose: Descriptive analytics – EO data processing. The component is able to assess the agriculture parcels from satellite data and farmers’ declarations in order to create a series of products like, Crop masks, Parcels used maps and Crop inadvertencies maps, based on SITS - Satellite Image Time Series.

**Clouds, Shadows and Snow Mask Tool** – Main purpose: Data curation - EO data preprocessing. The tool produces Sentinel-2 Clouds, Shadows and Snow Masks, based only on raw data, improving the results of the genuine quality assessment band. The results are raster maps (GeoTiff) with 4 label codes: 0 – for no data, 1 – for uncontaminated/ free pixels, 2 – for snow, 3 – for shadows and 4 – for clouds.

All the components used within the pilot are EO-related.

### 3.12.3.2 Data used

The pilot used three main categories of data that are common, with National specific differences, for the two cases:

- **External data** – farmers’ declarations: Pilot C2.1 CAP Support used farmers’ declaration regarding crop types and areas covered as input data. These data have been provided by the Romanian National Paying Agency, as well as its regional offices. For the 10,000 sqkm area of interest, more than 150,000 plots of different sizes have been analysed during the 2019 agricultural season. The analysis performed included parcels of over 0.3 ha, regardless of shape. Of course, the 10-meters spatial resolution made the narrower parcels difficult to properly label. Very related groups of cultures, which have synchronous phenological evolutions and similar aspect have been grouped into crop classes.

- **Optical Earth Observation (EO) data**: Landsat-8 OLI and Sentinel-2 MSI - both Sentinel-2A and Sentinel-2B have been downloaded for the area of interest, for a time interval between March and September 2019. The 10-meter spatial resolution of the Sentinel-2 data enables the survey of the smaller plots that in Romania represent a significant
number of CAP applications. The spectral resolution provides all the necessary information (visible, NIR, SWIR) for observing the crop phenology. On a more general note, TERRASIGNA’s technology uses both Copernicus Sentinel-2 and Landsat 8 imagery for a maximum of information availability and time series density compared to using only Landsat 8 or Sentinel 2 images separately.

- **Field data**: Field data have been collected according to a field tracking plan. Also, this category includes different datasets provided by the Agency for Payments and Intervention in Agriculture (APIA), based on the annual on-site compliance verifications of the farmers that applied for subsidies. All the field data have been used as independent validation data.

### 3.12.4 Business value and impact

The value created by the CAP pilot / EO Crop Monitoring component lays in the increase of efficiency that the payment authority would experience in using the satellite monitoring and Big Data technologies.

A further development of the Crop Monitoring Service is able to provide products tuned in order to fulfil the requirements of the 2015-20 EU Common Agricultural Policy. The developed technique is replicable at any scale level and can be implemented for any other area of interest.

The addressable customers of the project results are:

- APIA (National Subsidy Agency for Agriculture, Ministry of Agriculture) holds responsibility in Romania of the implementation of CAP mechanisms for direct payments. The entire procedure is handled by the Integrated System of Administration and Control (IACS) that also deals with the verification of the compliance of the declarations submitted by the farmers. Currently, a minimum of 5% from the applications is crossed-checked either by field sampling or by remote sensing;
- Italian National and Regional Paying Agencies;
- Similar authorities from other countries;
- Companies developing EO services and applications that could use Big Data technologies.

The potential added value from the CAP supporting services in C2.1 can be qualified with the following KPIs:
### Table 10: Pilot C2.1-Romania KPIs

<table>
<thead>
<tr>
<th>KPI short name</th>
<th>KPI description</th>
<th>Goal description</th>
<th>Base value</th>
<th>Target value</th>
<th>Measured value</th>
<th>Unit of value</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2.1_1 (Values measured for the Italian AOI)</td>
<td>Percentage of LPIS area processed vs global LPIS coverage in terms of hectares</td>
<td>Agricultural territory coverage</td>
<td>N/A</td>
<td>50%</td>
<td>71%</td>
<td>%</td>
</tr>
<tr>
<td>C2.1_2 (Values measured for the Italian AOI)</td>
<td>Percentage of parcels &gt; 0.5 hectares that are processed</td>
<td>Small parcel size capability</td>
<td>N/A</td>
<td>80%</td>
<td>98%</td>
<td>%</td>
</tr>
<tr>
<td>C2.1_3 (Values measured for the Italian AOI)</td>
<td>Parcel anomalous that are not re-classified</td>
<td>Re-classification performance</td>
<td>N/A</td>
<td>10%</td>
<td>2%</td>
<td>%</td>
</tr>
<tr>
<td>C2.1_4 (Values measured for the Romanian AOI)</td>
<td>Processed surface</td>
<td>Agricultural territory coverage</td>
<td>N/A</td>
<td>10 000</td>
<td>Trial 1: 10 000 km²&lt;br&gt;Trial 2: 130 000 km² (whole country)</td>
<td>sqkm</td>
</tr>
<tr>
<td>C2.1_5 (Values measured for the Romanian AOI)</td>
<td>Number of crop types addressed</td>
<td>Diversity. Ability to recognize different crop cultivation patterns</td>
<td>NA</td>
<td>5</td>
<td>Trial 1: 5 crop families&lt;br&gt;Trial 2: 32 crop types</td>
<td>crop types</td>
</tr>
</tbody>
</table>

### 3.12.5 How-to guidelines for practice

Pilot C2.1 CAP Support uses multi-temporal series of Earth Observation data, consisting of Sentinel-2 and Landsat-8 imagery. Earth Observation (EO) data provide wide and repetitive homogeneous coverage, translated into an unprecedented amount of information referred generally as “Big Data”. The technologies benefitting from the data volumes represent a solid solution for a continuous monitoring of CAP compliance. The Sentinel-2 satellites, part of the EU Copernicus data stream, hold an enhanced revisiting time, thus delivering regular coverage over large areas and allowing a uniform observation of the agricultural plots. The superior spectral resolution allows the identification of the phenological growth stages and the distinction between various crop types or classes.
However, the pilot also had to overcome some major drawbacks. The most serious problems that had to be solved and that served as lessons were:

- the use of data S2 and L8 together - which have a different format and resolution;
- correction of the geographical positioning (georeferencing) automatically - which deeply affects the quality of the classification for small or narrow plots;
- selecting the areas of interest from each image - which are not, as it might seem, the areas uncontaminated by clouds and shadows, but the areas where there is vegetal "activity";
- the construction of an algorithm that takes into account the matrix of semantic confusion between cultures - which required finding the natural classes of cultures that can be followed simultaneously, without serious mutual confusion.

Geospatial services together with Copernicus data can provide a really powerful tool for monitoring agricultural dynamics. The end-users, the National Paying Agencies, are able to benefit from the modern and effective near real-time service, based on the principles of sustainable agriculture and saving effort both in terms of costs and time.

A continuous agricultural monitoring service based on the processing and analysis of Copernicus satellite imagery time series is not just a CAP compliance tool but can also offer a great range of supplementary information for both public authorities and citizens.

### 3.12.6 Summary and outlook

CAP policy and activities from National and Regional Paying Agencies can radically benefit from the use of continuous satellite monitoring instead of random and limited controls. This market is one of the more promising for exploiting Copernicus data in the Agriculture domain. In the pilot the same objective has been faced by using two different approaches, one in Romania and the other one in Italy. Both approaches demonstrated their potentiality for the final users.

Terrasigna's proposed methodology has undergone continuous development and improvements over the last 4 years. e-GEOS proposed approach based on markers that demonstrate to be applicable not only for the CAP monitoring opening up the street for future innovation in the market. The highly-automated fuzzy-based proposed approach developed by Terrasigna for the Romanian AOI and from e-GEOS for the Italian case, used within the C2.1 CAP Support pilot allows the performing of Big Data analytics to various crop indicators, being reliable, cost- and time-saving and allowing a more complete and efficient management of EU subsidies, strongly enhancing their procedure for combating non-compliant behaviours.

The free and open availability of Earth Observation data is bringing land monitoring to a completely new level, offering a wide range of opportunities, particularly suited for agricultural purposes, from local to regional and global scale, in order to enhance the implementation of the Common Agricultural Policy (CAP).
3.13 Pilot C2.2: CAP Support (Greece)

3.13.1 Introduction, motivation and goals of the pilot

This highly ambitious pilot launched by NEUROPUBLIC and GAIA EPICHEIREIN in Northern Greece in an area covering 50000ha, targeted the evaluation of a set of EO-based services designed appropriately to support specific needs of the CAP value chain stakeholders. The pilot services relied on innovative tools and complementary technologies that sustained the interconnection with IoT infrastructures and EO platforms, the collection and ingestion of spatiotemporal data, the multidimensional deep data exploration and modelling and the provision of meaningful insights, thus, supporting the simplification and improving the effectiveness of CAP. The pilot activities aim at providing EO-based products and services designed to support key business processes including the farmer decision-making actions during the submission of aid application and more specifically leading to an improved “greening” compliance. The ambition of the current pilot is to deal effectively with CAP demands for agricultural crop type identification, systematic observation, tracking and assessment of eligibility conditions over a period of time. The pilot activities are fully aligned with the main concepts of the new agricultural monitoring approach which will effectively lead to fewer controls, will facilitate and expand the adoption of technology to the farmer communities, will promote the penetration of EO deeper into the CAP line of business and raise the awareness of the farmers, agronomists, agricultural advisors, farmer cooperatives and organizations (e.g. groups of producers), national paying agencies (e.g. Greek OPEKEPE) on how new technological tools could facilitate the crop declaration process.

3.13.2 Pilot set-up

The pilot has completed two rounds of trials in the greater area of Thessaloniki, Greece. The pilot mainly focused on annual crops with an important footprint in the Greek agricultural sector (rice, wheat, cotton, maize, etc.). NP is leading the pilot activities while its main stakeholders are the farmers that benefit from the services and GAIA EPICHEIREIN (end-user) that has a supporting role in the farmers’ declaration process. CSEM and FRAUNHOFER are also involved in the pilot providing their long-standing expertise in the technological development activities.

The goal of the pilot is to support the farmer in the crop declaration process and specifically in the “greening” aid application, so as to reduce errors that might lead to financial losses. Technology used in the pilot.

3.13.3 Technology used

3.13.3.1 Technology pipeline

Data collection: To provide CAP Support services that assist the farmers in their “greening” aid application, a set of heterogeneous data is required in different spatial and temporal resolutions. EO- and parcel-related historical data is needed in order to train crop models for crop classification. To collect all this data several data collection modules are used:
• In-situ telemetric stations provided by NP, so called GAIATrons, that collect ancillary weather data
• Modules for the collection, pre-processing of earth observation products, the extraction of higher-level products and assignment of vegetation indices at parcel level,

**Data processing:** The collected data was processed in different combination through several complementary data processing components (both Pilot components and DataBio components) provided by different partners. The DataBio components that supported data processing are the following:

• GAIABus DataSmart Machine Learning Subcomponent (NP): Supports EO data preparation and handling functionalities. Supports multi-temporal object-based monitoring and modelling and crop type identification,
• GAIABus DataSmart Real-time streaming Subcomponent (NP): This component supports:
  o Real-time data stream monitoring for NP’s GAIATrons Infrastructure installed in all pilot sites
  o Real-time validation of data
  o Real-time parsing and cross-checking.
• Neural Network Suite (CSEM): This component was used as a machine learning crop identification system for the detection of crop discrepancies.
• Georocket, Geotoolbox and SmartVis3D (FRAUNHOFER): This component has a dual role: It is a back-end system for Big Data preparation, handling fast querying and spatial aggregations (data courtesy of NP), as well as a front-end application for interactive data visualization and analytics.

**Data visualisation and presentation:** After the data is processed it needs to be provided in an understandable and decision relevant way suitable also for end users. The main component in this category is NeuroCode (NP). Neurocode allows the creation of the main pilot UIs in order to be used by the end-users (FSCs of GAIA EPICHEIREIN). An additional DataBio component providing information visualization functionalities is Georocket (FRAUNHOFER).

### 3.13.3.2 Data used
The specific pilot makes use of the following data assets that can be acknowledged for their Big Data aspects (in terms of volume, velocity, etc.):

1. **Sensor measurements (numerical data) and metadata (timestamps, sensor id, etc.):**
   This dataset is composed of measurements from NP’s telemetric IoT agro-meteorological stations (GAIATrons) for the pilot sites. More than 20 GAIATrons were fully operational at the area of interesting, collecting > 30MB of data per year each with current configuration (measurements every 10 minutes).
2. **EO products in raster format and metadata**: This dataset is comprised of ESA’s remote sensing data from the Sentinel-2 optical products (2 tiles for the area of interest). High volumes of satellite data were processed in order to extract the necessary information for identifying crop type and potential declaration discrepancies.

3. **Parcel Geometries (WKT), alphanumeric parcel-related data and metadata (e.g. timestamps)**: A dataset comprised of agricultural parcel positions expressed in vectors along with several attributes and extracted multi-temporal vegetation indices associated with them. The volume of this dataset is about 1 GB/year. The update frequency depends on the velocity of the incoming EO data streams and the assignment of vegetation indices statistics to each parcel. Currently, new Sentinel-2 products are available every 5 days approximately and the dataset is updated in regular intervals.

### 3.13.4 Business value and impact

CAP is a framework that defines the operation of farmers in EU. It aims to guarantee minimum levels of production, ensuring constant food supply in the EU at affordable prices for the consumers, as well as to ensure a fair subsistence level for those that depend on agriculture. On top of that, it provides the means for improving the competitiveness of the EU agricultural products. The CAP includes a number of policies related to farming, the environment, rural development and agricultural markets. Direct payments to farmers are considered a major pillar towards that direction. It helps them to stabilise their incomes and support the sustainability of their farms, as long as they meet the predefined criteria.

Direct payment schema includes the basic payment support (per-hectare) and a series of others supports targeting specific objectives or type of farmers. Those are the greening, young farmers, voluntary couple support, areas with natural constrains, single area and redistributive. In order to access the payments, farmers have to submit an aid application declaring, inter alia, all the agricultural parcels on the holding every year. Each year the Farmer, the beneficiary, must provide evidence to document his/her eligibility. His/her choices during the one-off submission process have great financing impact and may lead to losses or, even worst, trigger penalties.

The Greek farmers (beneficiaries) have direct access to the aid application system through the Web app of OPEKEPE (the Greek Paying Agency), which can be used by the farmer without cost. However due to the complexity of the process and the risks involved, the majority of the farmers prefer to benefit from the collection and advisory services offered by certified offices that help them complete the aid application using the OPEKEPE Web app.

GAIA EPICHEIREIN, through its associated network of Farmer Service Centers (FSCs) provides collection and advisory services to the Greek Farmers concerning the submission of the aid application for direct payments, including eligibility pre-check mechanisms for error reduction and proof provision. The total number of holdings in Greece for 2016 where 686,818. GAIA subsidy services are mainly oriented to aging small-sized farmers, which own 80% of the
holdings in Greece. Over the last two annual periods, GAIA EPICHEIREIN provided collection services and cross compliance checks to 76% of the holdings.

Even if GAIA EPICHEIREIN has a market share of 76%, the ongoing CAP changes and trends, the differentiations in the internal market and the new business plans for smart farming (driven by the evolution in sensor and space technology) indicate that GAIA EPICHEIREIN needs to evolve its services in order to keep its competitive advantage and sustain its market share.

Apart from the business value for the partners involved, the pilot introduces concrete benefits for the farmers and the agri-food sector as well. The results of the pilot effectively exhibit that EO-based crop identification services, tailored for monitoring greening compliance, offer a layer of protection against errors in the declaration process which could lead to a significant financial impact for the farmer. Additionally, and from a higher level, agricultural monitoring approaches could contribute to more efficient funding absorption, thus securing investments and progress in the agri-food sector.

The KPIs used in the specific pilot are listed in the following table along with the final DataBio results (measured values) that support the exploitation potential of the pilot.

Table 11: KPIs of the pilot C2.2

<table>
<thead>
<tr>
<th>KPI short name</th>
<th>KPI description</th>
<th>Base value</th>
<th>Target value</th>
<th>Measured value</th>
<th>Unit of value</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2.2_1</td>
<td>Decrease in false crop type declarations following the supporting services vs what would be expected based on historical data</td>
<td>10</td>
<td>8</td>
<td>9.4 of initial declaration were identified as potentially problematic</td>
<td>%</td>
<td>A 9.4% of the initial farmer declarations exhibited potential errors based on the followed methodology. The farmers were notified and received follow-up information. The offered advisory services allow the farmers holding parcels of &gt;10ha and more (prerequisite for the greening aid application) to be compliant to the greening requirements in respect to crop diversification, thus, favouring a further reduction to the percentage of erroneous declarations that threaten funds absorption.</td>
</tr>
<tr>
<td>C2.2_2</td>
<td>Accuracy in crop type identification</td>
<td>No prior information</td>
<td>&gt;80</td>
<td>98.5</td>
<td>%</td>
<td>The overall accuracy of the crop classification methodology used in the pilot reached 98.5%. Respectively, precision reached 99.1%, recall 94.6% and f1-measure 94.7%. Some crop types seem to be more easily</td>
</tr>
<tr>
<td>C2.2_3</td>
<td>Number of crop types covered</td>
<td>Initially no crops were being covered by the system</td>
<td>7</td>
<td>7 crop types supported in the greater region of Thessaloniki, Greece</td>
<td>identifiable (maize, cotton, rice, cereals) whereas others appear to be more challenging (rapeseed and tobacco)</td>
<td></td>
</tr>
</tbody>
</table>

### 3.13.5 How-to guidelines for practice

The offered DataBio solutions will allow the farmer (beneficiary) to deal effectively with the greening requirements. More specifically, DataBio solutions will be a valuable tool within the suite of digital CAP Support services offered by GAIA EPICHEREIN’s and its FSCs that support the crop declaration process. During the process and usually after the declaration period closes and error-checking tools are applied, the FSC would be able to check the farmer’s claim for the greening requirements, examine the results and inform the farmer for follow-up activities that better serve his/her interests.

### 3.13.6 Summary and outlook

DataBio offers new business opportunities and aims to directly improve GAIA EPICHEREIN’s position in providing advisory services for the farmers that lead error reduction during the crop aid declaration and protect their best interests.

Within DataBio, NP, together with other technological partners, has initiated a series of CAP Support activities for providing supporting tools and services, in line with the commands of EC’s new agricultural monitoring approach. This effort is expected to continue in the next years (contributing to the sustainability of the projects outcomes) as part of another high-profile research project, H2020 NIVA ([https://www.niva4cap.eu/](https://www.niva4cap.eu/)) where NP is a key partner and is close collaborating with the Greek paying agency (OPEKEPE). This will allow evolving/further validating the DataBio-enhanced services, so that they progressively improve the suite of CAP Support tools offered by GAIA EPICHEREIN.

### 3.14 Summarizing Analysis of Agricultural Pilots

The DataBio agricultural pilots provided some insights about the BDT market in agriculture. As DataBio pilots show, there are already providers of BDT for the European agricultural sector on the market. Technology providers that are already present on the market offer typically an end-to-end BDT pipeline consisting of components for data collection, data processing and visualisation of decision-relevant results as well as alert services.
Depending on their ownership and goals, the following types of BDT providers can be distinguished:

- Commercial providers of BDT pipelines are privately owned companies and operate for profit. The DataBio partners NEUROPUBLIC and GAIA EPICHEIREIN are one example of commercial BDT providers.
- State-owned research institutions do not operate for profit but should support farmers and achievement of societal goals, as for example better use of social common resources as irrigation water, use of less fertilizers and similar. DataBio partners that belong to this group are e.g. Tragsa, VITO, and others.
- Commercial providers providing additional functionality, services or components for BDT in agriculture (i.e. IBM, ATOS and others).

The typical business model for providing BDT for agriculture on the market is “DaaS”, i.e. Data-as-a-Service. This business model entails that data services, i.e. decision relevant data based on BDT are provided through a cloud in a per use manner. While commercial providers typically apply a certain payment model as for example a specific price per ha for using the technology, state-owned providers offer at present the data services for free (i.e. VITO’s services are for free in Belgium).

