Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

# The Bank of France datalake[1]

Renaud Lacroix,

Bank of France

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

BANQUE DE FRANCE
EUROSYSTÈME

Directorate general Statistics

# The Banque de France Datalake


## Renaud Lacroix

Director, Statistical and IT engineering division


« Building Pathways for Policy Making with BigData »

BI – IFC/BIS Seminar
Bali, 26 July 2018

BANQUE DE FRANCE
EUROSYSTÈME

# A FEW DRIVERS

## 1. New age of statistics

- ✓ Growing appetite for statistics…while
- ✓ General public more skeptic with regard to numbers (including official ones!)
- ✓ Legitimate requests for granular data
- ✓ New and very powerful competitors (GAFA)

## 2. Strategic challenge for Central banks

- ✓ Central banks must be able to **deliverer rapidly reliable intelligence at both micro and macro levels.**

## 3. Large consequences for data management

- ✓ Functional silos are not adapted anymore
- ✓ A clear guidance and an orderly process is a key to manage wide volumes of diverse data
- ✓ An innovative and scalable technology is crucial

BANQUE DE FRANCE
EUROSYSTÈME

# A POSSIBLE WAY FORWARD: BUILDING A DATALAKE

**1. Building a coherent and unique set-up**: for data

- ✓collection,
- ✓quality management,
- ✓pooling,
- ✓analysis,
- ✓dissemination

**2. Integrating the Big Data techniques**

**3. Delivering both economies and better work :** it is possible to do more with less spending

BANQUE DE FRANCE
EUROSYSTÈME

# OBJECTIVES OF THE DATALAKE PROJECT

- **The Datalake project consists in the building of a multidisciplinary granular data platform supplying flexible and innovative services to internal users.**

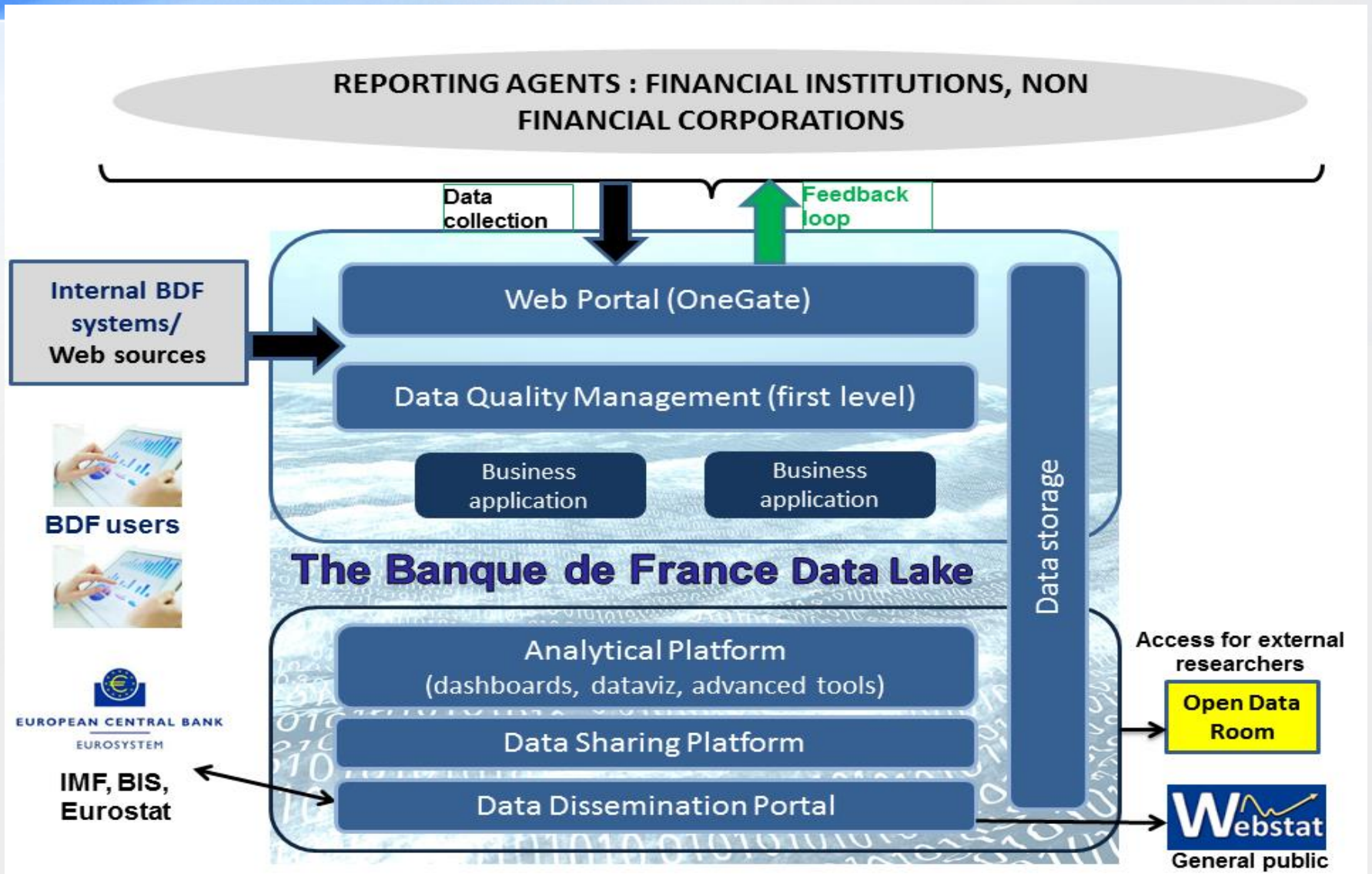
- The platform provides key services pivotal to operational activities of the Banque de France: data collection, data supply, data quality management, data storage, data sharing, analytics, and dissemination (external data exchange platform).

