THE ROLE OF TECHNOLOGY IN PRIMARY REPORTING[1]

December 2004

The primary reporting streamlining is a requirement represented by the European banking system in many fora. The role of technology in the development of an efficient primary reporting system is very important, although organisational aspects should not be undervalued. The technological support of a primary reporting system is made up of many components, each of which refers to specific aspects of the overall information supply chain.

This note aims to point out what would be, according to Banca d'Italia vision and experience, the main features of a possible technological support for primary reporting by the banking system. As technology is an answer to real needs, the note starts from the business drivers, i.e. what banks and central authorities would expect from the system (paragraph 1). The description of the information supply chain follows, i.e. what would be the various steps of the process that starts from banks and ends to the central authority (paragraph 2). Finally the solution drivers are clarified, i.e. what would be the main aspects to be taken into account in the design and development of a primary reporting system (paragraph 3). An outline of Banca d'Italia activity in primary reporting context can be found in annex 1.

1. THE BUSINESS DRIVERS

The business drivers from the banks and the central authority perspectives are outlined in Table 1-1.

Banks perspective	Central authority perspective
Information usefulness and essentiality	Minimum time-to-market
Co-ordination between authorities	Merits-and-costs evaluation
Clear definition of concepts and content	Minimum costs per unit of information
Help in understanding production criteria	Consistent definition of concepts and content
Fair time-to-market	Data quality and exploitation

Table 1-1 – Business drivers

The banks perspective

It is very likely that banks would aim to keep their primary reporting costs to a minimum. In order to achieve this goal, banks are deemed to require that:

- information to be reported is really essential, or at least very useful, for authorities that request it;
- there is some co-ordination between authorities to whom banks report, or at least between the most demanding ones (i.e. supervisors and central banks)[2], in order to avoid an excessive "fragmentation" of the reporting framework;
- authorities clearly define the various economic phenomena of their interest (e.g. "loan", "credit", "receivable", "default", "retail" and so forth) also ensuring, to the possible extent, that there is just one definition for a specific phenomenon;

[1] Written by P. Milani.

[2] Co-ordination is facilitated when a single "physical" authority plays the role of more "logical" authorities (e.g. Banca d'Italia, that plays the role of Italy's central bank and banking supervisor). However, also in this case a co-ordinated approach should not be taken for granted.

- authorities clearly define content to be reported, according to given definitions;
- authorities give some help to clearly understand criteria to be applied to produce the reported information, starting from internal information systems of the banks themselves;
- new requirements are presented sufficiently in advance with respect to their actual introduction.

According to what precedes, banks are focused on total costs, but also on minimum costs per unit of reported information. Major drivers for total costs are deemed to be information essentiality and co-ordination between authorities, while content description and production are reckoned to be the major drivers to obtain minimum costs per unit of reported information.

The central authority perspective

The central authority perspective would be to:

- have a minimum time-to-market[3] in case of new reports or variations to existing reports;
- carefully evaluate merits (for central authority) and costs (for banks) of the information to be reported;
- keep to a minimum the costs per unit of information requested;
- consistently define the information of interest;
- set up a proper data quality assurance and control system, exploit the reported information e.g. by means of business intelligence applications;
- have a level of automation compliant with the size of the reported information.

According to this description, central authority attention is deemed to be focused on time-to-market, on merits and costs of the requested information, on data quality and on costs per unit of reported information. In this sense, there is some correspondence between what is deemed to be useful for banks and for central authority. Moreover, there is an evident link between minimum time-to-market and minimum costs per unit of reported information.

2. THE INFORMATION SUPPLY CHAIN

The overall primary reporting information supply chain starts from new information requirements and ends to the central authority's data warehouse and related business intelligence applications (see figure 2-1). As one can see, the chain represents a recursive process.

[3] Time-to-market is determined by the time banks need to produce the new information and by the time central authority needs to adapt its reporting system. The two adaptations could and should of course proceed in parallel.

