Business Intelligence Exploitation for investigating territorial Systems, methodological Overviews and empirical Considerations

Mario Mezzanzanica, Mirko Cesarini, Roberto Boselli

[1]University of Milan Bicocca, Department of Statistics

Civil servants and service managers need accurate and up-to-date information about the population to improve the service provision models and therefore to meet the citizens higher and higher expectations in the current dynamic economic and social context. Public Administrations discovered the potential of using administrative archives to obtain accurate information about the population. Administrative archives contain a valuable information asset which describes accurately and extensively the population. The exploitation of such asset requires Public Administrations to integrate information spread across several departments, to address data quality issues arising when administrative data is used, and to develop analytical and reporting models. Public Administrations started using the Data Warehouse / Business Intelligence approach which has extensively been used in the private sector for accomplishing similar tasks. This paper will investigate how existing methodologies for building Data Warehouses can be applied to the public sector scenario, some public sector specific issues will be explored, and some case studies highlighting possible solutions will be presented.

1. Introduction

The tough economic conditions and the more and more dynamic market and social conditions require policy makers and civil servants to take care of the population both improving existing services and by implementing new active policies. To reach these goals, Public Administrations have built complex services, networking several service providers, creating partnership among public and private organizations as well. The leading or coordination of such networks increased the information needs of civil servants and policy makers. According to (Golfarelli et al., 2004) “During the last ten years the approach to business management has deeply changed, and companies have understood the importance of enforcing achievement of the goals defined by their strategy through metrics-driven management”. This approach is spreading in Public Administrations as well, as advocated by some public management doctrines like “New Public Management” (Dunleavy et al., 2005) and “Digital Era Governance” (Skalen, 2004).

The design and successful implementation of active policies or supporting services requires decision makers and public servants to obtain useful insight about local territorial systems where interventions will take place. Furthermore, information related to individual and collective needs plays a large and relevant role, therefore a deep knowledge of the population is required.

The economical and social settings have been undergoing deep changes in the last years leading to very complex scenarios. The job career lifecycle is a significant example from this point of view. Currently people change work type and employment more frequently with respect to, for example, 30 years ago, where a person frequently spent her/his whole working life in the same company. At that time, economical and technological evolution life cycle was synchronized with the workforce life cycle, therefore skill updates were mostly provided by generational replacement. Nowadays the technological and economical life cycle is significantly shorter than the duration of the working life of an individual, requiring people to learn continuously. This scenario required policy makers to develop active policies, e.g. providing qualification courses for unemployed people. However, information about job market trends is paramount to successfully implement support services and active policies in the job market place.

Furthermore the introduction of ICT to support the delivery of new and traditional services, and the easy customization capabilities provided by new technologies, outline new scenarios where services provided to the citizens can undergo a process of customization leading to the provision of multiple, target oriented version of similar services. In such a scenario, the service managers and the politicians strongly need information describing the service consumers and their reactions to changes in service delivery.

Data deriving from administrative sources (e.g. government registries, tax registries, etc.) assumes a basic value to gather information concerning the community. Administrative archives managed by PA nowadays contain data describing the whole population with fine grained details about each individual. Moreover, the introduction of the information and communication technologies has enlarged and empowered the availability and fruition of administrative databases, making information accessible to Organizations and Institutions for further surveys (Sundgren, 1993). Administrative archives are a rich source of information for statistical analysis and allow to perform meaningful and detailed analysis both on the whole population and on small subsets. Furthermore, by using heterogeneous archives coming from different administrations and describing different domains, it is possible to analyze the population from many points of view, achieving in this way a deep knowledge of the observed phenomena.

However, turning large amounts of data into useful and insightful information is not a trivial task. Data Warehouse, Business Intelligence tools and methodologies have been developed and used in the private sector to extract synthetic information (supporting decision making) from large firm archives. Data Warehouses and Business Intelligence have been rarely used for investigating Public Administration archives. Public Administrations traditionally have been refrained from investing funds on gathering information on their citizens and service customers (in some countries it was prohibited by law, and in general the topic has been considered difficult to justify in front of the whole population). Only in recent years the benefits of designing and improving services relying on fresh information over the population has been accepted as a positive contribution. Furthermore, the cost of developing Business Intelligence /

Data Warehouse projects has reduced significantly in the last years. Since many obstacles have vanished, decision makers and civil servants have raised their attentions to Data Warehouse and Business Intelligence methodologies.

Considering Business Intelligence and Data Warehouses, the public sectors however is a quite different scenario with respect to the private one, for example data archives (about the population) have huge dimensions and complexity, such archives have rarely been used for aggregate analysis (their typical use is for transaction support), therefore data quality issues can be very challenging. Furthermore, the (public sector) user requirements (i.e. civil servants and policy makers) are more difficult to find out with respect to the private sector (see Sec. 3 for further details), and collecting the requirements is a paramount step in Business Intelligence.

The development of a Business Intelligence project still is very challenging and resource consuming in the private sector, even if a consolidated practice is spread among the practitioners. Selecting the most appropriate development methodologies is paramount in the public sector, considering that Public Administrations are experimenting strong budget and resource reductions. The research question investigated in this paper is whether the existing methodologies and practices (for building Business Intelligence and Data Warehouse projects) developed for the private sector can be applied to the public and which changes are required to deal with the public sector typical issues. The paper is structured as follows: Sec. 2 will provide an overview of the Business Intelligence and Data Warehouse sector, Sec. 3 will analyze the differences between Business Intelligence exploitation in the private and public sectors, Sec. 4 will focus on exploiting public sector administrative data, Sec. 5 will illustrate two case studies and will highlight some hints for the public sectors, and finally Sec. 6 will draw the conclusions and will illustrate the future work.

