Santander

Data citizenship: The next stage of Data Governance

Lidia Crespo
Chief Data Steward
Santander UK
Santander group

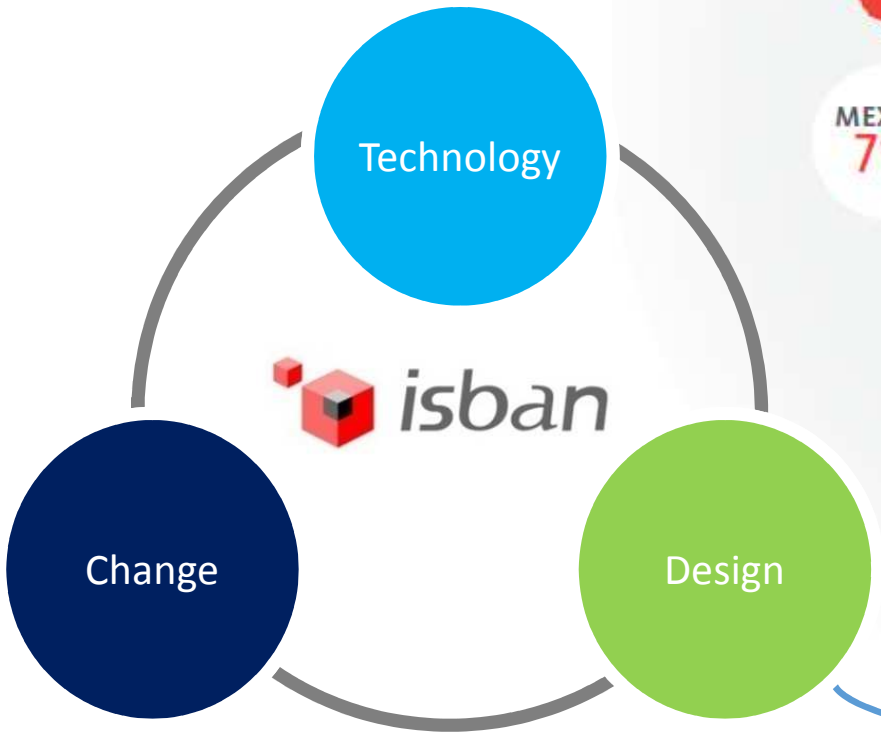
Antonio Alvarez
Head of Data innovation
Isban UK
Santander group

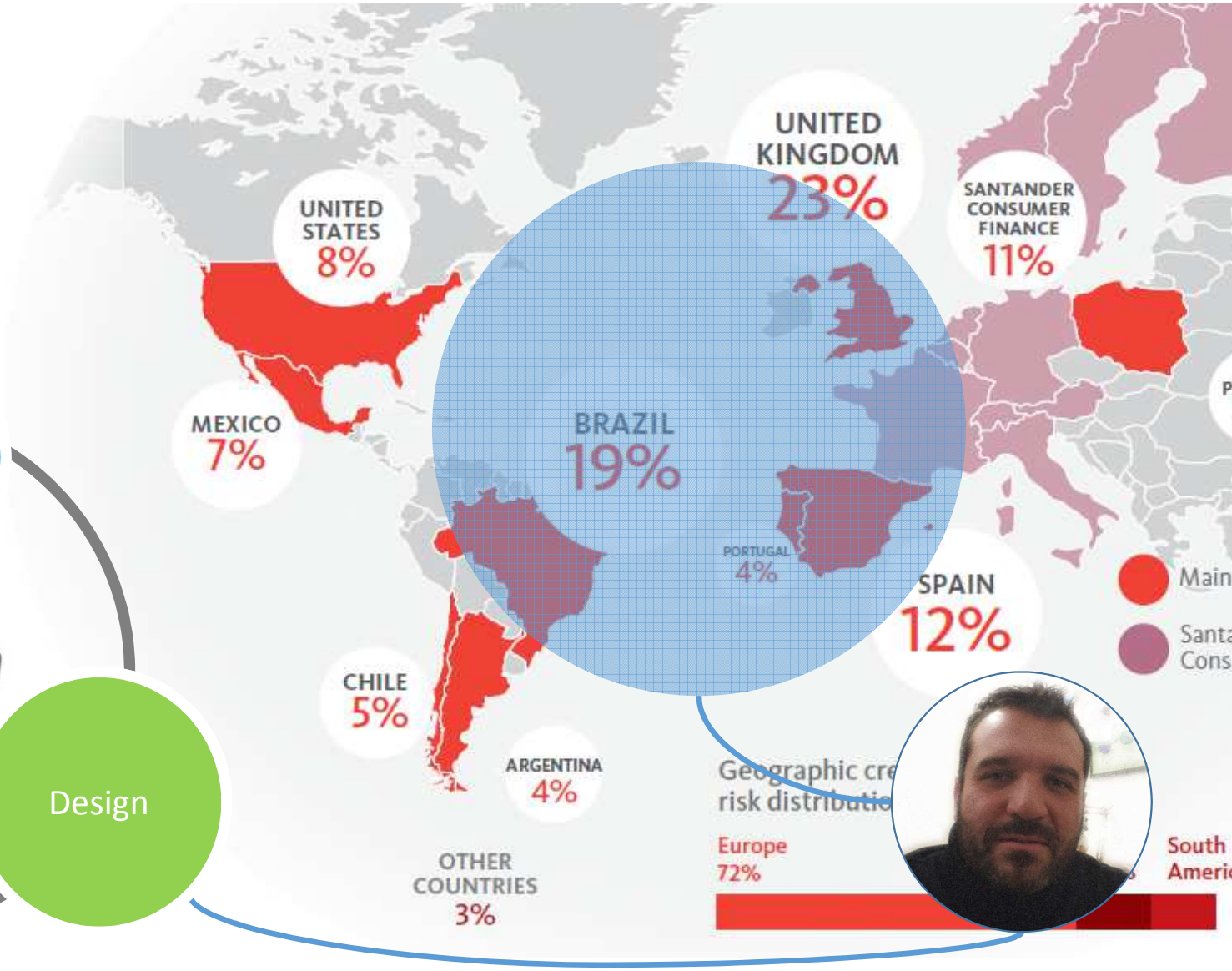
A. Introduction
B. Data for innovation
C. Governance and security
D. Santander UK Big Data Journey
E. Issues and lessons
(Data citizenship)
F. Yes, but how?

Data for innovation and value needs governance. The implicit tension can only be balanced through distributed responsibility.

Santander

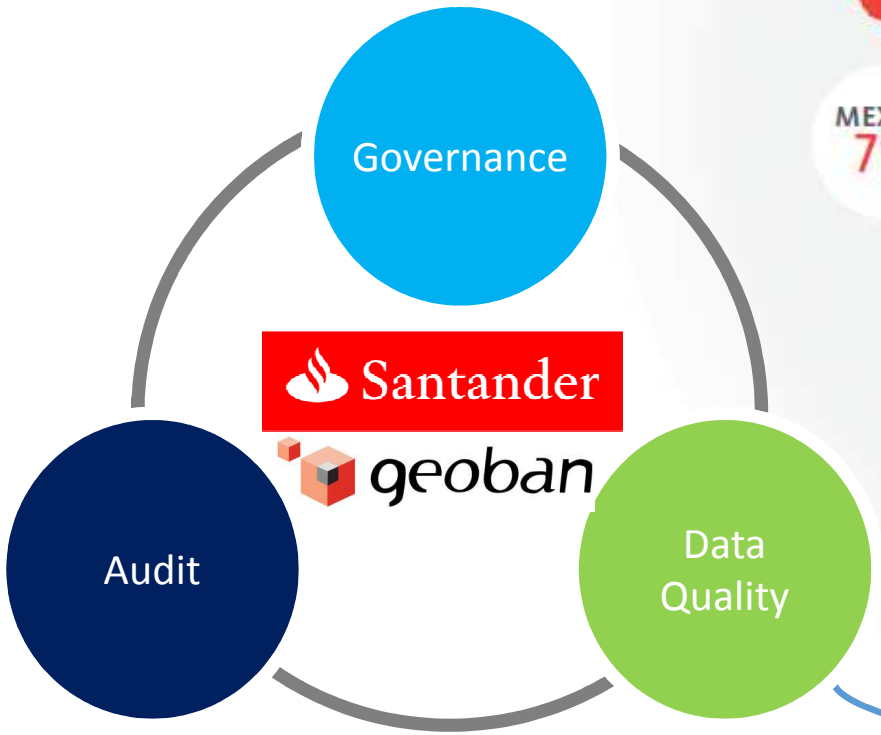
Antonio Alvarez
Head of Data innovation
Isban UK | Santander UK

