

IFC Satellite Seminar on "Post-crisis data landscape: micro data for the macro world", co-organised with the Central Bank of Malaysia and the European Central Bank

16 August 2019, Kuala Lumpur, Malaysia

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.



BANCO DE
PORTUGAL
EUROSYSTEM

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

59	11	85	1	68	58	72	43	95	45	27	57	14	24	89	84	48	71	5	94	18	75	89	38	75	21	43	42														
94	42	36	33	69	21	30	33	58	45	72	47	66	36	9	93	77	51	45	47	32	62	84	16	48	64	10	63	47	32	62	84	16	48	64	10	63					
51	17	6	92	38	86	0,8	9	16	89	38	60	1	71	73	80				19	77	51	83	87	59	97	31	0,6	19	77	51	83	87	59	97							
1,0	89	93	74	3	97	43	55	40	54	79	82	27	41	0							49	65	41	38	32	3	2	56	63	49	65	41	38	32							
45	14	62	82	96	87	25	70	0,6	0,8	0,3	0,8	0,1	0									0,8	0,0	0,7	0,6	76	55	60	0,4	0,9	0,8	0,0	0,7	0,6							
85	88	31	0,0	82	42	3	32	93	36	25	2	55												66	4	27	80	16	88	39	64	2	66	4	27						
33	90	49	38	68	69	10	44	52	71	75	83													28	38	76	3	64	76	9	36	31	88	38	76						
71	85	62	0	88	79	45	59	19	92	94	8														14	27	36	41	47	35	16	50	11	14	27						
71	9	25	20	41	90	27	1,0	85	22	50															12	49	60	74	0,0	62	55	27	77	12	49						
19	58	70	31	43	13	12	36	63	84	9																49	55	84	30	36	21	94	5	83	49						
50	83	94	0,6	21	43	27	98	97	15																	46	77	96	52	99	46	30	60	36	46						
71	9	59	32	75	13	71	39	63	95																76	34	38	87	24	12	35	66	95	76							
25	0,6	57	33	18	53	91	34	20	87																1	56	58	11	64	64	5	91	58	44							
79	43	26	74	98	0,6	9	0,3	60	60																	0,4	63	0,7	85	3	95	14	2	35							
6	73	42	77	47	5	0,8	26	11	7																	33	0,3	24	28	65	8	25	66	99							
76	41	85	55	32	91	86	6	70	3																	15	36	52	42	8	20	0	33	15							
80	100	87	13	14	90	57	28	54																		25	0	46	6	32	39	2	25	10							
23	64	20	74	0,4	37	93	95	44																		44	22	21	7	12	95	45	29	58							
59	11	85	1	68	58	72	43	95																		94	18	75	89	38	75	21	43	42							
94	42	36	33	69	21	30	33	58																		47	32	62	84	16	48	64	10	63							
51	17	6	92	38	86	0,8	9	16																		31	0,6	19	77	51	83	87	59	97							
1,0	89	93	74	3	97	43	55	40	5																	3	2	56	63	49	65	41	38	32							
45	14	62	82	96	87	25	70	0,6	0,8	0,3	0,8	0,1	0												8	76	55	60	0,4	0,9	0,8	0,0	0,7	0,6							
85	88	31	0,0	82	42	3	32	93	36																	54	80	16	88	39	64	2	66	4	27						
33	90	49	38	68	69	10	44	52	71																	60	3	64	76	9	36	31	88	38	76						
71	85	62	0	88	79	45	59	19	92	9																14	26	36	41	47	35	16	50	11	14	27					
71	9	25	20	41	90	27	1,0	85	22	50																12	36	60	74	0,0	62	55	27	77	12	49					
19	58	70	31	43	13	12	36	63	84	9																	5	83	64	55	84	30	36	21	94	5	83	49			
50	83	94	0,6	21	43	27	98	97	15	43	62															60	36	92	77	96	52	99	46	30	60	36	46				
71	9	59	32	75	13	71	39	63	95	61	45	90														35	66	95	90	34	38	87	24	12	35	66	95	76			
25	0,6	57	33	18	53	91	34	20	87	93	68	57	42													64	5	91	58	74	56	58	11	64	64	5	91	58	44		
79	43	26	74	98	0,6	9	0,3	60	60	86	57	38	11	0													85	3	95	14	2	48	0,4	63	0,7	85	3	95	14	2	35
6	73	42	77	47	5	0,8	26	11	76	33	24	57	94	74	74	4	5	5	5	0,3	24	28	65	8	25	66	72	33	0,3	24	28	65	8	25	66	99					
76	41	85	55	32	91	86	6	70	30	89	74	8	5	67	49	64	44	20	15	36	52	42	8	20	0	33	20	15	36	52	42	8	20	0	33	15					
80	100	87	13	14	90	57	28	54	29	12	96	69	49	11	44	86	31	65	25	0	46	6	32	39	2	25	65	25	0	46	6	32	39	2	25	10					