2. Business Intelligence and Data Warehouse Overview

A Data Warehouse is defined as “subject oriented, integrated, non time-volatile, management-supporting collection of data” (Inmon, 2009). Data Warehouses address key issues affecting Decision Support Systems: reduce the multiplicity of data sets being created in large organizations (which often are inconsistent and represent data according to different metrics, …); preserve historical data (operational information systems mostly store current values and do not record historical data, except in few cases and for little time); shift analysis activities (which may require huge computational and data storage resources) from the overloaded operational systems to systems specifically devoted to data warehousing. Data Warehouses historically have been created to store in a single place and in an integrated manner the information assets spread across the several information systems of an organization.

In (Kimball et al., 2008) the authors argued that they “had always referred to the overall process of providing information to support business decision making as data warehousing, the term Business Intelligence initially emerged in the 1990s to refer to the reporting and analysis of data stored in Data Warehouses after that many organizations had built Data Warehouses as data repositories without any regard to getting the data out and delivered to the business users in a useful manner”.

According to (Golfarelli et al., 2004) “Business Intelligence (BI) can be defined as the process of turning data into information and then into knowledge … Business Intelligence was born within the industrial world in the early 90’s, to satisfy the managers’ request for efficiently and effectively analyzing the enterprise data in order to better understand the situation of their business and improving the decision process”.

According to (Lonnqvist and Pirttimaki, 2006) Business Intelligence has several related terms including competitive intelligence (CI), market intelligence, customer intelligence, competitor intelligence, strategic intelligence, and technical intelligence. In North American literature, the term CI is frequently used and the external environment and external information sources are emphasized, e.g., (Cottrill, 1998), (Fuld, 1994), (Kahaner, 1996), (Vibert, 2004). In European literature, the term BI is considered a broad umbrella concept for CI and the other intelligence-related terms mentioned above. Nevertheless, almost all the definitions share the same focus, even if the term has been defined from several perspectives (Casado, 2004), and they all include the idea of analysis of data and information. Business Intelligence presents business information in a timely and easily consumed way and provides the ability to reason and understand the meaning behind business information through, for example, discovery, analysis, and ad hoc querying (Azoff and Charlesworth, 2004).

Different Public Administrations have started projects for integrating the content of several administrative archives into comprehensive repositories (e.g. the “Enterprise Data Warehouse” according to Inmon terminology) for statistical and analytical purposes, however the “Business Intelligence” portion of the task often lag behind. The delay of Business Intelligence / Data Warehouse exploitation is only one of the differences about BI in the public sector with respect to the private sector. Some more differences will be explained in Sec. 3.

The distinction between Business Intelligence and Data Warehouse systems is blurry in the scientific literature and in the communities of practitioners, furthermore there are no commonly shared definitions. Henceforth the article will focus on comprehensive Business Intelligence / Data Warehouse systems if not explicitly stated otherwise. The aim is to focus on the overall process of providing information to support decision making in the public sector.

2.1. Business Intelligence / Data Warehouse Architecture Overview

In this section an overview of a Business Intelligence / Data Warehouse architecture will be provided. Describing extensively the several types of architectures would require a lot of space and it is outside of the scope of this paper. The aim of this section is to provide a brief sketch of a Data Warehouse / Business Intelligence system from the logical point of view, abstracting from the technical and development detail. For a deeper investigation on the several architectures and the technical details, an extended literature is available, e.g. (Kimball et al., 2008), (Inmon, 2009), (Golfarelli et al., 1998). A synthetic Data Warehouse / Business Intelligence logical architecture is sketched in Fig. 1.

