World Journal of Social Science Research
ISSN 2375-9747 (Print) ISSN 2332-5534 (Online)
Vol. 3, No. 2, 2016
www.scholink.org/ojs/index.php/wjssr

Big Data Analytics National Educational System Monitoring and Decision Making

Catia Cantini[1*], Claudia Chellini[1] & Maria-Teresa Sagri[1]

[1] National Institute of Documentation, Innovation and Educational Research (INDIRE), Via M. Buonarroti, Florence, Italy

[*] Catia Cantini, E-mail: c.cantini@indire.it

Abstract

This paper reviews the applications of big data in supporting monitoring and decision making in the National Educational System. It describes different types of monitoring methodologies and explores the opportunities, challenges and benefits of incorporating big data applications in order to study the National Educational System. This approach allows to analyze schools as entities, which included in a local context with specific social, economic, and cultural development features. In addition, the paper attempts to identify the prerequisites that support the implementation of data analysis in the national educational system. This review reveals that there are several opportunities for using big data (structured and unstructured information) in the educational system, in order to improve strategic multidimensional knowledge for decision making and developing educational policies; however, there are still many issues and challenges that need to be addressed so as to achieve a better use of this technology.

Keywords

Big Data, semantic metadata, Open Data, interoperability, automatic word-processing, Linked Open Data, monitoring, decision-making, mixed methods

1. Introduction

Much of the information produced or collected by state and local administrations, as well as by other public agencies for their own institutional ends, can offer interesting and relevant databases for developing a more accurate study of a number of socially significant phenomena. The polyhedral nature of this information is of particular strategic importance when it comes to contributing to the construction of a more comprehensive understanding of any system, such as is the case with the educational system, which may be connected to a wide range of socio-economic and cultural dimensions. Offering a multidimensional analysis is then functional to a broader observation of such a phenomenon. It also aids the individuation of any social factors, which may play a determinant, or at any rate highly influential role in shaping phenomena, which take place in the field of education. Furthermore, such a multidimensional analysis can prove instructive in view of coming up with

concrete proposals for policy intervention, as well as for supporting and promoting more innovative and more carefully calibrated local formation networks.

New technological achievements of ICT may represent a strategic instrument to enhance and increase the possibilities for analysis and management of information in Public sector decision support (Sivarajaha et al., 2016). There is a clear coincidence between the exponential increment in the production of Big Data that we have been witnessing over the past few years, and the heightening of the digital world's attention towards the semantic inter-operability of different sets of information, which has been promoted by the Linked Data (LD) initiative, which has proposed to make huge amounts of data available, interconnected, and identifiable, through the use of Uniform Resource Identifiers. This coincidence makes it possible for us today to experiment with the development of new methods of analysis of social phenomena which take into higher consideration the complexity of relations between factors that are seemingly far removed from one another.

The technological model offered by EIF—European Interoperability Framework (Note 1) have led to the creation of instruments which allow for the aggregation of the datasets produced by each administrative entity. These datasets can then be freely combined, with the aim of reducing, the duplication of information and of valorizing already, held information by enriching our datasets with the information contained in those assembled and published independently by others and be they public agencies or third parties. Furthermore, many applications developed in the field of Natural Language Processing (NLP) allow to extend information assets. The NLP tools allow the use of unstructured data automatically documents extracting information that would otherwise remain buried under a sea of words.

The interaction between these different methodological approaches allows us to widen the information spectrum and monitor a system's state and transformation in a manner that is more complete and better suited to the needs of both policy-makers and agents within the system itself, thus creating more efficient services and innovative applications even out of datasets generated for completely different ends.

2. Data from the National Educational System: From Quantity to Value

In terms of size, speed, and variety, the volume of data produced and collected by ICT, has over the past few years moved from mainly representing an important business opportunity, to also gaining a determinant role in supporting processes of decision making (Provost & Fawcett, 2013). However, the data governance both managerially and analytically, of these immense sources of information, constitutes a necessary precondition for extracting from this potential treasure trove the value that one is looking for, and for contributing to the development of reliable studies and innovative applications.