**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

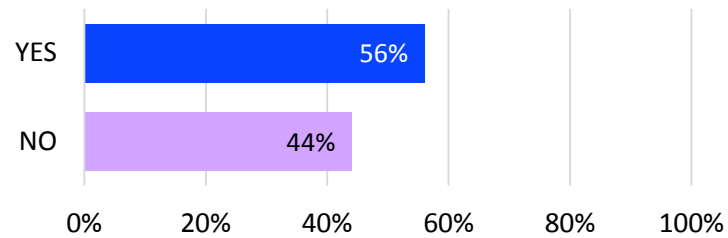
**THE NEW DATA
LANDSCAPE OF THE
POST-CRISIS HAS
PERMEATED
CENTRAL BANKS IN
MANY WAYS**

**SOME OF THESE
TENDENCIES CAN BE
QUANTIFIED**

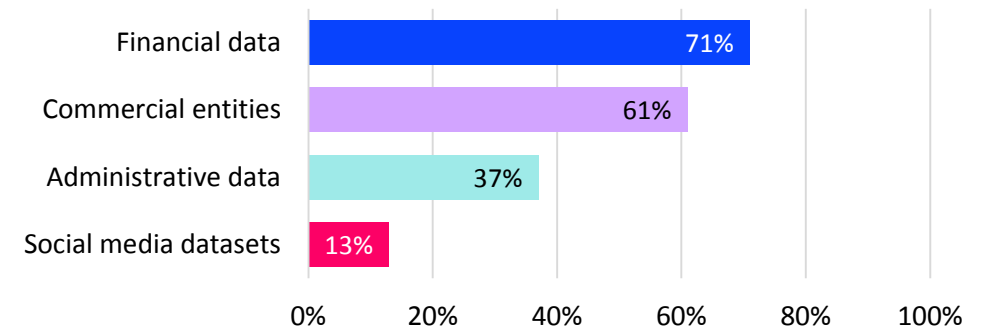


“ Work in big data can now be considered a mainstream activity for central banks ”

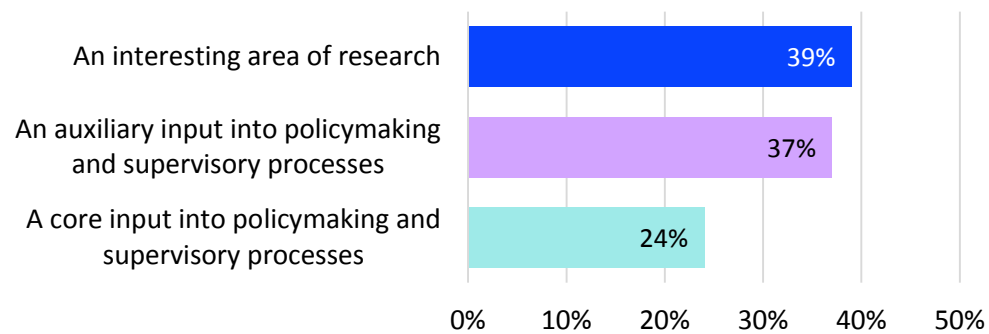
Central banks working on a project involving big data



