A GFBI Hub at the University of Lleida
From data to knowledge

Albert Sorribas
Vice-chancellor for Scientific and Technology Policies
University of Lleida (Spain)
Our findings reveal a consistent positive concave-down effect of biodiversity on forest productivity across the world, showing that a continued biodiversity loss would result in an accelerating decline in forest productivity worldwide.
A GFBI Hub must be more than just a repository of data.

Methods for extracting **meaningful information** are paramount for the GFBI community.

**Collaborative expertise** will be determinant in reaching the long term capabilities of a GFBI Hub.
Up-to-date maintenance of information

Integration of data from different sources

Project oriented data

Data validation and standardization

Secure data sharing

Needs

Coordination!

Automatic Follow-up of critical information

Coordination!

Speed-up collaborative projects

Log term vision
• Keep a secure access to DB
• Explore new opportunities (*there are no dragons!!*)
• Incorporate complementary information
• **DB Marketplace**: promote collaboration and projects
• **Computational issues**: methods, models, parallelization, ....
• Promote interdisciplinary work
• Consensus on data organization and identification.
• Distributed solutions.
• **Governance** (access control, coordination, data consolidation, etc...).
• Data for large GFBI projects
World-wide service & replication
GFBI Hubs coordination
To start a Data Center with added value services for the GFBI scientific community.

To become a reference in data management, integration, and processing so that the GFBI can undertake its (big) projects.

To become a reference for developing new data analysis techniques that can help the GFBI to achieve its goals.
UdL-GFBI Hub

Basic features

• A **data center** for the use of the **GFBI** members and authorized collaborators.

• **Secure storage of information** with an appropriate **control of ownership and information retrieval**.

• Added services for **preprocessing** information according to the specific requirements of a given project.

• Develop a **strategy** for data integration, visualization, and analysis.
External data sources integration aimed at incorporating, mirroring and/or syncing, when possible, information of external data sources.

- Transparent and user friendly
- Easy management and synchronization
- Project oriented protocols
- Data preprocessing and curation
- **Frontend** aimed at exploring information and showing indicators dashboards.
- **Backend** aimed at managing platform users, roles, permissions, datasets accesses, etc.
- **Coordination and Governance module** aimed at giving access to resources to privileged requests.

- Follow-up important parameters
- Access control
- Prevent information misuse
- Facilitate group interactions
Communication module focused on publisher/subscriber architecture pattern, where consumers may subscribe to certain topics, so that they are informed automatically when an event occurs.

- Facilitate up-to-date information and data.
- Alerts.
- Hybrid IoT/Big Data architecture, able to run advanced machine learning processes, but also with classical analytics support
- **Analytical engine**: made up of stream analytics, batch processing, alerts manager, and reporting
- **Storage** based on big data (HBase), in memory layer (REDIS or IGNITE) and relational (PostgreSQL)

- Data analytics is fundamental
- Implement advanced methods
- Explore new techniques
- Data visualization
Copyright Ontology
(Copyright example)

Data Access

Blockchain

Store
(tamper-proof and auditable)
Data Access

Action: republish (governed by Making Available Right)
- **Who**: any research institution
- **What**: datasets/0Hu1cY (dataset fingerprint)
- **When**: from 2017-07-30 (start), during 12 months (duration)
- **How**: non-exclusive
- **If**: attribute
  (recipient: GFBI, what: “Source: GFBI 2017”)

Can we re-publish dataset 0Hu1cY

Agree

Disagree

No, because it matches policy which prohibits it
The University of Lleida (UdL) is one of the leading institutions in Spain in forest research and is an active member of GFBI.

The Scientific and Technological Park of Lleida is dedicated to computation and agrifood industry.

EURECAT is the leading technological center in Catalonia, with headquarters in Lleida.

The Forest Sciences Centre of Catalonia (CTFC) is a research center of reference in forestry management and related industries in the Mediterranean area.

We lead the Agritech BIG DATA platform which provides support for the agrifood industry in collecting, processing, and analyzing data from many different sources.
Data Science at Lleida
Added Values

- Office space for temporal stay of researchers in Lleida (shared space in Agritech BIG DATA headquarters).
  - A controlled computer room for students.
  - Access to UdL services.

- Technical meetings
  - Coordination with other GFBI Hubs
  - Explore new analytical solutions
  - Incorporate expertise from technical disciplines

- Project planning
  - Prepare data for specific projects
  - Coordinate world-wide initiatives

- Support for mid term stays of a GFBI designated person as affiliated to UdL
UdL-GFBI Hub

activities

- Focus on exploring new techniques
- Open to complementary disciplines
- Project meetings
- Data Science collaborative courses
- Annual UdL-GFBI Hub meeting
- Hands-on Ph.D. students mentoring
- Open to society
See you soon in the opening of our GFBI Hub in Lleida!!!