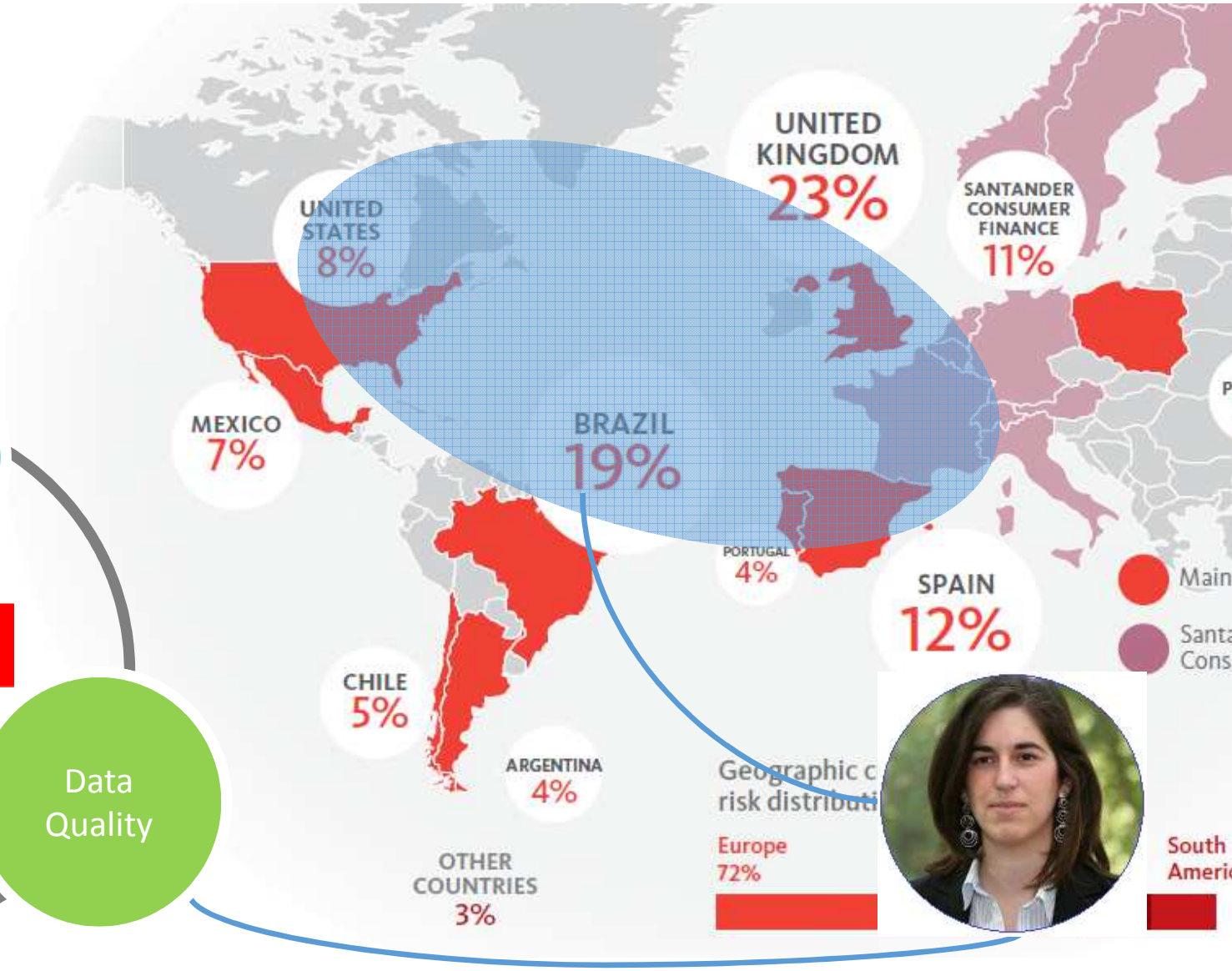




Santander

Lidia Crespo
Chief Data Steward
CDO area | Santander UK





Santander UK is a scale challenger

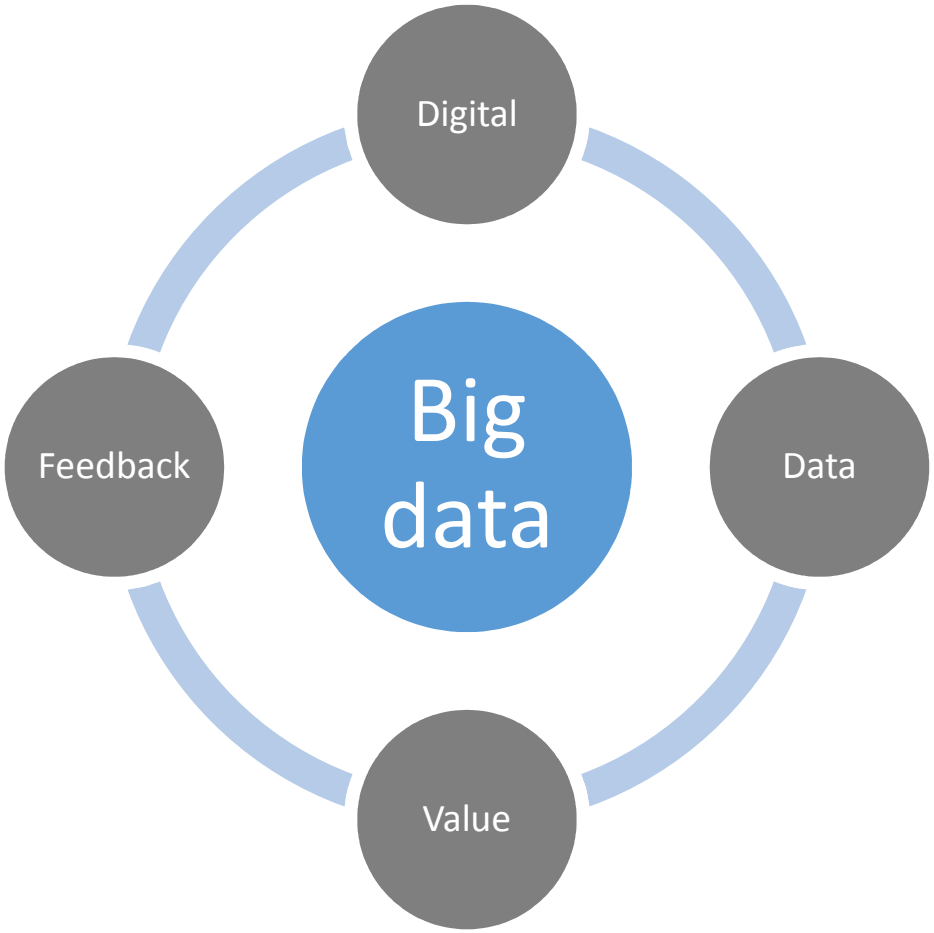
150 years of banking in our DNA. Culture of retail banking. Moderate risk. A very robust and traditional bank… with a culture of technology.

5th biggest bank in the UK. Growth by successful mergers. From 2004 to 2013, Santander bought Abbey, Bradford&Bingley, Alliance&Leicester and other smaller portfolios.

Now transforming into a Data driven organisation.


Data for
Innovation
B

{Sustainable value from data}



{Create a feedback loop}
{Make the feedback loop as quick as possible}
{Use the feedback to get to the why}
{In a digital world feedback is always in the form of data}
{Big data technology to understand customers}


“
If I had asked my customers what they wanted they would have said a faster horse
”
Henry Ford

“It is not the strongest of the species that survives, nor the most intelligent that survives. It is the one that is most adaptable to change.”

Leon C. Megginson, attributed to Charles Darwin

C

Governance and Security

Data chaos

From **Divergence**

- Decentralised data model
- **Decentralised user model**
- **Decentralised tool management**
- Distance from the original data
- Ineffectively governed model
- Increasingly unsecured model
- Silos
- Conflict for resources

To **The Big Data Platform**

- Centralised data model
- **Distributed user model**
- **Centralised tool management**
- Closeness to the original data
- Effectively governed model
- Increasingly secured model
- Multi-tenancy allows for a single platform
- Shared performance

Governance, security, privacy

From multiple inconsistent methods to one policy, one framework, one platform

From application based to domain driven data architecture

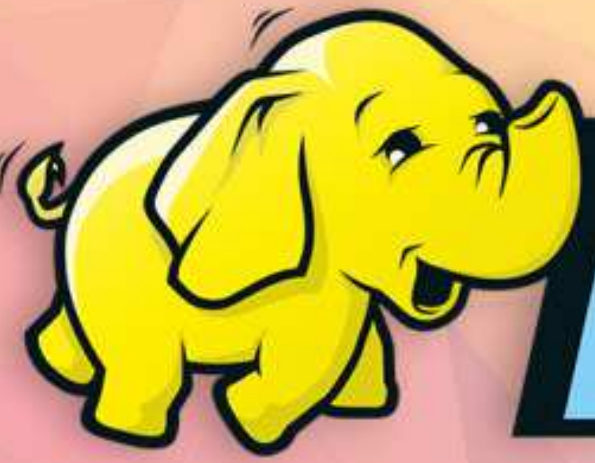
From applications to certified data products