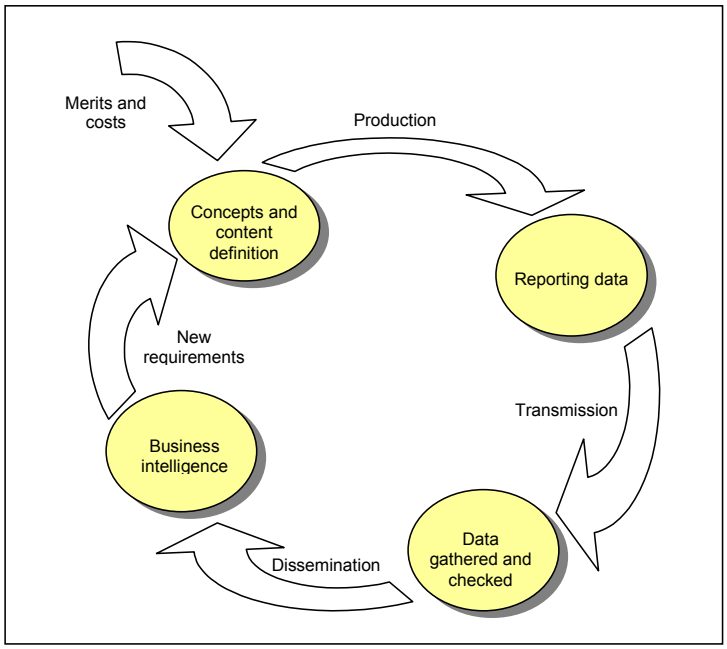

Figure 2-1 – The information supply chain

The banks activity

Banks have to produce the requested information. In order to do this, the activity seems similar to the creation of a data warehouse[4] at the bank site, because information to be reported is mostly built from internal data.

The production of the information to be reported is often not a trivial activity for banks, both for the quantity of the information itself and because of the necessity of a clear understanding of the calculation criteria.

After being produced, the information is reported to authorities using a specified protocol. In this context the term "protocol" means the overall reporting agreement with a specific authority, e.g. the time schedule, the reporting criteria (e.g. whole sending, updates and revisions) and frequency, the way the authority communicates remarks on data quality, the transmission media and format.

The central authority activity

The central authority is deemed to be very interested in time-to-market and data quality. In order to achieve these goals, it is essential to design a "robust" primary reporting process, possibly with a low level of manual intervention, where all steps of the process concur and quality is assured not only by a proper control activity but also by an effective production of the reported information.

When there are new information requirements the central authority would activate a merits and costs procedure, in order to evaluate the "cost-effectiveness" of the information itself. This procedure can be done in various ways, but the banking system should be consulted someway, also to be aware of the time needed to satisfy the new requirements. Then, the central authority would

[4] A data warehouse is a data store that contains information useful to understand the business of the whole enterprise or of a significant part of it (e.g. sales, production, clients and so forth). A data warehouse is mostly, although not exclusively, built from files belonging to operational procedures, i.e. software that operationally supports those sectors of the enterprise, or the enterprise itself, whose business is desired to be better understood. A typical data warehouse project is based on an Extraction, Transformation and Load (ETL) activity, where data is read (extracted), transformed according to previously defined rules and then stored (loaded) into the warehouse. In case of complex transformation rules, there could be the need for more than one extraction-and-transformation step.

consistently and formally describe definitions and content to be reported, present the results to the banking system and collaborate with it in order to identify criteria to be applied to calculate the reported information.

During the current business, the central authority would gather data according to a specified protocol. Data would be subject to quality controls, where various criteria can be applied. Remarks coming from quality controls are sent to banks, that correct the wrong data[5].

When data are deemed to have an acceptable quality, they are stored into the warehouse, ready to be disseminated and used in business intelligence applications.

3. THE SOLUTION DRIVERS

To put banks and central authority perspectives into practice according to the aforementioned information supply chain, proper organisational and technological arrangements would be needed.

The organisational profile

The organisational profile is deemed to be focused on the two following levels:

- an "authority" level, where there is a co-ordination between different authorities (e.g. supervisors and central banks) that have some definitions and content in common, in order to describe them in an integrated and unified way, e.g. able to ensure essentiality and non-redundancy to the requested information;
- a "co-operation" level, where authorities collaborate with banks in order to facilitate the production of the reported information.

The "authority" level

The most part of reporting burden for banks is deemed to be due to reports to supervision and central banking authorities; disclosure too has its own impact, together with the reporting to other authorities. Those reporting obligations are due to national authorities by all banks; cross-border banks have a slightly more complicated situation, because they have to report to authorities of various countries. So, in order to lower primary reporting costs co-ordination between national authorities is important for every bank, while supra-national or European co-ordination is crucial for cross-border banks. A supra-national example of the authority level is the Basle Committee, a European example is the Committee of European Banking Supervisors (CEBS), a national example is the Italian one[6].

The co-operation level

Co-operation level is very important too. In fact, also after a thorough description of overall definitions and content to be reported, banks usually need a help in order to understand the precise relationships between their internal data, mostly available in the operational procedures, and the required information. If implemented, co-operation can be *surprisingly effective*, lowering banks' primary reporting costs and shortening time-to-market very much. On the other hand, it implies a

[5] The overall data quality control and correction activity is usually called "editing".