- 2 new components :
  - data storage and **automated data quality management** for first-level quality checks
  - analytical platform (dashboarding, statistical treatments covering a wide range of user needs)

- **The implementation phase of the project will be completed by end2018**

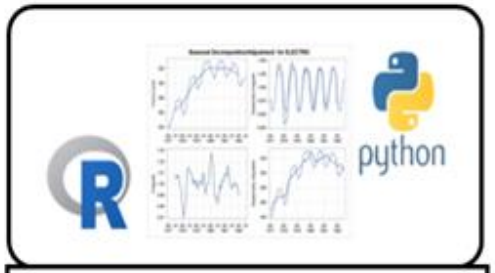
BANQUE DE FRANCE
EUROSYSTÈME

# AN OVERVIEW OF THE DATALAKE ECOSYSTEM



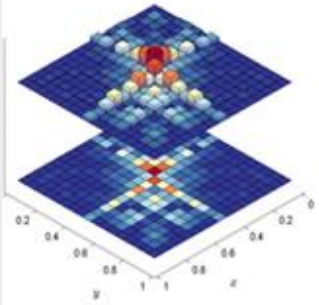
# Implementation of the Datalake core structure: an analytical platform for all users


Interactive query tools




Libraries of statistical functions, programming tools


Data science: machine learning, neural networks, deep learning



From prototyping ....

.... to large-scale production

- Specifications - Samples - Large-scale production - Final product


**Final product designed by end-users**

**« Industrialized » final product with the support of the project team**

BANQUE DE FRANCE
EUROSYSTÈME

# Early BDF customers of Datalake services

## ❑ Banking Services Directorate

- Gathering payments for private and corporate customers monitored by the BDF in a unique decision-making system.
- Provide the analytical tools to contribute to the fight against money laundering and terrorism financing.

## ❑ DATAGAPS project

- G20 recommendations on gathering information reported by systemic banks on financial risks by type of exposures

## ❑ Anacredit project (implementation of the ECB regulation)

- Collection of detailed information on both credit lines and credit risks from credit institutions

## ❑ Reengineering of the French NSA's information system

BANQUE DE FRANCE
EUROSYSTÈME

# A FEW FIRST CASE STUDIES IN MACHINE LEARNING

## ➢ Nowcasting / forecasting French GDP

- **Use of a rich database (more than 200 explanatory factors)** in the framework of adaptative LASSO with automatic selection of relevant variables for forecasting purpose
- **Valuable complement to more traditional approaches for forecasting**

## ➢ Improving tourism statistics

- Web scraping of accommodation platform (Airbnb, Booking.com,..) and machine learning to anticipate future demand relying on meteorological data and future events
- Comprehensive Use of credit card data for the estimation of both the spending of French residents abroad and the spending of foreign residents in France
- First attempts to use mobile phone data

# KEY CHALLENGES OF THE PROJECT 1/2

- **From an organizational perspective** :

  - **Top-down strategy, but bottom-up implementation :**
  - **Strong involvement** of the management at all levels

  - **Adapt to users requests** : business lines contribute to the definition of Datalake services while the Datalake team defines standards for mutualized services

  - Move **step by step** in order to deliver the best services to the customers

  - **Reshape the organization of work to benefit from automation and new techniques**

  - **Need for data-skills** in big data technologies (computer science, data science, IT experts) : develop internal training

  - **Leave no-one by the wayside** : everyone should benefit from the Datalake services and understand what we are trying to achieve and why

BANQUE DE FRANCE
EUROSYSTÈME

# KEY CHALLENGES OF THE PROJECT 2/2

- **On the IT side** :
  - large changes on IT infrastructures
  - a large number of components (softwares) are required, including some brand new
  - Do not seek innovation at any price !
- **On the statistical side** :
  - Learn to work with new data sets from Google, social media, wesites
  - Learn to work with new data analysis algorithms and enrich the traditional toolbox
  - Develop (interactive) visualizations tools and techniques for data dimensionality reduction
  -

BANQUE DE FRANCE
EUROSYSTÈME

# BIG CHALLENGES AHEAD FOR THE PUBLIC SECTOR

- **Public authorities are now in direct competition with the private sector in the sphere of economic information**
  - the appetite for real time intelligence can be inflated by the Big Data Revolution
  - GAFA and other global players are not at all regulated
  - technological progress in Big Data is permanent, rapid and difficult to anticipate

- **Risks therefore that "bad data chases good ones"**

- **Central banks must be more innovative and user friendly**
  - Make the publications more readable and visible
  - Supplement Bigdata strategy by open data initiatives