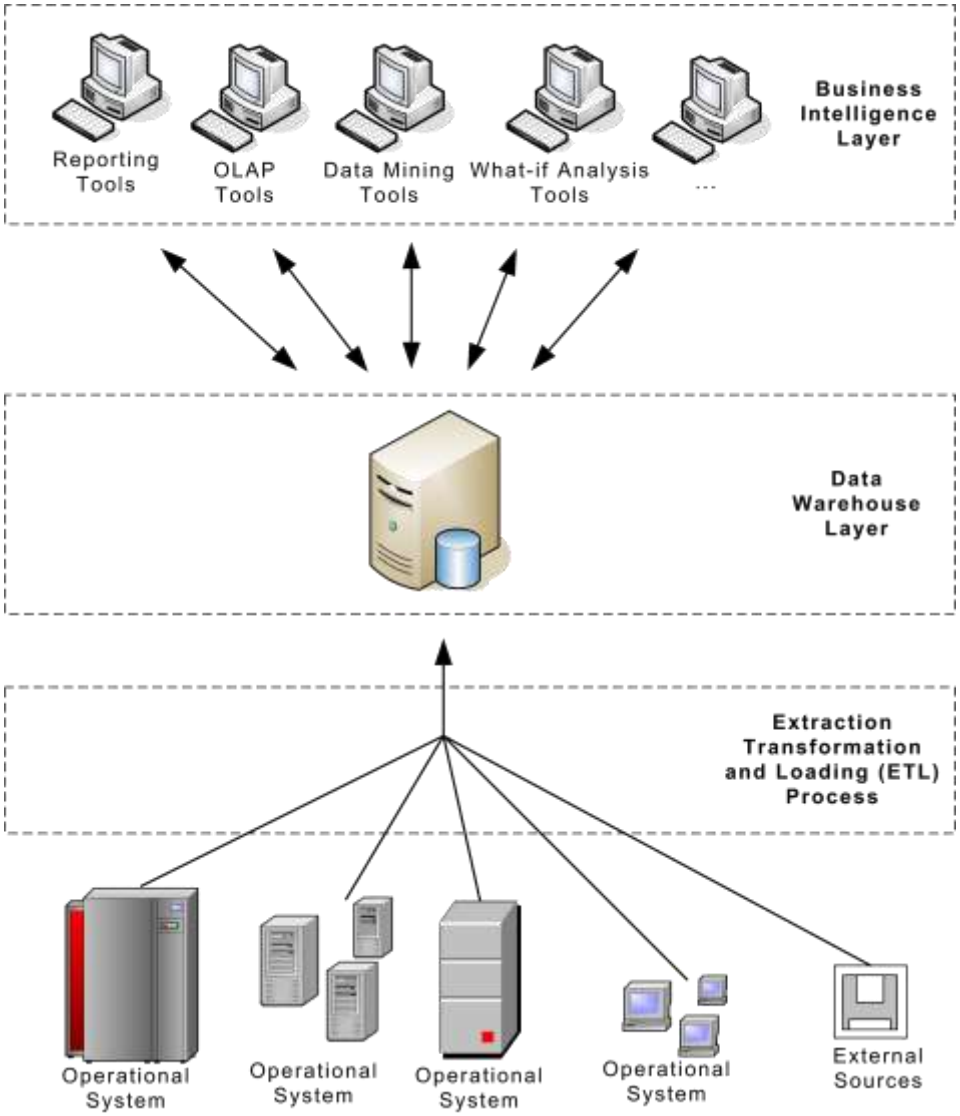

Fig. 1: Data Warehouse / Business Intelligence logical Architecture

The Information Systems depicted in the bottom represents the sources alimenting the overall Data Warehouse / Business Intelligence system (DWBI System henceforth). Sources may be data sets extracted by the information systems supporting the operational processes, data coming from the administrative processes, and data derived from external resources (e.g. market trend descriptions, budgeting processes). The Data Warehouse layer contains an integrated snapshot of the data produced by the sources, data before entering the Data Warehouse has undergone a reconciliation processes (where inconsistencies are resolved), a transformation process (where the several data formats are merged into a single consistent data format, a quality improvement process (where errors affecting the data are addressed), and a merge process (where the same data coming from different sources is stored into a single version). Here, some examples of errors or inconsistencies described in literature (Surkyn, 2006) are reported:

- NACE-sectors of economic activity can get consciously miss-registered because administratively they only serve to determine the social security regime of the employees. In that case, any other NACE-code linked to the same regime can be

used as a substitute. As a result, more “exotic” NACE-sectors can get underrepresented in statistics.

- Variable details that has no administrative use tend to be generally neglected.
- Different administrative archives may detect events with longer or shorter time-lags and this leads to inconsistency: people in the workforce may unduely still be found in school registers, retired persons may still occur in the registers of the active population. Such mistakes may be hard to detect and correct, as some of these situations may reflect reality (students can have a job, people receiving pensions can have a job).

The Data Warehouse layer may be physically implemented in several ways, to cite a few: through a middleware software layer, through a single enterprise-wide Data Warehouse, i.e. the enterprise Data Warehouse approach (Inmon, 2009), through several Data Warehouses sharing conformed dimensions, i.e. the bus matrix approach (Kimball and Ross, 2002). For further details see (Sen and Sinha, 2005).

Although the Data Warehouse layer may exist without a Data Warehouse tool, often a software implementing a Data Warehouse can be found on that layer. Data Warehouses manage data framed according to a multidimensional model composed of facts and dimensions. A fact is (one of) the main topic of interest in a set of data; topics or analysis perspectives that are connected via associations to facts are called dimensions, and they are usually the focus of Data Warehouses analysis. For further details see (Kimball and Ross, 2002).

2.2. Business Intelligence Development Steps

The development of a Business Intelligence / Data Warehouse system can be summarized in two logical steps: 1) building a comprehensive and conciliated repository from the overall data sources (the Data Warehouse layer in Fig. 1), and 2) developing the Business Intelligence layer that turns the huge data sets available from the Data Warehouse into synthetic information and presents those information to the decision makers. Each of these steps should be addressed by a specific development phase, although the two steps are not independent and influences each other.

Building a Data Warehouse requires assessing the information sources data contents, in terms of available data, structure in which data is managed, and quality. When multiple data sources are involved (i.e. when data is extracted by several information systems) data should also be merged into an integrated archive which has to be designed and implemented. As an example, suppose two archives should be integrated: the tax-income archives (which hold information on citizens income declarations and tax payments) and the labour contract type archives (which hold information about the job performed by a person and the kind of contract). The two source archives have some common information (e.g. name, family name, address) and some specific information (tax related information on one side, job related information on the other).

The task of obtaining insights about the information sources structures and contents may require a lot of effort, especially when there is no documentation on the source archives or if the source archives are maintained by legacy systems (the two cases occurs often in Public Administrations). When no documentation is available the ar-

chives contents and structures should be guessed by reverse engineering activities, which is a task that may require a lot of resources. Next the source data quality should be evaluated and quality improvements activities should be established. As reported in Sec. 4, a lot of effort may be required to improve the quality of administrative archives to be able to exploit them for decision support.