[6] Co-ordination in Banca d'Italia is put in practice by a high-level Committee, whose participants are the representatives of all institutional functions interested in banking industry data (supervision, economic research, markets oversight, payment systems and so forth). This forum decides about the information requirements that merit to be satisfied by the banking system; accepted requirements are clearly described and integrated into the overall reporting schema. Data administration duties are accomplished by some organisational units that put in practice Committee decisions in a co-ordinated way. See also annex 1.

non trivial effort for banks and central authorities. Supra-national or European examples do not seem to exist. A national example is the Italian one[7].

The technological profile

The technological profile refers to the primary reporting process support and will be described taking into account the two aspects of design and development.

a) design

The design aspect is deemed to be focused on the following components:

- the description of the information to be reported;
- the transmission protocol;
- the software support.

The description of the information to be reported, including related definitions, is an activity usually articulated in two parts:

- ✓ an administrative task that does the job making use of "traditional" instruments, like e.g. Microsoft Word and/or Excel;
- ✓ a someway technical task that describes the content to be reported into a dictionary, modelled according to some rules that, taken all together, are often called "*information model*". This meta-information can be used in many parts of the information supply chain, i.e. in the data production and transmission (banks perspective) and in the handling of the reported data (central authority perspective).

The information model is a crucial component of the overall system. It defines the rules describing reported data, represents the base for the conceptual layer of the dictionary data model and inspires the design of the software support[8].

The information model to be adopted is decided by the central authority, according to the characteristics of the reported data. In case of supervision and central banking the reported information is *multi-dimensional*[9] (see annex 2), both in case of primary reporting, i.e. reporting from banks to national authorities, and in case of secondary reporting, i.e. reporting from national to European or supra-national authorities (e.g. from European national central banks to the European Central Bank, or from world-wide national central banks to the Bank of International Settlements)[10].

In order to better understand what an information model is, two possible examples of different sets of rules follow:

---

[7] The mention is for the Italian PUMA2 inter-bank group, where there are representatives of Banca d'Italia and of Italian banks. This group (see also annex 1):

- evaluates the new requirements and puts them into relationship with the banks internal data, in order to facilitate data production for supervision and central banking. The technical result is a set of parameters (i.e. metadata) called "decisional table". This method can also be used by banks to facilitate data production for disclosure and other purposes;
- collaborates with Banca d'Italia in the merits and costs procedure: in case of new information requirements, banking representatives of PUMA2 group are asked to give a sufficiently precise evaluation of the production costs and time for the new data.

[8] The more metadata are used to "actively" drive software support ("active metadata"), the more is their value in the automation of some part of the information supply chain.

[9] Multi-dimensionality means that the information can be identified by more than one "dimension". Multi-dimensional models include of course the one-dimensional approach, as a particular case.

[10] In this sense one could say that information handling follows the "fractal" model, i.e. it roughly reproduces itself at every level.

1. each reported data has a name, a description and a "context". In the context it is possible to define the entity that the data describes, the reporting period and the measurement criteria. Different data can have links between them. Data is reported in the form of reports, i.e. batches of data that are sent together[11];
2. all "concepts" that make part of the reported information are defined (e.g. the economic phenomenon, the geographic location, the economic activity sector, the unit of measure, the methodology and so forth). Concepts have a name and a description and they can be plain text or can take their values from a code list. Each data is identified and qualified by two specific sets of concepts; the overall data structure identifies a key family. Data is reported in the form of data sets, that are batches of data that share the same structure (i.e. key family) and the same code lists[12].

There are a lot of information models all over the world. Every organisation adopts, "consciously" or "unconsciously"[13], one of them, because reported data is to be modelled someway. The ICT market proposes e.g. the "star schema" (or "snowflake") model, adopted by all data warehouse products; there is some world-wide activity in the field of the information model design[14]. According to a vast experience, it is important to assess the possibility to *map* the information from one model to another one, in order to ensure interoperability between different organisations using different data modelling criteria. In this context, the presence of a formal or semi-formal (i.e. narrative) description of all involved modelling rules lowers the risk of possible errors, shortcomings or misunderstandings.

The transmission protocol takes into account all aspects referring to the way banks and central authority interact, e.g.:

- ✓ the reporting criteria (e.g. whole sending, updates and revisions), including the reporting frequency (daily, monthly, quarterly, annually and so forth);
- ✓ the criteria with which data quality remarks are sent to banks (e.g. network messages, faxes, phone calls and so forth);
- ✓ the transmission media that can be used (e.g. private network, internet, diskette, CD-ROM, cartridge, paper and so forth);
- ✓ the transmission format, i.e. the format with which data is sent;
- ✓ the time schedule.

All previous information is metadata that describe the exchange process. These metadata can be useful for the automation of the process.