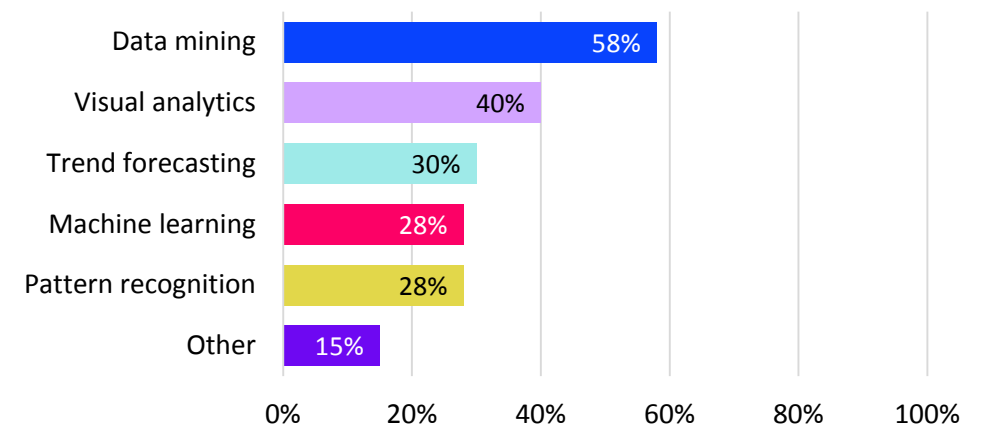
External sources used to obtain big data



Central bank's view of big data

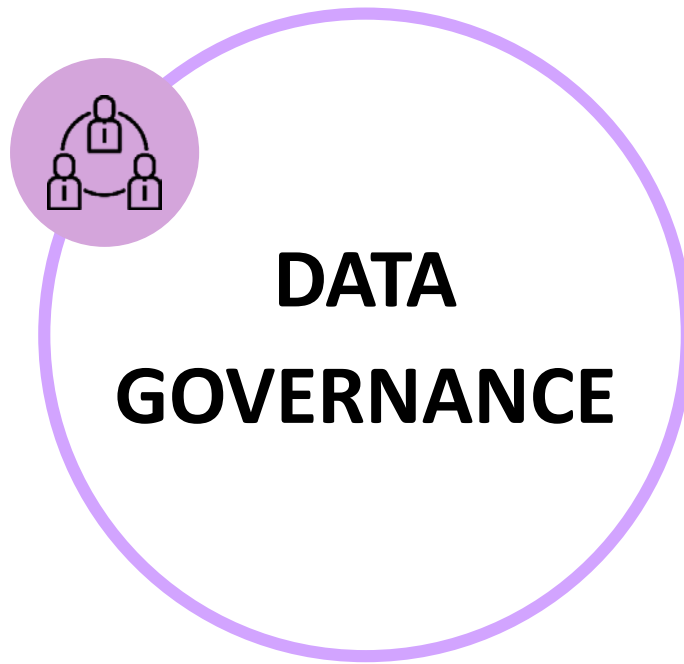


Approaches used to analyse big data



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**:

