BANQUE DE FRANCE
EUROSYSTÈME

Directorate General Statistics

Pooling and Sharing Statistical Series (P3S)

An on-going data management initiative at the Banque de France

Renaud Lacroix

Director, Statistical and IT engineering division

CEMLA-FIF, 8 June 2015

Outline

- Goals of the new set-up and key principle
- Data and confidentiality issues
- Governance
- Technical solution
- Project plan
- Access to individual data for external users

Goals of the new set-up

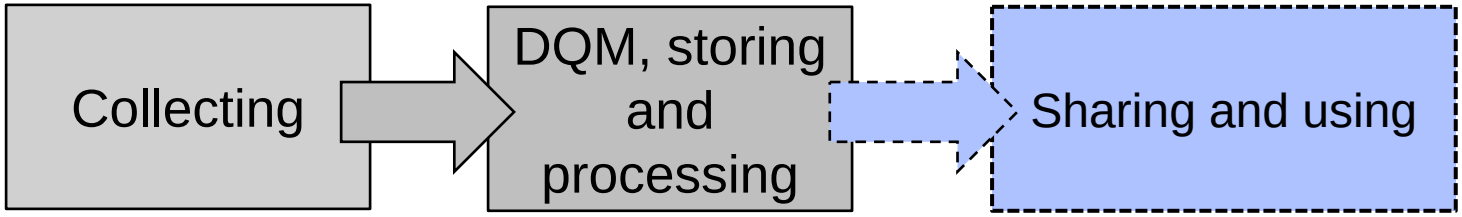
- **Pooling data ….**
  - To gather data on financial institutions and non-financial corporations
  - Collected by the CB and by the Banking and Insurance Supervisory Authority
  - While respecting confidentiality rules

Goals of the new set-up

- **…to allow enhanced analysis for all involved departments and for the supervisory authority**
  - Offering access to internal users on a ‘need to know’ basis
  - Fostering synergies and economy of scale

Goals of the new set-up

- **Create a value chain between all supervisory and CB functions**



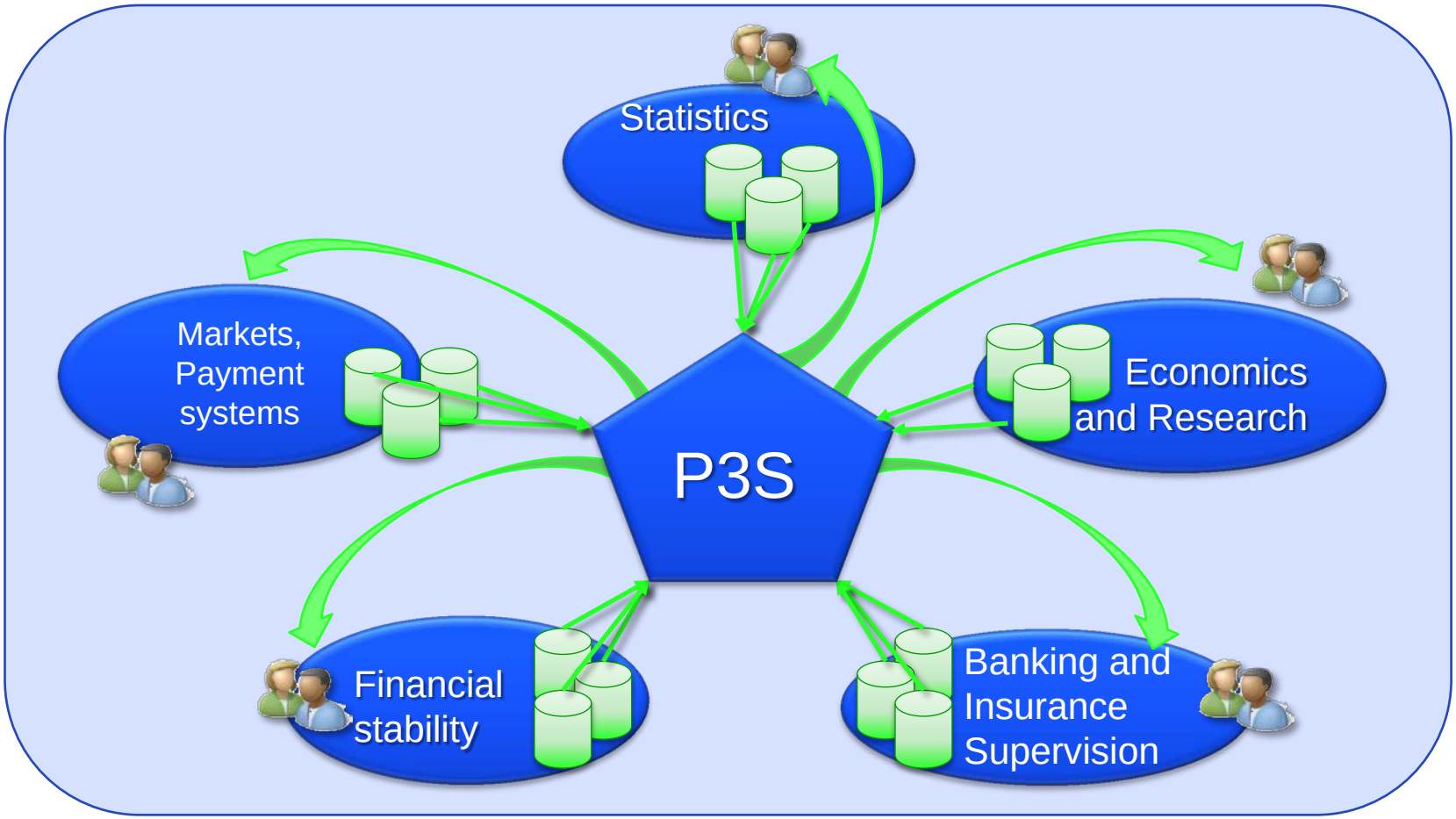
- **Contribute to the Banque de France Digital Plan implementation**