Alongside this phenomenon, and complementarily to it, we have witnessed the development of the Linked Data (LD) paradigm, an initiative which proposes the formation of well-structured datasets that

are interconnected, semantically inter-operable, and distributed across different repositories, both inside and outside individual organizations. The use of Linked Data for ends such as the acquisition, enrichment, disambiguation and publication of data, allows for the reduction of data integration costs and for the improvement of the quality of the data themselves (Note 2).

LD are now considered the highest form of publication for a data producer's resources and digital content. They thus constitute a valid option for the governance of data. For this reason as well, a strong international campaign of sensitization to the diffusion of linked parameters in the sector of public administration has resulted in the deployment of the paradigm in the context of public data, in the form of the so-called Linked Open Data (LOD) (Note 3).

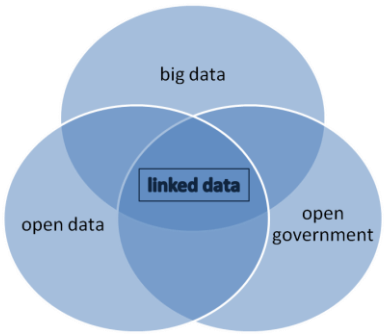

Figure 1. Relationship between Open, Government and Linked Data

Public administrations collect, organize, and manage vast quantities of public data; however, over the years, the autonomy of individual administrations, and the modalities of management of such data, have contributed to the creation of information islands, with little systemic vision. An isolated field of information holds a rather limited usefulness, whilst the combination of more sets of information can generate a notable added value within institutional praxis, developing a pool of datasets that are intercomparable comparable between their and can be mutually integrated. According to this perspective, public administration data, which had until now only held a role functional to the pursuit of institutional duties, now takes on—through the processes of data analysis—a whole different value in terms of stimulating the innovation of the entire social context, a crucial stepping-stone in any innovative process of e-government (Martelli & Bellini, 2013). In order to bolster this process of information sharing, National law, in keeping with the general EU orientation, promotes national policies which aim towards making public data assets more valuable (Note 4). Italian law, that is, points the country's administrations towards a process of production and publication of public data in a standardized format which is inter-operable at a national scale. It thus supports state organizations in their capacity to publish their information sources in common digital formats which make them easily reusable.

Scientific and academic research, following this methodological trend, is striving to demonstrate how

managing, analyzing and valorizing as best as possible the data generated by the national educational system, and which can be found across a number of institutional datasets and databases, constitutes an interesting opportunity for the development of an internal governance of ministerial information policies (Song & Ma, 2008). At the same time, this can offer the technical and methodological requirements for enriching already held information with what is being published in other contexts where the same technological paradigm in information management is followed.

The principal aim of an enterprise of this sort is actually twofold. On the one hand, the latter advocates an advanced monitoring of the educational system as an instrument for supporting the institutional decision-making process, contexts. On the other hand, it aims at giving schools a picture of their activities which takes into account the complexity of both their actions and of the contexts in which they act, allowing for a more careful reflection on their work and so as to aid the bettering of the services they offer the community.

From the methodological point of view, precisely so as to widen the information horizon, this project intends to focus on the integration and reuse of structured and unstructured information. Whilst the management of structured data and of business intelligence processes (data integration, data warehousing, reporting, data visualization) has matured considerably over the past decade, the management of non-structured data and of the their analytics has been lagging considerably behind, with the result that the majority of non-structured data are not used in decision-making processes. The documentary part of the educational system constitutes an unexplored source from which to draw information useful for integrated analyses, which may monitor the main socio-cultural changes taking place in the educational system, and support key systemic action. The increment in experimentation on the use of semantic technologies (Natural Language Processing) for the analysis of non-structured information and their integration with conventional statistical analysis can produce a better quantitative and qualitative interpretation of data and thus of the analysis itself. A synergic analysis of these two different sources of information (structured and non-structured data) allows for the investigation of certain aspects of the educational system which crisscrosses what is declared and put forward by the decision-maker (top-down process) with what is perceived and experienced by agents within the system (bottom-up process). This aspect is in fact particularly strategic in promoting a systemic analysis of a sociological phenomenon, in order to guide institutional reforms of such a phenomenon, as well as their consolidation in various social contexts. It is in fact often the case that an individual context's normative framework may be show a discrepancy between policy programs and the limits and necessities of lived experience.