From disperse knowledge to a searchable connected Knowledge hub

The next frontier: privacy by design

Santander UK Big data journey

Proof of Concept
T+3months



1

- 3 months
- Test, find potential, identify limits
- Plan the future

Analytics
T+5months

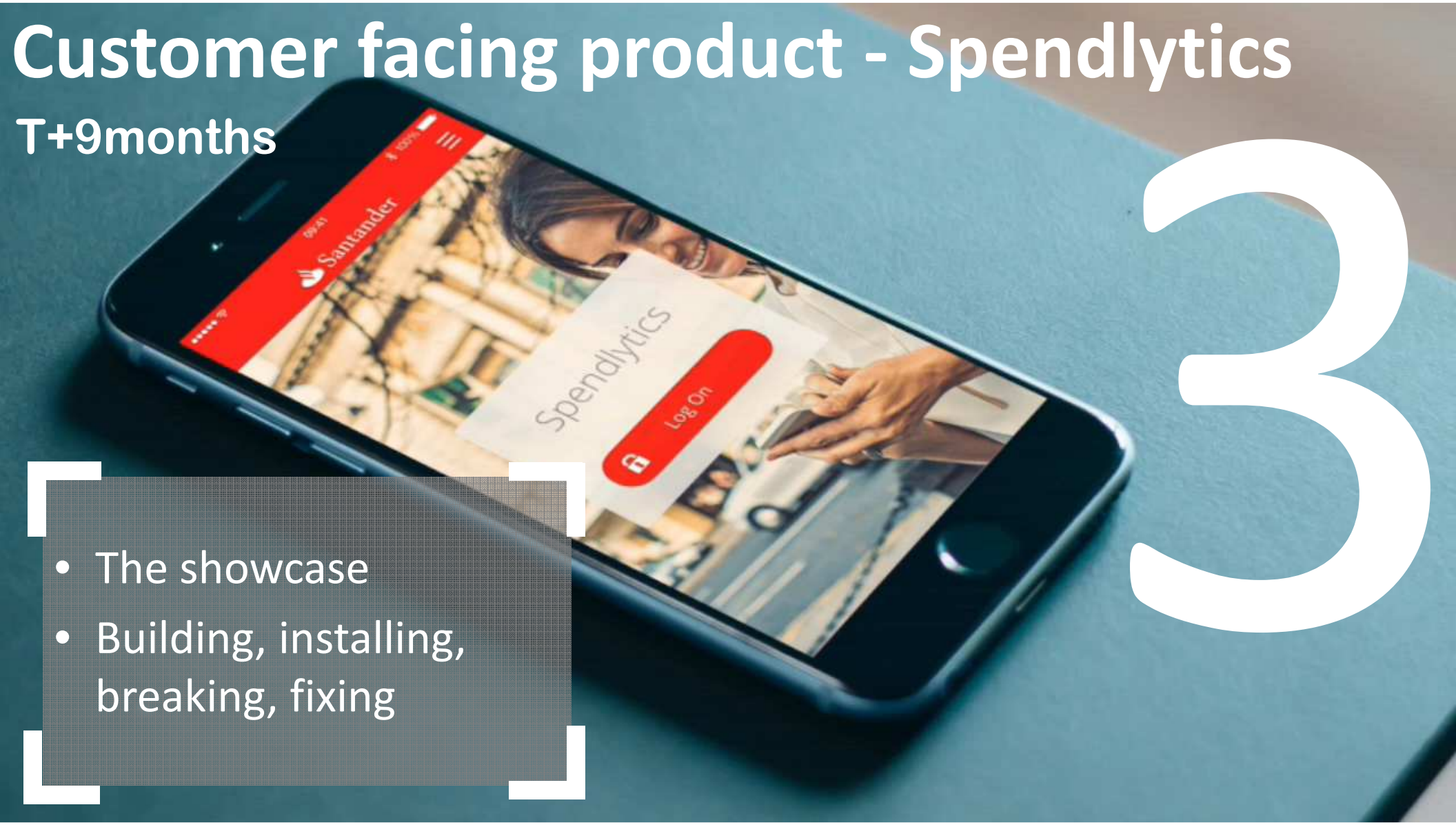
- Get a tool immediately
- Data, business and technology in collaboration