Providers of components that enable extensions of basic pipelines and additional services use different business models as license or open source-based business models. During the DataBio project commercial and state-owned providers of BDT pipelines were able to experiment and extend their pipelines with additional modules and functionality.

Providing the BDT services in a DaaS manner, means that BDT pipeline providers invest in the BDT pipeline and infrastructure, while end-users (farmers) entail variable costs as they pay per use. From the perspective of technology providers, to refinance the high upfront investment costs and achieve scaling effects (i.e. decreasing costs per user), it is necessary to acquire as quick as possible a high number of end-users. Thus, in this early stage of the market for BDT in Europe, an approach of “land grabbing” as a market entrance can be observed.

The European market for BDT in agriculture has some specific characteristics:

- In Europe smaller farms prevail (see Figure 11).
- Farmers might be sceptical against technology and it is necessary to build up trust for the technology
- The application of BDT is mostly based on algorithms that need to be trained and calibrated and require suitable historical data of high quality. Thus, it is not possible to enter the market and to offer the whole potential of the technology without establishing a collaboration with farmers.
Because of these specific characteristics of the market for BDT in agriculture, BDT providers target farmer associations and collaboratives as main customers. This is also the case in most of the agricultural pilots of DataBio that are connected to various farm collaboratives. Acquiring farm collaboratives enables bigger areas where the technology is applied, higher income and the opportunity to establish a trustable relationship to agricultural experts that can promote the technology to the farmers. Such experts can also help to interpret results into decision relevant information. Furthermore, farm collaboratives might already have or organise the collection of necessary data from farms. They might also dispose with historical data.

Overall, as also the DataBio project shows, BDT provides new opportunities for farmers and technology providers. However, for farmers BDT results in a high lock-in effect with the technology provider.
4 Business Analysis of Forestry Pilots

4.1 Introduction

In most of Europe forest management follows management plans that are typically updated every several years. These plans often lack effective implementation and monitoring methods to allow forest owners, managers and regulators to validate the progress in achieving the target objectives. In DataBio integrated tools were developed to support forest management planning that maximizes timber production and economic yield, while been capable to consider non-wood products and conservation areas. The DataBio forestry pilots aimed to optimize the use of tree resources, improve the identification of forest damage and provide forest health data. The pilots were carried out in four countries (Belgium, Czech Republic, Finland and Spain) and under three tasks (Multisource crowdsourcing services, Forest Health and Forest Data Management Services). They are listed in the table below. There were initial six main pilots, two of which were split into two sub-pilots, for a total of eight pilots.

Table 12: Overview of Forestry pilots

<table>
<thead>
<tr>
<th>Task</th>
<th>Pilot ID</th>
<th>Pilot title</th>
</tr>
</thead>
<tbody>
<tr>
<td>T 2.2 Multisource and data crowdsourcing /e-services</td>
<td>2.2.1</td>
<td>Easy data sharing and networking</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.2.2</td>
<td>Monitoring and control tools for forest owners</td>
</tr>
<tr>
<td>T2.3 Forest Health / Remote/Crowd sensing, Invasive species/damage</td>
<td>2.3.1-FI</td>
<td>Forest damage remote sensing (@Finland)</td>
</tr>
<tr>
<td></td>
<td>2.3.1-ES</td>
<td>Forest damage remote sensing (@Spain)</td>
</tr>
<tr>
<td></td>
<td>2.3.2-FH</td>
<td>Monitoring of forest health</td>
</tr>
<tr>
<td></td>
<td>2.3.2-IAS</td>
<td>Invasive alien species control and monitoring</td>
</tr>
<tr>
<td>T2.4 Forest data management services (forecast/predict)</td>
<td>2.4.1</td>
<td>Web-mapping service for the government decision making</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.4.2</td>
<td>Shared multiuser forest data environment</td>
</tr>
</tbody>
</table>

4.2 Pilot 2.2.1: Easy data sharing and networking

4.2.1 Introduction, motivation and goals of the pilot

This pilot aimed to develop and pilot standardized procedures for collecting and transferring data utilizing the Wuudis Service and DataBio platform from silvicultural activities executed in the forest. The Wuudis Service and the Wuudis Networking features were applied in the pilot.
Data sharing and a collaborative environment enable improved tools for sustainable forest management decisions and operations. Data becomes accessible to forest owners and other end users interfacing with e-contracting, online purchase and sales of timber and biomass (e.g. Metsaan.fi eService and Kuutio.fi). Higher data volumes and better data accessibility increase the probability that the data will be updated and maintained.

In this context, Wuudis Solutions launched the new version of Wuudis Service on 26 October 2018, which enables easy data sharing and networking particularly between forest owners and forest authority personnel. In this pilot, the Wuudis Service worked as a data sharing platform between authorities and end users providing mobility and data modification tools for the users.

4.2.2 Pilot set-up
The participants in the pilot included Wuudis Solutions (MHGS) and the Finnish Forest Centre (METSAK). The goal was to develop and pilot standardized procedures for collecting and transferring data from silvicultural activities executed in the forest, to support more sustainable forest management decisions and operations.

4.2.3 Technology used

4.2.3.1 Technology pipeline
The key elements in the piloted technology pipeline were 1) the online Wuudis Service, 2) the Wuudis mobile application and 3) the Metsaan.fi eService by the Finnish Forest Centre. The mobile application was developed to enable work quality monitoring in a standardized way (sample plots “kemera”).

All current real estate data is integrated from the Metsaan.fi eService to the Wuudis platform for DataBio pilots. Data is transferred via the Finnish forestry standard XML format. This initial forestry data is very crucial for the pilot, because every update affects the initial data directly. To enable this data connection, an interface to the authority system was developed in the Wuudis Service, applying strong user identification.

Additionally, a feature was developed to allow the user to decide what data he/she wants to send back to the authority.

The monitoring data consisted of the following information: forest estate, geometry of compartments, type of the forest work, sample plot locations, measured data per sample plot, measurement averages per compartment, measurement date and user information. The quality control data was added to the forest data standard during 2018.

4.2.3.2 Data used
The data generated and used in the pilot can be summarized as follows:

- Wuudis crowdsourced data collected by users and available through cloud APIs in the Finish forest information standard
- Open forest data provided by the Finish Forest Centre
- Open forest data provided by the National Land Survey of Finland
4.2.4 Business value and impact

The exploitable results of the pilot for the Wuudis service consist of European wide (or even beyond) commercial use of Wuudis as data sharing and networking tool across all forest stakeholders (Finnish, English, Spanish and French languages available).

Based on the experiences in the pilot and verified in two customer surveys, the following savings were measured due to both use of Wuudis instead of the old procedures based on individual laptops and software applications and having more accurate forest data:

- 83.13% decrease in operating costs of stakeholders using Wuudis for data sharing and networking
- 70.45% savings in working time of authority experts for daily routines related to forest data management
- 72.23% improvement of the quality of forest data (i.e. data for forest resources and stand information)
- Increased customer satisfaction
- Improved information about forest biodiversity

The standardized procedures and methods developed in the pilot can help in customization and scaling solutions such as Wuudis globally. Country and industry wide standardized procedures and methods developed in the Metsään.fi service and between Metsään.fi and Wuudis integrations will help to build similar solutions even globally in the forestry sector.

4.2.5 Summary and outlook

Overall, the pilot was successful in enhancing easy data sharing and networking. The pilot specified the requirements for refining and showing the crowdsourced forest data to METSAK’s IT system. The implementation of the new functionalities was carried out in collaboration with the METSAK’s development team and other METSAK’s projects.

Wuudis data sharing and networking tool are validated for use across all forest stakeholders. The standardized procedures and methods developed in the pilot can help in customization and scaling similar solutions globally.

4.3 Pilot 2.2.2: Monitoring and Control Tools for Forest Owners

4.3.1 Introduction, motivation and goals of the pilot

This pilot aimed to develop standardized procedures and applications for forest owners for collecting, monitoring and transferring data utilizing the Wuudis Service and the DataBio platform. The Wuudis Monitoring feature was applied in the pilot. The data collected through the Wuudis Service can be exported to third party IT systems through standard interfaces. In this pilot, an end-to-end data transfer solution was developed between the Wuudis Service and METSAK’s Metsaan.fi eService.
Based on the Wuudis Monitoring feature, we developed and used a work quality monitoring application (available for iOS and Android mobile platforms) in order to feed the forest inventory master data in real time operations into the METSAK’s databases, Metsaan.fi and METSAK’s forest resource data. High quality updates were provided for strategic planning through the Wuudis platform and for paying subsidies for cleaning and treating young seedling and young forest stands in a controlled way by METSAK. The collecting methods were to improve work quality and customer satisfaction and increase competition between contractors, resulting in decrease of care work costs of forest owners.

4.3.2 Pilot set-up
The key participants in the pilot included Wuudis Solutions (MHGS) and the Finnish Forest Centre (METSAK). The goal was to develop and pilot standardized procedures and applications for forest owners for collecting, monitoring and transferring data. In particular, the plan was to develop a work quality monitoring application in order to feed the forest inventory master data in real time operations into the METSAK’s databases, Metsaan.fi and METSAK’s forest resource data.

The overall target was to improve work quality and customer satisfaction and increase competition between contractors, resulting in decrease of care work costs of forest owners.

4.3.3 Technology used

4.3.3.1 Technology pipeline
Wuudis launched a work quality monitoring application (Laatumetsä in Finnish) during November 2018 to enhance better work quality monitoring while processing subsidy applications. This application, made available for iOS and Android mobile platforms, was taken into use by the METSAK personnel, forestry service providers and forest owners.

Forest damage (such as storms, snow, pests and diseases) monitoring through standardized procedures was developed together with METSAK, as well as easy-to-use mobile tools for these damage monitoring needs and non-wood product monitoring needs. Finally, the data was integrated with METSAK’s forest resource data systems. This allows forest owners and forest specialists willing to monitor and report forest damage information to authorities through a direct access to METSAK’s master database.

Forest damage crowdsourcing is a feature of the Laatumetsä (work quality monitoring) application.

4.3.3.2 Data used
The data generated and used in the pilot can be summarized as follows:

- Wuudis crowdsourced data collected by users and available through cloud APIs in the Finnish forest information standard9

---

9 The Wuudis Service data model is based on the Finnish forest information standard. All development activities during the DataBio project that will affect to the Wuudis data model are based on the Finnish forest information standard.
• Open forest data provided by the Finish Forest Centre including forest resource data as well as GIS data
• Quality control data for young stand improvement and tending of seeding stands
• Storm and forest damages observation and possible risk areas

4.3.3.3 Reflection on technology use
The work quality monitoring application was taken into use by over 500 users during the pilot. In addition, French and Spanish language were implemented into the work quality application.

During the year 2018, campaigns were run by METSAK to activate private forest owners and citizens to report forest damages via the Laatumetsä mobile application. Snow damage observation campaign was launched in the beginning of 2018 as the Eastern part of Finland was suffering from a very heavy snow load. The crowdsourcing campaign that was run in early 2019 by METSAK after the Aapeli storm hit heavily the western part of Finland was very successful. Many observations were received from the citizens and this helped to analyse the magnitude of the storm damages and to react faster for avoiding the possible larger damages for instance regarding the insect invasion for this specific area by activating the wood procurement actors. The successful campaigns demonstrate the adequateness of the technology choices made.

4.3.4 Business value and impact
The observations METSAK received via the Laatumetsä application helped to analyse the possible impacts in harvesting potentials for the snow damage impacted areas and thus prevent damages and growth losses.

The most logical approach was to evaluate these innovations through real customer use cases in forestry operations and via the end user customer survey. The final survey was completed in Q3/2019 and the results can be summarized as follows:

• Savings in working hours and authentic reporting due to the one-click subsidy application
• Increase in number of threats detected on time and faster forest thread detection
• Increased updates of forest data by each of the involved stakeholders (forest owners, contractors and authority representatives) resulting in overall higher quality forest data.
• Faster forest damage management lead time
• Increased number of forest experts using Wuudis (over 500 users at the end of the project)
• Improved and increased forest data coverage
• Improved customer satisfaction

Furthermore, the results of this pilot included:

• expansion of the data sharing and networking features on Wuudis
As the Wuudis solution for networking is already available on the market, it was possible to enter foreign markets as follows:

- closing a business deal with Galician Wood Cluster for Wuudis training to Galician stakeholders (4 events),
- and finally, Walloon forest authorities and forest management associations having shown interest towards the SPACEBEL-Wuudis concept implementation.

### 4.3.5 How-to guidelines for practice

The lessons learned from this pilot were related to marketing and dissemination activities, as slow start in marketing activities resulted in initially rather small number of users of forestry-care work quality-monitoring application (Laatumetsä). Increased marketing activities through several champagnes run by METSAK resulted in a faster increase of users. Overall, crowdsourcing initiatives for forest data require at least at the launch of the app higher marketing efforts.

### 4.4 Pilot 2.3.1: Forest Damage Remote Sensing

#### 4.4.1 Introduction, motivation and goals of the pilot

The goal of this pilot was to develop the Forest Inventory system for damage identification on the Wuudis Service based on remote sensing (satellite, aerial, UAV) and field surveys. In the DataBio project, selected Big Data partners integrated their existing market-ready or almost market-ready technologies into the Wuudis Service and the resulted solutions were piloted with the Wuudis users, forestry sector partners, associated partners and other stakeholders.

Earth Observation (EO) data from multispectral optical aerial, unmanned aerial vehicles (UAV) and satellite sensors present the optimal way to timely collect information on land cover over areas of various sizes. Particularly the availability of the Copernicus Sentinel-2 data and the applicable free data policy present a great opportunity for developing low cost commercial applications of EO downstream services in monitoring the environment. Online platforms, such as the Forestry TEP\(^{10}\) and the EO Regions\(^{11}\), enable creation of services for efficient processing of satellite data to value-added information.

#### 4.4.2 Pilot set-up

The consortium for this pilot consisted of: 1. VTT Technical Research Centre of Finland (VTT), 2. SPACEBEL, Belgium, 3. Technical University of Denmark (DTU) and 4. Wuudis Solutions, Finland. In addition, Forest Management Institute (FMI/UHUL), Czech Republic, coordinated their own pilot activities with this pilot. Wuudis Solutions was the original pilot leader, while for the last three months of the project VTT took over the pilot leadership. All activities were linked with the Wuudis platform, and inter-platform connections were developed between

---

\(^{10}\) [https://f.tep.com](https://f.tep.com)

\(^{11}\) [http://www.eoregions.com](http://www.eoregions.com)
Wuudis and two other platforms coordinated by consortium members: Forestry TEP coordinated by VTT and EO Regions! coordinated by SPACEBEL.

The pilot test sites were located 1) at the Hippala forest estate in Southern Finland, 2) in Wallonia, Southern Belgium, and 3) at a forest property “Barbanza, Enxa, Xian, Dordo, Costa de Abaixo e O Sobrado”, located in the municipality of Porto do Son, A Coruña province in Galicia in North-western Spain. This site is owned by the rural community “Comunidade de Montes Veciñais en Man Común (CMVMC) de Baroña” and managed by ASEFOGA.

In the pilot sites, VTT, SPACEBEL and DTU conducted demonstrations and further development of their forest monitoring applications and services. In parallel, FMI was developing a new methodology for forest health assessment, which allows assessment of forest health in the entire area of Czech Republic. The results of the FMI work can be linked to the online platforms used in the pilot through OGC WMS/WMTS interfaces.

### 4.4.3 Technology used

#### 4.4.3.1 Technology pipeline

In the first stage of the pilot, VTT established a service chain for forest inventory utilizing Sentinel-2 data in the Hippala test site in Southern Finland. The chain utilizes the Forestry TEP platform for data sourcing and processing, and the VTT software Envimon and Probability in processing on the platform.

The generated forest variable estimates include: stem number; stem volumes for pine, spruce, broadleaved and total; diameter; basal area; and height.

For easy integration of satellite maps and analysed (highlighted) theme maps, standard OGC WMS or WMTS interfaces were used as a starting point. The output was made available for integration in the Wuudis end-user system via WMS interface from the Forestry TEP.

In addition, Wuudis tree-wise monitoring MVP (minimum viable product) service was launched in June 2018 and sold to leading forest management associations (forest management associations of Pohjois-Karjala, Savotta and Päijänne) and forest industries in Finland. Over 4000 hectares were monitored during summer 2018 by Wuudis network of service providers.

In the second stage of the pilot, delivery of the VTT forest variable estimates in the XML-based Finnish Forest Information Standard format was enabled. This allowed forest management plan geometries to be used to extract remote-sensing pixel-based information and to store it back in the forest management plan, which could be used in Wuudis or any other online platform.
Figure 12: Forestry TEP

Figure 12 presents Forestry TEP, an online platform that enables efficient exploitation of the Copernicus Sentinel satellite data in forest monitoring and analysis. The satellite data is sourced from ESA and made available on platform, via the underlying infrastructure and data services of CREODIAS. Forestry TEP provides processing services and tools and serves also as a platform for new services, such as the Envimon and Probability tools of VTT that were used in this pilot.

Inter-platform connections between EO Regions! and Wuudis were also developed, enabling numerous possibilities to feed the Wuudis service in geographical and dendrometric content. This allows several scenarios for combined use of EO Regions! and Wuudis. Users can e.g. work independently on the platform to import their data, or they can use mobile applications to encode dendrometric data, or they can order forestry services from either platform. In all of these cases, the users will benefit from the increased offering and functionalities provided by the connection between EO Regions! and Wuudis.

In addition, a study on the usability of Senop hyperspectral camera for Boron deficiency mapping was performed at the Polvijärvi test site in Finland. Finally, several demonstrations of the transferability of the technical capabilities were performed in a test site in Galicia, Spain, where teams from DTU, FMI and VTT applied their methods in coordination with Wuudis platform.

4.4.3.2 Data used

The pilot utilized several different types of remotely sensed datasets as well as field data. Remotely sensed datasets included Sentinel-1 and Sentinel-2 satellite data and airborne hyperspectral remote sensing data. In the Finnish test site, sample plot data by the Finnish
Forest Centre was used as reference in the estimation model training. In the Galicia test site, Wuudis Solutions staff conducted field work, collecting forest variable information from ten forest stands. The measured information included six different forest structure variables: 1) species, 2) age, 3) basal area, 4) stem count, 5) mean diameter and 6) mean height. The field data was recorded in the Wuudis platform, together with photographs. In addition, all available information from the forest estate stands were recorded into the Wuudis system.

The Sentinel satellite data was found to be very useful for operational forest monitoring applications in online platforms. The systematic acquisition scheme and high temporal frequency (i.e. short revisit time) provides high amounts of data suitable for high temporal resolution service provision. The high number of spectral bands usable for forest monitoring purposes (10) in the Sentinel-2 satellites, combined with the 10-20 m spatial resolution, enables development of high-quality forest monitoring applications based on Sentinel data. Furthermore, Sentinel-1 and Sentinel-2 data is stored e.g. in Copernicus Data and Information Access Service (DIAS) platforms and can be accessed directly with processing platforms such as Forestry TEP.

For many forest monitoring applications, field data has a crucial role. In the pilot, two different types of datasets were used. The national coverage sample plot data by the Finnish Forest Centre used as reference in Finland was confirmed to be very suitable for the online applications demonstrated in this pilot. However, such field datasets are not available in all countries. For example, in Spain, in the Galicia test site, field data collected by Wuudis staff needed to be used. The small amount of data available limited the usability of the applications demonstrated. This highlighted the importance of the availability of operationally collected field data for forestry applications.

4.4.3.3 Reflection on technology use
Overall, the pilot demonstrated well the benefits of technology use in forest monitoring through a range of forest inventory applications utilizing Big Data and online Big Data processing approaches. These applications and services were further developed to better suit inter-platform operations and improve user experience. The services were integrated with the Wuudis platform, demonstrating the possibilities and benefits of inter-platform interactions. The resulting solutions were piloted with Wuudis users, forestry sector partners, associated partners and other stakeholders. The experiences from the pilot confirm the value of Big Data in forest monitoring and encourage further development of Big Data approaches for forest monitoring purposes.