In this sense, scientific and academic research, by proposing to experiment with an integrated analysis of statistical and documentary data, has as its objective the development and experimentation of a research methodology aiming at innovating the monitoring of the educational system through the use of mixed methods in a qualitative perspective.

3. "System-Monitoring" in the Education Sector

The fundamental objective of "system-monitoring" is to observe and represent a system as it evolves over time, so as be able to provide both interpretations and reading tracks of its configurations and dynamics, and directions for its improvement. Monitoring was first developed in the technical-industrial context, with the initial aim of checking on the correct performance of production chain activity, and later also of measuring business performance (Masoni, 1996, 2008). Today, monitoring has also found its proper place in the educational sector. In this context its objectives have broadened, spacing from measuring the efficiency of educational programs, to observing the changes brought on the school system by those same programs, to the support and promotion of improvements in the educational service. Initially employing the instruments and methods pertaining to quantitative research, monitoring has made available to policy makers and schools a number of ample and detailed representations of phenomena which have emerged over time in the wake of reforms of the educational system. At the beginning of the 1990s, monitoring acquired a heightened importance in Europe: the EU investment programs charged monitoring with the fundamental role of systematically controlling and verifying the carrying out of the investment programs themselves (Commission Européenne (CE), 1999). Over time, and with the succession of the various program cycles (EU, 2000, 2006, 2014), monitoring has tightened its focuses, indicating what data should be considered indispensable and defining adequate output and outcome indicators for the evaluation of the efficacy and efficiency of programmatic action (Iannacci et al., 2009). In this context, monitoring is to be understood as a form of system action, aiming at the structural improvement of governance capacities and policy execution (Ministero dello Sviluppo Economico, 2007), though it is still seen as mostly as a secondary instrument with respect to evaluating policies sanctioned by political decision-makers (Kusek & Rist, 2004; Mazzeo Rinaldi, 2012).

During the same years in which European attention to monitoring was heightened, Italy saw the diffusion at a local and national level of "system-monitoring" practices, experimented with in various forms. The development and rapid evolution of digital technology of the web furnished precious instruments to, on the one side, increment the number of subjects reached by monitoring practices, as well as the amount and kind of surveyed information, and on the other, to devise new monitoring models which would allow to survey process data "just in time". Monitoring, in its role as a support action for the construction or transformation of the educational system in its entirety or just in some of its facets, emerged in this way much strengthened. At the same time, monitoring began to acquire the status of an autonomous object of methodological study in its own right, although we have not yet sufficiently explored the full terrain of its functions and potentialities both for the policy maker invested in the process of decision making, and for subjects who are instead invested in the process of improving and innovating its practices.

Two main kinds of "system-monitoring" have been carried out in Italy over the past decade: one which

we will call descriptive, and another which we will instead call process-oriented.

The first kind of monitoring periodically takes a snapshot of a given phenomenon and thus offers the decision maker both a synthetic picture of the overall situation at a national level, and an understanding of the detailed configuration of the phenomenon at a number of local levels (geographical macro-areas, regions, provinces) down to even a single school. In this sense, monitoring results constitute also an instrument for supporting schools, which are enabled to carry out diachronic comparisons of their own activities with other schools' experiences. Though it is not specifically oriented to an in-depth analysis of specific aspects of the observed phenomenon, and it surveys part of the data in an aggregate fashion, descriptive monitoring has an important economic advantage: it allows for the use of limited resources in the acquisition of information that is necessary to gain a general understanding of a phenomenon and of its evolution, with the further aim of supporting the programming of activity on the part of the policy maker. In this kind of monitoring, data collection activities take place ex post, that is, after the conclusion of the formative process, which begins with the schools' yearly planning of their programs.