2

Customer facing product - Spendlytics

T+9months

3



- The showcase
- Building, installing, breaking, fixing

Broadening

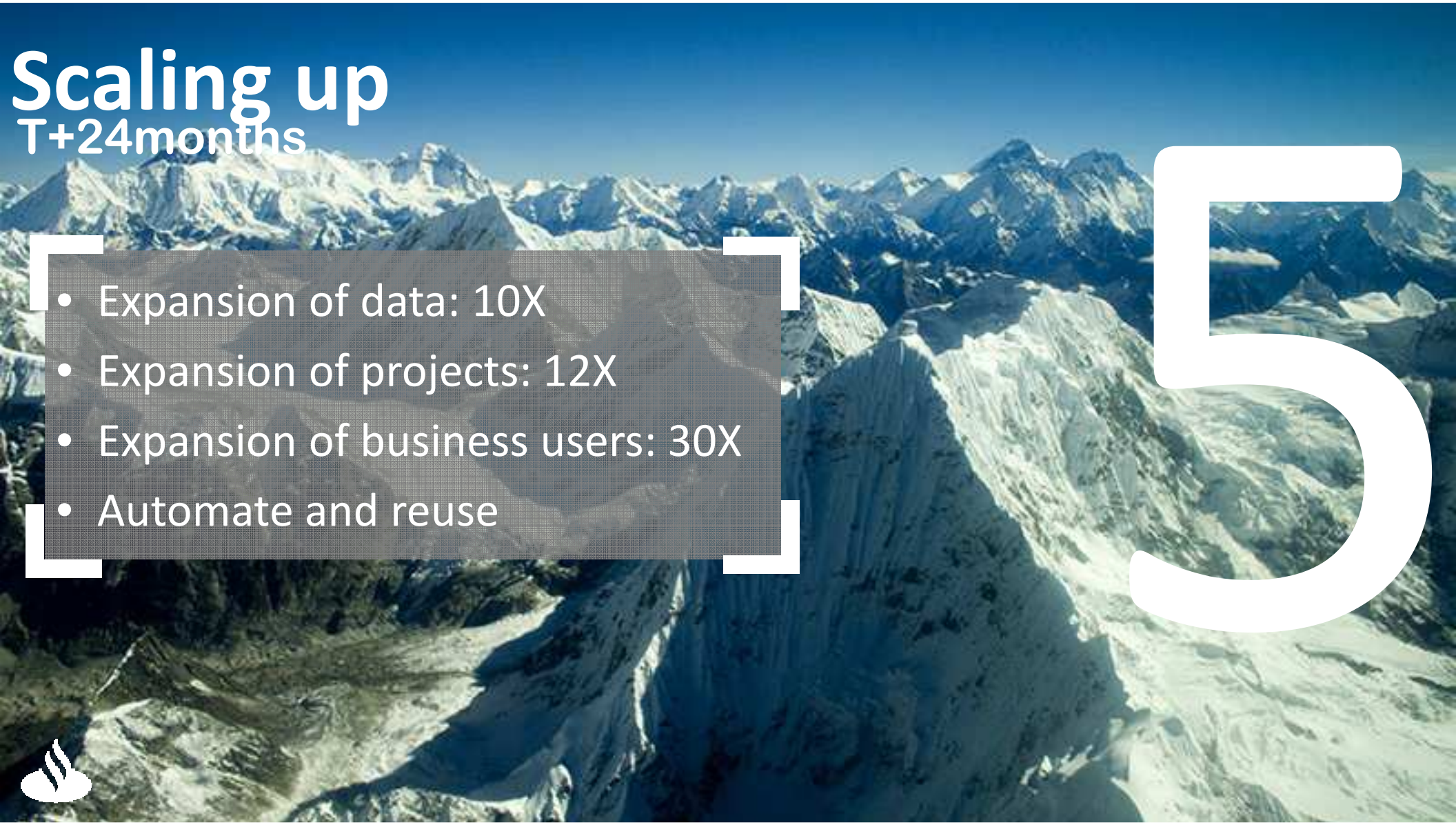
T+12months

- Early adopters
- Feedback and improve
- Data operating model

4

Scaling up

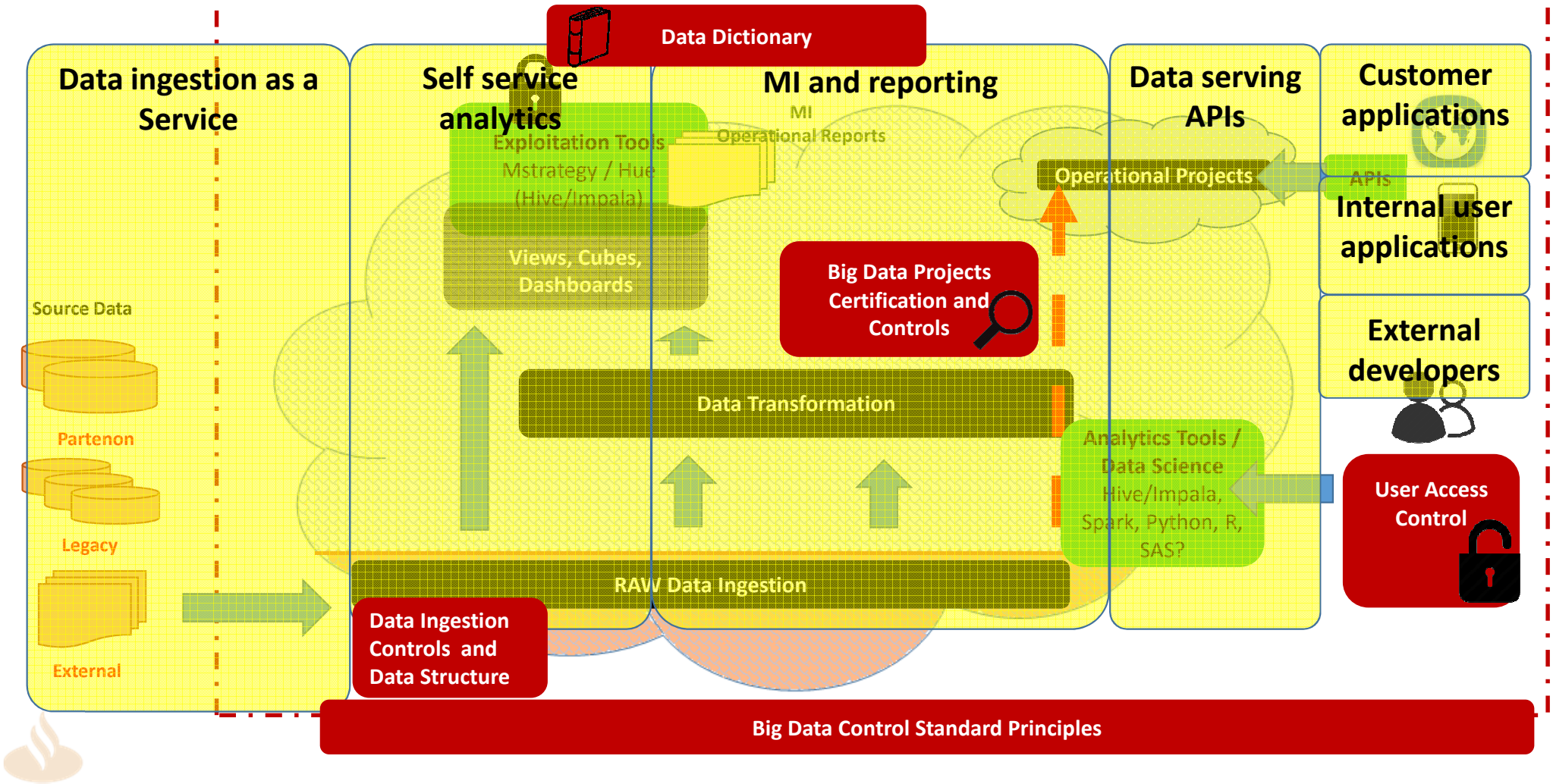
T+24months



- Expansion of data: 10X
- Expansion of projects: 12X
- Expansion of business users: 30X
- Automate and reuse

5

{Santander UK Big data framework}



Issues and lessons

E


“
Life is like riding a bicycle.
To keep your balance, you must keep moving.
”
Albert Einstein

Bumps in the road

Governance as a stopper for Innovation

No clear solution to Scale up the Governance

Agile projects and Self-service with less clear goals and checkpoints

Resistance to integrate Silos difficults effective data sharing

Applying privacy regulations to Self-service

Maintain the water fresh

Speed of Innovation and Self-service bypasses governance

Users eager to gain more freedom but not so eager to accept more responsibility. Self-service, Self-Responsible

Cultural transformation slower than technology transformation

Data duplication and inconsistencies

{Lessons from the wounds of the journey}

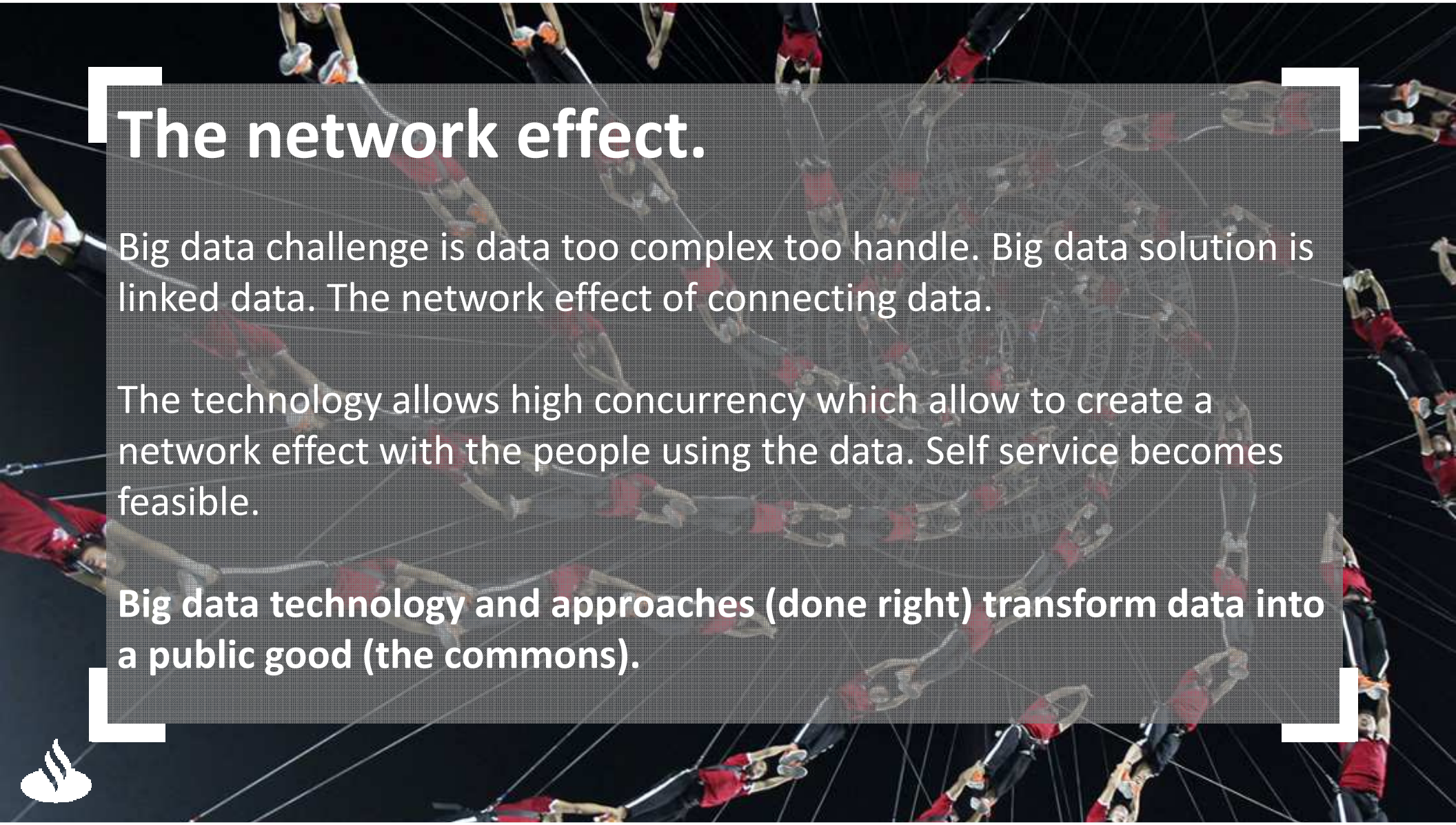
Sinthesis

DEMOCRACY AND CITIZENSHIP

Data ownership

Ownership *is the act, state, or right of possessing something.*

Data ownership is the act of having legal rights and complete control over a single piece or set of **data** elements. It defines and provides information about the rightful **owner** of **data** assets and the acquisition, use and distribution policy implemented by the **data owner**.



The network effect.

Big data challenge is data too complex too handle. Big data solution is linked data. The network effect of connecting data.

The technology allows high concurrency which allow to create a network effect with the people using the data. Self service becomes feasible.