4.4.4 Business value and impact
The entire pilot focused on development and integration of marketable forest inventory services into the Wuudis platform and other related platforms. Overall, the pilot results were successful in demonstrating the usability of a range of forest inventory applications on the platform. In general, the pilot demonstrated the possibility of inter-platform connections and for service provision, which enable wider exploitation of the services developed in this pilot
(and other services). The services are applied on the respective platforms and exploitation of the services is growing.

Table 13 presents the KPIs measured during the pilot. In addition to the measurable KPI’s, the pilot aimed at testing and demonstrating new services for forest damage monitoring. Several services were successfully tested and demonstrated in Belgium, Finland and Spain, utilizing several online platforms and inter-platform connections. This will increase the service offering in all the involved platforms (Wuudis, Forestry TEP and EO Regions) and enable higher revenue in the future.

**Table 13: Pilot 2.3.1 KPIs**

<table>
<thead>
<tr>
<th>KPI description</th>
<th>Goal description</th>
<th>Base value</th>
<th>Target value</th>
<th>Measured value</th>
<th>Unit of value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usability of tree-wise monitoring service MVP</td>
<td>Goal is to sell the tree-wise monitoring service MVP to forest management associations in Finland</td>
<td>0</td>
<td>no target value</td>
<td>3</td>
<td>n.a.</td>
</tr>
<tr>
<td>Surface processed with MVP</td>
<td>The goal is to increase the area processed using the MVP service</td>
<td>0</td>
<td>4000</td>
<td>5000+</td>
<td>ha</td>
</tr>
</tbody>
</table>

The pilot is a good example on how research results are used in business development. The pilot has brought together new commercial partners for added-value services on top of Wuudis platform. Business agreement between Wuudis Solutions and SPACEBEL regarding the distribution of the Wuudis services to the forest users of the EO Regions! platform and the commercialization of SPACEBEL’s earth observation forest products in the Wuudis platform is already in place. Negotiations are ongoing with VTT concerning inter-platform connections between Wuudis and Forestry TEP, as well as with the Finnish state forest enterprise Metsähallitus in order to scale the boron deficiency analysis concept to state forests in 2020.

Due to the pilot, Wuudis Solutions is now able to better understand the needs of Spanish market through real pilots. Wuudis Service has been tested in real business environment and the results were encouraging. Wuudis Solutions expects to acquire more customers in Spain through intensive dissemination activities together with new partners like Agresta and Cotesa. Wuudis Solutions has already secured new implementation and R&D projects in Spain/Galicia (e.g. TEMPO, ICEX, Galician Wood Cluster), and finally established a daughter company, Wuudis Solutions S.L., in Galicia in October 2019. From SME point of view, this is a good example how research results are used in business development in foreign territories.

### 4.4.4.1 Technology reflection on the general level

On a general level, more effort is needed to increase interest in platform services in the forestry community and ensure smooth user experience. In many parts of Europe, the forestry sector has a long history with strong traditions on practices. In order to increase interest from
the user side, the service providers now need to be able to 1) convince the forestry stakeholders on the concrete benefits of online service provision utilizing Big Data and 2) further polish practices for cooperation between service providers to ensure smooth and effortless user experience for increased interest.

The importance of local promotional activities and local components in the service chain cannot be overemphasized. This can be achieved by strong local involvement, which enables direct connection to local datasets and stakeholders. This, in turn, allows fine-tuning of the provided services according to local needs.

4.4.5 How-to guidelines for practice

For forest monitoring stakeholders (whether it be private forest owners, forestry companies or public entities), the best avenue for Big Data utilization for forest monitoring purposes is through online platforms. There are several levels of online platforms enabling utilization of EO data for forest monitoring purposes. These include e.g. the DIAS platforms providing data access to forestry related Big Data, and several platforms providing forestry related applications and services. These platforms include e.g. the Forestry TEP, EO Regions! and Wuudis platforms used in this pilot.

The application platforms provide direct access to satellite data and auxiliary datasets, and ready-made applications for the utilization of the data for forest monitoring purposes. In addition, e.g. Forestry TEP offers application development interface, where users can develop their own applications utilizing the Big Data available on the platform. Inter-platform connections bring further benefit to the users through wider service offering.

4.4.6 Summary and outlook

Overall, the pilot demonstrated well the benefits of technology use in forest monitoring through a range of forest inventory applications utilizing Big Data and online Big Data processing approaches. In addition, the pilot highlighted 1) the technical transferability of online platform-based forest inventory services and 2) importance of local involvement in fine-tuning services to meet local needs. All of the services tested in the pilot areas were technically implemented successfully. However, stronger local involvement in service definition and field data provision would be needed to provide more reliable and meaningful results for the users.

The pilot was very successful in further developing capabilities to perform comprehensive and near real-time quantitative assessment of forest cover over the project pilot areas. This type of near real time forest monitoring allows monitoring of forest damages, deforestation and forest degradation.

The pilot was also generally successful in creating the inter-platform connections. However, the challenges of integration of services between platforms and service providers became clear during the pilot. Best practices for inter-platform cooperation between service providers (both technical and financial) need to be developed to enable smooth and effortless user experience.
experience receiving the maximum benefit from the range of service providers working together.

4.5 Pilot 2.3.2-FH: Monitoring of Forest Health

4.5.1 Introduction, motivation and goals of the pilot

The goal of the pilot was to set up a methodology and tools based on remote sensing images (satellite + aerial + UAV) and field data for the early warning and monitoring of the health status of forests in large areas of the Iberian Peninsula. The tools should support decision making by public bodies and should be specifically adopted the final users’ needs and requirements. The work was focused in monitoring of the health of Quercus sp. forests affected by the fungus Phytophthora cinnamomi Rands and of the damage in Eucalyptus plantations affected by the coleoptera Gonipterus scutellatus Gyllenhal.

The first motivation for this pilot was to set up an optimised EO-based system for monitoring the health of Quercus ilex and Q. suber forest areas (mapping + assessment tools), so that authorities and forest owners will be able to optimise forest management resources and decision-making processes. The second one was to develop an efficient mapping and assessment tool for monitoring and assessing the damages caused by Gonipterus in eucalyptus plantations, in order to adapt management procedures and minimize economic losses.

Phytophthora cinnamomi affects very severely several tree species, among them Quercus ilex and Q. suber, in different areas in Spain (Extremadura, Andalucia, Castilla y León, Castilla La Mancha, Madrid), causing a great ecological and economic problem. Detection is currently performed through direct observations or through data sampling and analysis in the laboratory.

Related to this problem, Orthophotos (RGB + NIR) can be used to identify dead trees and locate possible affected areas of Quercus forests, and their evolution. Detailed RPAS (RGB + multispectral) and field data will be collected in the selected sampling plots and analysed (in-situ or in the lab); RPAS will be employed for intensive monitoring to establish the relationships between satellite-derived indexes and biophysical parameters from field data. A more general health status monitoring for big areas could be developed based on Sentinel 2-derived vegetation indices. Ancillary information about environmental conditions and management would be combined with EO data to detect areas under stress and, consequently, more prone to be affected by Phytophthora and/or other plagues and diseases.

Gonipterus scutellatus defoliates eucalyptus plantations very severely. Eucalyptus is one of the main commercial species in the North of Spain (Galicia, Asturias and Cantabria), where Gonipterus produces huge economic losses by impeding the development and growth of trees. Authorities (Xunta de Galicia), industrial companies from the paper sector (ENCE) and forest owners need an economic, systematic and objective tool for affected areas identification and damage assessment.
Sentinel 2-derived vegetation indices can be used for a systematic monitoring of the health status in the selected study areas; anomalies will show areas where Gonipteruous can be defoliating, which will be checked on the field, either visually or using RPAS. This information will be combined to define an optimal methodology for data acquisition and analysis. The rate of defoliation that can be detected must be analysed and established for different EO data.

Therefore, a mapping and assessment tool has been developed for monitoring the damages caused by Gonipterus in eucalyptus plantations, as well as for damage assessment, in order to adapt management and minimise economic losses. Besides this, EO-based solutions provide Public Bodies with valuable information and tools to help decision-making. An EO-based system for monitoring the health of big forest areas has been set up (mapping + assessment tools), so authorities are able to optimise forest management resources.

4.5.2 Pilot set-up

These works have been developed in collaboration with the Spanish Ministry of Agriculture and Environment MAPA (Área de Recursos Genéticos Forestales) in the framework of the “Phytophthora Working Group” coordinated by MAPA. All Spanish Autonomous Communities affected by Phytophthora participate in this Working Group, as well as research centers and Universities. Field data from Castilla – La Mancha Autonomous Community have been used, from Buenaventura y Oropesa (Toledo), where massive deaths of Quercus ilex trees caused by Phytophthora outbreaks have been registered. The orthophotos employed come from the Spanish National Plan PNOA (https://pnoa.ign.es/), which is produced every 2-3 years and covers the whole country. PNOA produces Visible and Infrared orthophotos with 0,25-0,5m spatial resolution (depending on the date) provided by the Public Administration. PNOA is led by the Spanish Geographic Institute (IGN). It was decided to employ the 2009 and 2018 orthophotos (available for the whole Spanish territory). In addition, as ancillary information, PNOA - LiDAR data were employed (density of 0,5 puntos/m2).

4.5.3 Technology Used

The works developed focused on acquiring data for establishing the relationships between RPAS and field data, and a correlation model so as to obtain a prediction algorithm for the early detection of decaying trees affected by Phytophthora based on spectral data. For more details regarding the methodology applied please refer to D2.3 [REF-02].

4.5.3.1 Technology pipeline

1. Acquisition of RPAS & Field data. Set-up of RPAS technology.
   - A field campaign was developed in July 2018 in Haza de la Concepción (Cáceres, Spain). 380 ha of sparse forest (‘dehesa’ type), from which 100 ha were selected for data acquisition. Vegetation compounded by Quercus ilex and Quercus suber showing different degrees of affection by Phytophthora. 81 Quercus ilex trees were sampled in 9 plots. Measured parameters were: mean leaf density (measured with LICOR 2200), mean leaf surface and biomass (green and dry), pigment concentration from leaves (chlorophyll and carotenes), crown and trunk
morphology, health status inventory and damage assessment, analysis of soil and roots for determining the presence/absence of *Phytophthora cinnamomi*.

- **RPAS data**: TRAGSA’s eBee+ platform acquired RGB (SODA camera) and multispectral images (Sequoia camera: Green, Red, RedEdge and NIR bands) over the study site. Different spectral indexes related with vegetation activity and pigments were calculated from multispectral RPAS data: NDVI, GNDVI, NGRDI, SAVI, OSAVI, ARI1, ARI2, TCARI y ratio.

2. **Development of a conceptual model for estimating the defoliation degree at tree-level according to the user’s requirements defined by Customers**

- The correlation model was obtained for the study site from the first campaign (100 has) so as to obtain a prediction algorithm for the early detection of decaying trees affected by *Phytophthora* based on spectral data. The extrapolation of the algorithm to the whole area (Haza ‘dehesa’, 380 has) was developed in stage 2.

### 4.5.3.2 Data used

The data used in the pilot can be summarized as follows:

- Sentinel-2 data
- RPAS data created by TRAGSA drones
- PNOA images provided by the Spanish national Geographic Institute
- Field data collected by TRAGSA staff

### 4.5.3.3 Reflection on technology use

Results obtained so far allow to state that it is possible to assess defoliation and assign treatment priorities by using RPAS remote sensing data. Nevertheless, some works must be finished, and improvements need to be made in the methodology so as to obtain an objective, operative and affordable service. Nevertheless, the low density of Eucalyptus crowns and the variety of the understory makes it impossible to automatically delimitate trees, which is key for extending the model at plantation-level and obtaining the risk maps.

### 4.5.4 Business value and impact

In what refers to the pilot methodology using RPAS & field data, the following results and conclusions have been reached:

- **Spatial resolution**: it is necessary to use VHR (Very High Resolution) images which allow the identification of individual trees (≤ 50 cm).
- **Spectral resolution** it is necessary to count on information on the InfraRed wavelengths at least, which allows to assess the status of vegetation
- **Temporal resolution**: the evolution of the disease does not require a very high temporal resolution à the frequency of ortophotos from the Spanish National Plan PNOA (2-3 years), with data available from 2005, is enough ([https://pnoa.ign.es/](https://pnoa.ign.es/))
- A radiometric normalization process is mandatory so as to work with PNOA ortophotos (RGB-NIR) with different acquisition dates (historic database).
In can be concluded that the use of RPAS is interesting for monitoring Phytophthora outbreaks at a local scale. For big surfaces (the area potentially affected is the half South of the Iberian Peninsula), the use of PNOA ortophotos is proposed. These are very interesting results from the management point of view, as PNOA is a free periodic product provided by the Spanish Public Administration.

With the IT and HPC tools already available, the methodology developed could be extended to a lot bigger ‘dehesa’ areas, so that it would be possible to estimate the number of Quercus ilex trees lost in a period of time in a certain province or region. The methodology is very interesting for the periodical monitoring of the vigour status of ‘dehesas’ (analysis of progression/regression of Quercus forests, detection of new outbreaks etc.).

Some of the problems encountered will be solved taking into account technologic solutions developed within the framework of DATABIO project.

Regarding eucalyptus plantations, in what refers to the pilot methodology using RPAS & field data, results obtained so far allow to state that it is possible to assess defoliation and assign treatment priorities at tree-level by using RPAS remote sensing data. The model obtained is properly adjusted to the criteria established by the final user ENCE.

The low density of Eucalyptus crowns and the variety of the understory makes it impossible to automatically delimitate trees, so works cannot be automatized. The goal of extending the model at plantation-level and obtaining risk maps has consequently not been reached.

In what refers to the pilot methodology using Sentinel & field data, no pattern was found which allowed establishing a correlation between Sentinel data and defoliation produced by Gonipterus.

4.5.5 How-To guidelines for Practice

According to the goals stated at the end of stage 1: (i) final adjustments of the assessment of the defoliation degree at tree-level using RPAS & field data were performed and (ii) a pilot methodology using Sentinel & field data was tested.

**Pilot methodology using RPAS & field data:** On the one hand, efforts were employed in simplifying the model based on RPAS data as much as possible, so as to make data acquisition and processing faster and more cost-effective. Results were satisfactory and the assessment obtained is reliable. On the other hand, several classification methods were tested in order to improve the methodology for the automatic extraction of tree crowns, which is necessary to automatize works and also for model extension (vertically and horizontally).

**Pilot methodology using Sentinel & field data:** NDVI derived from Sentinel 2 were used for monitoring *Eucalyptus* plantations. The main limitation encountered is the low number of images free of clouds in the study area. Sentinel 1. VH (linear) and ratio VH/VV (dB) polarizations were also evaluated to monitor *Eucalyptus* plantations.
4.5.6 Summary and outlook

In the monitoring of the health of *Quercus* sp. forests affected by *Phytophthora*, the following business areas could be pursued:

- The use of RPAS proved interesting for monitoring *Phytophthora* outbreaks at local scale. Then, maps of vigour/decay status from field & MS data from RPAS flights can be pursued, providing a reliable solution for small forest managers.
- At regional scale, the use of PNOA orthophotos is proposed for monitoring *Phytophthora*. Then, maps of surviving/dead trees can be produced, providing a reliable solution for forest managers from the Public Administration. These works have actually been developed in collaboration with the Spanish Ministry of Agriculture and Environment MAPA. All Spanish Autonomous Communities affected by *Phythopthora* participate in the Working Group, as well as research centres and Universities. A trial has been presented to the Working Group, and it has been considered of great interest. It is being assessed in order to apply the methodology to different study areas in Spain. It could be extrapolated to the whole area affected by *Phytophthora* in Spain and Portugal.

In the monitoring of the damage in eucalyptus plantations caused by *Gonipterus*, the following business areas could be pursued:

- It was found feasible to assess defoliation and assign treatment priorities at tree-level by using RPAS and field data. Then, Maps which show the ‘degree of defoliation’, and the ‘treatment priorities’ at ‘tree level’ can be produced.
- The low density of *Eucalyptus* crowns and the variety of the understory makes it impossible to automatically delimitate trees. Also, so far, no pattern was found which allowed establishing a correlation between Sentinel data and defoliation produced by *Gonipterus*. Then, no exploitation is feasible at this moment for these.

4.6 Pilot 2.3.2-IAS: Invasive alien species control – plagues

4.6.1 Introduction, motivation and goals of the pilot

Invasive Alien species (IAS) are a big threat for biodiversity in the Iberian Peninsula, the Canary Islands and the Balearic Islands, and cause significant economic losses.

Those invasive species are animals and plants that are introduced accidentally or deliberately into a natural environment where they are not normally found, with serious negative consequences for their new environment. They represent a major threat to native plants and animals in Europe, causing damage worth billions of Euros to the European economy every year. As invasive alien species do not respect borders, coordinated action at the European level will be more effective than individual actions at the Member State level.
Resources are limited and those actions and eradication measures can be complex and very expensive. Therefore, early warning and monitoring are key points for Spanish Public Bodies to be more efficient. The model developed by DataBio project regarding IAS identifies the geographic origin of the biological invasions that could likely affect Spain to provide a detailed spatial assessment of invasion risk and, eventually, to improve IAS management increasing the efficiency of preventive measures.

This DataBio tool has developed a Big Data model for assessing invasion risk in Spain based on a set of factors that strongly influence the geographic pattern and level of invasion as (i) environmental similarity, calculated from bioclimatic variables, (ii) Biodiversity similarity, approached through biogeographic information, (iii) Propagule pressure, estimated from data on trade, tourism, immigration, population and terrestrial transport network and (iv) Ecosystem disturbance, measured from land use and fire frequency.

It has been demonstrated that prevention is the most effective way to face the problem of biological invasions. Consequently, it is important to know the invasion risk in different areas and ecosystems for a better management of the problem. Moreover, targeting in areas and ecosystems at highest risk makes preventive measures most efficient. This technical case, by providing a detailed spatial assessment of invasion risk of IAS (independent from the species), expects to increase the efficiency of preventive measures and so to foster a better IAS management. Sensitive areas, i.e. those with the highest risk of invasion, have been identified and mapped.

Invasion risk is assessed by quantifying Big Data sources as climatic similarity (based on climatic distances and biogeographic information) and ‘propagule pressure’ (addressed by information on tourism, immigration and trade) between Spain and the rest of the world and, finally, the arrival ecosystems level of disturbance will be also incorporated, as well as the biogeographic regions of the territories that act as sources of IAS.

The IAS topic is linked to prevention, which is both an effective and efficient way of dealing with the problem of biological invasions. Indeed, the pilot identifies the areas in Spain at greatest risk of invasion providing crucial information for resources prioritization and for a better preventive control. For example, more resources and sensible measures could be addressed to the monitoring of areas at highest risk and goods from specific countries or regions.

4.6.2 Pilot set-up

It has been developed a risk of IAS invasion model in Spain, based on several factors related that define the geographical distribution pattern and risk of invasion level. To carry out this task, Big Data sources as climatic data, biogeographical data and socio-economic data have been processed to define the following functions at an unprecedented scale of 1 km x 1 km:

1. Environmental similarity, calculated from bioclimatic variables
2. Similarity of biodiversity, estimated through biogeographic information
3. Propagule Pressure, estimated from trade, tourism, immigration, population and land transport networks
4. Disturbance of ecosystems level (forest fires, for example)

This is a generic invasion risk assessment model, which takes the invasion as a universal process and not as a specific situation. Consequently, the model does not take into account species, but some factors that have proven to be linked to the geographical pattern and the level of risk, that is: it is a territorial model.

4.6.3 Technology used

4.6.3.1 Technology pipeline
The processing of the data, initially collected, required a transformation to integrate and to map data from different sources in a common system. In addition, this target system must use or display the data in a previously defined manner. In this way, depending on the specificity of the data and its suppliers, it was defined a workflow process to update the data and to define its updating frequency.

Big Data processing components using R were used, to create an ecosystem similarity models. The level and distribution of the risk of invasion in continental Spain is derived from environmental dissimilarity and beta diversity (the dissimilarity in species composition).

