



BIG DATA AND MACHINE LEARNING AT THE BANQUE DE FRANCE

JACQUES FOURNIER

DIRECTOR GENERAL STATISTICS

**« HARNESSING BIG DATA & MACHINE LEARNING TECHNOLOGIES FOR
CENTRAL BANKS »**

**BANCA D'ITALIA
ROME, 26/03/2018**



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



A POSSIBLE WAY FORWARD: BUILDING A DATALAKE

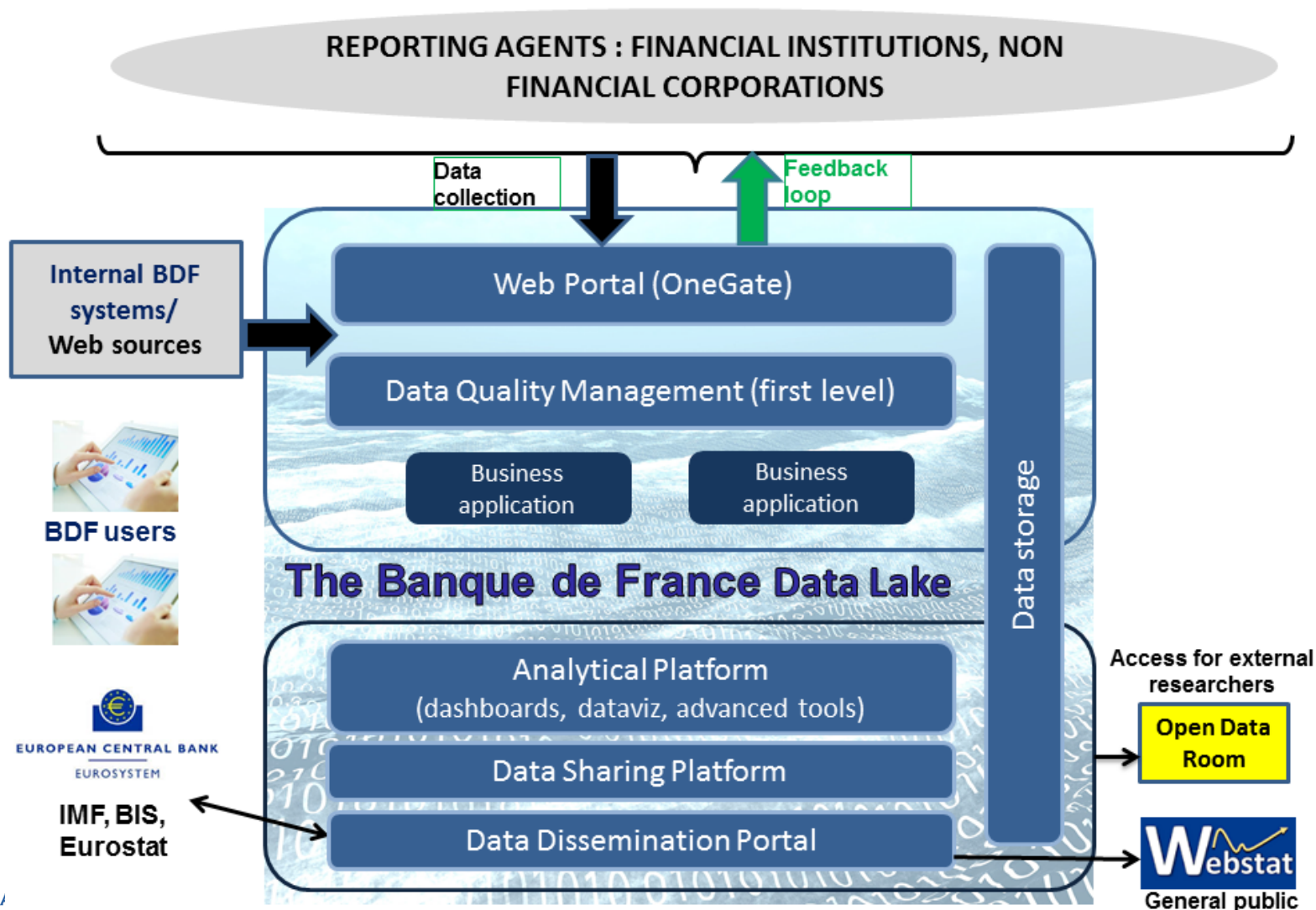
1. Building a coherent and unique set-up: for data