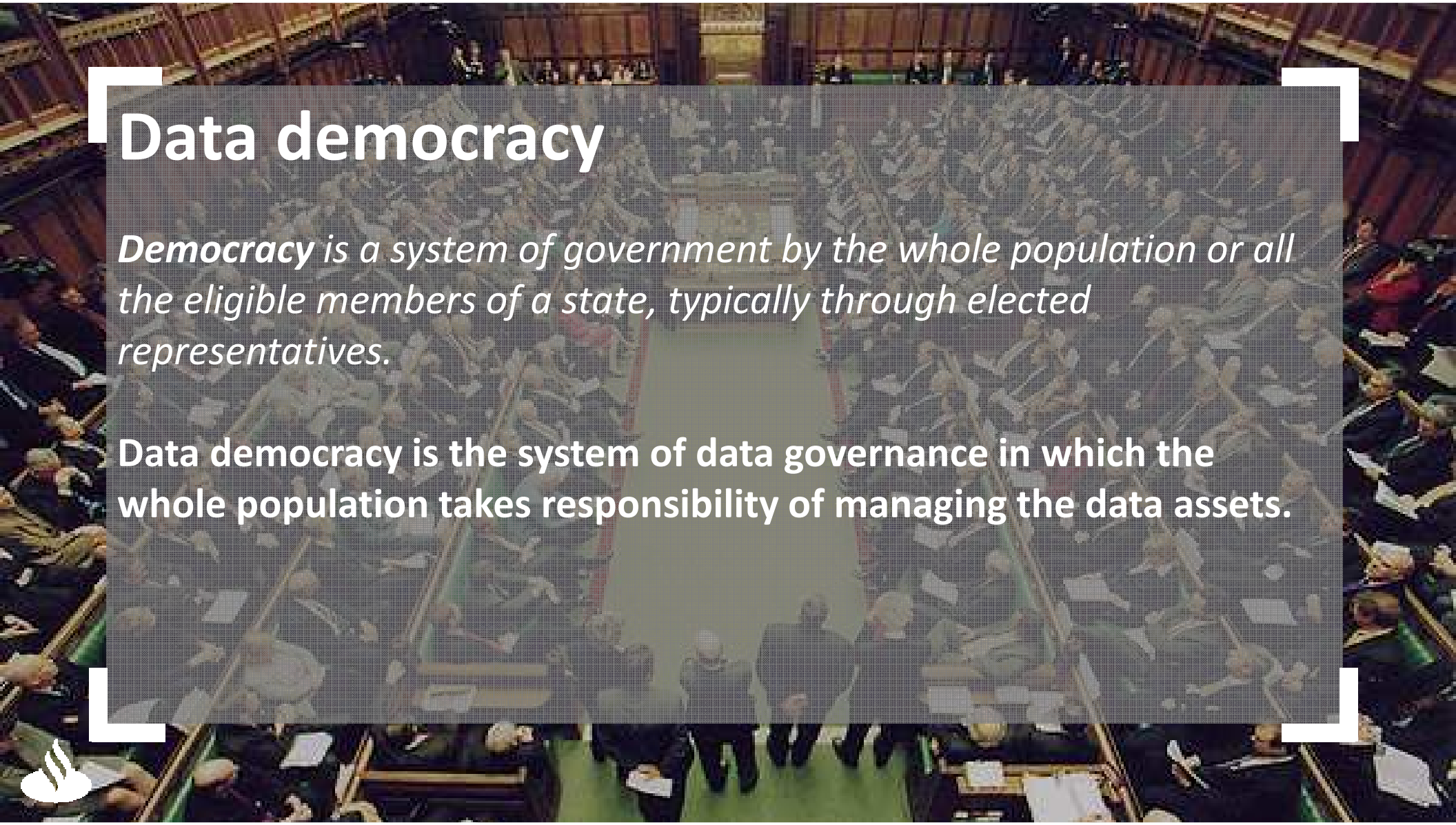
Big data technology and approaches (done right) transform data into a public good (the commons).

Data is a public good

*The **commons** is the cultural and natural resources accessible to all members of a society, including natural materials such as air, water, and a habitable earth.*

Data in a big data platform becomes accessible to all users and the consumption of one users does not impact the consumption of others.

This requires responsibility from everyone.



Data democracy

Democracy *is a system of government by the whole population or all the eligible members of a state, typically through elected representatives.*

Data democracy is the system of data governance in which the whole population takes responsibility of managing the data assets.

Separation of powers

A fundamental principle of the democratic government, whereby powers and responsibilities are divided among the three different branches of Government (legislative, executive and judicial).

In Data democracy, within a framework (Constitution) and a structure of data ownership, policy, enforcement and conflict resolution are delivered by different areas.

Data citizenship

Democracy is at its best when citizens become involved and active in the government.

In data citizenship, users' rights (the right to use the data, the right to value, comment and annotate the data) are also responsibilities that are rewarded. The properties of the data (quality, timeliness, provenance…) are defined by the community of users.

Transparency is the principle for sharing data, insights and software.

Warning: Avoid the tragedy of the commons

The tragedy of the commons is an economic theory of a situation within a shared-resource system where individual users acting independently according to their own self-interest behave contrary to the common good of all users by depleting or spoiling that resource through their collective action.

Put value to your data. Reward those that add value to the community.
Document, organise and architecturise your data and insights. Make them searchable and re-usable.

Change of culture

Shared ownership and responsibility requires a significant shift in the mind set of the organisation.

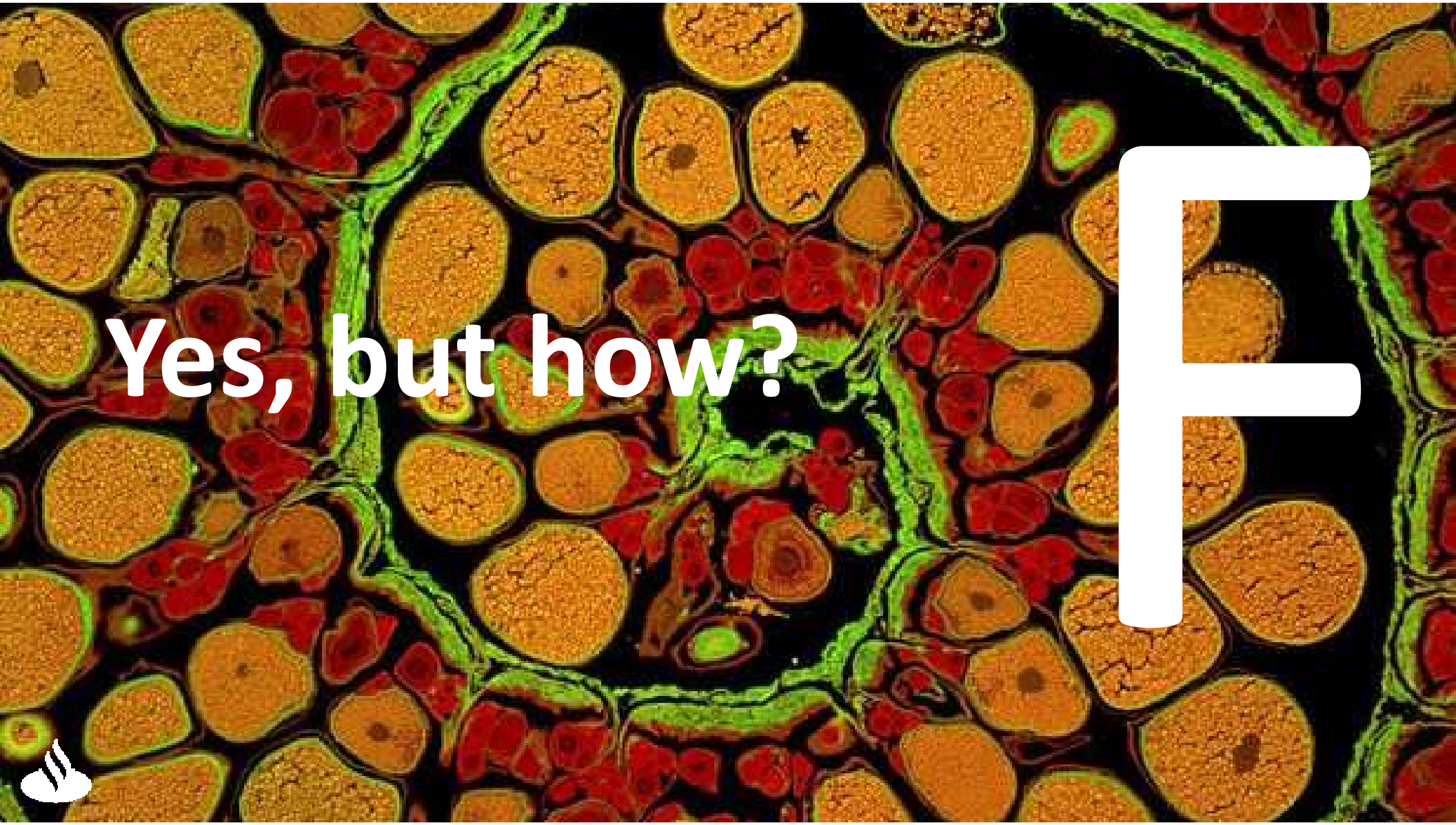
Servant leadership: focus on the needs of others, give support, acknowledge other people's perspectives, involve others in decisions and build a sense of community.
This leads to higher engagement, more trust, stronger relationships and to increased **innovation**.

Democratisation

T+30months

- Guided self service
- Data is a public good
- Love your data
- Care for your data

6


Yes, but how?

{Data democracy in Santander UK}

Constitution

Data Strategy and Data policy
Written by CDO on behalf of the whole organisation with input from all areas

Legislative

Community of Centres of Excellence
Represent the main forces in the generation, management and usage of data.
CoEs are like political parties that represent the hundreds of data users.