After the data sources have been studied and an integrated repository has been designed, the "Extraction, Transformation, and Loading" (ETL) process should be designed and implemented. The ETL process is in charge of extracting the data from the sources, correcting the errors, merging data, aligning them to a single codification when the original ones are different, and loading the results to the destination archive (e.g. the Data Warehouse). The ETL process should be designed twice: for the initial loading and for the periodic incremental loading through which the new data is periodically reversed into the destination archive. The overall process has been shortly described for the sake of simplicity. According to the complexity and quality of the source archives, the ETL process design can be very cumbersome, requiring the development and maintenance team to perform additional activities not illustrated in this paper. For further info see (Kimball and Caserta, 2004). Once the Data Warehouse layer is available and populated with data, there is a single source of integrated and trusted information about the business carried out by the organization (which may describe the selling process of a private organization or the status of a population served by a Public Administration).

The Data Warehouse layer is the foundation upon which the Business Intelligence layer can be built. The development of a Business Intelligence layer requires a multidisciplinary team able to understand the organization processes, the service provided, the information managed by the organization, which are the key performance indicators, and in general the synthetic information that can be useful for the decision makers. The Business Intelligence development team should have a multidisciplinary approach since: should be able to collect the "business requirements"[1]; should be able to find out the statistical and economical models that can be used to synthesize information consistently with the domain requirements; should be able to successfully communicate with the Business Intelligence final users in order to get feed-back on prototypes; should be able to convince business users to reach an agreement when terminology mismatches occurs (e.g. different departments of an organizations may use different terminology for the same concept, or use the same terminology for different concepts); should have the ICT expertise necessary to deal with the hardware and software of Business Intelligence solutions.

Knowledge sharing related issues emerge several times during the development of a Data Warehouse / Business Intelligence project. The BI team should exchange knowledge with the ICT team managing the operational information systems, at the same time should exchange knowledge with the business users and guess their requirements. Furthermore, once a Business Intelligence system has been developed and the business users start exploiting it, some more additional desiderata will emerge. Cause the knowledge sharing issues, the waterfall lifecycle based method-

[1] In the Business Intelligence literature, the term business requirements refers to the requirements of non technical users, who are frequently business experts or decision makers in the private sectors, although several more typologies of workers make use of a Business Intelligence system.

ology (which is often used for developing ICT systems) is not well suited for Business Intelligence and Data Warehouse development, rather iterative lifecycle methodologies should be preferred (Inmon, 2009). An iterative methodology solicit ICT from one side and business users from the other to identify further requirements once a prototype has been delivered. However iterative lifecycle methodologies make resource consumption and project progress difficult to estimate. This is a serious issue when lower budget are available for Business Intelligence projects.

2.3. Business Intelligence / Data Warehouse Development Methodologies

The development of a Business Intelligence / Data Warehouse system can be logically split into the development of the "Data Warehouse layer" and the development of the Business Intelligence layer of Fig. 1.

There is a wide literature on Data Warehouse development methodologies (Mattison, 1996), (Edwards, 1998), (Kimball et al., 2008) . Without going into details, all the methodologies can be classified according to the following approaches: data-driven, goal-driven and user-driven, whereas the approaches are not mutually exclusive.

- Data driven. According to (Inmon, 2009) the Data Warehouse user has a mindset of "Give me what I say I want, and then I can tell you what I really want", consequently users new requirements usually are the last thing to be discovered in the Data Warehouse development life cycle. Therefore in (Inmon, 2009) it is recommended to focus on the analysis of the corporate data model and relevant transactions, to design and populate a Data Warehouse and then to collect user requirements after the business users have evaluated the Data Warehouse content.

- Goal driven. Data Warehouse design is driven by organization business goals according to the goal driven approach. (Kimball and Ross, 2002) propose a four-step approach where business processes are selected, analyzed, and then the Data Warehouse contents is designed accordingly. (Boehnlein and Ulbrich vom Ende, 2000) propose to find out the goals and services that an organization provides to its customers, then to analyze the business processes, to analyze the relationships between business processes and customer transactions and after that to design the Data Warehouse.

- User driven. This approach proposes to develop a first prototype based on the business users needs. Business people have to be interviewed to define goals, to gather, prioritize, and define business questions supporting these goals. Afterward the Data Warehouse has to be designed accordingly.

In (List et al., 2002) it is reported a comparison and an evaluation of the three groups of methodologies.

The necessity of using an iterative lifecycle and the difficulty of obtaining requirements are common issues to all the approaches.

Considering the Business Intelligence development, development methodologies are described in (Moss and Atre, 2003), (Reinschmidt and Francoise, 2000) and (Gilad and Gilad, 1985).

A key aspect of building a Business Intelligence system are the methodologies used to find out the indicators that have to be provided to the decision makers. Such methodologies are: Management Accounting (MA) (Polimeni et al., 1981), Critical Success Factors (CSF) (Rockart, 1979), Key Performance Indicators (KPI) (Beatham et al., 2004), Balanced Score Card (BSC) (Kaplan and Norton, 1996). Finding the synthetic indicators, although being an important part of a Business Intelligence project, it doesn't cover all the required effort. The remaining is related to carrying out reporting (e.g. dashboards) (Few, 2006), multidimensional / on-line analytical processing (OLAP) (Chaudhuri and Dayal, 1997), data analysis, and data mining activities (Han and Kamber, 2006). These topics will not be described in this paper for simplicity. For further information, see (Kimball and Ross, 2002).

A common aspects of the several Business Intelligence design and development methodologies is the Evolutionary Development (i.e. a system evolves through an iterative process of design, development and use) which has always been a key concept in the Decision Support System theory (from which stem Business Intelligence systems). In (Courbon et al., 1978) the evolutionary approach is described as a non linear development process which is made by evolutive cycles which periodically end and that involve final users. Every time a cycle ends the final goal get closer. (Courbon, 1996) describes the cycles as a sequence of actions (the design of a new version) and reflections (user feedback) and states that the cycles are "learning processes". It is worth to mention that (Sage, 1991) has identified the driver of DSS evolution in the discovery of new information requirements in an evolutionary development phase.