- **Add a new dimension to the Banque de France information system**

- **Enable a data sharing that spans ‘silos’ and is crucial to analyse and better mitigate risks in the future - the ultimate aim being strategic and not technical.**

Key principle

Users use P3S data in their own Information System.
P3S data (and their metadata) are available in alternative formats (SAS, CSV)



Data and confidentiality issues

- **A collaborative work involving 5 DGs**
- Work-streams on confidentiality issues with representatives of all stakeholders and of the Legal department
- Data “offering” and “demand” expressed by each DG
- P3S data typology compliant to legal constraints

Data and confidentiality issues

- **15 main datasets**

Data from credit institutions
Data from securitization bodies and investment firms
Insurance data
Consolidated prudential data
UCITS
Household over-indebtedness
International banking data
Money and interbank market
Data from payment institutions and electronic money issuers
International activities of firms
Business survey
Securities holding and issue
Data from corporates
National data in TARGET 2
Means of payment

Data and confidentiality issues

- **Most (individual) data can be shared** by personal accreditations updated every 6 months *(Generally shareable data)*
  - *Monetary statistics, credit register, Financial reports from banks,…*
- **For a subset of more sensitive data,** access will be precisely limited and granted on a need-to-know basis *(Non generally shareable data)*
  - *Components of the solvency & liquidity ratios, information on business managers,….*
- **Access rights are defined** through an “Accreditation Matrix” which crosses sub-family and users

Data and confidentiality issues

- **Accreditations will be defined for subsets of data**
  - 202 subsets of data
  - 35 subsets « generally non shareable »
  - 167 subsets « generally shareable »

- **Functional matrix (supply/demand in datasets terms)**
  - Generally shareable data pooling rate ≈ 3

Governance : P3S Validation and Monitoring Committee (PVMC)

- **Committee at DG level, co-chaired by DGS and IT, including all stakeholders and the Legal department**
- Validating the lists of accredited agents
- Monitoring P3S functioning
- Addressing unforeseen issues
- Seeking for consensus

Governance : main tasks undertaken by the PVMC

Definition of access policy

Updates of the white list
(cartography of individual access to each dataset)

PVMC

High level monitoring

P3S PSG

Annual report to Governor and Deputy Governors with a presentation to BdF Executive Board

Governance : operational process for user accreditation to P3S for *generally shareable data*

User request

Management approval (Head of the unit)

P3S Service Desk

Status of the request

Person in the white list for the dataset required

Yes : access granted

No : access is temporarily denied and the request is submitted to the PVMC (update of the white list)

Governance : operational process for user accreditation to P3S for *restricted data*

User request

Management approval
(Head of the unit)