4.6.3.2 Data used
The Big Data Sources selected and processed have been:

- Imports data from any country and territory to Spain in 2013-16, classified by year, product, amount, number of operations and destination province.
- Tourism data from any country and territory to Spain in 2013-16 period, classified by year, country of origin, number of visitors and province of destination.
- Immigration data from any country and territory to Spain during the 2013-16 period, classified by year, country of origin, number of people and province of destination.
- Road network data: Roads GIS (Geographical Information System) layer classified by type of road (highway or highway, national road, provincial road, local or regional road, track, road, rail).
- Population data: Cities and villages GIS layer and number of inhabitants
- Disturbance Data: GIS layer showing different levels of disturbance (for example, land uses)
- Fire data: Classified by municipality and/or coordinate, date, year and magnitude

4.6.3.3 Reflection on technology use
Although there is no doubt, on a theoretical level at least, that the environment plays an important role in the establishment and expansion of exotic species, exerting a filter of the species that can be established in a new territory, there are more and more studies that reduce the role of the environment as a mediator in the invasion process against other factors, fundamentally of an anthropic nature and those that point to the interaction between
different factors above the individual role of each of them. Although they are not shown, the preliminary results we are getting by adding the propagule pressure and the disturbance point in the same direction. Our result adds, therefore, to the most recent evidence on the secondary role that the environment plays in the invasion process, at least for terrestrial species and at the scale studied.

4.6.4 Business value and impact
The pilot monitored the whole Iberian Peninsula (Spain and Portugal). Risks maps were produced at resolution 1km x 1km for Spain and Portugal. Areas in France and Morocco were also studied. No study was performed for the economic impact if the pilot methodology is used, but this impact is estimated in the millions of euros.

4.6.5 How-to guidelines for practice
Regarding hardware resources, the performance of the standard hardware servers has been proved not enough, since the use of R for the generation of maps is an extraordinarily resource-consuming process. Among the difficulties encountered, we must highlight the large number of temporary files that are generated, with a very high size, when analysing the processed information using R. To solve this problem, an outsourcing solution of these operations could be needed. Establishing collaboration with Super Computing Centers is a solution for this issue.

4.6.6 Summary and outlook
A first conclusion of the analyses completed in the pilot is that the risk of invasion derived from the (di) environmental and biogeographic similarity contrasts sharply with the current distribution of the wealth of invasive alien species in Spain, which suggests that the environment is not the main ruler of the geographical configuration of biological invasions in Spain. The risk estimation in the Canary Islands is currently being refined, considering not only the effects of the world on the Islands, but also the risk of invasion derived from the rest of Spain on them. Models have also been tested including only climatic dissimilarity and biogeography (DC + BG) and the first results of Propagule Pressure and Biogeography (PP + BG). The second, (PP + BG) closely resembles the patterns of wealth of exotic invaders detected in a work developed for Spain.

The complete model is being developed integrating propagule pressure, climatic dissimilarity, and biogeography and ecosystem disturbance. In the first analyses it is observed that the complete model is very similar to the propagule pressure model, highlighting the importance of this factor in the model. This is due to the fact that the pressure of the propagule presents great differences between the Spanish provinces, these being much less marked in terms of climate dissimilarity. These first definitive model analyses are also giving results that are consistent with the current geographic distribution of the richness of vascular plants and of exotic birds. The coastal provinces appear more threatened than the central ones (with the exception of Madrid). Anthropic factors seem to play a leading role in determining the geographic patterns of biological invasions.
The analysis of disturbance of receiving ecosystems in the models is pending communication networks, population density, land uses (SIOSE), fires, etc. At this point we have some complications because we want to have the latest information update (SIOSE) that is not yet officially published, but which we hope is shortly, and need to homogenize how this information is presented with the existing one similar in the border countries (Portugal and France). It is also necessary to implement the result maps obtained and test the information with real data of invasive species, checking the similarity in the potential distribution and the actual distribution. On the other hand, the optimal surface is being studied to perform the analysis and check how they fit.

4.7 Pilot 2.4.1: Web-Mapping Service for the Government Decision Making

4.7.1 Introduction, motivation and goals of the pilot

Our main motivation for this pilot was to develop a new methodology for forest health assessment based on Copernicus satellite data. This allows us to assess the forest health of the entire area of Czech Republic, while reducing costs for field surveys and highly effective identification of forest owners eligible for subsidies / tax relief. This pilot was focused on technological development of the processing of Sentinel-2 optical data. Although the first pair of Sentinel-1 and Sentinel-2 satellites has been launched already in 2014 and 2015, there was no settled methodology for satellite data processing and interpretation available for the Ministry of Agriculture and its associated organizations. Utilizing the great potential of high-spatial and temporal resolution satellite data for forestry, with special focus on forest health trends was thus the main goal of the pilot. In addition, the forest owners have benefit from public-available mapserver, where all forest health status maps are published to allow pro-active management of their forest properties.

This pilot focused on assessing the development of forest health from the Sentinel-2 data and, since of 2018, also from commercial high spatial and temporal resolution PlanetScope data, uniformly for the entire country. Pre-processing and interpretation of Sentinel-2 data resulted in outputs for the assessment of forest health. Processing of these outputs can be divided into two stages:

- pre-processing of Sentinel-2 satellite data in the form of seamless Czech Republic according to a user-defined time period. L3 outputs with minimization of cloud influence in given period and taking into account pixel value for maximum phenological stage of vegetation.
- selection of the most appropriate vegetation index based on the level of correlation with ground survey (forest health status defined as changes in defoliation).

This evaluation resulted in map products classifying in particular areas of significant deterioration in health, where attention was directed and, above all, support to forest owners
by the state and the regional administration. These documents were of great importance for monitoring the situation of bark beetle calamity, which hit the Czech Republic fully since 2018.

4.7.2 Pilot set-up
The pilot involves pre-processing and interpretation of satellite data, collection of in-situ data for forest plots, development of statistical prediction models of forest health and publication of the results through various channels - as specialized map output for experts (Mapserver of FMI, open WMS services) and mapping portals for broader audience.

4.7.3 Technology used
Country-wise forest health trends are obtained in two independent steps:

- Sentinel-2 satellite data pre-processing and cloud-free mask synthesis
- Retrieval of forest leaf area index (LAI) absolute values and its trends

The processing chain is implemented in three follow-up processes:

- Batch downloading
- Atmospheric corrections of raw images (so-called L2 process)
- Automated synthetic mosaic generation (so-called L3 process, or space-temporal image synthesis.

4.7.4 Business value and impact
The exploitation of pilot results was to a large extent achieved:

- Web mapping service for on-line publication of forest health layers in GIS and web environment (http://geoportal.uhul.cz/wms_dpz/service.svc/get)
- Specialized web portal “Kurovcovamapa.cz” (https://www.kurovcovamapa.cz/)

Moreover, based on the timely detection of recent salvage logging and dead wood, the Ministry of Agriculture issued a public decree to apply different forest management regime in the areas with ongoing beetle calamity. The areas are updated regularly, at least yearly, and are based on the outcomes of Stage 1 (cloud free image generation from source Sentinel-2 data, forest health maps) and Stage 2 (multi-source remote sensing) of the pilot.

Web-mapping service for government decision making will work for the future to allow publication of pilot results to broad forestry community of the Czech Republic as 1) dedicated WMS services, 2) specialized map portals “Kurovcovamapa.cz” and “Trendy”. According to those maps the Ministry of Agriculture of Czech Republic issued a „Public decree“ (legislation instrument to help forest owners by reducing the regulation of their obligations under the Czech forest law so that they can manage the bark beetle calamity) in April 2019 and updated on September 2019. According to this map and the officially published legislation instrument,
the forest owner can modify forest management regime on determinate area (optimization of timber harvesting and processing resources, partial release of regulation for the transfer of seeds etc.). All these measures will help reduce the overall loss to forest owners due to climate change and the ongoing bark beetle calamity in the Czech Republic. The overall loss may be close to hundreds of millions of euros.

4.7.5 How-to guidelines for Practice
The pilot describes successful applications of both public open remote sensing data (Sentinel-2) and commercial PlanetScope data. The deployment of the pilot required the use of Big Data approaches, especially for the interpretation of dense time series of Sentinel-2 satellite data. This includes both the requirements of large data storage volumes and computational power to yield high quality satellite images used for interpretation of forest health and its trends. The FMI has chosen development its own processing chain of satellite data, which was not feasible on available in-house IT infrastructure. Instead, the FMI opted for renting the available resources on IT4I super computational facility. If such know-how and/or finances are not available, several commercial solutions for Sentinel-2 data processing are currently available on the market - e.g. the Sentinel Hub, or wide range of Copernicus DIAS services. An alternative would be to deploy the processing entirely on the cloud platforms, e.g. the Google Earth Engine. However, the data archive of level 2 (atmospherically corrected data) is not complete and starts in 2017 season.

Development of prediction models between forest health and satellite observations requires a collection of massive volumes of in-situ data, covering forest plots of different species compositions and health conditions. The FMI has available manpower of skilled foresters. Yet, the collection of approximately 200 forest plots took two seasons and involved three field groups. From our view, the in-situ data are crucial aspect of interpretation of any remote sensing data which should not be omitted.

The FMI is a government organization directed by the Ministry of Agriculture of Czech Republic. Its primary goal is to provide timely and accurate information about the status of the forests to the ministry. In this sense, remote sensing data were proven to be valuable data source about ongoing bark beetle calamity, allowing mapping the extent and progress of the beetle spread in near real time. Based on this data, a renewal of Czech forest law was issued. Here, the calamity bark beetle zones are identified (based on remote sensing data analysis) and regularly updated. This was a great impact and success story of the pilot. For the daily work of foresters however, even bigger impact had the publication of our remote sensing analyses in the form of specialized web-mapping portal “Kurovcovamapa.cz” as simple maps of forest loss and unprocessed dead standing wood. This highlights the importance of dissemination of the results, not only in the form of specialized outputs for the experts, but also as simple mapping tools for broader audience.

4.7.6 Summary and outlook
The FMI developed a successful pilot in the forestry sector, targeting the aspects of forest health monitoring in longer perspective (trends of Sentinel-2 satellite data) and timely
detection of bark beetle calamity (high spatial and temporal resolution PlanetScope satellite data). The outcomes of the pilot had a direct impact on Czech forestry sector, offering results in various forms - the FMIs map servers, open WMS services and specialized web-mapping portals for broad audience. The renewal of Forest Law introduced the identification of bark beetle calamity zones based on the analysis of satellite data.

Results:

- Web-mapping service for government decision making is operational.
- According to this map, the Ministry of Agriculture issued “Public decree” to help forest owners by reducing the regulation of their obligations so that they can manage the bark beetle calamity (on 4/2019, updated on 9/2019).
- Based on this map and the legislation instrument, the forest owner can modify forest management (optimization of timber harvesting and processing resources, partial release of regulation for the transfer of seeds etc.).
- All these measures will help reduce the overall loss to forest owners due to climate change and the ongoing bark beetle calamity in the Czech Republic. The overall loss may be close to hundreds of millions of euros.

4.8 Pilot 2.4.2: Finnish Forest Data based Metsään.fi services

4.8.1 Introduction, motivation and goals of the pilot

Private forests are in a key position as raw material sources for traditional and new forest-based bioeconomy. In addition to wood material, the forests produce non-timber forest products (for example berries and mushrooms), opportunities for recreation and other ecosystem services.

In Finland, private forests cover roughly 60 percent of forest land, but about 80 percent of the domestic wood used by forest industry. Today, the value of the forest industry production is 2.1 billion euros, which is a fifth of the entire industry production value in Finland. The forest industry export in 2017 was worth about 12 billion euros, which covers a fifth of the entire export of goods. Therefore, the forest sector is important for Finland’s national economy.

The Finnish Forest Centre (FFC) is a public organisation and operates under the steering of the Ministry of Agriculture and Forestry Finland. Gathering the forest resource data from privately owned forests in Finland is one of the FFC’s statutory tasks and today around 1.5 million hectare of private forest inventories are annually updated. The inventory cycle for all of the private forests in Finland takes around 10 years and covers 14 million hectares of privately owned forestland.

Remote sensing and airborne laser scanning based forest resource data gathering and maintenance was started in the beginning of 2010 by FFC. At present, the forest resource data covers almost 90 percent of the surface area of productive forest land in private forests. The forest resource data is utilized by forest owners and forestry actors. The forest resource data
is constantly updated and maintained with the subsidy applications, forest use declaration notifications as well as with the update requests provided by the forest owners via the Metsään.fi service. Furthermore, the stand growth is added to all forest stand compartments in the forest resources database annually and the forest management or felling proposals are simulated for the compartments accordingly.

The monetary benefits of this forest resource data ecosystem have been estimated by Natural Resources Institute Finland as well as by Metsäteho Oy and they are annually over EUR 26 million. The potential monetary benefits are annually around EUR 110-120 million. Furthermore, the forest resource data provides additional and indirect benefits for the forest service providers and via the investments around EUR 1.95 billion.

The objectives of the Finnish forest data ecosystem are to ensure the high-quality and comprehensive forest inventory, which is standardized, up to date and easy to use. Furthermore, the forest data is an enabler for FFC to produce the public services as well as data products based on the forestry sector demand.

The Metsään.fi service is based on forest resources data that has been collected by remote sensing since 2011. Forest data can be utilised in, for example, the regional planning of forests and commercial forestry, to support the assessment of wood use possibilities and generally for developing forest businesses.

The Metsään.fi service included in the Metsään.fi website is a free e-service for forest owners and corporate actors (companies, associations and service providers) in the forest sector. The service aims to support active decision-making among forest owners by offering forest resource data and maps on forest properties, by making contacts with the authorities easier through online services and to act as a platform for offering forest services, among other things. In addition to the Metsään.fi service, the website includes open forest data services that offer the users national forest resource data that is not linked with personal information.

Figure 13: An example of Metsään.fi map layer consisting of multiple datasets
The Metsään.fi service was launched in November 2012 as a version that was subject to charge and was changed to a service free of charge for forest owners in 2015. By the end of 2018, about 110,000 forest owners had logged into the service. The forest owners that use the service own forest properties that are larger than average. The Metsään.fi service’s usage activity was increased in particular by forest owners, who experienced that the presented recommendations for forest management matched their own objectives.

A central challenge in developing the website is to integrate several different sources of information into one entity that offers forest owners and actors all forest and nature data simultaneously. From the perspective of both forest owners and actors, the up-to-datedness of forest resource data and improvement of quality is one of the most important development objects.

It is inherent for a service that is maintained with public funds that it is seen to be necessary and that it is being used. By the end of 2018, already over 100,000 forest owners had logged into the service. This is about a third of forest properties measuring over two hectares. The forest owners and other industry actors see the service useful in many ways, but there are also areas that need improvement. It is important for future use and usefulness of the service to improve it and its content continuously.

The Metsään.fi website was also further developed through the DataBio project, where the objective was to improve the use of forest resource data and Metsään.fi service. The pilot focused on Metsään.fi databases and e-service integration to the national service architecture of Finland (based on X-Road approach) where important features were for example data and user security, single-login and easy user role based authentication and data access permissions. Furthermore, the launch of open forest data service, as well as related crowdsourcing services, were included in this pilot. These new types of data gathering methods were also expected to increase the availability of FFC’s forest resource data.

The two recognized areas for crowdsourcing solutions were as follows: showing quality control data for young stand improvement and early tending for seedling stand, and storm damage data. Other possible crowdsourced data, such as other forest damage than storm damage data, were also evaluated during the project. Another pilotable topic was the open-data interface to environmental and other public data in Metsään.fi databases. This topic was highly dependable of development of the Finnish forest legislation.

In these pilots the requirements were specified for refining and showing the crowdsourced forest data to Metsään.fi users. The implementation of the new functionalities and data-presenting was carried out in collaboration with Metsään.fi’s development team and other FFC’s projects.

4.8.2 Technology used

4.8.2.1 Technology pipeline
The technology pipeline was specifically tailored for this pilot, however the Suomi.fi based data transfer service enables the data transfer in standardized way between the FFC and
other partners. Also standardized forest data can be utilized for other purposes and on different scenarios. Suomi.fi service is also applied for the user identification and authentication by Metsään.fi-service and many other public organizations in Finland.

The technology pipeline related components consisted of Metsään.fi-service, open forest data service and Wuudis solution for mobile data gathering as follows.

![Figure 14: Example of pilot data processing pipeline on a high abstraction level](image)

4.8.2.2 Data used

The following data assets were utilized in the pilot:

- Forest Resource Data
- Open forest data
- Customer and Forest Estate data
- Storm and forest damages observation and possible risk areas

4.8.3 Business value and impact

The pilot deliverables consisted of integration of the Metsään.fi-service with the national service architecture of Finland (based on X-road approach). This phase consisted of important features such as for example data and user security, single-login and easy user role-based authentication and data access permissions. Open forest data service was launched in March 2018 and related crowdsourcing services, including Wuudis based Laatumetsä mobile application for the forest damages as well as quality control monitoring, were published in the end of 2018.

In the beginning of 2019, the required XML standard schema version was released and, after that, the X-road approach was applied also for the crowdsourcing solutions regarding the forest damages reported by the Laatumetsä mobile application. This activity was successfully implemented and finalized in September 2019 and it was mainly a technical solution improvement activity and therefore not visible for the end users.
In the beginning of the project, top-level evaluation criteria for the pilot was agreed and this was preliminary based on the Finnish Forest Act at the time being. However, the Finnish Forest Act was revised in March 2018 and the project evaluation criteria was updated accordingly. Additionally, more detailed key performance indicators were chosen to evaluate the results more precisely on the pilot level. The updated top-level evaluation criteria with achieved results was as follows:

In the beginning of the project in 2017, the amount of FFC’s forest resource data was around 200 GB. The amount was expected to increase by approximately 100 GB per year during the project, amounting to around 500 GB by the end of 2019. The result in the end of October 2019 was 574 GB.

The coverage of forest resource data in Metsään.fi-service was in the beginning of 2017 around 11 million hectares. The amount was expected to increase by 800 000 hectares per year, amounting to around 13.4 million hectares by the end of 2019. The result in the end of October 2019 was 12.5 million hectares. The target was not completely achieved due to the fact that the data was getting outdated for the areas where the laser scanning was done over 10 years ago.

The amount of data available for downloading for forestry operators' own information systems was in the beginning of the DataBio project around 1.5 million hectares. The amount was expected to increase by one million hectares per year, amounting to around 4.5 million hectares by the end of 2019. The result in the end of October 2019 was 8.2 million hectares.

The amount of forest owners as Metsään.fi end users was in the beginning of DataBio project around 70 000. The amount was expected to increase as follows: 85 000 in the end of 2017, 100 000 in the end of 2018 and 110 000 in the end of 2019. The result in the end of October 2019 was 119 046 forest owners.

The amount of forestry service providers, i.e. so-called actors using the Metsään.fi service, was in the beginning of the project around 380 pcs. The amount was expected to increase as follows: 550 in the end of 2017, 650 in the end of 2018 and 750 in the end of 2019. The result in the end of October 2019 was 794 users.
Based on the above top-level evaluation criteria and achieved results can be stated that the pilot targets were well achieved and exceeded. The results of the pilot were very promising and they clearly indicate that by standardized solutions i.e. with standardized data and data transfers as well as application programming interfaces, it is possible to build a completely new type of ecosystem, which is utilizing multiple data sources. In this type of ecosystem, the data sources can be scalable from closed datasets to open data as well as the data can be further enriched with crowdsourcing solutions, where citizens are acting as observers. This type of ecosystem consisting of the pilot specific pipelines is fully scalable and exploitable for the European forestry sector or even globally. By applying the same data standards also, the forestry sector businesses could be expanding their business opportunities across country borders.

The pilot specific business impact and benefits were further analysed during the pilot with technical KPIs (Key Performance Indicators), which were identified in the beginning of the pilot as follows. Most of the indicators are indicating very positive business impacts based on the pilot findings. These can be summarized as follows:

- Increased user satisfaction regarding the e-services flexibility and quality: a Net Promoter Index of 48 was measured. To increase user satisfaction, it is of great importance for the adoption of the products among users and also to turn users into promoters of the service.
- Improvement of data quality
- Increased the number of e-applications processed from 26% (baseline) to 35% after the pilot execution. Based on the fact that utilization of the e-Service and especially e-application will save 75% costs compared to the traditional way of working, the increased use of e-application results in high operative costs savings for all involved stakeholders.
- Increased productivity of employees, measured by the amount of the contacted (phone meeting) forest owners or service providers (users of the Metsään.fi services) by the same number of employees
- Improved sustainability by the amount and coverage of the data related to nature objects.
- Increase of the overall data amount of open forest data from 0 to 439.3 GB
- Increase of the total amount downloaded data via the Mesään.fi open forest services implemented during the DataBio project. This KPI increased from 0 to 16295 GB.
- Number of visits in open forest data service reached 10 928 529.

