Post-crisis skills landscape: the emergence of “purple people”¹

Luis Teles Dias,

Bank of Portugal

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.
Post-crisis skills landscape: the emergence of “purple people”

IFC-BNM-ECB Satellite Seminar “Post-crisis data landscape: Micro data for the macro world”

Kuala Lumpur, 16 August 2019

Luís Teles Dias
Deputy Director | Statistics Department
In numerous central banks MICRODATA is becoming the nucleus of the statistical function.

But in recent years things have gone much far...
THE NEW DATA LANDSCAPE OF THE POST-CRISIS HAS PERMEATED CENTRAL BANKS IN MANY WAYS

- Microdata have become more granular
- Projects involving big data are more frequent
- The interest of central banks in big data is evolving from a mere interest area of research to an auxiliary input (or core input) for policy making and supervisory processes
- Very big structured datasets coexisting with very large unstructured data
- Integrated reporting schemes addressing multiple purposes within the central bank are already in place and will continue to be developed
- New sources supplementing (or partially replacing) established reports are becoming commonplace – commercial data, administrative data, social media datasets
New approaches to analyse and explore the new data landscape – data mining, visual analytics, machine learning, pattern recognition, etc.

New data architectures are being implemented to address the need to:

- **Combine data** from multiple datasets
- Promote **data-sharing** throughout the institution
- Allow the **regular** data exploration and also **experimental** activities with the data
- Increase the **usefulness** of the impressive wealth of data now available not only to the internal users but also, e.g., to the research community

THE NEW DATA LANDSCAPE OF THE POST-CRISIS HAS PERMEATED CENTRAL BANKS IN MANY WAYS
SOME OF THESE TENDENCIES CAN BE QUANTIFIED
Work in big data can now be considered a mainstream activity for central banks.

Central banks working on a project involving big data

- Yes: 56%
- No: 44%

Central bank’s view of big data

- An interesting area of research: 39%
- An auxiliary input into policymaking and supervisory processes: 37%
- A core input into policymaking and supervisory processes: 24%

External sources used to obtain big data

- Financial data: 71%
- Commercial entities: 61%
- Administrative data: 37%
- Social media datasets: 13%

Approaches used to analyse big data

- Data mining: 58%
- Visual analytics: 40%
- Trend forecasting: 30%
- Machine learning: 28%
- Pattern recognition: 28%
- Other: 15%

Source: Findings of the 2018 survey on “Big data and Central Banks” by Central Banking
TWO BIG CHALLENGES

These trends prompt for significant investments in IT infrastructures and software. However what we have experienced at Banco de Portugal is that the greatest challenges stand in two particular facets:

- **DATA GOVERNANCE**
- **STAFF SKILLS AND TRAINING**
In 2017 Banco de Portugal launched the INTEGRATED DATA MANAGEMENT (IDM) programme

The IDM is a major transformational initiative of the Strategic Plan 2017-2020.

The goal of the IDM is to strongly contribute to a better use of the available data in the Bank by means of rationalisation of the processes associated with its collection and processing and to promote its effective sharing throughout the whole organisation.

The IDM is **jointly coordinated** by 2 departments:

- STATISTICS
- IT
SUCCESS FACTORS FOR THE IMPLEMENTATION OF THE IDM

- Strong sponsorship from the Board
- Dedicated teams (both in Statistics and IT)
- All departments must recognize themselves in the programme and should be involved in the decisions
- The level of understanding of the programme by the various departments is not homogeneous. Adequate expectations management is vital
- Pursuing the global vision supported by iterative objectives
- Setting an enterprise-wide DATA GOVERNANCE

THE MAJOR CHANGE INDUCED BY THE IDM IS NOT TECHNOLOGICAL BUT CULTURAL/ORGANISATIONAL
1st challenge: **THE GOVERNANCE MODEL**
Centralized coordination with decentralized roles and responsibilities across the organisation