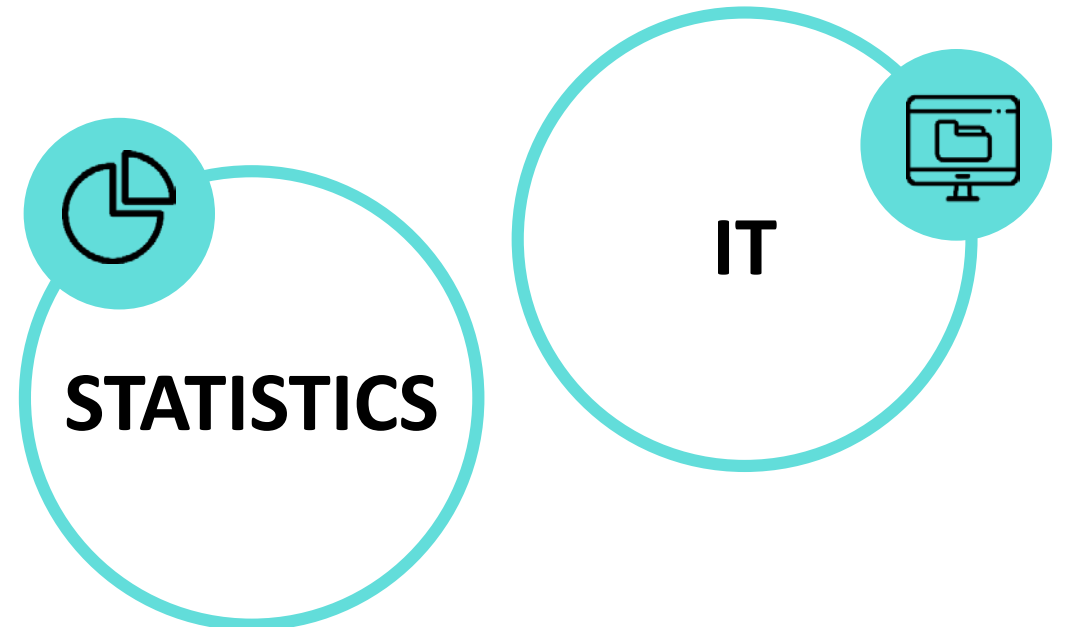


In 2017 Banco de Portugal launched the INTEGRATED DATA MANAGEMENT (IDM) programme

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:



**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**

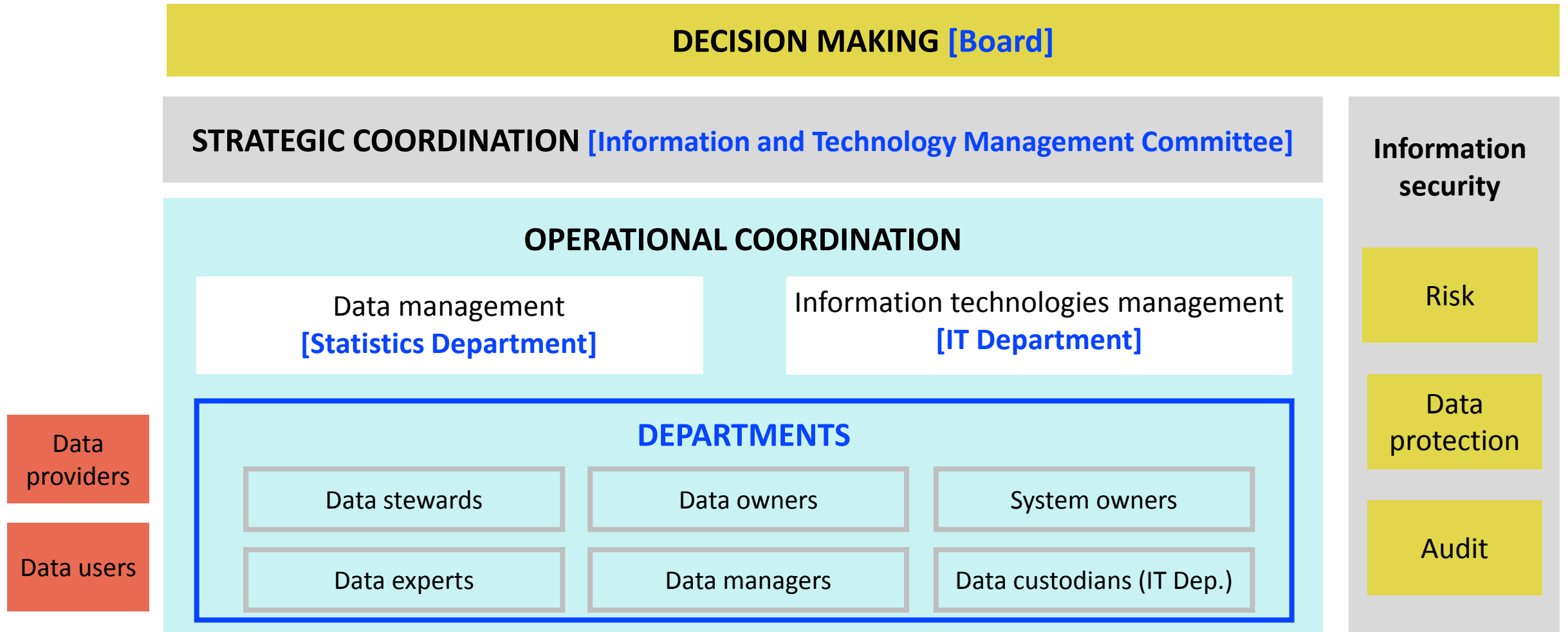
Setting an enterprise-wide **DATA GOVERNANCE**

