Box 2.7

SUPTECH DEVELOPMENTS (TECHNOLOGICAL INNOVATION APPLIED TO SUPERVISION)

The Banco de España included the need to boost technological innovation in its 2024 Strategic Plan. To this end, among other initiatives, in the supervisory sphere it designed a roadmap for incorporating new technologies into supervisory activity.

One of the pillars on which this roadmap is based is the internal development of SupTech tools. To undertake this task, it is essential to combine people with training in new technologies and knowledge of and experience in the banking supervision function, and access to rich and quality data sources.

Three data sources can be accessed for the development of SupTech tools: i) information from confidential returns; (ii) granular loan information, mainly from the Banco de España's Central Credit Register (CIRBE); and (iii) unstructured data in the form of text: internal capital and liquidity adequacy assessment reports, news, contracts, etc.

The combination of both factors —appropriately trained and experienced staff and quality data sources— should take us to the highest level of the well-known "Data, Information, Knowledge, Wisdom" pyramid (Figure 1), and thus have the capacity to make well-informed decisions based on an understanding of the underlying information, in order to improve the effectiveness and efficiency of supervision.

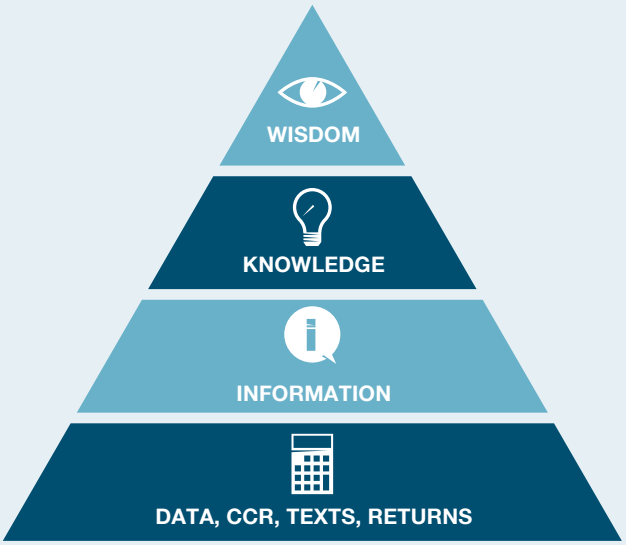
In the development of SupTech tools, four lines of work can be distinguished, which, based on the available data sources, interact in some way with each other (Figure 2):

1 **Improving the quality of the information reported in the CIRBE.** One of the techniques used to identify shortcomings in data quality is the reconciliation of the data reported to the CIRBE with the corresponding information in the confidential returns reported.

2 **Detecting changes in trends and anomalous data based on machine learning algorithms.** An example of this line of work is the tool called Avalanch, which analyses on a monthly basis, by using machine learning models, possible significant changes in the information received on new mortgage loans. The tool applies the adversarial validation technique, which is widely used in artificial intelligence to identify possible differences between two datasets.

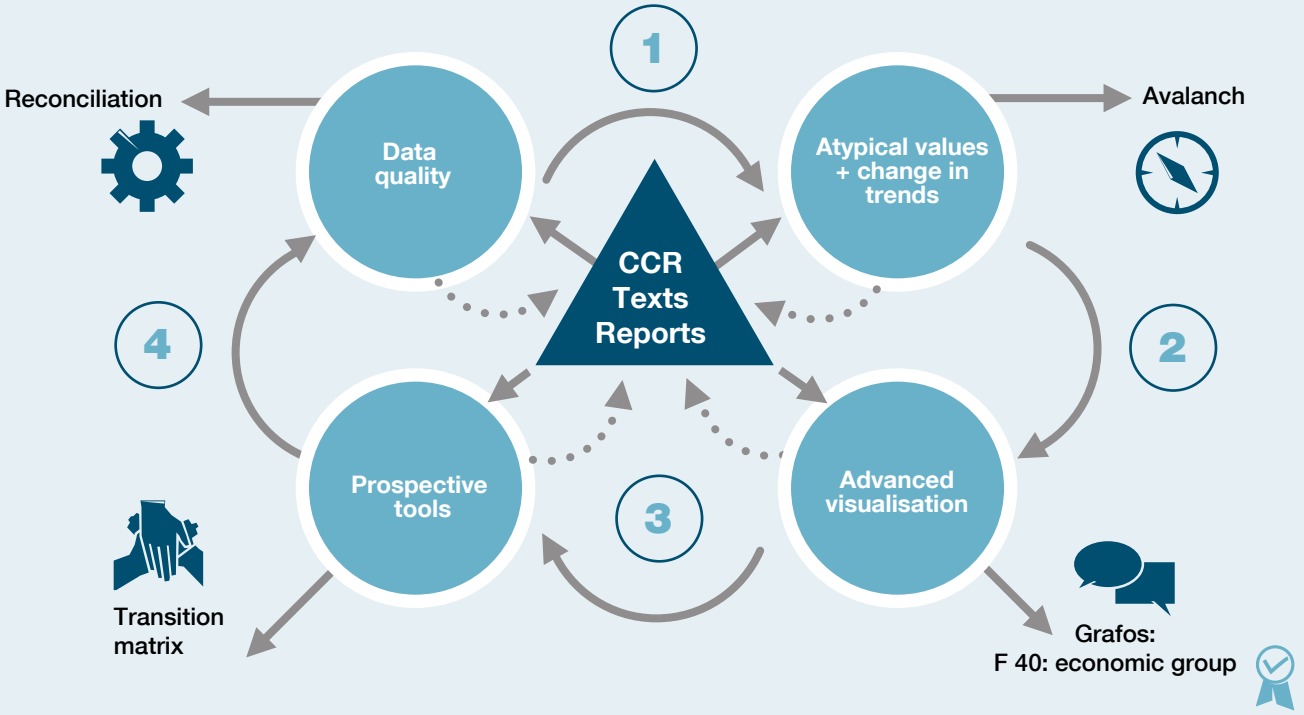
   Using an algorithm with a hierarchical decision tree structure (Figure 3), the most predictive variables are recursively searched and datasets are compared on a monthly basis, looking for significant changes or outliers. For example, between loans of a certain type granted by one institution on different dates; between loans granted by one institution versus another; or