Protocols can vary from central authority to central authority and they can be very articulated, because of the transmission media influence. For example, criteria considered valid in case of network usage could be different in case of paper usage, while within a national primary reporting system there could be some banks that use the network, some others that use electronic media and some others that send data on paper. The same applies to the transmission format. In case of paper

[11] These rules more or less describe the XBRL information model.
[12] These rules more or less describe the SDMX information model.
[13] In this context, "conscious" modelling means that the information model has been formally described, e.g. using the "entity-relationship" technique or the "Unified Modelling Language" technique, and/or semi-formally described, e.g. using a narrative form, while "unconscious" modelling means that a formal or semi-formal description is not available.
[14] Referring to the exchange of experiences, it is worthwhile citing the "Metanet" project, an international initiative of applied research funded by European Commission that took place from 2001 to 2003. Participants - essentially coming from Europe, United States, Canada, Australia - have been universities, research centres, software houses, statistical institutes, central banks, Eurostat and the European Central Bank. During the project various experiences have been presented; among them, the model used for the data exchange within the European System of Central Banks, forerunner of the SDMX information model, and the Banca d'Italia primary reporting information model.

usage, there would be some prospects; in case of electronic media usage, "traditional" flat files are likely to be found; in case of network usage, the format[15] could be based on syntaxes (e.g. XML) able to favour interoperability between heterogeneous technological environments[16] and supported by a large number of off-the-shelf tools. In any case, the format should be capable to handle data modelled accordingly to the chosen information model.

The software support for banks and central authority activities depends on the decided level of automation. At the same time, the level of automation is deemed to depend on the dimension of the reported information and/or on the investments that involved organisations want or can afford.

In the following a high-level automation choice will be outlined, i.e. a choice that stresses the role of the software support both at the banks and at the central authority site. In that case the software support would be represented by a *metadata-driven software platform*, whose role:

- *minimises time-to-market* in the central authority perspective, because of its ability to handle (i.e. acquire, control, store, aggregate, disseminate) every reported data, modelled accordingly to the information model and whose definitions, including exchange process information, are hosted in the dictionary (see figure 3-1). In this context, any variations to the reported information would ideally imply almost only corresponding variations to the dictionary itself;
- *minimises costs per unit of reported information* in the banks perspective. In fact, in the Banca d'Italia experience it is possible to define an internal layer, close to the information systems of the banks and fed by their internal procedures, and a set of rules with which it is possible to automatically calculate the information to be reported, starting from the internal layer itself, and possibly check it for quality purposes (see figure 3-2). In this context, after the internal layer feeding, reported data are automatically produced using a software platform based on those definitions (i.e. metadata); moreover, variations to the reported information that can be totally produced by the present content of the internal layer itself can be satisfied very easily and at almost no costs for banks, because they only imply some metadata changes.

As one can see, the high-level architecture of the two platforms is similar. Although conceptually complex and difficult to implement, a metadata-driven software platform is deemed to be *very effective* in lowering primary reporting costs and shortening time-to-market. Moreover, there is an evident link between an integrated approach to data administration and a software platform as the technological response to these integration requirements.

It is worthwhile noting that a properly automated information supply chain also collaborates in *data quality* enhancement; in fact, it allows the design of a "robust" primary reporting process, thus minimising the risk of accidental errors due to manual intervention. Data quality controls are likely to be made using different techniques. They usually involve some non trivial calculations. The major attention is normally on that part of the reporting schema whose quality assurance is deemed to ensure a sufficient quality to the overall reported information (so called "selective editing" strategies).

The *organisation of the data* into the central authority warehouse, like the information model, is influenced by the characteristics of the reported information, that is multi-dimensional. A good data organisation should not be overly influenced by changes in the reported information and should favour the development of business intelligence applications.

[15] Examples of "standard" transmission formats suitable to network usage are GESMES/TS, XBRL and SDMX-ML.
[16] Although the regular execution of a primary reporting and the presence of clear agreements between all "actors" could also suggest the adoption of more efficient syntaxes.