4.8.4 Summary and outlook

Related to the launch of the open-data interface to environmental and other public data in Metsään.fi databases the main finding was that simple solutions do work, however it is good to plan and reserve enough resources, not only for the development activities but also for the maintenance, end user support as well as training.
Regarding the shared multiuser data environment and Metsäään.fi services, certain purpose limitation factors were hindering to apply similar authorisation processes for all of the end users. The backend service provider Suomi.fi could not provide the needed option for the user role specific authorisation profiles. This type of factors could have been perhaps identified and mitigated during the pilot’s risk management planning phase.

The findings related to the crowdsourcing solutions was that due to the available technologies it is easy to implement and launch new type of data gathering solutions. However, the difficulty is in motivating the citizens to produce the information with new type of tools especially when the information is not necessarily fully integrated with the processes of the public authorities.
5 Business Analysis of Fishery Pilots

5.1 Introduction
This section will explain the fisheries pilots in a business context. To do so, fisheries and their challenges will be briefly explained, as well as how the DataBio project is related to these challenges. Then the individual pilots are explained, focussing on the business value they provide. The technical parts are more thoroughly described in other DataBio deliverables and are therefore handled only briefly in this document.

5.2 The commercial fishing business
Fisheries stocks are a renewable food resource, which is under considerable pressure worldwide. Fisheries also provide jobs and income to coastal communities, and the EU therefore requires fishing to be sustainable not only environmentally, but also economically and socially. The sector is also expected to contribute to long-term European food security and economic growth. The application of Big Data technologies in fisheries to maintain sustainable fishing stocks and competitive industries can be divided into three areas:

1. Management of fisheries resources and estimation of stocks.
2. Predictions of market conditions of the harvested resource.
3. Support fisheries trip planning and operations to harvest with effort and cost reduction.

Fisheries is a very complex sector due to the international context of the seas and oceans, the monitoring difficulties of a 3D environment more difficult than the atmosphere and complex international trade of seafood and other derived products (e.g. fishmeal). On the other hand, there is little coordinated use of Big Data technologies in the sector which is extremely competitive and where confidentiality seem to be important as important as some degree of information sharing for healthy stocks management.

Fuel consumption is a challenge for most fisheries, as it represents 60-70% of total annual cost of a vessel activity [REF-35][REF-36][REF-37][REF-38]. Ocean going pelagic fishing vessels apply both energy efficient gears, such as purse seines, and energy intensive gears, such as trawl. The vessels are frequently searching for fish between fishing operations, since pelagic species migrate and exhibit predominantly schooling behaviour. The vessels therefore have a diverse need for energy, from low during loading of fish from the purse seine, through intermediate when cruising and searching, to high when fishing with trawl gears. The vessels have been engineered to become very flexible in the production, routing and consumption of energy on board, while the crew of the vessels are operating the vessel based on habits and preference for configuration of the power system.

Fisheries planning and routing are important factors both for achieving high prices and for reducing the fuel consumption within fisheries. Decisions about when, where and how to harvest are taken by expert fishermen based on their own experience and information gathered from industry contacts (formal or informal) and available data (public and private). In most cases, such information is limited to meteorological forecasts, catch reports and
communication with a small number of collaborating fishermen. For the oceanic tuna fisheries, information from the company's previously deployed fish aggregation devices (FADs) is also available. FADs are small, floating structures, deployed to attract fish and are often equipped with communications technology to report fish aggregations in the immediate vicinity of the device. The subjectively perceived market development is an important factor for fisheries planning, but there are no tools to assist the fishermen in this respect. Furthermore, it is a very dynamic and changing activity with changes in fishing quotas and management restrictions that require fast adaptation of an activity that need long term strategies (e.g. expensive ship equipment).

Fish stock assessment are traditionally carried out based on yearly measurement campaigns following a predefined pattern to sample the distribution of fish in the ocean and by country fishing monitoring systems. These oceanographic campaigns apply both test fishing and hydroacoustic observation to sample the distribution of fish in the ocean. The data from these campaigns are used in statistical models for stock estimation and resource management. The International Council for the Exploration of the Sea (ICES) determines quota recommendations for the national authorities, which have jurisdiction over the fish stocks. Great effort is expended in the collection of this data, but the spatial and temporal coverage is limited by the associated costs.

Fisheries activity monitoring is based on Vessel Monitoring System (VMS) and dedicated reports on vessels and catch. This allows environmental and fisheries regulatory organizations to track and monitor the activities of fishing vessels both in a country's territorial waters and in its Exclusive Economic Zone extending 200 nautical miles from each country's coasts. EU, including Norway through EEC, requires VMS and Electronic Report Systems (ERS) onboard all fishing vessels longer than 15 meters (above 12 meters since 2012). Novel approaches are being underdevelopment to have faster and more integrated fishing effort activity such as Automatic Identification System (AIS) particularly in high seas. Fishermen and landing sites are required by law to report catch data for monitoring purposes. It is common practice for the fisheries for the small pelagic species in the North Atlantic to report the catch volume, species and quality to the sales association while at sea to auction the catch to a buyer which will specify the landing port.

5.3 Summary of the fisheries pilots

The fishery pilots focus on two separate types of fisheries in two countries: Oceanic Tuna fisheries in Spain (operating in international waters at high seas) and small Pelagic fisheries in Norway. The areas encompassed by these pilots have an annual capture production above 13 million tons.

Six separate pilot cases have been defined, addressing key concerns as the cost of fuel and vessel maintenance as well as overfishing and fisheries planning. The pilot cases cover these three separate viewpoints: i) immediate operational choices, ii) fishing vessel trip and fisheries planning and iii) fisheries sustainability and value.
Figure 16: Overview of fisheries pilots

The six fishery pilots are divided into three thematic groups (fisheries operation, planning and sustainability) and two fisheries (oceanic tuna and small pelagic). Although the tasks are broken down thematically in the project plan, e.g. as illustrated by the task status summary in the following section, it makes more sense to present the pilots and the intermediate results organized per fishery after the introduction as these share more implementation commonalities than the general thematic groups.

5.3.1 Datasets and challenges for fisheries pilots

In addition to its volume, data collected on a large scale form a diverse set of sensors, published record and regional observation systems also exhibits other unique characteristics as compared with data collected for a single purpose. It is commonly unstructured and require more real-time analysis [REF-41]. Many of these aspects are present in the fisheries pilots. The pilots are likely to end producing over 10 TB of data per year, coming from many different sources. Such sources include earth observations, sensors onboard fishing vessels (acoustics, machinery, operations), simulations (meteorological, oceanographic and marine biology) and human annotations. The update frequency, regularity and volumes of these sources are on very different levels, affected by simulation times, vessel communications and satellite orbits. The lack of standardisation of data acquisition on board vessels and data structuring poses another challenge for these pilots.

Table 14: Data production by DataBio fisheries pilots

<table>
<thead>
<tr>
<th>Dataset type / Variety</th>
<th>Dataset</th>
<th>Volume (GB)</th>
<th>Velocity (GB/year)</th>
<th>Start date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fish transactions</td>
<td>Catch reports, economic figures</td>
<td>&lt; 1GB</td>
<td>&lt;0,07</td>
<td>20120101</td>
</tr>
</tbody>
</table>
### Operational data from vessels and buoys

<table>
<thead>
<tr>
<th>Source</th>
<th>Operational data</th>
<th>TRL</th>
<th>TRL</th>
<th>Date start</th>
<th>Date end</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESAS Eros KingsBay LiegFi Echebastar A1</td>
<td>coworkers</td>
<td>86</td>
<td>2,3</td>
<td>20140901</td>
<td>20140903</td>
</tr>
<tr>
<td>pilot data Echebastar UE Fleet data</td>
<td></td>
<td>30</td>
<td>30</td>
<td>20140826</td>
<td>20180731</td>
</tr>
<tr>
<td></td>
<td></td>
<td>18,8</td>
<td>17,1</td>
<td>20141010</td>
<td>20170101</td>
</tr>
</tbody>
</table>

### EO, CMEMS, Met

<table>
<thead>
<tr>
<th>Category</th>
<th>Data</th>
<th>TRL</th>
<th>TRL</th>
<th>Date start</th>
<th>Date end</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ocean physics (current/temp)</td>
<td>Wave data (open ocean)</td>
<td>310</td>
<td>99,8</td>
<td>20160101</td>
<td>20161209</td>
</tr>
<tr>
<td>Biogeochemistry</td>
<td>Wind</td>
<td>283</td>
<td>325</td>
<td>20160101</td>
<td>20161209</td>
</tr>
<tr>
<td>Coastal waves (Met)</td>
<td>Oceanic Tuna EO + research data</td>
<td>938,8</td>
<td>1203</td>
<td>20160101</td>
<td>20160918</td>
</tr>
<tr>
<td></td>
<td>115,5</td>
<td>165,7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>37,1</td>
<td>121,2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>802</td>
<td>625,9</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### SINMOD

<table>
<thead>
<tr>
<th>Category</th>
<th>Data</th>
<th>TRL</th>
<th>TRL</th>
<th>Date start</th>
<th>Date end</th>
</tr>
</thead>
<tbody>
<tr>
<td>Historic oceanographic data</td>
<td>Operational oceanographic data</td>
<td>3244</td>
<td>2245</td>
<td>20110101</td>
<td>20180614</td>
</tr>
<tr>
<td>NOAA atmospheric+SINMOD input</td>
<td>N/A</td>
<td>374</td>
<td>2408</td>
<td>20180614</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N/A</td>
<td>0,7</td>
<td>0,7</td>
<td>20180614</td>
<td></td>
</tr>
</tbody>
</table>

### Hydroacoustics

<table>
<thead>
<tr>
<th>Category</th>
<th>Data</th>
<th>TRL</th>
<th>TRL</th>
<th>Date start</th>
<th>Date end</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIMRAD EK80 series echosounder (SIMRAD SX90 Sonar Simrad SN90 Sonar +echosounder)</td>
<td>121,2</td>
<td>5402</td>
<td>Per cruise</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N/A</td>
<td>317</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N/A</td>
<td>20180612</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### WP3 Total

<table>
<thead>
<tr>
<th>Category</th>
<th>Data</th>
<th>TRL</th>
<th>TRL</th>
<th>Date start</th>
<th>Date end</th>
</tr>
</thead>
<tbody>
<tr>
<td>All fishery pilot data assets</td>
<td></td>
<td>10167</td>
<td>11174</td>
<td>N/A</td>
<td></td>
</tr>
</tbody>
</table>

#### 5.3.2 Overall progress and benefits from the pilots

The pilots started at different TRL levels due to the large variation in pilot areas. Several pilots started at a TRL level lower than usual for an innovation action. This implies that effort was needed to build the pilot infrastructure before the end users could make use of the results. The TRL levels for central technologies for each pilot before and after the DataBio project are shown in Table 15. As the table shows, large increases in the TRL levels have been achieved in many areas covered by the pilots. The average TRL level of these technologies have increased from 2.2 to 4.5, which is quite significant.
### Table 15: Development of TRL levels in the fisheries pilots.

<table>
<thead>
<tr>
<th>Pilot</th>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pilot A1</td>
<td>4-5</td>
<td>4-5</td>
</tr>
<tr>
<td>· Intelligent sailing and loading</td>
<td>2-3</td>
<td>4-5</td>
</tr>
<tr>
<td>· Engine failure prevention</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pilot A2</td>
<td>5-6</td>
<td>6-5</td>
</tr>
<tr>
<td>· Onboard acquisition of heterogenous data</td>
<td>4-5</td>
<td></td>
</tr>
<tr>
<td>· Onshore management and analyses of vessel data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pilot B1</td>
<td>1-3</td>
<td>1-3</td>
</tr>
<tr>
<td>· Fusion of environmental data and tuna industry data</td>
<td>1-4</td>
<td></td>
</tr>
<tr>
<td>· Tuna fisheries strategies based on private/public data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>· Tuna abundance estimation based on private/public data</td>
<td>1-4</td>
<td></td>
</tr>
<tr>
<td>· Probabilistic forecasting of tuna distribution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pilot B2</td>
<td>3-5</td>
<td>5-7</td>
</tr>
<tr>
<td>· Service for investigating biomarine properties and fish catches</td>
<td>1-6</td>
<td></td>
</tr>
<tr>
<td>· Forecast service for biomarine systems</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pilot C1</td>
<td>1-3</td>
<td>1-3</td>
</tr>
<tr>
<td>· Assimilation of catch data into spatio-temporal fish model</td>
<td>3-4</td>
<td></td>
</tr>
<tr>
<td>· Fish abundance classification from hydroacoustic data streams</td>
<td>1-3</td>
<td></td>
</tr>
<tr>
<td>· Model for fish migration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pilot C2</td>
<td>2-5</td>
<td>5-7</td>
</tr>
<tr>
<td>· Price predictions</td>
<td>3-7</td>
<td></td>
</tr>
<tr>
<td>· Historical catches and transactions services</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In addition to increasing the TRL level of necessary technologies and pipelines in the pilots, the fisheries pilots have improved the cooperation and collaboration between participating institutions. This combination has triggered several new projects and project initiatives based on the use of Big Data within the fisheries sector. This is expected to further increase the use of Big Data and Big Data technologies within fisheries significantly, which show the impact of the DataBio project on future developments in this domain of the bio economy. Another sign of the importance of the DataBio project for the increased use of Big Data technologies within fisheries, is the great stakeholder engagement. In addition to less formal contacts and project initiatives, this has resulted in many dissemination events. These are more fully described in the D3.3 deliverable.
5.3.3 The Tuna Fisheries Pilots

5.3.3.1 State of the Art

The EU requires fishing to be environmentally friendly, economically viable and socially sustainable to provide long-term European food security. As the catches of tropical tuna might have reached their limit, this fishery needs to improve its management of some stocks, reduce its costs and carbon footprint to achieve such objectives.

Fuel consumption may represent 50% of the total operational costs of the tuna vessels, being one of the main concerns of the fishing companies [REF-42]. Moreover, world fishing industry emissions per landed tonnes of fish have increased by 21% recently. Large pelagic fish, such as tuna species, are highly migratory. Because of this, vessels targeting tuna species tend to have higher and more variable fuel consumption costs than other fishing of coastal species [REF-44]. For example, considering the yearly fuel oil consumption and IMO (International Maritime Organization) proposed emission factor for CO2, a total amount of 3 billion litres of fuel were consumed in a year by these fleets. The type of fuel used in tuna freezers fishing vessel’s engines is mainly distillate diesel oil, namely Marine Diesel Oil (MDO). According to IMO, the burning of 1 tonne of distillate fuel oil (diesel oil) involves 3.026 tonnes of CO2. For instance, about 7.7 billion tonnes of CO2 is emitted in a year by the tuna purse seine fleet in fishing operations.

It is worthwhile to highlight that this sector provides 25,000 direct jobs and 54,000 indirect jobs in the EU. Fish distribution and human behaviour modulate fuel consumption, which in turn influences both the profitability and sustainability of the fisheries industry; as well as the impact on the ecosystem, through greenhouse emissions.

The tropical tuna fishing industry uses Earth Observation (EO) data, to characterize the environmental conditions of the surrounding areas in order to locate fishing grounds with less effort (i.e. time, fuel and costs). High digitalization of tuna vessels means that their capacity to record data and to use existing EO data has increased. However, due to the large volume, diversity of sources and quality of recorded data, they are sparsely used for further analysis, and remain intact and unstructured requiring lots of resources for their integration and analysis.

Big data methodologies seem to be the solution to deal with such large volume of data and exploit it efficiently to turn it into useful information. Solving these problems demands new system architectures for data acquisition, transmission, storage, and large-scale data processing mechanisms [REF-41]. Big data processing techniques, enhanced by machine learning methods, can increase the value of such data and their applicability to society, industry and management challenges. Machine learning has already proved its potential in marine sciences applied to fisheries forecasting [REF-37][REF-38][REF-39]. However, its use by the fishing industry is behind the state-of-the-art and day to day applications, as compared with the other shipping industries.
The aim of oceanic tuna fisheries pilots is to improve economic sustainability of oceanic tuna fisheries while reducing their emissions footprint. This double objective would be through reducing fuel use and therefore the cost. The system aims also to provide advice on potential strategies they can follow. Purse seine is the surface fishing gear that contributes most to the catch of yellowfin and skipjack globally.

5.3.3.2 Importance of Big Data for oceanic tuna fisheries

The goal of the pilots targeting the oceanic tuna fisheries is to provide the crew and shipowner with information that benefits fisheries planning. The information will be provided based on extensive historical datasets within fisheries activity (AIS, VMS, GPS tracking), catch statistics (logbooks), oceanographic conditions (SST, salinity, chlorophyll), meteorological conditions, data FADs (GPS data, echosounder data, SST), engine data (engine parameters such as RPMs or fuel consumption). The hypothesis is that large amounts of historical data combined with machine learning updated may be able to forecast species distribution and vessel efficiency. This will in turn lead to reduced fuel consumption through targeted effort and more efficient engine operation.

![Figure 17: Plot of the active buoys deployed by the whole Basque fleet for one month of 2009, each black dot is the position sent from the buoy to the vessel (n=1.250.000).](image)

These pilots use two types of data. Historical data are used for model development, while near real-time data facilitates operational use. Important data for these pilots are:

**Historical data:**

- Data from FADs (GPS and echosounder)
- Data from vessels engine sensors (fuel consumption, RPM...)
- Data from AIS, VMS and vessels own navigation system (vessel routes)
- Data from observers on-board (catch and by-catch data, visits to FADs without fishing)
• Data from logbooks (daily catch)
• Oceanographic conditions (remote sensing: SST, chlorophyll, height of sea level, currents)

Real-time or close to real-time data:
• Data from buoys (GPS and echosounder)
• Data from vessel engine sensors
• Position of the vessels
• Remote sensing
• Meteorological conditions

The historical operational ships’ data of the Echebastar shipping company involved in the project are being collected on board and uploaded to the cloud (google drive). The historical data consist of recorded parameters on board with automatic recording system and manually recorded data by the crew (like fish catches and fuel oil bunkering). The monitoring system is recording 117 parameters of ship’s propulsion performance every 10 seconds and uploads the data to the cloud every 24 hours (propulsion engine parameters, propulsion parameters, electrical energy consumption on board etc.) or when internet connection is available.

The fishing operations are carried out in harsh environments and in areas far from the fishing companies base. The monitoring and recording systems have suffered from malfunctions during the project. The physical solution of problems on board has been difficult, because these ships call harbour only about once per month, in remote areas and in unpredictable manner (depending on fish captures). The logistic difficulties to solve the technical problems of the monitoring systems on board have supposed a big challenge. Despite difficulties, the historical data have been collected and analysed in collaboration with the industry.

5.3.4 The Small Pelagic Fisheries Pilots

There are several nations engaged in fisheries after small pelagic species in the North Atlantic Ocean. The fishermen of the various countries share common biological resources in the form of fish stocks and primary production. The fish stocks are regulated by international agreements, and each year the UN agency FAOs ICES group assess the stocks, give advice on the amount of each fish stock may be harvested in order to preserve the resources for future generations. The national governments then set the yearly quota for their fishermen (mostly) in accordance with the advice from ICES and the international agreements governing trans-national fish stocks. The fishermen then receive an allotted quota from their national government and must harvest this in accordance with national laws and regulations. The fishermen therefore do not control how much catch they are permitted to catch each year, their only method of ensuring profitability and continued investment is to plan, operate and harvest the resources in a fashion that is efficient and gives good prices for the catch.