THE MAJOR CHANGE INDUCED BY THE IDM IS NOT TECHNOLOGICAL BUT CULTURAL/ORGANISATIONAL

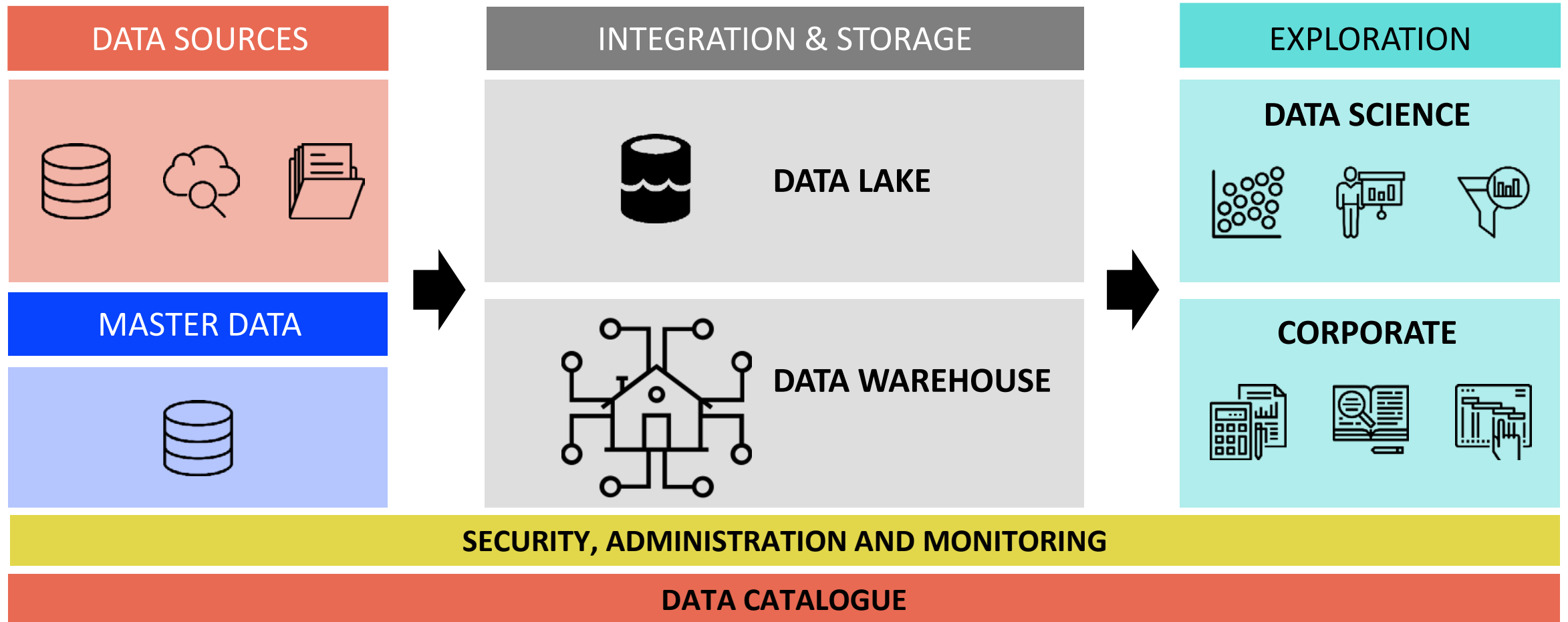


1st challenge: **THE GOVERNANCE MODEL**

Centralized coordination with decentralized roles and responsibilities across the organisation