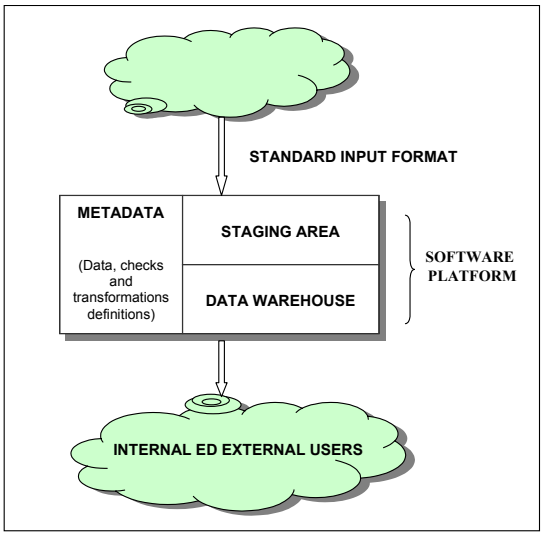

Figure 3-1 - software platform in the central authority perspective

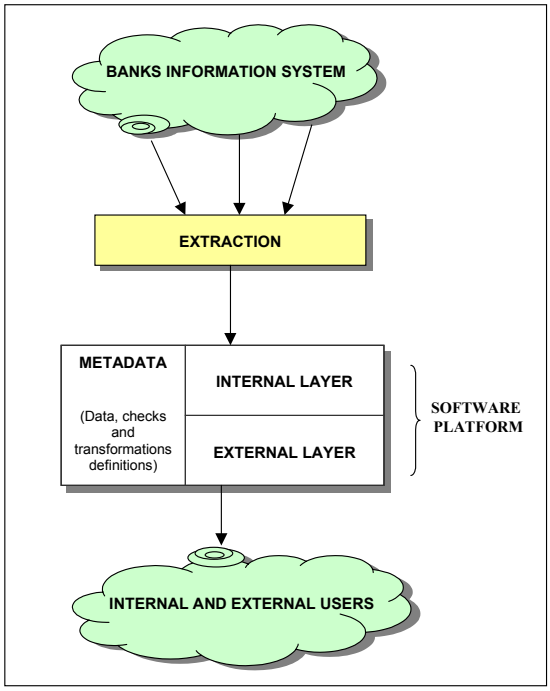

Figure 3-2 – software platform in the banks perspective

b) development

According to what has been said in the design aspect, also this aspect will be handled taking into account the most complex case, i.e. the development of a software platform. It will be described taking into account the "buy" and "make" options, i.e. according to a strategy based on the almost exclusive acquisition of ICT market products ("buy") or to a strategy that does not exclude the acquisition of market technologies but that is essentially based on the in-house development of a significant quantity of software ("make").

The "buy" option

This option takes into account "market" solutions, i.e. solutions that are based on ICT market products and are characterised by a non significant effort in terms of custom software development.

The "buy" option can follow the "one source" approach, where all components are acquired by just one supplier, and the "best of breed" approach, where the overall system is made by the almost seamless integration of products bought by various suppliers, that represent the best of each interested market segment. The "one source" approach is of course simpler and less risky, while the "best of breed" approach is more demanding in terms of required skills but promises higher returns in terms of quality of the overall service. In any case, an organisation should be careful on arguments like suppliers dependency, suppliers overall quality, maturity and market depth of the chosen technologies.

There are some consolidated technologies that can support the automation of the primary reporting system.

Information handling by both banks and central authorities can be supported by data warehouse products and solutions. These solutions are usually oriented toward data production, quality control, storage and dissemination; data definition is made according to a standard information model called "star schema". Data warehouse solutions require the presence of a relational data base management system (DBMS)[17] and of some other software products specific for the data definition, production and storage, quality control, dissemination, business intelligence. In case of "one source" approach, these products are often found as add-ons of the DBMS itself ("back-end" products, for data definition, production, storage and quality control) or of the Application Server ("front-end" products, for dissemination and business intelligence). The "best of breed" approach has been recently facilitated by an Object Management Group (OMG) initiative, called Common Warehouse Metamodel (CWM), where a reference model has been defined in order to standardise metadata exchange between data warehouse and business intelligence products of different vendors, thus ensuring interoperability between those products that vendors declare CWM compliant.

Data warehouse solutions are generally costly and, especially for banks, they are deemed to be economically justifiable only if a bank already uses them for satisfying its specific business needs. These solutions someway lack data "versioning", i.e. the smooth handling of variations to the data definitions that have occurred overtime. In fact, their common approach to data production[18] is "CASE-based"[19], i.e. they help technicians to generate software able to handle a specific reporting schema; in this context, every new reporting schema, or significant variations to a previous reporting schema, could imply the generation of a new software, based on the last version of data definitions, and this means some impact on software configuration management.

An efficient data transmission usually makes use of a network. To best solve security concerns the network could be a private one, either "physical" or "logical" (e.g. "virtual private network"); in case of internet usage, the "Public Key Infrastructure" (PKI) technology could be adopted. At the central authority site, the link between the network and the internal information system can be supported by an integration middleware. The format to be chosen should be compliant with the adopted data modelling rules.

A possible example of a high-level automation of the overall information supply chain would be both banks and central authority adopting data warehouse solutions, the internet as the transmission medium, an integration middleware at the central authority web site in order to connect incoming messages with the internal information system, a proper solution for security concerns. The transmission format would be decided by the central authority, but it should probably adhere to the adopted information model ("star schema"). A perhaps simpler architecture, possibly adopted in

[17] Relational DBMSs examples are IBM DB2, Oracle 9i Database, Microsoft SQL Server, open source MySQL.
[18] So called ETL (Extraction, Transformation and Loading) tools.
[19] CASE: Computer Aided Software Engineering.

case of non-systematic reports and/or for small banks, can envisage the usage of data entry solutions, based on market tools.