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

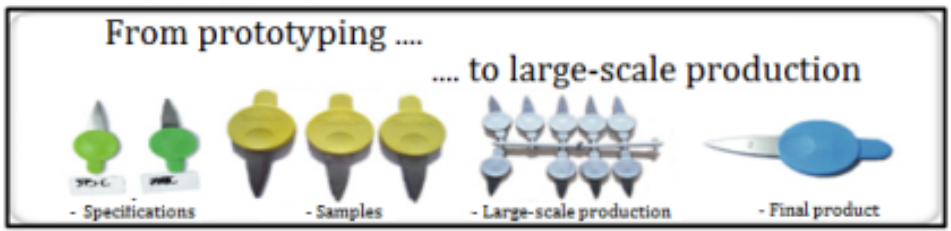
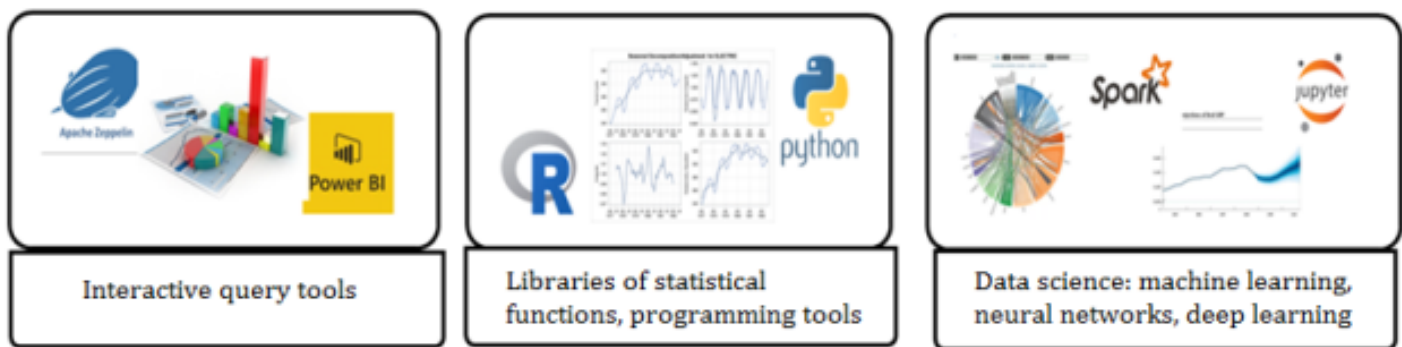
2. Integrating the Big Data techniques

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

THE BDF DATALAKE GOES-LIVE AS OF 2018 (ALL LAYERS AVAILABLE BY END 2018, PHASED INCLUSION OF BUSINESS LINES)



THE DATALAKE PROJECT : AN ANALYTICAL PLATFORM FOR ALL USERS



Final product designed by end-users



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



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

➤ **Estimation of households green financing (when embedded in non earmarked loans)**

- Google Trends data allow for the assessment of the popularity of queries related to such loans
- Calibrations linking those Google Trends data to other official statistics



BIG CHALLENGES AHEAD FOR THE PUBLIC SECTOR

- **Public authorities are now in direct competition with the private sector in the sphere of economic information**
- **GAFAs and other global players are not at all regulated:**
 - Do not deliver easily their data to official statisticians
 - ... and when they are available, these are raw data which do not fall within the standard framework of statistical analysis
 - Use marketing techniques that can shadow official statistics
 - Are often obscure on their methodology
- **Risks therefore that “bad data chases good ones”**
 - Central banks can be more innovative and user friendly...
 - Will it be sufficient?
 - Need for a European/global regulation?