THE LOGICAL DATA ARCHITECTURE



2nd challenge: **STAFF SKILLS**

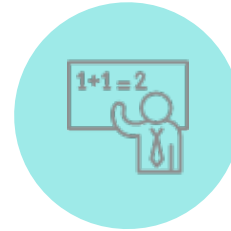


ECONOMIC

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**

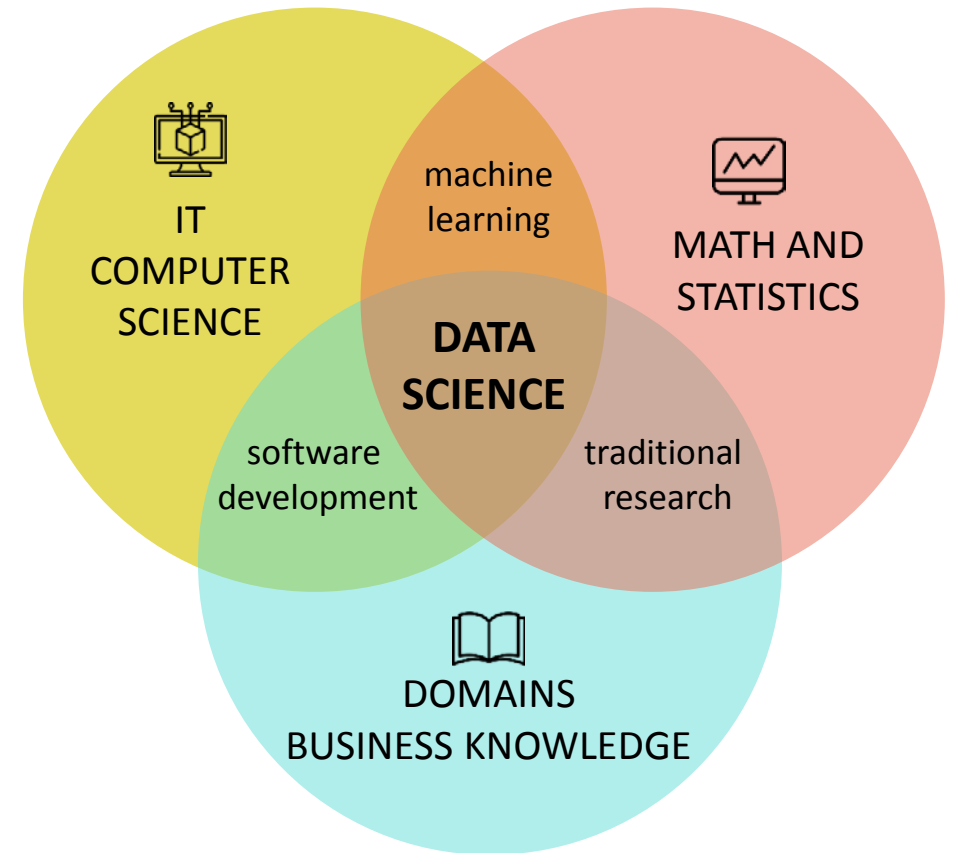


DESPITE NUMEROUS INITIATIVES (road shows in universities, open days, recruitment announcements, ...) ATTRACTING DATA SCIENTISTS TO THE BANK PROVED TO BE AN UNSUCCESSFUL TASK

- 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





MATH AND STATISTICS

Data analysis

- Data for decision-making
- Calculations and problems
- Statistics for decision-making
- Data exploration
- Complex analyses

Advanced analytics

- Statistics inference
- Regression analyses
- Predictive analyses and machine learning
- Neural networks and deep learning
- Data mining
- Big data analytics

Data visualisation

- On-screen data models
- Dashboarding
- Storytelling
- Infographics

(...)