The "make" option

This option takes into account "custom" solutions, i.e. solutions that require a significant part of in-house software development. The option usually does not start from "scratch" but follows the "integration" approach, where the overall solution is based on a mix of custom software components and ICT market products[20].

The technologies mentioned before for the "buy" option can also be useful in the "make" scenario. For example, the DBMS plays an important role in the automation of the central authority supply chain. The same applies to integration middleware and security solutions if internet is the preferred transmission medium. Data entry solutions based on market tools can be used. The "make" option could also implement some "public" specifications, developed by co-operation initiatives[21], e.g. in the field of transmission formats. Proper open source products could be taken into account. The in-house development of a software platform is deemed to require sophisticated technical skills and an information model of fully defined semantics.

A "make" option could evolve overtime into some sort of "buy" scenario: for example, an organisation using a DBMS could decide in the future to adopt data warehouse products instead of custom software; or some "niche" software could evolve toward a more comprehensive and consolidated solution; or a co-operative approach could be followed, where some organisations that share the same requirements decide for a joint development of the software solution[22].

An example of a "make" option in action is the Italian solution for primary reporting (see figure 3-3). Banks normally produce required information using market tools that incorporate the "decisional table", designed and fed by a co-operation initiative between Banca d'Italia and the Italian banking system, called "PUMA2 group"[23]. In this context banks have to develop the "extraction" software only[24], i.e. the software that extracts the data from their internal systems, while the quality control and the production of the information to be reported are made by the market tool supposed to have been bought, starting from those data and applying the control and production rules defined into the "decisional table". The reported information comes out according to a specific protocol[25] and is sent using the Italian banking system private network[26]. When arrived in Banca d'Italia, the information itself is captured by an in-house middleware and transferred to a software platform[27], developed in-house and based on a dictionary, that performs data quality controls, including remarks for banks, aggregates data and stores them into the warehouse. Other in-house tools have been developed for dissemination (internal and external) and business intelligence purposes; also this software is

[20] For market products and vendors what has been said in the "buy" option applies.
[21] These specifications are sometimes accompanied by some software tools, with a cost or free of charge.
[22] This approach could be interesting for organisations willing to pursue the software platform approach but unable, for whatever reasons, to develop it by themselves.
[23] From a software perspective, the "decisional table" is a set of parameters (i.e. metadata) that guide the production of the reported information and its quality control at the banks site. The value of parameters, i.e. the content of the "decisional table", is set-up by the PUMA2 inter-bank group, using a tool developed by Banca d'Italia. After its update following an evolution of the reporting schema, this content is downloaded and transferred to software firms that incorporate it into their market tools, that support Italian banks in primary reporting for central banking, supervision and also other (e.g. disclosure) purposes. The "decisional table" is modelled according to the Banca d'Italia primary reporting information model.
[24] It is worthwhile noting that the integrated approach followed by Banca d'Italia in the definition of the reported information minimises the banks' effort in the development of the "extraction" software.
[25] The adopted format is proprietary and its syntax is "traditional", neither EDIFACT nor XML.
[26] In the future the use of internet is envisaged.
[27] The platform is almost completely driven by the dictionary definitions and is unique, i.e. the same software handles every data modelled according to the same rules (Banca d'Italia primary reporting information model), being it able to smoothly handle variations to the data definitions that have occurred overtime.

completely driven by the dictionary definitions. The dictionary, multilingual, can contain information that refers to more than one reporting schema, managed by the same or by different organisational functions; it can also handle the reporting schemas evolution overtime. The Banca d'Italia primary reporting information model is the base for the conceptual layer of the dictionary data model. Banca d'Italia has also developed some data entry software, for non-systematic reports; the overall architecture makes use of the internet, of an integration middleware and, in the next future, of a PKI solution.

