Prof. Caj Södergård,
Technical Coordinator,
VTT

European Big Data Value Forum
Helsinki, 14.10.2019
What do we mean with the DataBio platform?

- A technical **environment**
  - where software is **developed**
  - to be **deployed**, e.g. as Docker modules, in hardware, operating system, middleware or a cloud

- We focus on **Big Data**, handling high **Volume**, high **Velocity** and high **Variety**.

- DataBio platform provides a big data **toolset** which offers **functionalities** primarily for **services** in agriculture, forestry and fishery.

- The functionalities enable new **software components** to be **combined** with both open source **and** proprietary components and infrastructures.

- The DataBio toolset supports the forming of reusable and deployable **pipelines** of interoperable and **replacable** components provided by several partners.
DataBio platform serves the 26 pilots
DataBio platform serves the 26 pilots

DataBio platform with big data components and datasets

WP4 Components & IoT datasets
DataBio platform serves the 26 pilots

Big Data Pipeline

WP4 Components & IoT datasets

WP5 Components & Earth Observation datasets

Agro Pilot 1
Agro Pilot 2
Agro Pilot 13
Forest Pilot 1
Forest Pilot 2
Forest Pilot 7
Fishery Pilot 1
Fishery Pilot 2
Fishery Pilot 6
Pipelines vs. Services

Pipeline

• A **pipeline** = a chain of processing components arranged so that the output data of each element is the input data of the next
• Has clear interfaces between components and to environment.
• Is a ”white box” showing internal wiring for the developer (configuration & deployment)

Service

• Provides usability, not inner wirings of components, nor where they are deployed.
• Accessed through API:s (web services, remote call).
• Refer to so called ”end points” (queries) activated remotely and executed in the cloud.
• Represents a ”black box”.

BDVA Reference Model
Platform development in DataBio in numbers

- **60 components** in two trial rounds
- **1-6** Components used in pilots
- **1,8** components /pilot
- **14** new user interfaces
- **59** new APIs
- **+2,7** in Technology Readiness Level (1-10):
Managing project assets: components, pipelines, datasets and reports

- **DataBio Hub** helps disseminating results by providing a catalogue of various public and private digital assets of a project with interdependency and ownership information.

- Describes 95 **components**, 39 **datasets**, 12 **pipelines**, and 25 **pilots** and provides links to project reports.
Example of pipelines in a Fishery pilot

DataBio platform with big data components and datasets

WP1-3

Agro Pilot 1
Forest Pilot 1
Fishery Pilot 1

Agro Pilot 2
Forest Pilot 2
Fishery Pilot 2

Agro Pilot 13
Forest Pilot 7
Fishery Pilot 6

Deliverables D4.1, D4.2, D4.3, D4.4, Milestone M7, M8

WP4 Components & IoT datasets

WP5 Components & Earth Observation datasets

Deliverables D5.1, D5.2, D5.3, D5.4, Milestone M9, M10
## Pilot: Oceanic tuna fisheries immediate operational choices

<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pilot leader</td>
<td>Zigor Uriondo, EHU/UPV, <a href="mailto:zigor.uriondo@ehu.es">zigor.uriondo@ehu.es</a></td>
</tr>
</tbody>
</table>
| Objective     | • Improve vessel energy efficiency  
                   • Predict machinery faults  
                   • Vessel loading to minimize fuel consumption |
| KPIs          | - Fuel vs kg catch  
                   - nautical miles sailed,  
                   - catch efficiency |
| Trial Focus   | Historical data collection, collation and visualization.                     |
| Datasets & Status | 3 Ships with 3 to 4 years operational data collected  
                                   • Fish catches  
                                   • Fuel oil consumption  
                   | 117 ship measurements every 10s  
                                   • main engine (1)  
                                   • flows to/from engine  
                                   • propeller (1)  
                                   • aux engines (5)  
                                   • flows to/from aux engine  
                                   • fish hold (20) |
Components from partners form pipelines
This document is part of a project that has received funding from the European Union’s Horizon 2020 research and innovation programme under agreement No 732064. It is the property of the DataBio consortium and shall not be distributed or reproduced without the formal approval of the DataBio Management Committee. Find us at www.databio.eu.
Event driven application
Deployment - location view
Installation on ship
OpenVA realtime user interface showing events generated by IBM PROTON
Conclusions

• DataBio platform is an *environment* for developing and deploying software

• It uses primarily components from partners collected into *pipelines*

• The DataBio platform toolset has 60 components supporting 27 pilots

• Components have been refined into higher TLRs and combined into multiple new pipelines for agriculture, forestry and fishery
77 years of innovations

Read more: www.vttresearch.com