These research findings emphasize the knowledge sharing and transfer issues that are involved in a Business Intelligence / Data Warehouse development processes.

Agile methods (Fowler and Highsmith, 2001) which are characterized by frequent delivery of prototypes to the final users in order to collect feed-back seems to be a pregnant way for addressing knowledge sharing issues. In (Conboy, 2009), (Greening, 2010), and (Hughes, 2008) applications of agile development methodologies to information systems, Business Intelligence, and data warehousing systems have been studied, but some more investigation is required.

3. Business Intelligence and Public Administrations

Data Warehouse and Business Intelligence exploitation in the public sector is far behind the private one. Several reasons can be added to explain this. In (Nutt, 2006) it has been investigated the differences between public and private decision making practices. Some of the differences found can also be used to explain the lag among the public and the private sector.

- private sector managers are more apt to support budget decisions made with analysis and less likely to support them when bargaining is applied. Public sector managers are less likely to support budget decisions backed by analysis and more likely to support those that are derived from bargaining with agency people.
- Legislative mandates constrain budgets, which limits or even prohibits public sector leaders from spending money to collect information for decision making. Many

public organizations are prohibited from diverting funds from service delivery to collect data on emerging trends in that service delivery. Even when information collection is possible, professionals are reluctant to take resources from service provision to collect such data.

- Public organizations have multiple goals, which can be vague, controversial, or both (Baker, 1969), (Bozeman, 1984). Goal ambiguity makes vital performance outcomes unclear for public sector organizations.

- The demands made by interest groups, flux in missions, and manipulation by important stakeholders and third parties create a complex and confusing set of expectations, which often conflict. Equity in dealing with clients and providing services is more important that efficiency in such organizations.

Although many of the reasons just introduced still hold in the Public Administration, the pressure for obtaining knowledge about the population (and having it almost in real time), the budget constraints, the need to offer better services with constrained resources have reduced the barriers. Furthermore the cost of the technologies necessary to implement a Data Warehouse / Business Intelligence project has dropped out significantly in the last years, making the development of such projects affordable by Public Administrations. The equity issue also hindered the adoption of Business Intelligence among the Public Administrations. Policy makers have been reluctant in providing optimized services to population subsets, being afraid of equity issues. Anyway, in last years civil servants and decision makers realized that optimizing the service delivery for specific population subsets is not a violation of equity, since resources spared by optimization can be used to provide customized services to the other population subsets.

4. Exploiting public Sector administrative Data: an Overview

According to (Surkyn, 2006), several advantages have raised interest in the use of administrative data as a source of statistical information by the national statistics institutes.

> Administrative data have become available on a growing number of subjects, and that the technical means for exploring them keep advancing. In most domains of administration paper documents have long been replaced with electronic ones, and information is transmitted using communication networks. Citizens, businesses and other administrations are offered the possibility of directly consulting or entering information to the administrations database. Sample based survey research is extremely costly, labor intensive and time consuming. Administrative censuses may be repeated more rapidly, reducing periodicity in census data. As a result of an integrated system of administrative functioning and statistical implementation, the full Finnish census dataset has been available for every single year since 1987. Exhaustiveness of administrative sources is also beneficial. In some cases the administrative universe even exceeds that of the classical census.

A marked tendency is reported towards re-using statistical data, notably administrative sources (Hoffmann, 1995), (Thomsen and Holmøy, 1998) (Buzzigoli, 2002). This, in turn, has sharply increased the demand for easy access to a variety of pre-existing data sources (Sundgren, 1993).

An attempt to use Administrative archives of the Public Administration as sources for aggregate analysis (e.g. population statistics) reveals errors and incompatibilities among each other that do not permit their usage as a statistical and decision support basis. These errors and incompatibilities are usually undetected during administrative use, since they do not affect their day-by-day use in the Public Administrations (Helfert and Herrmann, 2005); however they need to be fixed before performing any further aggregate analysis. The data sources need to undergo a quality improvement process before being used for analysis, see (TDQM, 2005) for further details on quality improvement. Data coming from different archives should also be integrated. The topic is very deep and extensive and has many connections with the archive integration activities described in Sec. 2.2. For further info, see (Cesarini et al., 2007), (Denk and Froeschl, 2000), (Hatzopoulos et al., 1998), (Papageorgiou et al., 2001).

The result of the cleansing and integration operations is an integrated and huge archive of data where the aggregate and statistical analysis can be performed. Such a huge amount of data (embracing different aspects of the same reality) is usually stored in a Data Warehouse which facilitates further analysis and the execution of queries aimed at computing synthetic information (Mezzanzanica et al., 2006).

The regular use of administrative data for statistical and decision support activities can provide useful information about the service assessment and about the target citizens in a very quick an inexpensive way. In Fig. 2 it is shown the time required to evaluate the effectiveness of a policy enactment with a Data Warehouse based system and with a traditional survey based process. (Cesarini and Mezzanzanica, 2007).