Executive

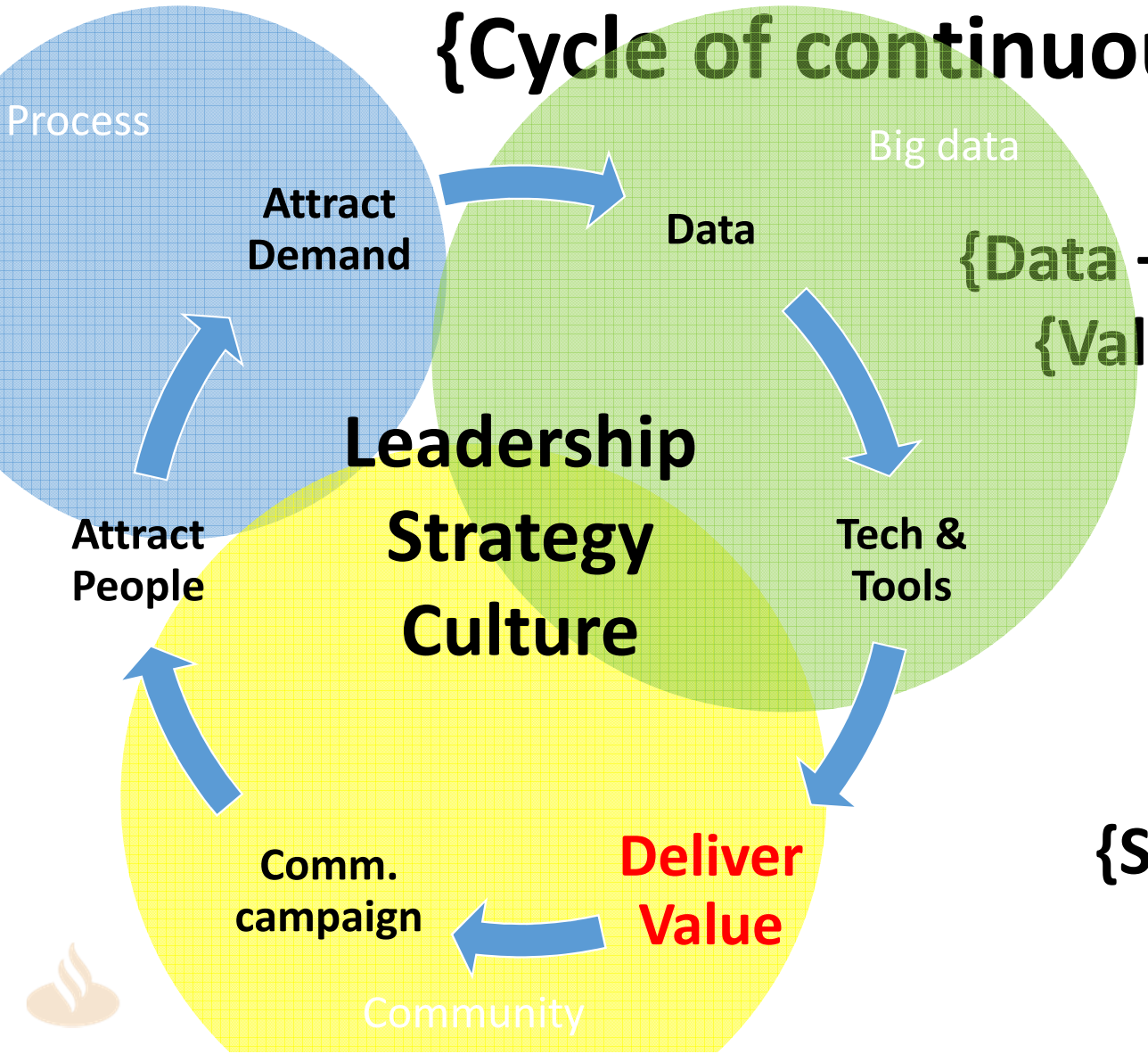
Operations
- Geoban for quality, certification, reconciliation
- Produban (IT Ops) for access, monitoring and support

Judicial

Chief Data Officer
Applies policy and referees when there is conflict.

But we are still a monarchy.
If the king, or the royalty, want to override the CDO, they still do it in many cases.

