



EUROPEAN FEDERATION OF
DATA DRIVEN INNOVATION
HUBS

Experiments in Practice: Poznan Supercomputing and Networking Center offer and example scenario

Marcin Plociennik

PSNC



The EUHUBS4DATA project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951771



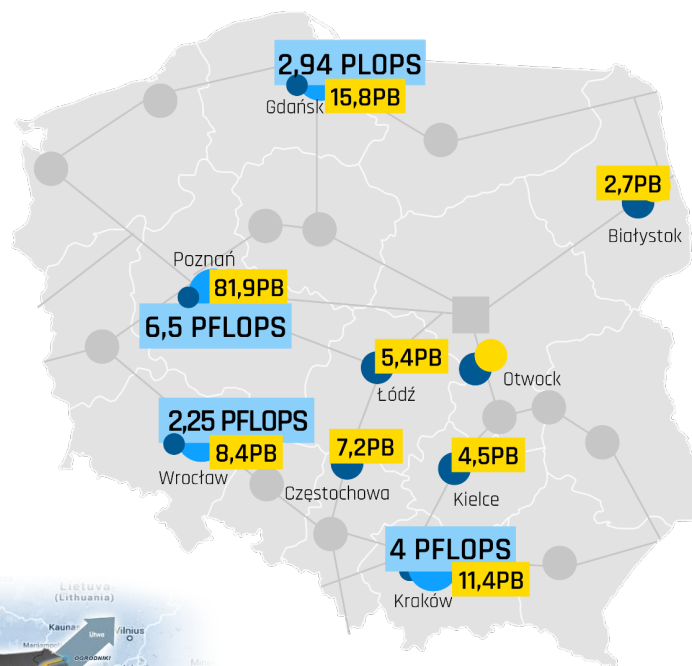
BIG DATA VALUE
PUBLIC-PRIVATE PARTNERSHIP

This project is part of **BDV PPP**

PSNC

- **Leading operator of Polish e-Infrastructure**

- National Research and Education Network – PIONIER
- Research Metropolitan Area Network - POZMAN
- HPC Center: upgrade to 15 PFlops
- 2 Data Centers: over 80 PB under upgrade → NDS



PSNC – data driven eInfrastructure

National Data Storage infrastructure:

Planned:

250 PB tape (long-term storage/archive)

250 PB disk (on-line storage, data lake)

10 PB SSD/NVMe (access acceleration)

Target: 1 EB of storage capacity

Partners: 5 HPC & 4 MAN sites:

HPC: PSNC, Cyfronet, TASK, WCSS, NCBJ

MAN: Białystok, Częstochowa, Łódź, Kielce



A complete environment for **data-driven experimentation**:

Data preparation & preservation:

- Data acquisition, cleaning, enrichment
- Discovery & search
- Sharing & publishing
- Data re-use, discovery and exploitation
- Data protection / preservation

Data processing & analytics:

- High performance computing
- Data Analysis, Data Science...
- Machine Learning, AI



PSNC as hub in EUhubs4Data

HPC⁴ Poland



SERVICE NAME	CATEGORY	DESCRIPTION	NETWORKS
GEOSPATIAL TOOLKIT	Big Data Platforms (PaaS)	PSNC geospatial toolkit brings together a set of tools provided b...	GeoTools, GeoSpatial, map creation, geo-data
LINKED DATA PIPELINES SERVICE	Product Development	The main goal of these pipeline is to define and deploy (semi-) a...	data integration, linked (open) data, Ontologies, pipelines
ROHUB	Data Management	ROHub is a holistic solution for the storage, lifecycle managemen...	fair objects, research management, research objects
SEMANTIC ANNOTATION	Product Development	The semantic annotation service provides a simple REST API design...	agriculture, Entity recognition, linked (open) data, text mining
SPARK	Big Data Platforms (PaaS)	Apache Spark is general purpose, high performance and horizontall...	spark,machine learning
SYMBIOTE	Data Management	Symbiote is an ecosystem for connecting, exchanging information a...	IoT interoperability, sensor data

Mission

Create and provide advanced high performance computing tools, addressing the real-life demand of Polish manufacturing companies