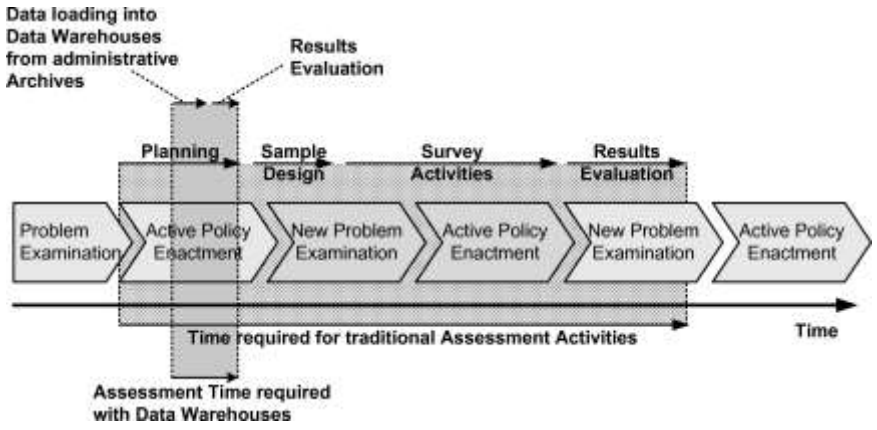

Fig. 2: Time comparison between a Data Warehouse based and a traditional survey based evaluation system.

Using information gathered from administrative archives to evaluate a policy enactment, allows to obtain feed-backs very soon and continuously, thus policies can be improved while they are enacted.

5. Not reinventing the Wheel

Last section highlighted the difficulties of carrying out a Business Intelligence projects on Public Administration archives due to the novelty and to the fact that large archives are involved. On the contrary, in the private sector there is a large community of practice and the development of a project in a completely new sector seldom occurs. Carrying out a project of large dimensions and for the first time in a specific sector will exacerbate the knowledge transfer issues among technical and business users. This is due to the fact that team components have to compensate huge knowledge gaps when they have different background and they start working on new fields.

Focusing on Public Administrations, the main advantage of agile and of iterative methods (described in Sec. 2.3) is that they facilitate knowledge exchange among business and technical users, which is a key issue in Data Warehouse / Business Intelligence development projects. However iterative (and agile) methods have a not-so-trivial drawback: they increase resource consumptions and make resource allocation not (easily) predictable. For large and innovative projects like the development of Business Intelligence over Public Administration archives, the resource allocation is not a trivial task. Resource termination before the end of a Business Intelligence project is very dangerous since BI projects are on/off projects: a BI project can be considered successfully ended when the Data Warehouse has been built and populated with high quality data, analytics have been built and accepted by the business users, business users have started using the system, and requiring new features. Should one of the elements not be ready at the end of the project (i.e. when the allocated resources will end), then the project has high chances to fail.

Budget consumption and resource constraints should be carefully evaluated since they are a sensible topic in modern Public Administrations. From this point of view, the innovative aspects - of developing Business Intelligence on Public Administration archives - contributes to increase the project costs, since the lack of experience of the people involved will reflect on the project costs.

A solution may be found by observing that Public Administrations are composed by several sub-organizations, most of them share similarities. E.g. a Public Administration is composed of several municipalities, that operate in similar context (although different), with similar goals, constraints, and facing similar expectations from their citizens. An interesting case study (described in the next subsection) show how Public Administrations can exploit their similarities to reduce the effort necessary to build Business Intelligence systems.

5.1. Case Study

The Milan Municipality (a Municipality located in northern Italy) has developed a Data Warehouse / Business Intelligence Project where several administrative archives have been integrated into a Data Warehouse. Business Intelligence has been used to derive information useful for civil servants and decision makers to improve services and to promote active policies supporting the populations. The archives integrated are the Registry archive (which holds information about people age, family

composition, and place of living), the Tax and Income archive (which holds information about people annual income), and the Job archive[2]. These source archives are managed by different Public Administrations, which have signed an agreement to exchange their data. The Data Warehouse that has been built contains information about the population from different viewpoints (economic, social, and geographic) and provides the decision makers with a powerful instrument useful for making decisions on how to ameliorate services to the citizens. The development of the project consumed a lot of resources (the data cleansing and integration tasks required more resources than expected), the development of analysis models useful to understand the reality of the municipality required also a lot of resources, privacy concerns had to be addressed extensively. For further information see (Mezzanzanica and Zavanella, 2010). The project has been carried out in collaboration with the CRISP Research Center (CRISP, 2010) and the Department of Statistics (Dep. of Statistics, 2010) of University of Milan Bicocca.

The Como municipality[3] started a similar project some years later: a data Warehouse has been built integrating the Registry archive and the Job archive. Como didn't integrate the Tax and Income archive (like the project in Milan) but added some other source archives to the project: the Scholar registry and the public transport archive (Como DW 1, 2010). The two additional archives allowed the Como municipality to improve the plan of the public transportation service used to carry children from home to school. The project has been carried out in collaboration with the CRISP Research Center. Although the Como archive contents were different from the Milan ones, the development of the Como Data Warehouse and Business Intelligence layer required less efforts since most of the models (especially the cleansing models and the analysis models) could be reused. Furthermore, since the two project have a lot of common aspects, relationships have been established among the civil servants and decision makers of the two cities using the Data Warehouse projects. This turned into the building of an (informal) relationship network where people share knowledge. The development of models for studying the job market place has increased thanks to the collaboration of the civil servants of the two cities, which have detected each other issues, provided feed-backs, and collaborated to the discovery of solutions. Furthermore, the Como municipality drove the development of additional features (e.g. the deployment of data on a Geographical Information System) which hace been shared with the Milan municipality[4]. The network is nowadays a community of practitioners who helped the Public Administrations involved to successful develop further Business Intelligence projects and sparing resources.