{Cycle of continuous value creation}



{Data + Tech + Talent = Value}

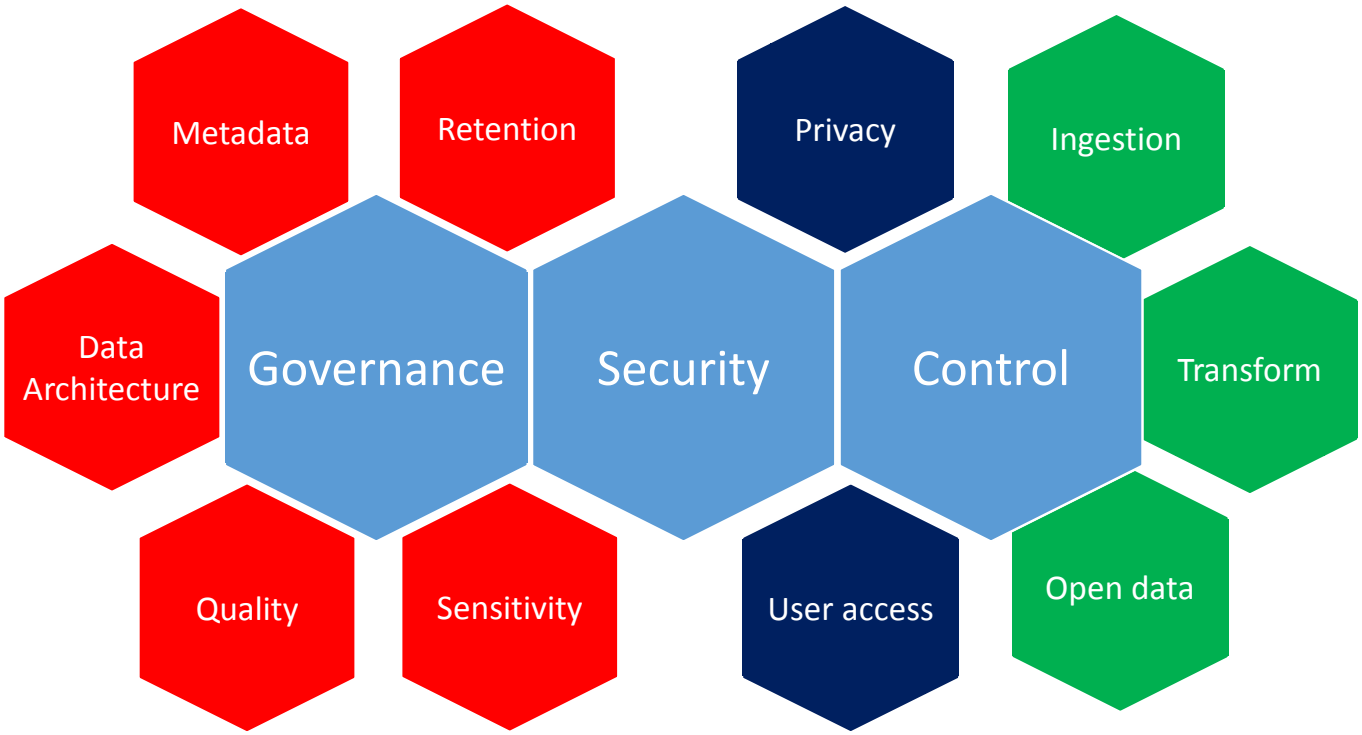
{Value + Communication = Gravity}

{Gravity + Leadership = Culture}

{Gravity + Processes = Community}

{Strategy + Community = Success}

{Don’t compromise on Governance and Security}



{I AM Risk}

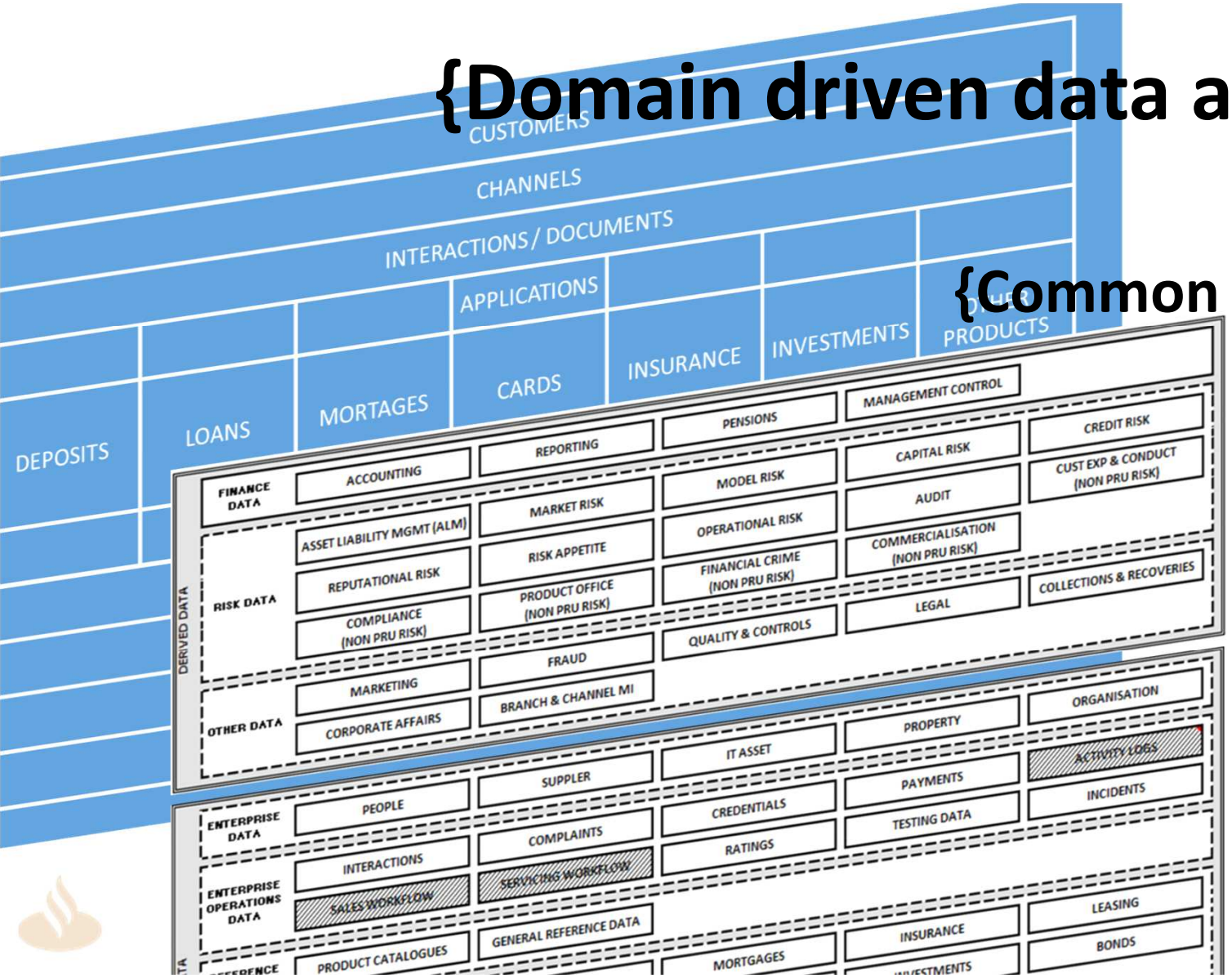
{Identify}

{Assess}

{Manage}

{Report}

{Domain driven data architecture}



{Common understanding}

{Shared}

{Comparable}

{Build trust in your data and platform}

{Automated}
{Centralised}
{Crowdsourced}

automatic process

BAU Control

users rankings

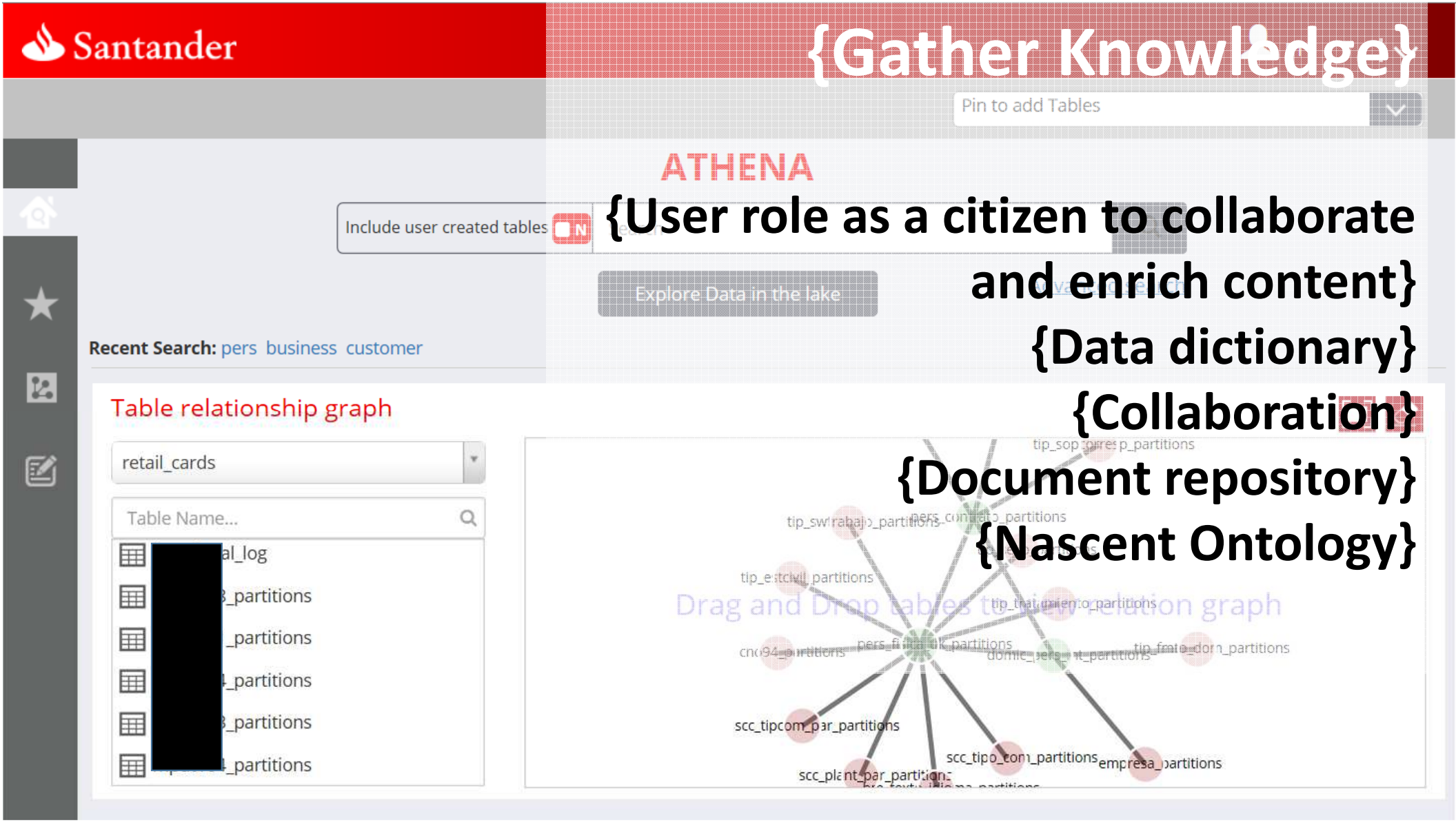
formal evaluation

STANDARDISATION		FORMAL CERTIFICATION	EVALUATION
VALIDATION	RECONCILIATION	INGESTION CERTIFICATION	COLLABORATIVE EVALUATION
FORMATTING	ENRICHMENT	TRANSFORMATION CERTIFICATION	FORMAL EVALUATION

Automatic process executed just after every data ingestion to validate the data being included in the repository, to homogenise the data formats and structures in the platform and to enrich the records with extra fields valuable for later analysis.

Manual procedure lead by the CDO and executed by several teams to certify that the data ingestion and transformation processes are properly implemented, through the execution of different testing scenarios with both valid and invalid datasets.

Manual evaluations made by users through subjective rankings and comments about the data quality, and quality distinctions given by the data quality authority to recognize some extraordinary value of certain datasets.


Santander
{Gather Knowledge}
Pin to add Tables
ATHENA
Include user created tables
Explore Data in the lake
Recent Search: pers business customer
Table relationship graph
retail_cards
Table Name...
_partitions
_partitions
_partitions
_partitions
_partitions
Drag and Drop tables to view relation graph
tip_sop_corresp_partitions
tip_estcivil_partitions
scc_tipcom_par_partitions
scc_tipo_com_partitions
empresa_partitions
{User role as a citizen to collaborate and enrich content}
{Data dictionary}
{Collaboration}
{Document repository}
{Nascent Ontology}

{Model of Governed Self Service}

More control

CENTRALISED

More access

DECENTRALISED

Single Centre of Excellence

A single area manages the usage of the Big data platform.

CONS
It reduces the potential of the Big data platform which is built for concurrency.

Other models

Guided Self Service

- Analysts have the tools to do self service analytics in the Big data platform.
- Tools for improved discovery are available too.
- Additionally a guide team supports the discovery, navigation and usage of the platform whilst advising on best practices, enriching the use cases and recommending techniques to make the most from the data and the platform.
- The central team monitors the usage to proactively add features and evolve the technology.

Full Self Service

Any individual can access the Big data platform.

CONS
It reduces the added value of collaboration and knowledge sharing whilst potentially overstressing the platform.