P3S Service Desk

If no consensus reached (power of veto from data producers)

PVMC

Governor, Deputy Governors

Yes / No

Decision communicated by the PVMC : access granted or denied

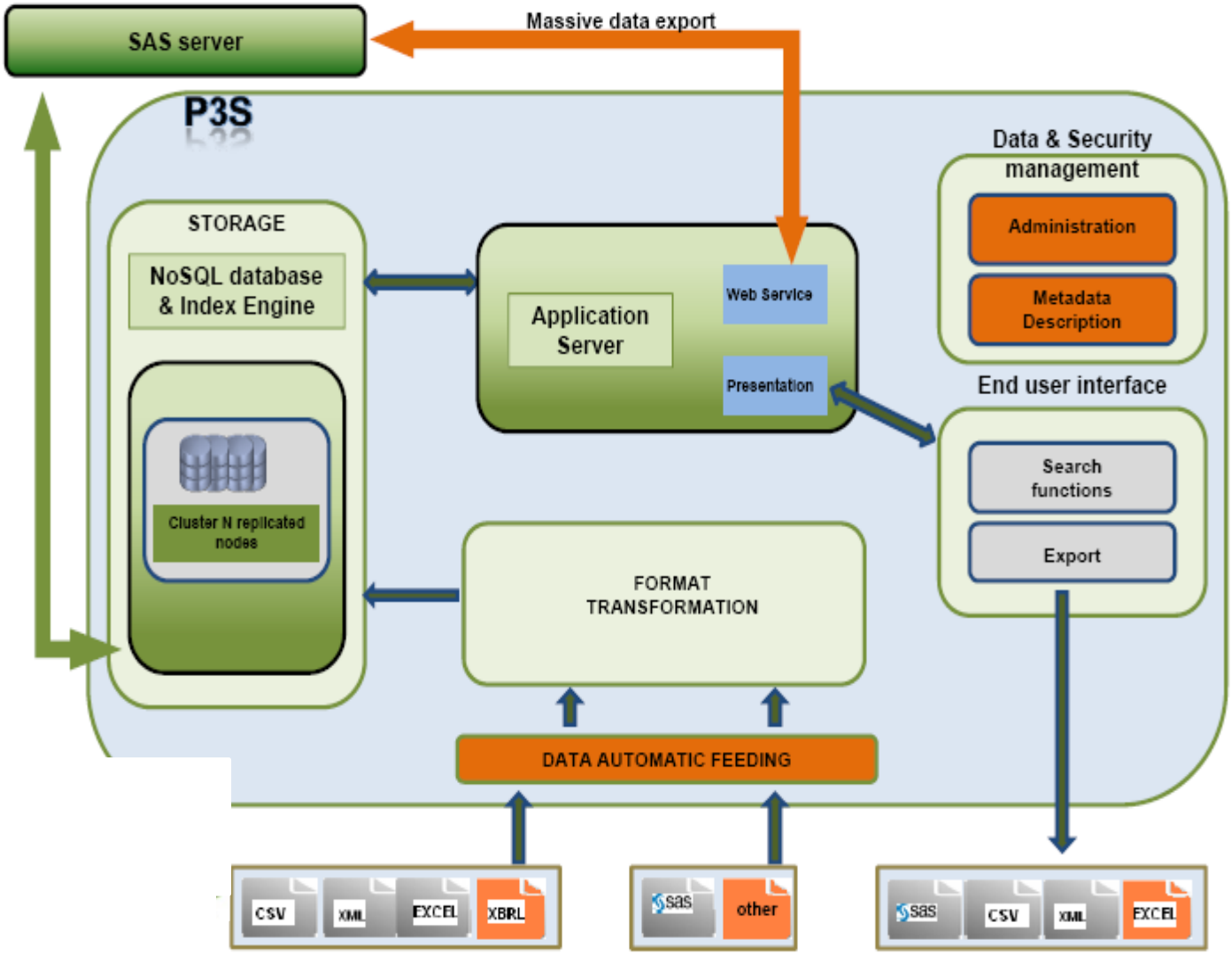
Technical solution : design

- Ability to manage large volumes of data
- Secured access rights
- Flexibility to integrate and handle heterogeneous data:
  - **Codification** of series is specific to each dataset
  - **Technical formats** : XBRL, SDMX, SAS,…
- Powerful research engine
- Scalability and interoperability (integration of new data sets, connection to additional analysis tools)
- A “Proof of Concept” has been performed and approved by key users

Technical solution : architectural frame

- **A solution based on open source « Big Data » technology**
  - A dedicated ‘BigData’ platform in the Banque de France Datacenter
  - An up-to-date search engine (ElasticSearch)
  - A Not Only SQL (NoSQL) DataBase well suited for the storage of heterogeneous data
  - Main differentiator : all formats are accepted (SDMX-ML, XBRL,..)