The creation of community of practitioners, or communities devoted to knowledge exchange among civil servants and public decision makers, as showed in the two case studies described above, seems to be a pregnant way of fostering the spread of Business Intelligence solutions using administrative archives. This could address the

[2] In Italy, firms and Public Administrations must notify the "employment state agency (centri per l'impiego)" every time a person is hired, fired, and when its contract changes, e.g. a fixed term contract is turned into an unlimited time contract. The archive of the employment state agency contains detailed information about a citizen current employment status and employment history.

[3] Como is located about 50 km away from Milan.

[4] The components have been developed using open source or copy-left technologies, which have fostered reuse.

knowledge sharing issues arising when Data Warehouse and Business Intelligence projects are to be carried out. The network would also favor reuse of cleansed archives (an archive could be used by several projects, sparing on the data cleansing archives), of software developed ad hoc, of analytic, economic, and social models. Furthermore, these communities foster the diffusion, among Public Administrations and civil servants, of a decision making culture based on data analysis. Further investigation on the topic is needed to understand which factors facilitate knowledge exchange in the scenario just introduced.

6. Conclusions and Future Work

This paper has presented the use of Business Intelligence (and Data Warehouse) tools and methodologies to obtain information over a population, starting from the contents of Public Administration (administrative) archives. Data Warehousing and Business Intelligence practices are well developed in the private sector. The public sector has started only recently to exploit such methodologies to analyze the contents of its archives. The paper has illustrated the reasons justifying the lag between public and private sector. The paper has also provided an overview over the development methodologies created for the private sector. The specific issues arising when moving existing methodologies to the public sectors have been illustrated: namely the general lack of experience in creating Data Warehouse and Business Intelligence solutions for the public sectors, the difficulties of collecting requirements, knowledge exchange issues that occurs when civil servants and policy makers should work with analysts and ICT people, the need of developing new social and economic models for analyzing data. Two case studies have been presented showing that the identified public-administration-specific-issues have been mitigated by the creation of a network of practitioners who share knowledge, feed-backs, and even software, and who have actively collaborated in developing Business Intelligence solutions, reducing the development time and effort. As a future work it will be further investigated the impact of networks of practitioners on the development of Data Warehouse and Business Intelligence solutions over administrative archives, looking for the factors facilitating the sharing of knowledge among ICT people and final users like civil servants and public decision makers.

References

Azoff, M.; Charlesworth, I. (2004): The New Business Intelligence. A European Perspective.

Baker, R. (1969): Organizational theory in the public sector. Journal of Management Studies, 6(1).

Beatham, S.; Anumba, C.; Thorpe, T.; Hedges, I. (2004): KPIs: a critical appraisal of their use in construction. Benchmarking: An International Journal, 11(1).

Boehnlein, M.; Ulbrich vom Ende, A. (2000): Business process oriented development of data warehouse structures. In Proceedings of Data Warehousing 2000. Physica Verlag.

Bozeman, B. (1984): Dimensions of "publicness": An approach to public organization theory. New directions in public administration, 46.

Buzzigoli, L. (2002): The new role of statistics in local public administration. In Proceedings of the conference quantitative methods in economics (multiple criteria decision making xi.

Casado, E. (2004): Expanding Business Intelligence Power with System Dynamics. In Raisinghani, M., editor, Business Intelligence in the digital Economy: Opportunities, Limitations and Risks. Idea Group Publishing.

Cesarini, M.; Mezzanzanica, M. (2007): E-government as decision support system to improve public services provision. In Proceedings of the European Conference on e-Government, Den Haag, The Netherlands.

Cesarini, M.; Mezzanzanica, M.; Fugini, M. (2007): Analysis-sensitive conversion of administrative data into statistical information systems. Journal of Cases on Information Technology, 9(4).

Chaudhuri, S.; Dayal, U. (1997): An overview of data warehousing and OLAP technology. ACM Sigmod record, 26(1).

Como DW 1 (2010): Available at http://www.crisp-org.it/?action=categoria&ID=27.

Conboy, K. (2009): Agility from first principles: Reconstructing the concept of agility in information systems development. Info. Sys. Research, 20(3).

Cottrill, K. (1998): Turning competitive intelligence into business knowledge. Journal of Business Strategy, 19(4).

Courbon, J.-C. (1996): Implementing systems for supporting management decisions: concepts, methods and experiences, chapter User-centered DSS design and implementation. Chapman & Hall, Ltd., London, UK, UK.

Courbon, J.-C.; Grajew, J.; Tolovi, J. (1978): Design and implementation of interactive decision support systems: An evolutive approach. Unpublished Manuscript. Institute d'Administration des Enterprises, Grenoble, France.

CRISP (2010): Crisp research center. www.crisp-org.it.

Denk, M.; Froeschl, K. (2000): The IDARESA data mediation architecture for statistical aggregates. Research in Official Statistics, 3(1).

Dep. of Statistics (2010): Department of statistics, university of Milan Bicocca. www.statistica.unimib.it.

Dunleavy, P.; Margetts, H.; Bastow, S.; and Tinkler, J. (2005): New public management is dead-long live digital-era governance. Journal of Public Administration Research and Theory, 16(3).

Edwards, K. (1998): SAS Rapid Warehousing Methodology. SAS Institute, White Paper.

Few, S. (2006): Information dashboard design: the effective visual communication of data. O'Reilly Media, Inc.

Fowler, M.; Highsmith, J. (2001): The agile manifesto. Software Development, 9(8).

Fuld, L. (1994): The new competitor intelligence: the complete resource for finding, analyzing, and using information about your competitors. John Wiley & Sons New York.

Gilad, B.; Gilad, T. (1985): A systems approach to business intelligence. Business Horizons, 28(5).