The Fishery pilots A2, B2, C1 and C2 focus on the fisheries for small pelagic species in the North Atlantic Ocean from a Norwegian standpoint. The topics for the pilots are vessel operation, fisheries planning, market prediction and estimation of fish stocks. The
stakeholders of these activities are in the entire value chain of the Norwegian socio-economy system; they are engaged in operation of fishing vessels, marketing, trading of fish, product processing plants, biological and technological researchers serving the fisheries sector, as well as the government bodies responsible for regulating and managing the sector. While the fisheries sector contains numerous stakeholders, this project only includes the fishermen and market perspective with the participating stakeholders:

- The vessel operator stakeholders are in this project represented the shipowning companies Ervik & Sævik, Eros, Kings Bay and Liegruppen. These companies operate large ocean-going vessels which combine purse seining and pelagic trawling in the North Atlantic.
- Norges Sildesalgslag (Norwegian Fishermen’s Sales Organization for Pelagic Fish) is a sales organization, owned and operated by fishermen. Due to Norwegian law, this association has monopoly of trade in pelagic species from Norwegian vessels regardless of catch location, and for all pelagic fish caught in the Norwegian EEZ regardless of vessel nationality.

Through its stakeholder partners, the project has easy access to both vessel-generated data and market data for the small pelagic species in the North Atlantic. The small pelagic fisheries pilots in DataBio rely on participation from the vessels and the sales association, and therefore share technical infrastructure.

The main research partner in these pilots is SINTEF Ocean, which in the form of being a research institute for ocean technologies, is both a stakeholder and participant. SINTEF Ocean has as a part of this project initiated the SINTEF Marine Data Center (SMD) to take advantage of Big Data tools in collecting, combining and providing value from a large variety of data sources. Such data sources include both vessel sensors, public data sources and oceanographic simulations.

Figure 18: Overview of data sources, stake holders and components in the pelagic fisheries
The small pelagic fisheries pilots are highly dependent on Big Data technologies, both for modelling the ocean environment and the fish stocks as well as processing harvested data. The datasets, stakeholders and analytic needs are illustrated in Figure 18. The data needed includes satellite data (meteorological and oceanographic), model data (predictions and hindcasts), local measurements (shipborne instruments) and reports on fish catches and sales:

- **Onboard measurements from** e.g. echo sounders, navigation, machinery and propulsion are provided by the shipowning companies Ervik & Saevik, Eros, Kings Bay and Liegruppen.
- **Meteorological data and hindcasts/forecasts** are collected from satellite-based oceanographic measurements from CMEMS, NOAA and other sources as input for daily predictions.
- **Oceanographic data and hindcasts/forecasts** are provided by the SINTEF SINMOD model.
- **All pelagic catches landed in Norway since 2012** are provided by Norges Sildesalgslag. They have developed an API for making these data available for the pilots. The data are continuously updated as new catches are landed. This provides locations, amounts and price for each catch. Each catch is typically defined in terms of approximately 70 variables, such as catch size, where it is caught, sale price, storage method, sales method etc.
- **Catch areas and other definitions** are provided by The Norwegian Fisheries Directorate, such as definitions of various codes representing fish species, catch areas, conservation methods, storage methods, seller, vessel and so on. These data are necessary to interpret the data from Sildelaget.
- **Historical currency exchange rates** are found from the commercial bank DnB ([https://www.dnb.no/en/currencylist/historical](https://www.dnb.no/en/currencylist/historical)).
- **World Bank, EMODnet, Comtrade, Eumofa, Eurostat, ICES and Statistics Norway** offer various data which are of interest when developing price forecasts for pelagic species.

A common architecture approach has been chosen for implementing the small pelagic pilots, starting from the project background related to the "A2 – Immediate operational choices" pilot (e.g. Ratatosk vessel logging system and STIM analytics software) and including the three new pilots as different aspects of fisheries planning. The goal has been to implement one overall architecture capable of handling the needs of all four pilots, adding specific dataset and analytics requirements from each pilot implementation. The resulting shared architecture is presented in the DataBio deliverable D3.3, while how it is used in the individual pilots is described in the individual pilot summaries.

5.4 **Classification of KPIs in Fishery Trials**

The DataBio project provides a contribution to the diffusion of BDT technology in fisheries by concentrating on fishing vessels and developing and providing BDT solutions that can be used by them. Given the fact that fisheries are a highly regulated area in particular in terms of
where, when and which fish in which quantity is allowed to be caught, fisher boats have less opportunities to optimize their productivity and profitability by increasing the catch quantity. The optimization efforts mainly concentrate on the resource and costs side of the profitability and productivity equation. The fisheries trials of DataBio concentrate on the use of BDT to improve catch efficiency and profitability and to minimize operational costs and risk of fisher boats. The specific focus and related KPIs of the fisheries trials can be summarized as follows:

- **Minimization of operational costs**: KPIs in this category include reduction of time spent on fish operations (e.g. steaming), improved vessel energy efficiency (propulsion modes/engine configurations and electrical energy production) as well as reduced vessel downtime and costs savings through condition-based maintenance.

- **Sustainability and reduction of environmental impact and operational risks**: time and energy savings by optimization of fishery operations, as well as preventive maintenance, will help reduce co2 and nox emissions and risks of downtime and accidents. Better usage of catch and fish observations from the fishing fleet in fish stock estimation will reduce risk of overfishing, and data integration and transparency will help in reducing illegal, unreported and unregulated (iuu) fishing. KPI transparency for the end consumer through certified sustainability fishery labelling of seafood like the blue msc label (marine stewardship council), raise consumer awareness for sustainable fisheries and help drive consumer preferences for protein food.

- **Catch efficiency, productivity and profitability**: KPIs calculated by comparing the outcome with the use of resources: examples include income from fish catch sales versus time (crew) and energy costs spent looking for, catching and delivering the fish, for example quantified as energy consumption (kwh) and distance sailed (nm – nautical miles) per kilogram fish. Profitability kpis include marked aspects as price achieved in the market per fish landing and in average per quota, as well as traditional cost versus income considerations.

To summarize, the DataBio fishery trials create value of BDT by optimizing data driven decision making and overall optimisation of the operation processes of fisheries ships. Given this, the subsequent business analysis will focus mainly on the analysis of the business cases based on the expected optimisation of the business processes of fishing vessels. For two pilots, P-A1 and P-B1, there is an option for establishing a spin-off company, in case the trials are successful. For these two trials and where appropriate also business models for the potential spin-off are developed.

### 5.5 Pilots A1 and B1: Oceanic Tuna Fisheries Immediate Operational Choices and Planning

Pilots A1 and B1 are both related to tuna fisheries and are based on the cooperation with the same shipping company and vessels. A short overview of the two pilots was already provided in Section 5.3.3 and more technical details in deliverable D3.3. In this section first the objectives of the two pilots are explained in more detail for each pilot separately. As both
pilots target the same vessels, it can be difficult to attribute measured business impact to the individual pilots. Thus, the business impact is discussed and summarized in the final section together.

5.5.1 Pilot A1: Oceanic tuna fisheries immediate operational choices

5.5.1.1 Introduction, motivation and goals of the pilot
The main targets of the pilot focus on energy efficiency on board to reduce fuel consumption and on condition-based maintenance of the propulsion system in order to reduce ship downtime and increase safety on board. This is done via optimization of propulsion system operation in order to minimize the fuel consumption.

To reach these goals, ships are recording energy performance data with on board systems and uploading the data periodically to cloud services. The data are available for analysis by on shore services, like company machinery superintendents. Data analytics have been used to analyse the recorded data and obtain ships energy consumption equations that are used for operational decision-making. The propulsion engine performance data has been analysed with machine learning techniques to develop models that inform of engine condition deviation from the healthy state. This deviation information is used to proactively participate in engine maintenance and inform in advance of forthcoming problems or inform of undetected problems by ship’s technical staff. In this way, minor faults can be detected in advance and be solved before coming big failures without compromising vessel safety and operation.

5.5.1.2 DataBio Solution for Oceanic Tuna Fisheries Immediate Operational Choices
Different solutions have been developed to be used by technical staff on shore and by crew on board. IBM has implemented their event-based prediction (PROTON, PROactive Technology Online) component on board two ships with a dedicated computer. VTT has employed their OpenVA component to develop the User Interface (UI) for IBM PROTON on board ships and for on shore analysis of data collected on board. VTT has developed and implemented a server-based visualization and analysis tool to be used by technical staff of fishing company on shore. EXUS has used their analytics framework to develop the engine fault detection tool based in historical engine performance data. EXUS has also developed the UI of the software and has employed some of the solutions developed by VTT for the data collection and treatment from GoogleDrive.

The solutions have been tested by Echebastar Fleet in their vessels while EHU (University of Basque Country) has coordinated the partners in the pilot and has also developed the fuel oil consumption equations based in the historical performance data of the vessels (fuel consumption model). The equations developed have been implemented in the B1 pilot for energy saving decision making.

5.5.1.3 Analysis of collected data for energy efficiency model development
The energy efficiency target has been pursued with a ship’s fuel consumption model that is used with the weather models to give a more efficient route to go from point A to point B.
The energy efficiency model also assists crew in deciding which propulsion mode (constant speed or variable speed) and which ship speed are most suitable from an energy efficiency perspective to go from point A to point B. The developed models use common parameters but have specific coefficients for each of the ships and offer great accuracy in fuel consumption prediction depending on ship speed. Skippers use this information for decision-making when they have to decide where to go in the fishing operations.

Offline software for monitoring ship performance has been developed and implemented. The offline monitoring software is used by ship owner’s technical staff from shore in order to collaborate with the crew on board for a more efficient fishing operation.

5.5.2 Pilot B1: Oceanic Tuna Fisheries Planning

5.5.2.1 Introduction, motivation and goals of the pilot
The aim of oceanic tuna fisheries pilots is to improve economic sustainability of oceanic tuna fisheries while reducing their emissions footprint. This double objective would be through reducing fuel use and therefore the cost. Historical environmental and fishing behaviour data have been combined to detect improvement strategies.

5.5.2.2 Datasets analysed that can be further exploited for research
The following datasets have been revised or used in this pilot:

- **VMS data (Vessel Monitoring System):** VMS data of 5 fishing vessel from Echebastar company and that operated in the Indian ocean.
- **AIS data (Automatic Identification System):** The AIS data provides detailed tracks of industrial fishing vessels, which have the potential to provide estimates of fishing activity and effort in near real time.
- **Logbooks:** Logbooks are records of catch and effort registered at the time of the catch operation. This information is available for each of the above 5 fishing vessels.
- **Observer data:** This data comes from the National On-board Observers Programme of the tropical tuna fishing. The aim of this programme is the estimation of the by-catch and discards of the tuna purse seines fishing in the Atlantic and Indian ocean. The minimum coverage of the programme is the 10% of the fleet in each ocean, however the Echebastar company fleet have a coverage of 100%.
- **Environmental data:** Physical and biogeochemical data provided by the Copernicus Marine Environment Monitoring Service.
- **Fuel consumption:** Fuel consumption data of the Echebastar company fleet.
- **Biomass data:** Tuna and bycatch biomass information measured by means of echosounders coupled in Fishing Aggregation devices (FAD).
- **Databsets of global catches (BlueBRIDGE and Global Fisheries Landing Database):** BlueBRIDGE database provide global tuna catches information from 1950 to 2015, whereas Global Fisheries Landing Database (GFLD) provides catches information of commercial, small-scale, illegal and unreported fisheries from 1950 to 2014.
5.5.3 Data processing flow

A summary of the data processing scheme is shown in the figure below. Each data processing step is represented in a rectangle, within which main step and the programming language used are detailed. Two data sources were processed: environmental variables and tuna fisheries data. Previous developed and tested scripts were used when possible. Then both data sources were merged into a geographical grid. The first step was to download the environmental data from Copernicus and JPL at time frame needed and by daily steps for the studied geographical area. After that, two derived variables were calculated, fronts of chlorophyll concentration (CHL) and fronts in sea surface temperature (SST). Finally, the environmental variables were merged with the grid template in a daily time step.

Tuna fisheries data processing also started with the collection of the raw data. Due to the different sources of fisheries data, different formats and some errors were present, making necessary to clean and reformat the different raw datasets. VMS and logbook data were combined to be able to calculate the fishing and cruising effort by vessel. Observer data came in two parts: Vessel activity and set information. The former has the trip information such as trip start and end date, speed, latitude and longitude among others. The latter has the information of catches, in our case species and kg fished. The last source of data comes from the buoys attached to fishing aggregating devices (FAD), these datasets provide the accurate information on the geo-location of the buoy and rough estimates of the biomass aggregated underneath. Finally, environmental variables and tuna fisheries data were merge with the possibility of use different time scale such as daily, weekly or monthly.

![Diagram of data processing flow](image)

*Figure 19: Scheme of the data processing flow*
5.5.4 Data and results visualization

5.5.4.1 Historical data Web-based visualization
The visualization is based on HSLayers, a JS library that allowed to read and display geospatial data available in various raster and vector format, through different services (WMS, WMS-T, WMTS etc.) at initial stages of the project.

![Web-based visualization of geospatial data](image)

*Figure 20: Example of web visualization of WMS-T services provided by CMEMS in the Indian Ocean*

5.5.4.2 Forecasts of species distribution visualization
Machine learning approaches that are characterised by having an explicit underlying probability model, which provides a probability of the outcome, rather than simply a forecast without uncertainty were used to tuna fisheries in Indian Ocean. Bayesian networks are a paradigm suitable to deal with uncertainty, providing an intuitive interface to data. These intuitive properties of Bayesian networks and their explicit consideration of uncertainties enhance the confidence of domain experts on their forecasts [REF-37][REF-38][REF-39][REF-40]. The methodology developed in [REF-37] is a pipeline of supervised classification methods which include feature selection, features discretization and the learning of a naive Bayes classifier (a type of Bayesian Network).

![Bayesian network](image)

*Figure 21: Conceptual diagram showing a Bayesian network to forecast tuna biomass based on satellite data and models combined with fisheries data*
The application of this methodology selected the following features (or predictors): Chl-a, net primary production, temperature, salinity, oxygen, nutrients and currents velocity. The following figures show the results of the Bayesian networks with identify tipping points between the predictive environmental factors as shown in the figures bellow. These figures show also the probabilistic relationship between environmental factors and the low or high tuna captures biomass.

Figure 22 shows the a priori probabilities in the data. No captures where achieved in 84% of the tries, in the rest of attempts half of the time low biomasses where achieved with a final result of successful fishing of more than 20 tonnes only in 8% of the fishing attempts. It also allows better understanding of the environmental conditions. For example, 53% of the times the chlorofile was high (> 0.105 mg/m3) and 82% of the times the non-surface temperature was low (<10.6ºC).cWe can use the model to understand the environmental condition when the captures were high (71% of the times; Figure 23) where the primary production, temperature and clorofile where higher than average. The network allows also to make forecasts where we can see the higher probability of high captures given specific combinations of environmental conditions (Figure 24).

**Figure 22: Bayesian network visualizing tipping points in the relationship between environmental conditions and the different levels of tuna captures**

**Figure 23: Bayesian networks showing environmental condition when the captures where high**
Figure 24: Bayesian network showing the higher probability of high captures given several favourable environmental conditions

It was first applied only to past captures from public sources with poor results, however the results improved significantly when additional data from the tuna company was used. This highlights the importance of working close with industry. The final model can forecast successfully the areas of lack of tuna 80% of the times, i.e. helps to identify areas to be avoided that would waste fuel. The model is also able to forecast areas of high biomass with only a 25% of false positives, so it is right ¾ of the times. The model was validated using 10-fold stratified cross-validation.

Figure 25: Scale show areas of higher probability of finding high biomass of tuna. Green circles show successful fishing attempts and in red circles failed fishing attempts. Thin lines at sea show Economic Exclusive Zones (EEZs) showing territorial waters where only the country fleets and authorized fleets can fish.
5.5.5 Business value and impact for oceanic tuna pilots

The historical data of vessel performance have been collected and analysed to calculate the Key Performance Indicator (KPI) values. For the analysis the newest vessels of the fleet have been used since these are better equipped for monitoring purposes. These are 3 vessels that have similar configuration, structure and machinery, making them very suitable for comparison and benchmarking. The vessels analysed started operation gradually between 2014 and 2015. Historical data since start of operation have been used for KPI calculation. The data available in the recording systems on board have been analysed together with the data available by the ship owners in their own records. The final KPI calculations of Trial 2 result have been calculated using captures and fuel consumption of 2019 (see table below).

Table 16: Fishery Pilot A1 assessment criteria.

<table>
<thead>
<tr>
<th>NAME</th>
<th>DESCRIPTION</th>
<th>BASE VALUE</th>
<th>UNIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>SFO_NM</td>
<td>Propulsion Engine Specific Fuel Oil volumetric consumption per sailed nautical mile while fishing.</td>
<td>77.64</td>
<td>L FO/Nm</td>
</tr>
<tr>
<td>LFO_kgCatch</td>
<td>Ship specific Fuel Oil volumetric consumption per kilogram of fish caught (total fuel oil consumption including auxiliary engines).</td>
<td>0.4424</td>
<td>L FO / kg Catches</td>
</tr>
<tr>
<td>FO_consumption</td>
<td>Total Fuel Oil Consumed by the vessel per year of operation.</td>
<td>3,756,580</td>
<td>L</td>
</tr>
<tr>
<td>SOGave</td>
<td>Average ship velocity in steaming condition.</td>
<td>8.00</td>
<td>knot</td>
</tr>
<tr>
<td>kgCatches</td>
<td>Total fish caught per year.</td>
<td>8,505,580</td>
<td>Kg</td>
</tr>
<tr>
<td>Sailed_NM</td>
<td>Sailed nautical miles per year.</td>
<td>44,932</td>
<td>Nm</td>
</tr>
<tr>
<td>LFO_day</td>
<td>Fuel Oil consumed by the vessel per day of operation.</td>
<td>15,433</td>
<td>L/day</td>
</tr>
<tr>
<td>Day_trip</td>
<td>Average value of days spent per fishing trip (from departure to return to harbour).</td>
<td>24.41</td>
<td>day/trip</td>
</tr>
<tr>
<td>NM_trip</td>
<td>Average value of sailed nautical miles per fishing trip (from departure to return to harbour).</td>
<td>4,648</td>
<td>Nm/trip</td>
</tr>
</tbody>
</table>

The year 2019 has not ended at the time of calculating the KPIs, but all vessel operations are finished and the vessels are moored in harbour. The vessels have stopped fishing because they have reached their maximum allowable catch quota and are not allowed to continue fishing operations. This means that values like catches and sailed miles are final for 2019. The total fuel oil consumption of the vessels will however increase towards the end of the year due to use of the auxiliary engines while moored. It is, however, estimated that this added fuel consumption would only increase the total consumption approximately 2% to 3%.
Figure 26: Total Sailed Nautical miles and fishing days (3 ships).

All the 3 ships used as reference to obtain the KPIs have been repaired within the period analysed (2017-2018) and one of them has been also repaired during 2019. When vessels go to repair works, they are usually stopped a time period between 30 and 60 days. In the repair period, the vessel is not sailing, hence, variables like fuel consumption and sailed nautical miles suffer a reduction in the year of the repair work that would be in the order of 10% of total fuel consumed. Anyway, this has to be considered within usual operation of ships and not likely to impact total fish caught by the ship and its fuel consumption by caught ton. Every 3 years maximum, a ship has to be repaired and will be stopped around 30 days. This means that basically every year one of the vessels will be repaired and KPIs will suffer the same impact every year caused by engine stop. Being this “repair effect” already considered in the analysis cycle.

When analysing the KPIs, it is necessary to consider the regulatory changes in the Indian Ocean in 2017 regarding tuna fishing. Quotas were established and ships had to discontinue their fisheries operations. This means that it is difficult to compare the energy efficiency of the tuna fishing vessels operating in the Indian Ocean before and after the new regulations were effective. Also, as the quotas are varying from year to year, their influence on the energy efficiency of the vessels will vary from year to year.

Due to this quotas policy, a clear decline in the total fishing days and total sailed nautical miles has occurred since 2016. This declining tendency is less in 2019, and 2019 values are basically similar to 2018 values. There is a very slight increase of sailed days during 2019, but that could
be because there were two ship repairs in 2018 and only one in 2019. All this suggests that the fleet has adapted to the new situation.