DOMAINS BUSINESS KNOWLEDGE

Information management

- Data catalogue
- Master and reference data
- The data warehouse
- Data governance

(...)

THEMATIC AREAS



COMPUTER SCIENCE / IT

Databases

- Databases concepts
- Introduction to SQL
- Data extraction, transformation and loading
- Data warehouses development

Big data

- Modelling massive amounts of data
- Big data tools

Methodologies

- Agile
- Rapid application development (RAD)
- Business event analysis and modeling (BEAM)
- Parallel Computation
- Distributed Computing

(...)

TECHNOLOGIES and TOOLS

DATA EXPLORATION

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

ADVANCED ANALYTICS

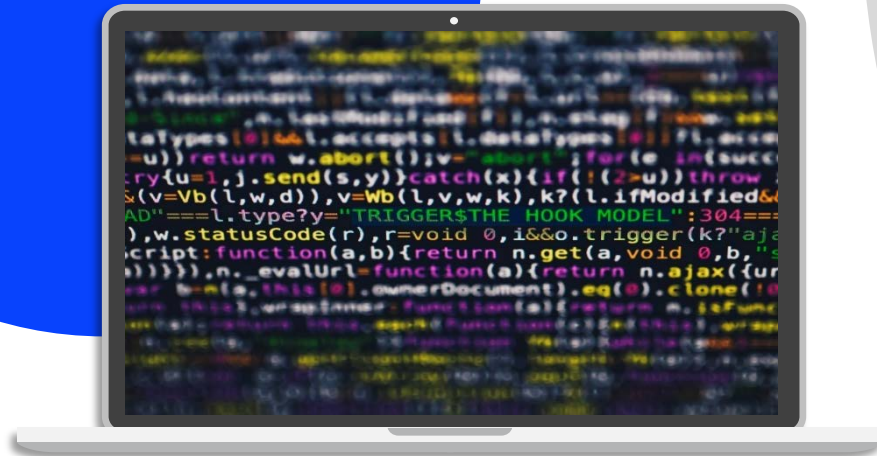
- Power BI
- Tableau
- SAS
- R
- Python
- Stata
- Matlab
- (...)

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

GENERIC

Relevant transversal concepts linked with the data organization in the Bank

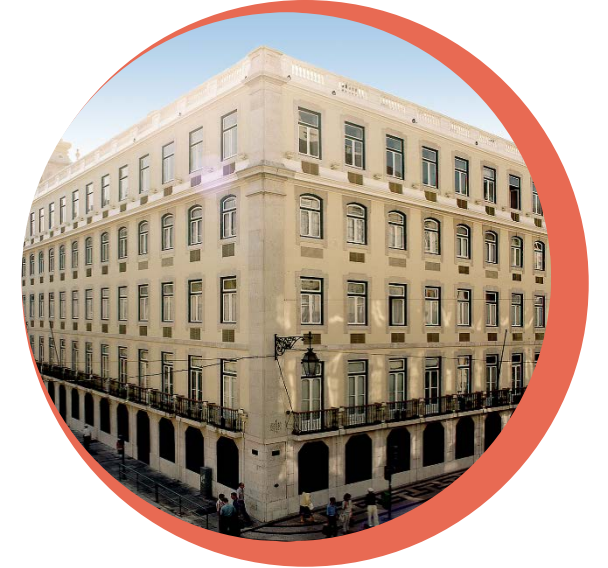
LEVERAGE KNOWLEDGE WHEREVER IT EXISTS ...



UNIVERSITIES



TRAINING COMPANIES



BANCO DE PORTUGAL STAFF

- 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