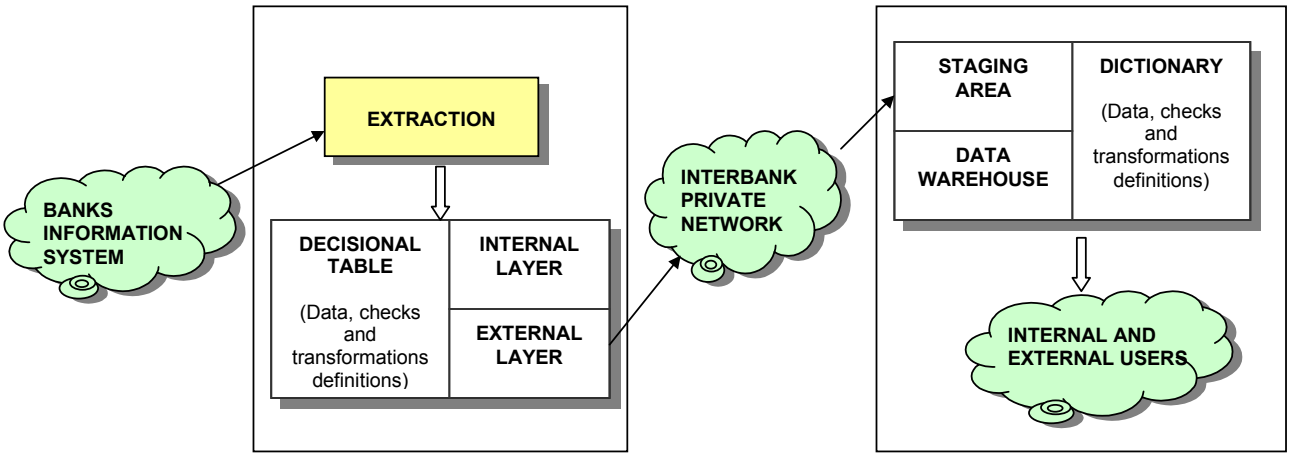

Figure 3-3 – The Italian solution

ANNEX 1

Banca d'Italia performs the following institutional functions: banking and financial supervision; market oversight; promotion and protection of competition in the credit sector; economic and institutional analysis, research and study; and, jointly with the ECB, payment systems oversight. Other central banking functions are performed by the Italian Office of Exchange (UIC). Within the ESCB, Banca d'Italia concurs to determine monetary policy for the Euro area, co-operates in the currency issue function and implements these policies in Italian money and financial markets. In the sphere of economic policy, Banca d'Italia performs high-level consultation for the legislative and executive branches. It also provides the treasury service for the central government.

The organisational model chosen for the management of primary reporting by the banking and financial industry is based on an *integrated solution* in terms of data administration, structure of the reporting schemas and production by the industry of the information to be reported. From the Banca d'Italia perspective, this choice brings to lower the time-to-market of new information requirements, to a higher level of data quality and to a better service for end-users. From the industry perspective, this choice is deemed to lower total reporting costs.

The core of the organisational solution is represented by the "Committee for the management of the institutional functions information system", whose representatives come from institutional areas that make the greatest use of primary reporting content. In short, the Committee:

- co-ordinates the various information requirements, ensuring the essentiality, completeness and absence of redundancy of reporting schemas;
- promotes concepts definitions for general use within the overall primary reporting system and documents them in a dictionary;
- activates a formal merits and costs procedure in order to assess the costs for the industry against the merits for the users of new information requirements, before they are introduced.

The Committee is co-ordinated by the Banca d'Italia Statistical Services Department, that is also in charge to manage primary reporting from an operational point of view and to develop and maintain the corresponding software support.

With reference to the production of the reported information by the banking industry, the inter-bank group PUMA2, co-ordinated by Banca d'Italia and with representatives of 15 Italian banks, is aimed at the preparation of a reference documentation, standard for the whole industry, that defines the production rules for the information to be reported, starting from the firms internal information systems. This documentation can be interpreted by proper ICT market tools, that run at the firms site and calculate, check and transmit the information itself according to the aforementioned rules. PUMA2 group also collaborates with the Committee in the merits and costs procedure, in order to provide useful elements for evaluating costs, time and, in general, the overall impact of new requirements.

Data quality is first of all ensured by a robust primary reporting process, supported by a high level of automation. It is checked at various levels, at the firms site and at the Banca d'Italia site. Generally speaking, data quality controls verify formal correctness, logical coherence, time series trend, comparison between trends of different intermediaries and so forth.

Once arrived in Banca d'Italia, incoming data is managed by a highly computerised system, articulated in two sub-systems. The first sub-system files the incoming information, monitors the data quality and sends any remarks to reporting firms. The second sub-system stores the output sent by the first sub-system into the data warehouse and calculates new aggregates. Other in-house tools have been developed for dissemination (internal and external) and business intelligence purposes.

Banca d'Italia exchanges a lot of data with other national institutions, e.g. the Italian National Statistical Office (ISTAT) and the market "watchdog" (CONSOB).

ANNEX 2

This note is aimed to point out the importance of multi-dimensional phenomena in supervision and central banking activity and, consequently, the necessity that the data representing those phenomena are transmitted and organised accordingly. The discussion will be based on an example (loans, deposits and banking branches, classified by geographic location) taken from Banca d'Italia supervision practice.