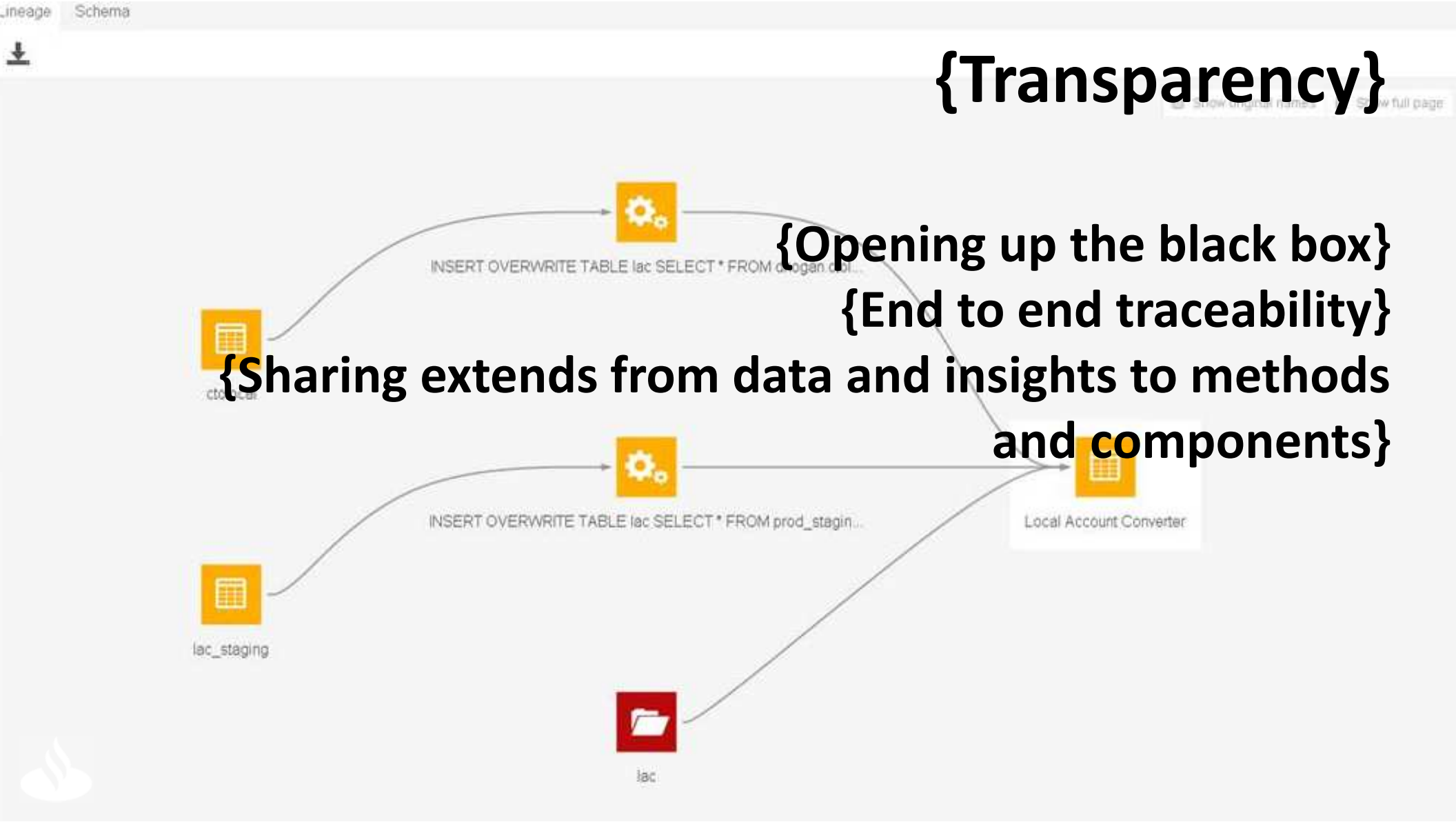
{Pragmatism through exception process}

{Business needs supported, not blocked}

{Quick delivery of PoC and MVPs}

{Commitment to specific controls and to progress Strategic plan}

{Logging of debt (technical, data or process) so it can be addressed}



{Transparency}

{Opening up the black box}

{End to end traceability}

{Sharing extends from data and insights to methods and components}

{Guidelines of the Data Community}

Inventory and Documentation	• All reports captured in a central Report Library managed by the COE representatives. • Metadata fed into the Data Dictionary: Report id, name, description, owner, Area, Frequency, Content and Process summary, Key data sources, Sensitivity and Criticality assessment.
Data Quality and Control Model	• The COE area should review and self certify the quality of the report as a minimum. • For shared processes or reports for multiple COEs customers a peer validation/certification is recommended. • Critical operational BAU processes handed over to central IT and DQ control teams for increased support and certification. • The COE should provide a mechanism for users to raise issues with the reports, the process or the data
Access / data privacy	• Where reports contain sensitive data (specially customer information) the data owner approval should be obtained for its publication. • Data owners may agree to grant a general approval for current and future reports with certain types of data for a specific purpose or target audience (this may include the review of Job roles – Access groups mapping). • Usage of new categories of sensitive data or significant changes in purpose require confirmation. • Data privacy and usage of customers data assessment will be incorporated into the reviews.
Existing Policies and Forums	• Self service reports/dashboards published should be reported in the area workbook including criticality assessment and controls established for audit reviews. • Regulatory reporting to be communicated to the relevant forum. • Key Incidences and Quality Issues to be presented at Data council forum for prioritization.

{Peer review for Continuous Improvement}

{Accountability in the Community}
{Community approval of Operating model}
{Peer review of insights generated by Self service model}
{Pull and push feedback model}
{Prioritisation}
{Collaboration}

{Privacy by design}
{Compliance with GDPR}
{Competitive advantage}

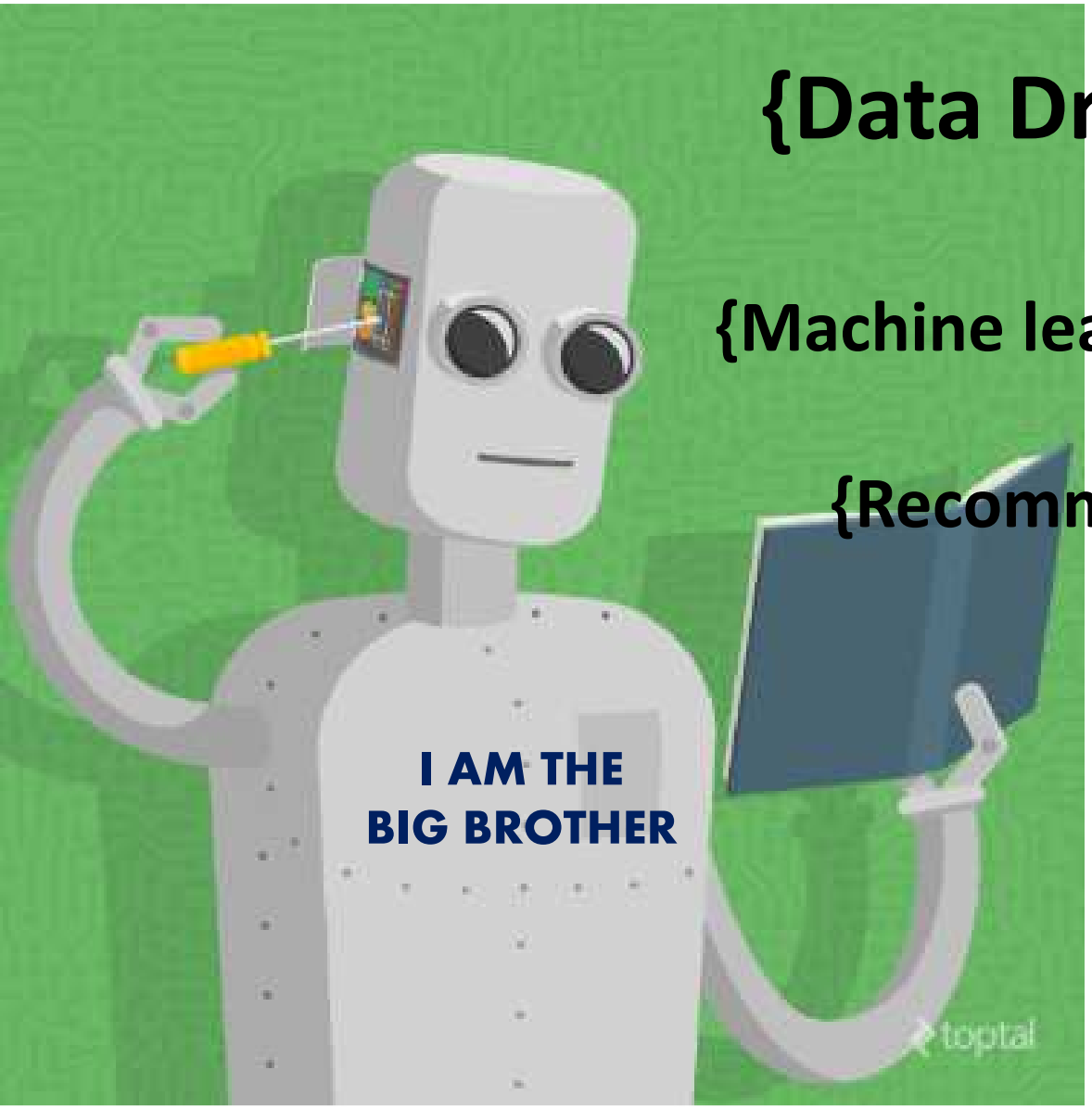
- **Customer data definition**
  - All sensitive data (and customer data) clearly identified and secured
  - Data lineage and traceability
- **Consent**
  - Define purpose and usage and get positive consent
  - Central view of customer and consent, published by API
  - All applications filtered by API
  - All interfaces filtered by API
- **Right to be forgotten**
  - Crawler dedicated to anonymising customers
- **Portability**
  - Interface with all available information of a customer

The Data Driven Organisation

T+?

- Scientific: ML, AI…
- Embedded
- Real time
- Cultural change

7



{Data Driven Governance?}

{Machine learning on Users behavior}

{Real time assessment}

{Recommendations for efficiency}

{Improved security}

Distributed data technology allows for distributed usage through self service analytics.

To sustain this distributed innovation, distributed data governance is necessary.

Thank you…

And get in touch:

antonio.alvarez@isbanuk.com
lidia.crespoparra@geoban.co.uk

Santander