Golfarelli, M.; Maio, D.; Rizzi, S. (1998): The dimensional fact model: a conceptual model for data warehouses. International Journal of Cooperative Information Systems, 7(2).

Golfarelli, M.; Rizzi, S.; Cella, I. (2004): Beyond data warehousing: what's next in business intelligence? In Proceedings of the 7th ACM international workshop on Data warehousing and OLAP. ACM.

Greening, D. R. (2010): Enterprise scrum: Scaling scrum to the executive level. In Hawaii International Conference on System Sciences, Los Alamitos, CA, USA. IEEE Computer Society.

Han, J.; Kamber, M. (2006): Data mining: concepts and techniques. Morgan Kaufmann.

Hatzopoulos, M.; Karali, I.; and Viglas, E. (1998): Attacking diversity in NSIS' storage infrastructure: The ADDSIA approach. In International seminar on new techniques and technologies in statistics, Sorrento.

Helfert, M.; Herrmann, C. (2005): Introducing data-quality management in data warehousing. In Wang, R. Y., Pierce, E. M., Madnick, S. E., and Fisher, C. W., editors, Information quality. ME Sharpe.

Hoffmann, E. (1995): We must use administrative data for official statistics–but how should we use them? Statistical Journal of the United Nations ECE, 12.

Hughes, R. (2008): Agile Data Warehousing: Delivering World-Class Business Intelligence Systems Using Scrum and XP. Iuniverse Inc.

Inmon, W. (2009): Building the data warehouse. John Wiley.

Kahaner, L. (1996): Competitive intelligence: how to gather, analyze, and use information to move your business to the top. Simon &Schuster, New York.

Kaplan, R.; Norton, D. (1996): Balanced scorecard. Harvard Business School Press.

Kimball, R.; Caserta, J. (2004): The Data Warehouse ETL Toolkit. Wiley Publishing, Inc.

Kimball, R.; Ross, M. (2002): The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling. John Wiley & Sons, Inc., New York, NY, USA.

Kimball, R.; Ross, M.; Thornthwaite, W.; Mundy, J.; Becker, B. (2008): The Data Warehouse Lifecycle Toolkit, Second Edition, chapter 1. Wiley Publishing, Inc.

List, B.; Bruckner, R.; Machaczek, K.; Schiefer, J. (2002): A comparison of data warehouse development methodologies case study of the process warehouse. In Database and Expert Systems Applications. Springer.

Lonnqvist, A.; Pirttimaki, V. (2006): The measurement of business intelligence. Information Systems Management, 23(1).

Mattison, R. (1996): Data warehousing. McGraw-Hill.

Mezzanzanica, M.; Mariani, P.; Zavanella, B. M. (2006): Statistical information systems and data warehouses for job marketplaces. In Proceedings of the XLIII Conference of the Italian Statistic Society, Turin, Italy.

Mezzanzanica, M.; Zavanella, B. M., editors (2010): I numeri della città: un quadro socio-economico del comune di Milano sulla base di fonti amministrative. Franco Angeli. Written in Italian, English translation of the title: The Numbers of the City: a socio-economic Landscape of Milan on the Basis of administrative Sources.

Moss, L.; Atre, S. (2003): Business intelligence roadmap: the complete project lifecycle for decision-support applications. Addison-Wesley Professional.

Nutt, P. (2006): Comparing public and private sector decision-making practices. Journal of Public Administration Research and Theory, 16(2).

Papageorgiou, H.; Pentaris, F.; Theodorou, E.; Vardaki, M.; Petrakos, M. (2001): A statistical metadata model for simultaneous manipulation of both data and metadata. Journal of Intelligent Information Systems, 17(2).

Polimeni, R.; Fabozzi, F.; Adelberg, A. (1981): Cost accounting. McGraw-Hill.

Reinschmidt, J.; Francoise, A. (2000): Business intelligence certification guide. IBM International Technical Support Organisation.

Rockart, J. (1979): Chief executive define their own data needs. Harward business review, 57(2).

Sage, A. P. (1991): Decision Support Systems Engineering. Wiley-Interscience, New York, NY, USA.

Sen, A.; Sinha, A. P. (2005): A comparison of data warehousing methodologies. Commun. ACM, 48(3).

Skalen, P. (2004): New public management reform and the construction of organizational identities. International Journal of Public Sector Management.

Sundgren, B. (1993): Making statistical data more available. International Statistical Review, 64(1).

Surkyn, J. (2006): Different census systems in Europe: lessons for the transition to a register-based census system in Belgium. Available at http://www.ucl.eu/cps/ucl/doc/demo/documents/Surkyn.pdf.

TDQM (2005): the mit total data quality management program. http://web.mit.edu/tdqm/www/index.shtml.

Thomsen, I.; Holmøy, A. (1998): Combining data from surveys and administrative record systems. The Norwegian experience. International Statistical Review, 66(2).

Vibert, C. (2004): Competitive Intelligence: A framework for web-based analysis and decision making. Thomson/South-Western.

Authors:

Mario, Mezzanzanica, Prof.

University of Milan Bicocca, Italy

Department of Statistics

8, via Bicocca degli Arcimboldi, 20126, Milan, Italy

mario.mezzanzanica@unimib.it

Mirko, Cesarini, Dr.,

University of Milan Bicocca, Italy

Department of Statistics

8, via Bicocca degli Arcimboldi, 20126, Milan, Italy

mirko.cesarini@unimib.it

Roberto, Boselli, Dr.,

University of Milan Bicocca, Italy

Department of Statistics

8, via Bicocca degli Arcimboldi, 20126, Milan, Italy

roberto.boselli@unimib.it