![Graph showing total fish catches and fuel oil consumed per kg of catch (3 vessels)](image)

**Figure 27: Total consumed fuel oil and fuel oil consumed per kg of catch (3 vessels)**

Although a clear decline in sailed miles and sailed days is clearly observed in the graph, there is no similar tendency in the catches. The catches have increased. And considering that sailing days and consumed fuel has been reduced, the fuel oil consumed per kilogram of catch has been noticeable reduced. Although a clear decline in sailed miles and sailed days is clearly observed in Figure 27, there is no similar tendency in the catches. Therefore, fuel oil consumed per kilogram of catch has been noticeable reduced. This is partially explained by a decrease in vessels’ speed (Figure 28). Both variables follow same tendency what is quote logical. A vessel’s energy consumption follows a pattern of energy consumption that roughly varies with the third power of velocity, i.e. energy consumption proportional to speed in the power of three (V3). This value is a rough approximation but is valuable for a tendency analysis.
Figure 28: Average fuel oil consumption and average vessel speed during sailing (3 vessels)

Although this analysis has been based on only 3 vessels, and they are sister vessels equipped with monitoring systems, it is worth mentioning that all the five ships of this fleet reduced their ratio between fuel consumption and catch during 2017. The reductions were from 4% to 30%, with a 19% reduction on average. During 2018, further reductions (16 % on average) can be observed in relation to 2017. This means an average reduction of 33% during 2017 and 2018 in relation to 2014-2016 averages. It is unclear if these reductions have been maintained during 2019 due to data delays (e.g. multiple countries authorities involved) and the impact of management changes which require further analysis and verification.

However, it is not possible to distinguish how much of this improvement is due to DataBio technologies or other continuous and ongoing initiatives to improve their operations and sustainability such as the MSC certification, bioFADs or new Indian Ocean fisheries management regulations. Certainly, the project has helped the company to consider a more holistic approach to their operations, looking not only at the catches, but also at the fuel consumption. The data collection done in the DataBio project has made it possible to identify some strategies that explain this reduction. On one hand, they have changed from a hunting strategy to harvesting strategy based on knowing better where the fish is, instead of searching for it. On the other hand, a reduction of speed while maintaining the amount of caught fish suggests an optimization of fishing operations. A possible reason for this is improved screening of data and improved data analysis by the crews. Although this cannot be directly linked to the DataBio project, we cannot neglect the contribution of DataBio to this improved analysis capability and increased fishing efficiency. Ship owner and crews have been involved
in the project and this always improves people involvement and consciousness with the subject of the projects, in this case, fish the same with reduced environmental impact.

As highlighted in the stakeholder engagement, other potential benefits are in terms of the datasets from different sources combined which is leading to further applications of DataBio outcomes for fisheries management. This is highlighted in the collaboration with Food and Agricultural Organization of United Nations (FAO) and the building of tuna indexes based on buoy and observer data. Similar collaborations for data sharing and collaborative research have started with the European DG Joint Research Center Directorate for giving advice to the European Commission. This improvement on stock estimation and management has a direct impact on both the Echebastar fleet and other tuna fishing fleets, since it is required for sustainable certifications and to avoid sudden closures of the fishing due to quota restrictions. A project proposal has also been submitted for integrating and improving the DataBio technology with real-time buoy data monitoring in partnership with a buoy company. Similar further exploitation of the knowledge developed during DataBio is expected to continue. For example, adapting some of the work from tuna pilots to small pelagics and vice versa or to other type of vessels and fisheries that operate differently but to which adaptation is possible.

5.5.6 Summary and outlook

The tuna pilots have accomplished some improvements in terms of tools available to monitor and reduce fuel consumption in the day by day operational on board of each ship. The tuna pilot has also provided tools to analyse and propose fishing strategies at the fleet level as overall year strategy. Furthermore, datasets revised or produced are being used to improve the fisheries management (e.g. buoy data or the AIS data revision with FAO). Further planning and actions are being taken to get these tools and datasets to be exploited beyond DataBio project with some of the current partners and additional actors identified during the project (e.g. FAO or buoys and other commercial companies).

5.6 Pilot A2: Small Pelagic Fisheries Immediate Operational Choices

5.6.1 Introduction, motivation and goals of the pilot

This pilot assists in the operation of relatively complex machinery arrangements on-board small-pelagic fisheries vessel through presentation of measurement of current state and historic performance. The energy needs of the vessel for propulsion power, deck machinery, fish processing and general consumption are met by the same power generation system, which on newer vessel can be configured to produce and distribute power in a variety of ways. The machinery systems of the vessel may meet the requirements in a variety of ways, but do not contain a feedback on efficiency or suggest actions to re-configure power production and distribution.

The main motivation of fishermen is the harvesting of fish, not operation of the vessels power production system. With increased system complexity and an increased number of sensor readings the potential for optimal operation of the vessels machinery systems may be of less
interest than the collection of revenue from catches. The data produced on-board the vessel from the increasing number of sensors may still give the crew valuable information, and automated collection, analyses and presentation based on BDT. A modern small pelagic fisheries vessel is equipped with numerous sensors for both navigation and power production on the bridge and in the engine room. These sensors are a part of the vessel automation system, which may consist of several separate systems with varying interaction. The data produced by these systems are used for the immediate operation of the vessels and inform the crew about the status of the system. The sensor readings are not stored on board and not utilized for long term operational efficiency. When the limited buffers and logs on the automation system are full, they are overwritten with never contents. It is therefore difficult for the crew to rate the operation of the vessel 'right now' towards the historic performance or capability of the vessel. In order to make the historic performance and capability of the vessel available, a large time history, or data horizon, must be established.

5.6.2 Pilot set-up
The location of this pilot is on-board vessels. The vessels are typically in the order of 30-80m in length and follow the target species across the North-Atlantic, Norwegian and Barents Seas. The pilot automated data collection and decision support on small pelagic vessels and integrated the on-board data collection system with an onshore data center.

The consortium involved in this pilot consists of:

- SINTEF Ocean is a contract research organization committed to technical research within marine applications. SINTEF Ocean leads the pilot and is also the main contributing research organization.
- The fishing vessel owners Liegruppen Fiskeri, Eros, Ervik & Sævik and Kings Bay conduct fisheries after pelagic fish species in the North Atlantic. Their role in this pilot is to contribute with access to their vessels for installation of data acquisition equipment and to test a decision support application.

5.6.3 Technology used
Machinery, navigation and energy consumption has been monitored by instrumentation and installation of logging equipment on board of the participating vessels. The collected data are analysed, and the vessels integrated into the SINTEF Marine Data Center infrastructure. The signals recorded on board the vessels are augmented with synthetic signals for decision support in order to cope with the inherent heterogeneous nature of data collected from different fishing vessels. Datasets are heterogenous due to different engine system layouts, different choice of suppliers for propellers, prime movers and auxiliary engines. The new synthetic signals enable the four vessels to slot into a data collection and processing pipeline in the SINTEF Marine Data Centre.

5.6.3.1 Technology pipeline
The following technologies have been found relevant for this pilot:
• **Saltstack** provides configuration management of data centre servers, on-board logging equipment and computers for decision support, facilitating remote access.
• **Docker** provides containerisation and facilitates version control of onshore systems.
• **DC/OS** provides container orchestration and communication.
• **GlusterFs** provides replicated and distributed storage of and access to collected data.
• **STIM** provides time series manipulation and analysis of historic data onshore and real-time data on-board
• **RATATOSK** provides distributed data acquisition, signal routing and real-time statistic summaries on-board vessels
• **PURSENSE** Decision support application that aggregates data transmitted by RATATOSK and visualises them on the bridge of the vessel.

### 5.6.3.2 Data used

Fishing vessels are continuously monitored, and have sophisticated communications equipment, including satellite-based data links. The usage of earth observation is limited for the immediate operation since the requirement of data links to obtain updated information from large datasets are constrained by the high costs of satellite data connections. The energy system of the vessels is therefore an isolated island in the sea. The data needed for the efficient operation of the vessel originate on the vessel and are needed on the vessel. The data streams used for this pilot are the vessels own sensor and automation systems:

1. Navigation ship speed, course, movements in waves and orientation
2. Energy power producers and consumers
3. Consumption engine instantaneous fuel consumption
4. Floating position loading condition of the vessel tanks

These data sources typically contain in the order of 100 signals which are sampled every second and higher frequency for motion data.

### 5.6.3.3 Reflection on technology use

The pilot applies the same system on three different vessels. The homogenous population of fishing vessels pose some challenges when trying to maintain a common pilot with different underlying technologies, even though the vessels appear quite similar from the outside. The collection of data is paramount to the pilot, and it is evident that additional systems must be inserted into processes which generate, but do not store data. In effect there are many operational systems which could, with time, take advantage of Big Data, but do not possess the backlog of data required in order to make use of the new developments in information technology such as machine learning and data mining.

The optimum, practical attainable, operational configuration of the powerplant on board of the vessel is contained in the historic data. Effort has therefore been spent at establishing this historic database. The individual design nature of the vessels also implies that there is limited transfer of knowledge between the datasets of the individual vessels. The first technological hurdle for the pilot is the implementation of data harvesting and retrieval of data from the vessels. The retrieved data is of high value for the future and must be kept securely stored, if
it is lost there is no way of recovering the data. The pilot has therefore integrated the measurement system on-board the vessels with a datacenter to store all collected data securely for future use and to establish the ability to curate data.

5.6.4 Business value and impact
The value created in this pilot is both the possible energy savings brought by a decision support system for optimizing operations, but also from the technology pipeline and the side-effects of the decision support system. The decision support system must monitor the vessel and creates a long-time-series dataset which has previously not been available. The decision support system is built upon a new technical infrastructure installed on-board which is flexible and allows for future expansion. Energy savings are inherently difficult to quantify in an industry with large yearly variations such as the pelagic fisheries following the migratory patterns of the target species. The collected data will enable the vessel owner to tailor new built vessels to their specific operational profile and will know in detail what is needed from energy systems. This will allow the vessel owner to optimize new designs, but will also enable the vessel owner to take informed decisions about novel propulsion and energy systems. The data obtained through this pilot may enable the vessel owners and ship design companies to lower emissions and fuel consumption by use of batteries, hydrogen or other energy sources in the future while knowing the energy demands of their current vessels.

5.6.5 How-to guidelines for practice
Fisheries for small-pelagic species is dominated by smaller companies, which may operate one or two vessels. The vessels are seldom similarly equipped sister vessels, but separate new building projects for each company with different choice of machinery configuration, suppliers of machinery, propulsion and control systems. The separate vessels must provide a common signal "package" if a similar decision support system is to be deployed on all vessels. Available data streams must be analysed during vessel operation and a common signal "package" deployed by the use of synthetic signals derived from a combination of available signals in order to complete the signal package.

The collected data has been analyzed and the vessels integrated into the SINTEF Marine Data Center infrastructure. Data collected from the vessels and processed on the datacenter allows the development of common combination of signals. The synthetic signals enable the four vessels to slot into a data collection and processing pipeline in the SINTEF Marine Datacentre. This integration of heterogenous vessel data, or sensor platforms, into a common system has highlighted the need for feedback of both analyses techniques and synthetic signal generation, but also of updated decision support databases to the vessels. The introduction of new signals, real or computed, may necessitate an update from the data center to the vessels of both signal definitions, analyses and the database which decision support is based on.

5.6.6 Summary and outlook
This pilot has installed decision support, monitoring systems on several vessels, established data streams from vessels to shore and collected operational data from small pelagic fisheries
vessels. The data collected and infrastructure is unique as it contains the most comprehensive overview of operations of small pelagic fisheries vessels, and the vessels are integrated with a data management infrastructure. The decision support application installed on-board the vessels shows the potential of using BDT on-line during operations to continuously build and maintain a statistical memory of the vessels past performance. The past performance which is built from the vessels own data gives more value than generic advice based on heuristics. The infrastructure is flexible and allows for future applications and usage of data to leverage value both for the crew, the vessel owner and ship design companies.

5.7 Pilot B2: Small Pelagic Fisheries Planning

5.7.1 Introduction, motivation and goals
One important challenge in pelagic fisheries planning is to determine which fishing ground location to seek for a targeted specie. The skipper and / or the chief of fishing operations need to collate and manually process various pieces of information both before and after departure in order to catch the assigned quota in an efficient manner. The main objective of this pilot was to evaluate the effect of utilizing Big Data technologies in the pelagic fisheries planning. The work in the pilot focused on developing services that could help improve the vessel operation planning with better fishing ground targeting and improved timing of the fishing execution.

The working hypothesis of the pilot was the causality between oceanographic parameters such as temperature and low-trophic organisms such as Calanus with the location and migration patterns of pelagic species. Therefore, a useful service would be to visualize oceanographic and biology parameters together with historical catch data of various species. The primary pilot goal was to create a web portal that enabled the end user to browse through this information in a map. This includes the ability to select a time period of reported catch data of specific pelagic species, which then are displayed in a map that includes oceanographic attributes. A playback feature lets the user see the time evolution of the selected attributes.

5.7.2 Pilot set-up
The fishing operations region for which the pilot was providing decision support includes large portions of the Norwegian Sea and the North Sea, totalling approx. 1.5 million square kilometres. Pelagic fisheries usually only operate in small subregions of this, depending on targeted species.

The consortium involved in this pilot consists of:

- SINTEF Ocean is a contract research organisation committed to technical research within marine applications. SINTEF Ocean leads the pilot and is also the main contributing research organisation.
• Norges Sildesalgslag (Norwegian Fishermen’s Sales Organization for Pelagic Fish) is a sales organisation, owned and operated by fishermen (a ‘coop’), selling fish at a first-hand basis from fishermen to buyers – for further sales / export. They contribute with knowledge and accumulated data about fish catches.

• The fishing vessel owners Liegruppen Fiskeri, Eros, Ervik & Sævik and Kings Bay conduct fisheries after pelagic fish species in the North Atlantic. Their role in this pilot is to contribute with their knowledge about fisheries planning and also serve as end user for the pilot web portal.

The pilot started from scratch in the sense that it is did not extend an existing service. To achieve the main goal of the pilot, it thus required the creation of a new service, which involved development in all aspects of the BDVA reference model. The physical location of the running service is set to SINTEF Ocean’s data center, but relies on both private / public earth observation data and reported catch data. Important activities in the pilot have been to identify Data Sources, select appropriate components / assets, and configure necessary Data Management and Data Processing Architecture. This work facilitated the primary goal of the project, namely provisioning of the web portal and its Data Visualization. Definition of key performance indicators that directly quantify the fishery operation performance were quickly dismissed, because any evaluation of such indicator depends on unmeasurable and non-deterministic factors. Any potentially improved measurement of fishery efficiency can only speculatively be attributed to the introduction of the pilot service. As a consequence, "key performance indicators" were instead defined as measurable progress / completeness of the technological components used in the pilot.

5.7.3 Technology used

5.7.3.1 Technology pipeline

The following technologies have been found relevant for this pilot:

• Saltstack provides configuration management of data centre servers, facilitating version control and remote access.

• Docker provides containerisation and facilitates version control of onshore systems.

• SINMOD provides biomarine simulations and simulation of fish migrations.

• DC/OS provides container orchestration and communication.

• CouchDB provides storage of and access to catch data.

• GlusterFs provides replicated and distributed storage of and access to collected data and the results of biomarine simulations.

• KRAKK provides data scraping functionality, especially for data from Sildelaget.

• GeoServer provides an open source server for sharing geospatial data.

• Python Scripts that make use of RESTful API and GDAL for ingesting SINMOD oceanographic and biology data rasters into GeoServer.

• Python Flask is used as a WSGI (Web Server Gateway Interface) web application framework to develop the web portal.

• uWSGI is used for serving the web portal.
Crossfilter, D3.js, dc.js and Leaflet are important JavaScript libraries for presenting data in the web portal.

5.7.3.2 Data used
The implementation of this pilot is based on a number of data sources:

- **Catch data** are made available by Sildelaget through an API developed by Sildelaget for DataBio. This API makes available all pelagic catches landed in Norway since 2012, and it is continuously updated as new catches are landed. This provides locations, amounts and price for each catch. The catch data from Sildelaget is proprietary datasets that will not be available after the project. However, the Norwegian Directorate of Fisheries recently Open Sourced the catch data histories.

- **SINMOD** oceanographic and biological hindcast and forecast data for the Norwegian Basin, including temperature, salinity, ice thickness and concentration, NO3, calanus finmarchicus, calanus glacialis, and chlorophyll. These parameters were provided both historically since 2012 and regularly with short-term forecasts two days into the future with a spatial resolution of 4 km in polar stereographic projection. The SINMOD data source relies on several satellite and buoy-based inputs, see the next pilot for details.

5.7.3.3 Reflection on technology use
The SINMOD operationalization produce netCDF4 files that largely follow the Climate and Forecast convention 1.5. Nonetheless, there have been several issues with making the automatic data processing pipeline robust and maintenance-free. In particular, some iterations were needed to settle on standardised naming conventions of the variables, consistent spatial resolution, as well as correct projection parameters between the historic and predictive datasets. The process of making SINMOD data available to the map service involves extraction of selected depths and timepoints so that only relevant data are being served by GeoServer. Instead of using the netCDF-plugin of GeoServer, we rather used GDAL to manually reproject netCDF files into the destination projection as GeoTIFF files. Large datasets compressed and uploaded as zip files, caused GeoServer to be unable to decompress and ingest the files on the server, which spawned the need for file handling logic accommodating this issue. GeoServer’s built-in colorbar legend currently lacks the necessary flexibility to show customized styling in a satisfactory manner, which again warranted manual customization. We experienced intermittent issues with newly ingested rasters, where it cached transparent tiles, probably because tiles were cached before its ingestion into the PostGIS database were done. This issue was not easily reproducible, nor did it produce any error messages, causing undetected issues with the web map service.

We chose tiled WMS to serve the raster data. The styling of the layers was done on the server side, so no styling configuration were needed in the web application Leaflet. Designing styles that work globally for the whole year for a single attribute is challenging, because of the span of interesting values changes can change throughout the year. WMS playback were achieved using a Leaflet plugin, but the flexibility in zoom levels with different tiles made it challenging for the plugin to buffer many timepoints in a manner that gave good user experience. Some
5.7.4 Business value and impact
We conjecture the impact of the new service provided by this pilot is minimal, that is, the pilot end users do not yet actively make use of the web portal in their fishery planning. The reason for this is multi-faceted. First, the time period at which the service has been available is very brief, with fair service reliability. The user experience in the web application can be frustrating, due to lack of responsiveness. There is a lack of fundamental features that can be of interest for the user to look for specifically interesting phenomena. As an example, a simple extension would be the ability to select a region and provide key information / analysis on demand. The portal was specifically designed for desktop application use, but in hindsight it should have been readily available on all platforms, including smartphones and tablets. The graphical user interface design could also have been more targeted to specific use cases; for instance, by providing several subpages, each designed to provide a very limited set of information. One such tailored design could lower the threshold for use. The key business impact is the introduction of this new web application, which can and will serve as a platform for fishermen to acquire federated information to help them do fisheries planning.

5.7.5 How-to guidelines for practice
The pilot was designed on top of systems and infrastructure that are designed for use in production. DC/OS are made ready for production use cases, which includes scalability, load-balancing, resource management et cetera. What the pilot technology design does not cover is the use case where the user sits on a low bandwidth network, which is often the case for ocean-going fishing vessels. Therefore, the web portal is more practically applicable in an onshore, by-the-computer setting. We believe that the concept of collating information and providing insight into multi-origin data in a clear manner still has great potential for improving fishery planning. Establishment of a minimally viable product, that the end user is interested in could spawn a foundation for future features that have an impact on how the fisherman make use of Big Data and technology in planning their operations.

5.8 Pilot C1: Small Pelagic Fish Stock Assessment

5.8.1 Introduction, motivation and goals
Pelagic fish stock assessments are traditionally based on research cruises with dedicated research vessels, catch statistics and non-spatial stock models. These methods are criticised for low cost efficiency, being based on too little measurements and unable to adapt to the effect of the rapid climate changes.

The objective of this pilot has been to demonstrate that the combination of information from many various assets can be used to produce better population dynamics estimates for pelagic species. Specifically, crowdsourced data collection effort from fishing vessels combined with
public/private data assets, biomarine modelling and data analytics are assumed to be able to increase both the accuracy and precision of fish migration and stock assessments.