Figure 1
DATA SCIENCE AS A PYRAMID, WITH DATA AT ITS BASE



SOURCE: Banco de España.

Figure 2
INTERRELATION AND INTERACTION PROCESSES



SOURCE: Banco de España.

differences between sub-populations of the same institution (e.g. between transactions granted through an agent versus those granted by the institution's branches). In this way, changes in loan origination policies, anomalous data, reporting errors, etc. can be detected.

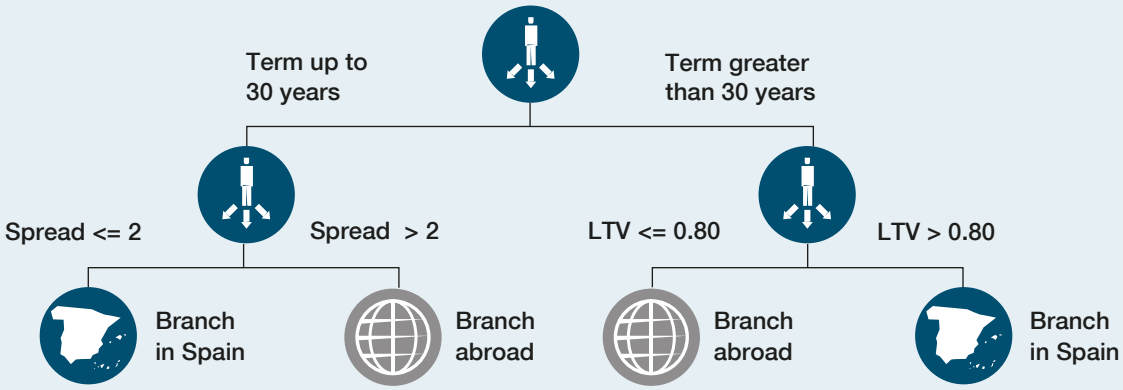
3 **Graphical representation of large amounts of data to facilitate their interpretation and analysis.** Noteworthy is the tool called Grafos —a tool that we are going to share with the SSM, in our capacity as SupTech centre— which allows the interactive visualisation of "multiple relationships (n to m)".

Certain complex relationships can be represented by graphs that facilitate analysis. For example: based on granular information on loan portfolios, relationships (n to m) between debtors, guarantors, collateral and/or economic groups of a set of loans can be represented by graphs; further, ownership relationships between companies forming an economic group can also be represented by graphs.

In the latter case, the visualisation of the multiple relationships of an institution's investee structure makes it possible, for example, to identify the main changes in an economic group between various dates or the structure resulting from mergers and acquisitions between credit institutions; to analyse the impact on solvency of regulatory amendments; or to analyse subsets of the economic group (e.g. the group in a given country).

The tool provides basic graphical capabilities for marking nodes and paths, interactive searches and filtering, or navigation functions through the graph. It also makes it possible to compare graphs (changes in a group between two dates), to perform graph calculations (e.g. to calculate the effective ownership interest percentages that a parent company has in its subsidiaries in complex structures), to identify ownership interests that are in loops (which may lead to capital inefficiency) or to show all paths from the parent company to each investee.

Figure 3
AVALANCH. CASCADING DECISION TREE



SOURCE: Banco de España.

It is also a lightweight tool, easily integrable into Office, which does not require the installation of complicated new environments or other components on the supervisor's basic workstation.

4 **Exploration of predictive capacity based on CIRBE information and extraction of information from unstructured texts.** An example of a tool that explores predictive capacity is the so-called Transition Matrices, which tries to study the payment behaviour of loans and credits through the monthly analysis of the number of months in arrears of each operation reported to the CIRBE.

The tool compares the status of each loan in two consecutive months and generates the percentages of the money exposure that migrate between one arrears level and the other (percentage of the exposure with "x" months in arrears reported in the month that moves to "y" months in arrears in the following month), resulting in a "transition coefficient matrix" between in arrears statuses.

The statistical treatment of these monthly transition matrices makes it possible to project an estimate of future losses of the portfolios (the time evolution of which is an indicator of their quality) or to identify anomalous behaviour in the portfolios.