Technical solution : architectural frame



Technical solution : metadata

- **Metadata are used** to describe the statistical series, support the search functionalities and link heterogeneous data
- **Metadata selection is performed by data scientists** who identify the most suitable information to be used as Metadata. This is an ongoing process, the Metadata list being enriched continuously. Furthermore Metadata can be queryable or not, depending on their nature
- **Non queryable metadata** also called “descriptive” metadata provide information on sub-families and can be visualised in the data catalog
  - *Data sensibility, applied methodology,….*
- **Queryable metadata** can be used as search criteria and address either common or specific attributes
  - *Family, sub-families, source, population,… (common attributes)*
  - *Accounting sector,… (specific attributes)*

Technical solution : search engine

- **ElasticSearch** is the selected search engine for the P3S
- **This search engine is well suited** for “BigData” contexts with large volumes as well as unstructured and heterogeneous information stored in a NoSQL database
- **Search can be performed** not only through pre-defined criteria such as the (queryable) metadata but also using “natural language”.

Technical solution : storage

- **2000 GB of data integrated into MUSES,** generating a storage need of 10000 GB after having taken into account the auxiliary tools required for the organization, research and data identification

- **Target : 400 - 500 million series**

- As a comparison, the BDF macro-economic database records 6 million series - i.e. 28 GB of data

Project plan

- **Project duration is at least 2 years**
- **Four steps are considered**
  - **Step 0 (end 2014 – early 2015):**
    Initialization of the technical base
  - **Step 1 (june 2015) :**
    Most functional and technical requirements
    Integration of a first significant group of subset of data, for all DGs
    Related control and restitution functionalities
  - **Step 2 (end 2015) :**
    All requirements
    Further integration of the subsets of data
  - **Step 3 (2016) :**
    Completion of the integration of the subsets of data

Project plan

- **The realization phase is on track and on schedule**

- **First go-live is scheduled for July 2015 :**
  - Integration of an highly significant group of data covering all business areas
  - Access granted to a first batch of 60 users from all business areas (BDF and Supervisory authority)

- **Starting in January 2016 :**
  - Progressive accreditations increase (300 users as a target)
  - Integration of additional datasets in P3S

Main datasets to be loaded in P3S by end 2016

- BSI and MIR Data from credit institutions
- FINREP reporting
- Other prudential data (solo & consolidated)
- Data from securitization bodies and investment firms
- International banking data
- Insurance data
- Investment funds data
- Business surveys
- Securities holding and issuance
- Credit Register
- Data from the balance sheet of corporates
- Granular data on 'Household over-indebtedness'
- Data on negotiable debt securities

Access to individual data for external users

- Since 2012, the Banque de France has reset the instruction procedures for individual data access;

- The Banque de France has chosen a pragmatic approach:
  - Prevent legal risks
  - Contain operational costs;

- The procedure aims at guaranteeing :
  - A level playing field among researchers
  - A better-governed process
  - A full transparency of data requests.

A new process implemented by the Directorate General Statistics (1/3)

- The applicant(s) fills in a detailed application form describing the research project and the team organisation
- A confidentiality agreement is signed by each member of the research team
- **Applications are collectively reviewed by a decision body (‘Secretariat for the instruction of data requests ’) chaired by the Director General Statistics.**

A new process implemented by the Directorate General Statistics (2/3)

- The committee is composed of representatives from all business areas of the Banque de France and the legal department ;
- The Secretariat decides on the approval of the request based on :
  - the relevance of the project
  - the legal framework under which the data have been collected

    Strict adherence to the European regulation for data collected according to an ECB regulation
  - the data protection measures indicated by the applicant.