Let us start remembering what multi-dimensionality means, using the aforementioned example. Assume that Banca d'Italia wants to know how loans, deposits and banking branches are distributed by geographic location, at the municipality level. There are here three economic phenomena (loans, deposits, branches) whose values banks have to give classified by municipality. In abstract terms, we could say that the information of interest can be defined in a two-dimensional, discrete space, whose axes are represented by the economic phenomenon and the municipality, and whose valid co-ordinates are given by the combination of loans, deposits or branches for the "economic phenomenon" axis and by the proper municipality, taken from the set of Italian municipalities, for the "municipality" axis. This kind of information is commonly called multi-dimensional.

Multi-dimensionality is very important in supervision and central banking activity. Information e.g. about banking system morphology is necessary to conduct banking supervision and to study the banking system itself. All this information is intrinsically multi-dimensional: typical classification variables are client or branch geographic location, economic sector and/or branch of the client activity and so on. Moreover, according to Banca d'Italia experience multi-dimensional data account for the most part of the data volume sent by the banking system.

Coming back to the example, Banca d'Italia asks, on one hand, the banking system to send this information and, on the other hand, its technicians to organise it into the internal information system. Let us focus on the way data are calculated at the sender site (banks) and organised into the information system at the receiver site (central authority).

Within the information system of a generic bank, data about e.g. a loan is likely to be organised together with, among the others, the information of the branch that supplied it. So, this data is likely to be extracted substituting the corresponding branch code with the code of the municipality where the branch itself resides (this information is likely to be taken from the branches registry). All extracted data are now aggregated by municipality. The same applies, *mutatis mutandis*, to deposits. In case of branches, one could extract the branch code and its municipality from the branches registry and aggregate by municipality, giving as a result the number of branches in each municipality. At the end of the work, the generic bank is likely to have generated an information articulated in three fields, the economic phenomenon (loans, deposits, branches), the municipality, the value: the multi-dimensionality of the information is represented by an identifier ("key") with more than one field (i.e. phenomenon and municipality).

When the data arrives to Banca d'Italia, it must be stored into its own information system and organised in such a way to facilitate analysis. With this respect, there is a wide-spread opinion within the market (i.e. software and client companies) that the best way to organise multi-dimensional data is to distinguish between economic phenomenon (if present) and other dimensions, to maintain the multi-field identifier structure and to link each field (i.e. economic phenomenon and municipality) to the corresponding list of possible values (i.e. loans, deposits and branches for the economic phenomenon; Roma, Torino, Firenze, and so on for the municipality[28]). The market calls "star schema" this way to organise data and OLAP (On Line Analytical Processing) the analysis that starts from data organised according to the "star schema" itself. So,

[28] Italian municipalities are more than 8000.

data and metadata are stored into a relational[29] data base management system according to the OLAP model and they form the so called data warehouse, from which end users start for their so called business intelligence applications. As a matter of fact, Banca d'Italia organises data coming from the banking system in DB2 (the IBM relational data base management system), according to a proprietary model that is close to that suggested by the OLAP approach.

To conclude, we have stressed the importance of multi-dimensional phenomena in banking supervision activity and we have said that, from a very technical point of view, the corresponding information is likely to come up multi-dimensional at the sender site (banks) and to be organised multi-dimensional at the receiver site (central authority). In this context, *one should ask for an information model, and consequently a transmission format, being able to properly handle multi-dimensional data*.

Multi-dimensional data can of course be modelled also in a one-dimensional way; this can be done e.g. assuming that the multi-field identifier is in fact a one-field identifier and consequently defining the data in the dictionary. However, in the data administration perspective this approach can be *disastrous* in some cases. According to the example, the approach that we could call "the one-dimensional representation of a multi-dimensional data" would mean to assign a single code to every combination of economic phenomenon and municipality. Remembering that the number of Italian municipalities is more than 8000, that would mean to define some 8000*3=24000 codes (one-dimensionality) in the dictionary, instead of some 8000+3=8003 codes (multi-dimensionality). In more abstract terms, the alternative is between the *multiplication* and the *summation* of the number of possible values for each variable that identifies the multi-dimensional data. It should also be noted that, in case of one-dimensional representation of a multi-dimensional information, when data arrives to destination multi-dimensionality should often be reconstructed in order to support business intelligence requirements, and this would imply the development of some specific software, based on tables that would transform the one-dimensional representation into the multi-dimensional one.

[29] Strictly speaking, the usage of a relational DBMS is not mandatory, because OLAP can in principle be organised also under other models (i.e. hierarchical or network). As a matter of fact, OLAP is implemented in a simpler way if data is managed by a relational DBMS; moreover, OLAP tools on the ICT market make use of a relational DBMS.