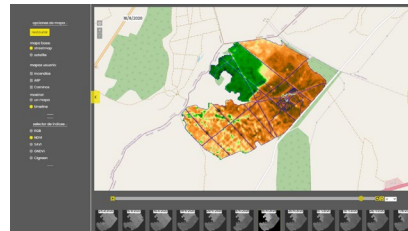
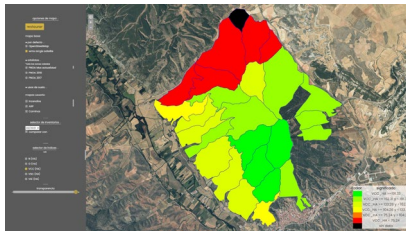
Datasets

OPEN LAND USE - OLU	Agriculture, Fisheries, Forestry and Food, Environ..	Creative Commons CCZero (CC0-1.0)	Czech Republic, Poland, Spain	Open Land Use Map is a composite map that is intended to create d...
OPEN TRANSPORT MAP - OTM	Agriculture, Fisheries, Forestry and Food, Transpo..	Creative Commons CCZero (CC0-1.0)	Czech Republic, Poland, Spain	The Open Transport Map displays a road network which - is suitabl...
POLISH LAND PARCEL INFORMATION SYSTEM (LPIS) DATASET	Agriculture, Fisheries, Forestry and Food, Environ..	Creative Commons CCZero (CC0-1.0)	Poland	Polish Land Parcel Information System (LPIS) data (land parcel an...
SMART POINTS OF INTEREST DATASET	Agriculture, Fisheries, Forestry and Food, Environ..	Open Data Commons Open Database License (ODbL-1.0)	EU-wide	The Smart Points of Interest dataset (SPOI) is the seamless and o...
URBAN ATLAS (LAND USE AND LAND COVER DATA FOR LARGE URBAN ZONES) DATASET	Agriculture, Fisheries, Forestry and Food, Environ..	Creative Commons CCZero (CC0-1.0)	Czech Republic, Poland, Spain	This dataset contains agriculture related lands (hilucs_code<200)...



Success story - EUhubs4Data experiment

- **Förecast** provides real-time Earth Observation (EO) geo-information products by leveraging and integrating data from satellite, airborne LiDAR, aerial orthoimages and forest information to create **advanced algorithms** for high accuracy, reliability, up-to-date and high resolution to forest assets across different **ecosystem services**.
- **New** the design and operative **architecture: BigData Infrastructure**
- **New methodology to get non-parametric models** based on field truth and auto labelled.: **Geospatial scalability**



- **PSNC services used:**
 - **Geospatial toolkit**
 - Hadoop platform used for Geospatial Toolkit: 384 CPUs, 768 GB RAM, 21TB storage
 - **Cloud:** 11 nodes x 8cores x16GB RAM, 19.5 TB storage



LONG RANGE PLANNER

Increase up to 160% the data acquisition capabilities of long-range drones in power line inspection thanks to developing an automatic mission planning

COMPANY	COUNTRY
FuVeX	Spain

AI4SAM

This experimentation will focus not only on the optimization problem of maximizing the reachability to the mobility means while minimizing the installation costs, but also on exploring incentivization models for fostering citizens engagement to soft ...

COMPANY	COUNTRY
KENTYOU	France

FÖRECAST 2.0

Förecast is a forest intelligence solution, designed to offer accurate and updated forest inventories in a multi-purpose platform. forecast provides real-time Earth Observation (EO) geo-information products by leveraging and integrating data from ...

COMPANY	COUNTRY
FORA FOREST TECHNOLOGIES, SLL	Spain

Federation value – PSNC perspective

Short term

- Common service and data catalogue
- Visibility and networking
- Wide offer of the trainings in federation
- Building blocks supporting data spaces

Long term

- Strong relations with other *DIH's*
- Access to wide range of services for the local ecosystem SME's
- Interoperability
- Trust building



EUROPEAN FEDERATION OF
DATA DRIVEN INNOVATION
HUBS

Thank you !

Marcin Plociennik

PSNC



The EUHUBS4DATA project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951771



BIG DATA VALUE
PUBLIC-PRIVATE PARTNERSHIP

This project is part of **BDV PPP**