The pilot has concentrated on three research questions:

- How can hydroacoustic data be cost-efficiently collected on a fleet of fishing vessels?
- How can a fleet of fishing vessels be part of a crowd-sourced data collection system?
- How can biomarine modelling and spatio-temporal modelling of pelagic species be used for stock assessments?

To cost efficiently collect hydroacoustic data from fishing vessels, the integration against existing hydroacoustic sensors were important. Due to the large variations in equipment and interfaces, as well as lack of interface possibilities for much of this equipment, this proved to be a serious challenge. The pilot created a preliminary interface against one type of equipment, but cost-efficient integration against the hydroacoustic equipment of a substantial part of the fishing fleet is not solved.

To make a fleet of fishing vessels part of a crowd-sourced data collection system, cost efficient installation and maintenance of the vessel installations are needed. The most important challenges are the variation in vessel systems, sensors and their setup, as well as how these changes over time. This pilot addressed these challenges by using configuration management systems using version-controlled configuration descriptions. This gave a way to perform remote maintenance, updating and reconfiguration, as well as simplify initial installations.

To model the fish stocks and their behaviour, both adequate biomarine models and correction of these based on measurements are needed. This pilot has developed a preliminary migration model of one pelagic species. Also, a preliminary method for correcting this model using data assimilation is developed, and this correction is performed on historical data. The results show that more data for correction is needed, and this is the focus of new research initiatives.

### 5.8.2 Pilot set-up

Pelagic fish stock assessments are traditionally based on research cruises with dedicated research vessels, catch statistics and non-spatial stock models. These methods are criticised for low cost efficiency, being based on too little measurements and unable to adapt to the effect of the rapid climate changes.

Even though dedicated research vessels give data of high quality, the surveys are costly and cover only a fraction of the oceans. Satellites provide better geospatial coverage, but have limitations in spatial resolution and available measurements. Various systems for automatic, unmanned, in situ marine data acquisition have been developed, based on drifting or stationary buoys, gliders, ships of opportunity, or a combination of methods.

The consortium involved in this pilot consists of:
• SINTEF Ocean is a contract research organisation committed to technical research within marine applications. SINTEF Ocean leads the pilot and is also the main contributing research organisation.

• INTRASOFT International offers IT solutions to a wide range of international and national public and private organizations. INTRASOFT has performed comparison of different methods for classification of hydroacoustic measurements.

• Norges Sildesalgslag (Norwegian Fishermen’s Sales Organization for Pelagic Fish) is a sales organization, owned and operated by fishermen (a ‘coop’), selling fish at a first-hand basis from fishermen to buyers - for further sales/export. They contribute with knowledge and accumulated data about fish catches.

• The fishing vessel owners Liegruppen Fiskeri, Eros, Ervik & Sævik and Kings Bay conduct fisheries after pelagic fish species in the North Atlantic. Their role in this pilot is to contribute with their knowledge about fish migration patterns and how this is observed from the fishing vessels, as well as the technical installations available onboard the fishing vessels.

This DataBio pilot has been aimed at assessing if and how stock assessments of pelagic fish species could benefit from low cost data collection during fishing vessels normal operations, combined with biomarine simulations and simulations of the migration patterns of pelagic fish species. To this end, this pilot has aimed at developing a demonstrator version of an infrastructure consisting of both vessels and shore systems.

5.8.3 Technology used

5.8.3.1 Technology pipeline
Relating to the above specified research questions, the following technologies have been found to be relevant for this pilot and its implementation:

• Saltstack provides configuration management of both shore servers and vessel equipment, facilitating version control and remote access.

• Ratatosk provides onboard data acquisition, data exchange and monitoring of these functions.

• STIM provides efficient analysis of collected data (except for hydroacoustic data).

• Docker provides containerisation and facilitates version control of onshore systems.

• SINMOD provides biomarine simulations and simulation of fish migrations.

• Ratacoustics provides integration between hydroacoustic equipment and Ratatosk.

• DC/OS provides container orchestration and communication.

• CouchDB provides storage of and access to catch data.

• GlusterFs provides replicated and distributed storage of and access to collected data and the results of biomarine simulations.

• KRAKK provides data scraping functionality, especially for data from Sildelaget.

5.8.3.2 Data used
The implementation of this pilot is based on a number of data sources:
• **Catch** data are made available by Sildelaget through an API developed by Sildelaget for DataBio. This API makes available all pelagic catches landed in Norway since 2012, and it is continuously updated as new catches are landed.

• **Hydroacoustic data** are found to be important for correcting the biobrane models and the fish migration model. Some data have been collected using ad hoc methods, but creating general tools for large scale deployment have proved to be challenging.

• **Vessel operational data** is important for determining what the hydroacoustic data represents in both time and space. Also e.g. ship motions can be important for interpreting the data. The vessels Eros, Kings Bay, Ligrunn and Christina E are contributing with such data.

• **Global ocean tidal components** M2, S2, N2, K2, K1, O1, P1, Q1, Mf, Mm, and SSa at the open boundaries of the SINMOD model is imported from the TPXO 7.2-model ([http://volkov.oce.orst.edu/tides/TPXO7.2.html](http://volkov.oce.orst.edu/tides/TPXO7.2.html)).

• **Boundary conditions** for the large scale 20 km model are acquired from the Mercator Global Ocean model system.

• **Atmospheric** input for the large-scale models is acquired from NOAA Global Forecast System

• **Atmospheric** input for the local scale models is provided by the Norwegian Meteorological Institute from the 2.5 km MetCoOp EPS system.

• **Sea surface temperatures** are downloaded from the product METOFFICE-GLO-SST-L4-NRT-OBS-SKIN-DIU-FV01.1 in the Copernicus database [http://nrt.cmems-du.eu/motu-web/Motu](http://nrt.cmems-du.eu/motu-web/Motu).

5.8.3.3 **Reflection on technology use**

The selected technologies seem to be adequate for the tasks, and there are no obvious benefits from making technology changes. But as there are possible alternatives for most of them, the final choice is as much dependent on preferences and existing tools as on the task itself. Without loss of benefits, one may e.g. replace Saltstack with Ansible, Puppet or Chef, Docker could be replaced by Mesos Containerizer, DC/OS could be replaced by Mesos or Kubernetes, CouchDb could be replaced by another database or file storage and GlusterFS could be replaced by Ceph. But as of now, no benefits are seen from making such changes in the choice of technologies.

One possible exception is within hydroacoustic data collection, where a Simrad echo sounder was used in the project. This echo sounder facilitates two main approaches for collecting hydroacoustic data in a systematic manner. One is to use the record functionality in the graphical user interface, the other one is through a subscription-based application programming interface. The first approach is simplistic in that a vessel crew member basically pushes a record button and the system will record data. The downside is that it requires human intervention from the crew, and real-time processing is cumbersome. At the beginning of the project it was deemed as a volatile approach. Therefore, it was decided that API-based data acquisition was a more robust and long-term investment and suited as an extension of the existing data acquisition system (Ratatosk). The subscription API is a comprehensive
implementation that enables access to processed and unprocessed data streams and parameters using Ethernet User Datagram Protocol (UDP). Our approach is to implement this subscription API and make the data streams available to the Ratatosk logging component, enabling both real-time processing and storage to file. Most of the functionality toward the subscription API is in place, but the adaptations to connect to the Ratatosk component is currently lacking.

The currently available hydroacoustic echo sounder dataset has been used as a preliminary comparison of classification methods. The dataset consists of five hydroacoustic frequencies, \{18, 38, 70, 120, 200\} kHz, which are computed into mean volume backscatter strengths. Four different algorithms have been tested on the dataset: Naïve Bayes, k-Nearest Neighbors, Support Vector Machine and Principal Component Analysis. The goal is divided into two tasks:

- Identify and remove seabed echoes and determine fish shoal presence;
- Discriminate plankton from fish, identify fish species, and perform biomass evaluation.

Accuracy is high for all tested methods, but this is due to the few positives of the dataset. Kappa is a more sophisticated metric that shows how much the algorithm improves the average expected accuracy. Kappa shows more varying results comparing the different methods.

For simulation of the marine ecosystem and the migratory behavior of selected species, the tool SINMOD was used. This tool fits perfectly to the task, as it is able to integrate the simulation of oceanography, low-trophic biology and how this affects higher order processes. For demonstration purposes, a preliminary fish migration model for herring (Clupea harengus) was developed, based on simple behavioral rules and corrected by reported catches. Even if very simplified, the model was able to recreate migration patterns. The model will need to be developed further before it can provide actual value for fish stock assessments, but the results are promising.

### 5.8.4 Business value and impact

This pilot only aims at general level business impact. The aim of this pilot has been to demonstrate that the combination of data collection, existing datasets and biomarine simulations can benefit pelagic fish stock assessments. The business value of the pilot will only materialise at that point in time when the developed methodologies and technologies gets implemented as part of the fish stock assessment process. At that time, the business impact of improved stock assessments and thereby improved management and production of the oceans can be very large. If, for example, the production (and thus the catch) of pelagic fish species could be increased by e.g. 10% as a result of this work, this would amount to approximately € 60 mill increase in first hand value of pelagic fish species in Norway alone.

### 5.8.5 How-to guidelines for practice

As stated above, alternatives exist for many of the technologies used in this pilot. Still, the combination of provided functionalities are a good fit for the pilot objectives. Most notably, the abilities of such a system is to:
• Adapt to the great variations of sensors and configurations onboard fishing vessels, as well as changes over time. This includes both hydroacoustic equipment and operational sensors, such as motion reference systems and GPS.

• Handle a large fleet of vessels of opportunity in a structured way, with respect to installation, configuration, maintenance and data collection.

• Simulate oceanography, marine biology and fish migrations, while assimilating available data for model and output corrections.

• Extract useful information from hydroacoustic equipment with respect to e.g. fish species and amount of fish.

• Provide systems for data flow, analysis and storage which are suitable for large scale deployment.

Most of the systems and infrastructure developed in the pilot are ready for use in production, and many of these are easily available. But for such a system to really have an impact on fish stock assessment, improvements are needed in the interpretation of hydroacoustic data and in the fish migration modelling.

5.8.6 Summary and outlook
This pilot has shown a way that enables fish stock assessment to take advantage of Big Data technologies, based on combining Big Data methods and tools with biomarine modelling. The main idea is to combine various new and existing data sources with biomarine modelling to provide new, continuously updated knowledge about the marine ecosystems in general and of the pelagic fish species in particular. The end result depends on the combination of multiple parts. Large scale data collection (crowdsourcing) onboard vessels of opportunity (primarily fishing vessels), focussing on hydroacoustic data, combined with new analysis methods, will provide in situ measurements of fish concentrations. Catch data are continuously provided by Sildelaget. Biomarine models are used for integrating available data sources, driving the fish migration models.

It is clear that more research and development will be necessary before the results from this pilot can contribute to an improved stock estimation, in particular within fish behaviour modelling, collection of hydroacoustic data and interpretation of such data. Still, the results constitute a framework in which the future stock estimation and marine ecosystem monitoring can be based.

5.9 Pilot C2: Small pelagic market predictions and traceability

5.9.1 Introduction, motivation and goals
Norwegian fishermen in the pelagic sector are performing fisheries for different pelagic species. The timing for these fisheries is to some extent determined by the availability of fish and the fish migrations, and to some extent the shipowners make strategic decision about when and where to do their fishing based on expectations of both market development and
fishing possibilities. These are important choices, but there is a lack of tools helping the fishermen to make the right one.

The goal of this pilot is to enable the fishermen to make the right strategic decisions, which can make a big difference in both profitability and landed quality. To this end, machine learning has been employed to try to predict the development of pelagic market segments, so that the fisheries may be targeted against the species that will give the highest yield given the predicted economic outlook. The Norwegian mackerel market has been used as a case, as this is an important pelagic species with large price fluctuations. The basis for the market predictions has been to combine different data sources relevant for price development, such as time, season, predicted catch volume and financial data. Machine learning and predictive analytics has been used to model the relation between market development and other factors. These models can then be used to provide predictions for how the market will develop in the future.

5.9.2 Pilot set-up

The Norwegian pelagic fisheries include a range of fish species, where the most important ones are mackerel, herring, blue whiting and capelin. The pelagic species typically display migratory behaviour, and this is important for when and where they can be fished and how efficient this fishery will be. In addition to their availability for fishing, the market processes, the size of the quotas and weather fluctuations is important for planning when and where the individual fishing vessels go fishing for the different species. These are nontrivial decisions, and they must be taken without firm knowledge about neither the premises (future development of e.g. the market) nor the outcome (such as the revenue of the vessel). The reason for this is that many of the underlying processes are inherently impossible to predict as they are either governed by psychological factors (market processes, economic processes, other vessels strategies) or unpredictable natural processes (long term ecology, meteorology and oceanography). Still, it is possible that new methods may reduce the uncertainty of these processes and their impact on the fish prices. The aim of this pilot has been to contribute with such methods.

The consortium involved in this pilot consists of:

- SINTEF Ocean is a contract research organisation committed to technical research within marine applications. SINTEF Ocean leads the pilot and is also the main contributing research organisation.
- Norges Sildesalgsslاغ (Norwegian Fishermen’s Sales Organization for Pelagic Fish) is a sales organization, owned and operated by fishermen (a ‘coop’), selling fish at a first-hand basis from fishermen to buyers - for further sales/export. They contribute with knowledge and historic and present data on mackerel catches and price.
- The fishing vessel owners Liegruppen Fiskeri, Eros, Ervik & Sævik and Kings Bay conduct fisheries after pelagic fish species in the North Atlantic. Their role in this pilot is to contribute with their knowledge about the mackerel fisheries and the pelagic market.
The goal of this pilot is to enable the fishermen to make the right strategic decisions, which can make a big difference in both profitability and landed quality. To this end, providing the fishermen with historical data and tools for analysing these data have been found to be the most promising method. Therefore, a web portal has been developed with this functionality. In addition, machine learning has been employed to try to predict the development of pelagic market segments, so that the fisheries may be targeted against the species that will give the highest yield given the predicted economic outlook. The Norwegian mackerel market has been used as a case, as this is an important pelagic species with large price fluctuations. The basis for the market predictions has been to combine different data sources relevant for price development, such as time, season, predicted catch volume and financial data. Machine learning and predictive analytics has been used to model the relation between market development and other factors. These models can then be used to provide predictions for how the market will develop in the future.

5.9.3 Technology used

5.9.3.1 Technology pipeline

Relating to the above specified research questions, the following technologies have been found to be relevant for this pilot and its implementation:

- **Saltstack** provides configuration management of shore servers, facilitating version control and remote access.
- **Docker** provides containerisation and facilitates version control of onshore systems.
- **DC/OS** provides container orchestration and communication.
- **CouchDB** provides storage of and access to catch data.
- **GlusterFs** provides replicated and distributed storage of and access to collected data.
- **KRAKK** provides data scraping functionality, especially for data from Sildelaget.
- **Python Flask** is used as a WSGI (Web Server Gateway Interface) web application framework to develop the web portal.
- **Scikit-learn and Keras** are important Python libraries used for training prediction models.
- **uWSGI** is used for serving the web portal.
- **Crossfilter, D3.js, dc.js and Leaflet** are important javascript libraries for analysing and presenting results in the web portal.

5.9.3.2 Data used

The implementation of this pilot is based on a number of data sources:

- **Catch data** are made available by Sildelaget through an API developed by Sildelaget for DataBio. This API makes available all pelagic catches landed in Norway since 2012, and it is continuously updated as new catches are landed. This provides locations, amounts and price for each catch. Each catch is typically defined in terms of approximately 70 variables, such as catch size, where it is caught, sale price, storage method, sales method etc.
• **Catch areas** and other definitions are provided by The Norwegian Fisheries Directorate, such as definitions of various codes representing fish species, catch areas, conservation methods, storage methods, seller, vessel and so on. These data are necessary to interpret the data from Sildelaget.

• **Historical valuta exchange rates** are found from DnB (https://www.dnb.no/en/currencylist/historical)

• **World Bank, EMODnet, Comtrade, Eumofa, Eurostat, ICES and Statistics Norway** offer various data which can be of interest when developing price forecasts for pelagic species. Data scrapers have been developed for these data sources for use into price prediction pipelines.

5.9.3.3 Reflection on technology use

The selected technologies seem to be adequate for the tasks, and there are no obvious benefits from making technology changes. But as there are possible alternatives for most of them, the final choice is as much dependent on preferences and existing tools as on the task itself. Without loss of benefits, one may e.g. replace Saltstack with Ansible, Puppet or Chef, Docker could be replaced by Mesos Containerizer, DC/OS could be replaced by Mesos or Kubernetes, CouchDb could be replaced by another database or file storage and GlusterFS could be replaced by Ceph. But as of now, no benefits are seen from making such changes in the choice of technologies.

In a case study, the possibilities for direct predictions of the mackerel prices were investigated. The focus was on long term predictions, aiming to enable the fishermen to do long term strategic decisions. As the market is greatly influenced by psychological factors, the results were not expected to be good. This can be compared to predicting the stock market, as understandably is a close-to-impossible task.

Preliminary exploratory analyses for mackerel show the expected seasonal variations, as well as variations so far not explained. Figure 29 shows the variations in daily average mackerel price and daily catch from 2012 until today for Norwegian mackerel landings. Only the second half of each year is plotted, as this is the main season for this fishery. The size of each point marker reflects the amount of daily/weekly catch. The seasonal variations are obvious, while the variations with other variables in this dataset than time is not.
5.9.4 Business value and impact

To enable the fishermen to understand how the market prices depend on other factors, a web portal has been developed. Using this portal, one can investigate how the prices have developed with factors such as species, landed quanta, year, time of year, moon phase and catch location. This web portal is based on providing the possibility to filter historical catch data along the relevant factors, such as e.g. selecting only catches of mackerel last year in a small time window, and then slide this window to see how the prices varied with time. Also, similar procedures can be employed to consider variation with moon phase. Or one can go in the opposite direction and select only the catches giving the highest prices to investigate under which circumstances high prices were achieved.
Figure 30: Screenshot from the web portal, where filtering of historical data is facilitated

The service developed in this pilot is, as far as we know, the first of its kind. It is very difficult to estimate the potential business impact. Even if one can investigate how the fisheries historically have performed and been times with relation to the market, any changes in timing of the fishery would influence the market, and we do not know how efficient the fishery could be performed for alternative timings. If one assumes that the market would not be affected by shifting the fisheries to the autumn, and that the fisheries could be performed in autumn without affecting other fisheries, shifting 10% of this fishery to autumn would approximately generate an extra 700 000 €.

5.9.5 How-to guidelines for practice

The developed technology makes it possible to analyse and filter large amounts of historical data in a web browser. It is possible to do such analyses without access to the detailed base data, which supports the use of data which should not be distributed except for in aggregated forms.

The developed services are targeting fishermen and shipowners, and aims in particular for assisting them in long term choices, such as deciding when the fisheries for different species should be performed.
5.9.6 Summary and outlook

The developed service enables fishermen to learn from history, possibly increasing their understanding of the market processes. As such, it can make them better suited for taking the correct decisions about where and when to go fishing for which species.
6 References

<table>
<thead>
<tr>
<th>Reference</th>
<th>Name of document</th>
</tr>
</thead>
<tbody>
<tr>
<td>REF</td>
<td>Reference</td>
</tr>
<tr>
<td>-----</td>
<td>------------</td>
</tr>
<tr>
<td>Reference</td>
<td>Source</td>
</tr>
<tr>
<td>-----------</td>
<td>--------</td>
</tr>
<tr>
<td>REF</td>
<td>Authors</td>
</tr>
<tr>
<td>-----</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>42</td>
<td>Suuronen, P., Chopin, F., Glass, C., Løkkeborg, S., Matsushita, Y., Queirolo, D., &amp; Rihan, D.</td>
</tr>
<tr>
<td>44</td>
<td>Parker, R. W. &amp; Tyedmers, P. H.</td>
</tr>
<tr>
<td>REF</td>
<td>Reference</td>
</tr>
<tr>
<td>--------</td>
<td>---------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>