**DECISION MAKING [Board]**

**STRATEGIC COORDINATION [Information and Technology Management Committee]**

**OPERATIONAL COORDINATION**

**DEPARTMENTS**

- Data management [Statistics Department]
- Information technologies management [IT Department]
- Risk
- Data protection
- Audit

- Data providers
- Data users

- Data stewards
- Data owners
- System owners
- Data experts
- Data managers
- Data custodians (IT Dep.)
THE LOGICAL DATA ARCHITECTURE

DATA SOURCES

INTEGRATION & STORAGE

DATA LAKE

DATA WAREHOUSE

EXPLORATION

DATA SCIENCE

CORPORATE

SECURITY, ADMINISTRATION AND MONITORING

DATA CATALOGUE
ECONOMIC
FINANCE KNOWLEDGE

Business competencies are required.

These are key to perceive what is
the relevant information and
potential relationship between data
sources.

MATHEMATICS

With the increased volume and
data variety, mathematical
knowledge is needed to be able to
perceive correlations and other
knowledge that may be hidden in all
the data now available.

COMPUTER SCIENCE

Data still need cleaning,
transformation, aggregation, etc.

New skills are also required to
process unstructured data.

THESE PROFILES ARE IMPORTANT ON THEIR OWN, BUT ARE
MUCH MORE VALUABLE WHEN COMBINED
DESPITE NUMEROUS INITIATIVES (road shows in universities, open days, recruitment announcements, …) ATTRACTING DATA SCIENTISTS TO THE BANK PROVED TO BE AN UNSUCCESSFUL TASK

● To respond to this unsuccess Banco de Portugal decided to launch its own internal training programme – the SCHOOL OF DATA SCIENCE

● The Board approved the programme at end-2018 to start in 2020 covering 3 thematic areas

GOAL
To offer a structured and continuous training program aiming at providing the staff in business areas with a blend of skills – TECHNICAL, ANALYTICAL AND COMMUNICATION
TECHNOLOGIES and TOOLS

DATA EXPLORATION
- Excel (VBA and excel for BI)
- Tableau
- SAS
- Reporting Services
- R Notebooks
- Jupyter
(...)

ADVANCED ANALYTICS
- Power BI
- Tableau
- SAS
- R
- Python
- Stata
- Matlab
 (...)

DATA VISUALIZATION
- Excel Power View
- Power BI
- Tableau
- PowerPoint
- D3.js
 (...)

DATABASES
- MS SQL Server
- Hadoop ecosystem
- Column store databases
 (...)

DIFFERENT PROFICIENCY LEVELS AND AN INCENTIVE FOR EACH STAFF MEMBER TO SHAPE ITS OWN TRAINING JOURNEY

ADVANCED
Strong specialisation in a thematic area

INTERMEDIATE
Content required for a proficient level of professional users

BASIC
Introduction to the different thematic areas

GENERAL
Relevant transversal concepts linked with the data organization in the Bank
LEVERAGE KNOWLEDGE WHEREVER IT EXISTS …

- Strategic partnership with a University to assist managing the programme
- Start with a MVP (minimum viable product) for 100 staff members and a subset of 12 courses (including on-line courses) from the thematic areas and from technologies/tools
CONCLUDING REMARKS

Extracting the most from the new post-crisis data landscape requires **organisational transformations**. Data governance [DG] and adequate staff skills were the major challenges that we have identified at Banco de Portugal.

Besides clear decision making, a successful DG strategy implies a **solid partnership between business and technology**.

DG policies will lose their value if they’re not followed in day-to-day operations across the organisation. **Data stewardship** (in all business areas) is now looked as a critical function for the success of the DG.

In a fast-changing technological landscape, **PURPLE PEOPLE**, who can combine data savviness with domain-specific business knowledge, are highly demanded.

In Statistics (but also in other business areas) the challenge is to move from a traditional “economist-statistician” profile to a “economist-data scientist” profile, thus valuing a virtuous blend of economic/business training and advanced technological expertise.