A new process implemented by the Directorate General Statistics (3/3)

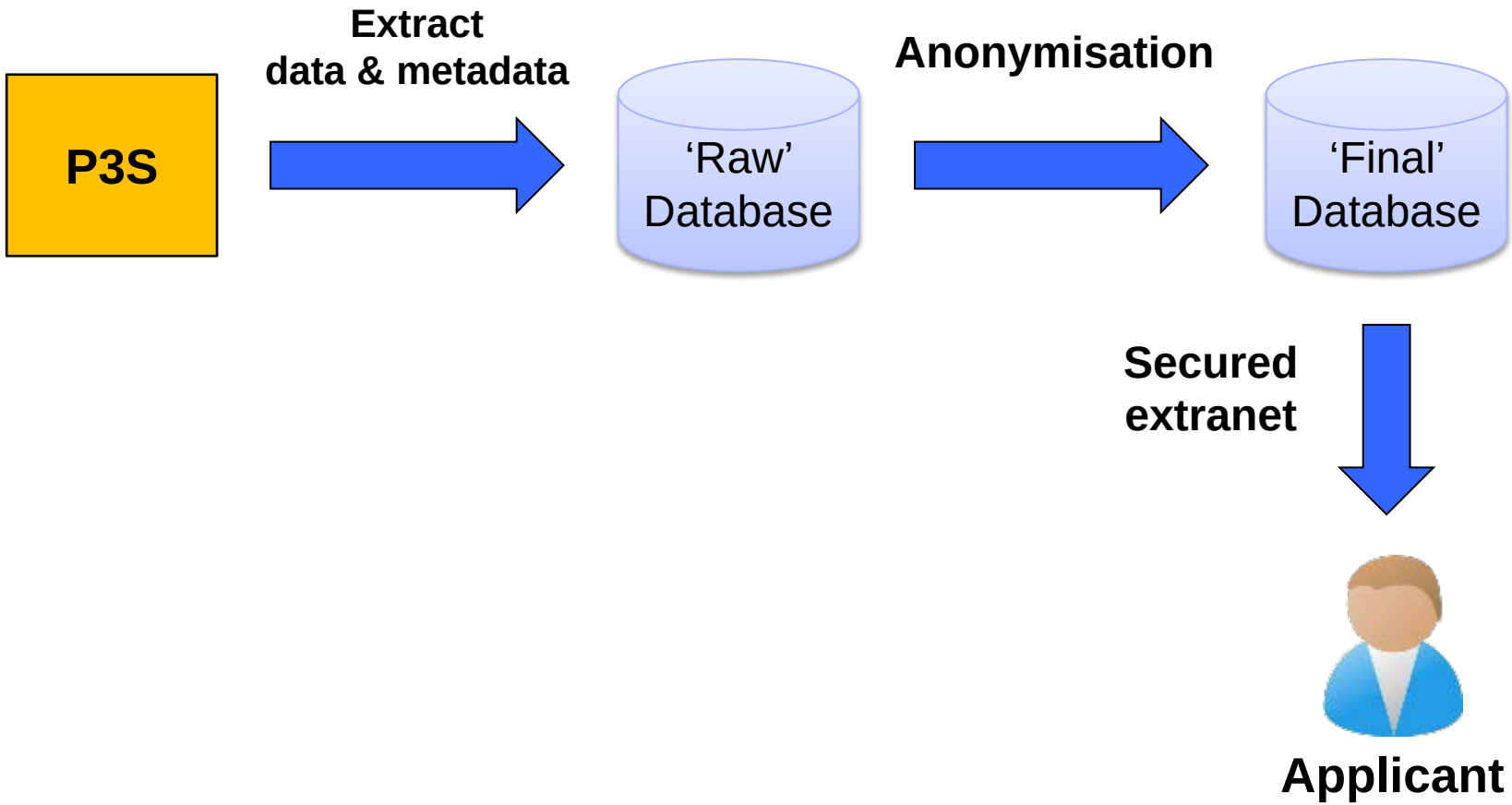
- In case of approval of the request by the Secretariat, the DGS is responsible for providing the data to the applicant (in close cooperation with the unit in charge of the production of the relevant data) ;

- The entire technical process may be lengthly and cumbersome :

  1. Data extraction and preparation : specific to each dataset
  2. Anonymisation : challenging – and difficult to automate – when indirect identification must be taken care of.
  3. USB drive hand-delivered to the applicant

Towards a more efficient process

Though an internal platform designed exclusively for BDF users (including supervisory departments), P3S will allow to speed up the first phase and enrich the data with metadata



Thank you for your attention