Process-oriented monitoring, instead, has as its aim the documentation of a whole cycle of a formative process, from planning, to execution, to its end results, collecting data punctually and in a disaggregated fashion. In this kind of monitoring, data collection activities are executed in the context of the carrying out of the educational programs, following each phase of the process. In this sense, process-oriented monitoring offers plenty of potential. The results of observation, both as it progresses and at its conclusion, serve to support the decision maker at different times during the programming and re-programming of his or her policies. Programming and re-programming can thus take place not only at the end point of a process, but also while the latter is ongoing. Furthermore, the instruments that are employed in process-oriented monitoring configure themselves simultaneously as both information surveying instruments and as operational instruments which support schools in managing even highly complex activities. An example of this kind of system monitoring is employed in Italy for the documentation and management of the educational programs promoted by the National Operational Program "Education" which is financed by European Structural Funds.

As techniques and methods are strengthened over time, together with a heightened consideration in scientific circles of qualitative research as a means to inquire into reality that is as reliable as quantitative research (Patton, 1997), the potentialities of monitoring have expanded further and the potential informational pool has extended to also include non-structured information, offering the possibility of conducting in-depth analysis on semantic dimensions which were otherwise difficult to comprehend. In this context as well, technological development is at an important stage, and is furnishing resources which enable us to imagine new qualitative methodologies which we may employ also in system monitoring, where in-depth analysis cannot but have a total value as well as the aim of aiding the development and innovation of the entire system. The integration of methods, techniques and instruments of quantitative and qualitative research that is typical of mixed methods (Caracelli et al.,

1989; Teddlie & Tashakkori, 2009; Creswell, 2011; Driscoll et al., 2007) has broadened the horizon of the possibilities for system monitoring. In Italy, over the past few years, there has been a heightened interest in mixed methods not just in practice but also in methodological study, whilst on the other hand there are as of yet still no system monitoring processes centered on mixed methods, which are at the heart of the model that we intend to build and put into execution.

The research presented here thus locates itself within a very fundamental educational domain about the Alternance. Since the 2006, the Public Education Ministry has been monitoring the Alternance paths with a quantitative approach to measure the extent of phenomenon in the several areas of the country. The quantitative approach does not allow to study the processes and dynamics involved in certain phenomena, this methodology does not allow to understanding the interconnections and to develop a more thorough overview, which instead is useful for to propose new strategies for system improvement. In order to describe the complexity of the phenomenon of Alternance training, the research aims to test a new monitoring, which look at the phenomenon in a wider perspective and according to different point of views. The Alternance paths proposed should be seen in the social, economic, cultural and business in which the educational project is contextualized.

It concentrates on the development and experimentation of a multi-dimensional analysis within a systemic view based on mixed methods. It argues, that is, for the integrated use of quantitative and qualitative methods within system monitoring, so as to relate with each other the various dimensions that can determine or have an effect on a phenomenon, with the aim of constructing an ampler vision of the totality of the phenomenon itself. The multi-dimensional analysis of a phenomenon compares large quantities of heterogeneous data, held in different databases. Accessing the datum in and of itself can thus constitute the starting point for the quantification and concrete study of the role played by certain socio-cultural, political, and economic factors. As already noted, the development of this new method of analysis is made possible by the coming together of the enhancement process of their informational assets carried out over the past few years by public institutions and of the technological research which has produced new digital resources and tools. Multidimensional analysis plays a strategic role for the world of research, offering unusual scopes for inquiry and allowing for the coupling of traditional monitoring methods with an approach more closely based on the integration of quantitative and qualitative methods derived from a systemic use of digital technologies in elaborating public data. This kind of methodology makes it possible to aggregate different information sources that are available throughout the web, and to make explicit the links and meanings implicitly held in—or in some cases outright absent from—web resources, such as data, pages, programs, etc. This is done on the basis of systemic relations, which allow to crisscross and combine these resources independently of their format and, most importantly, of their producers. This method thus appears particularly interesting with an eye to reconstructing the reticular structure of a given relevant phenomenon in the most complex and comprehensive manner possible.

4. The School as a System and as a *Scenario* of Social Phenomena

As has already been said, the objective of the present research is the development and experimentation of a new model of "system-monitoring" applied to the Education Sector, with an eye especially to observing the existing relationship between schooling and the workplace. To fully comprehend its potential, it may be useful to highlight the importance of the existing relation between the following three aspects: monitoring as a surveying tool, the phenomenon to be monitored and the reference system within which that phenomenon is inscribed. In this sense, it is necessary to highlight how in research epistemology the success of a monitoring process depends first of all on its capacity to furnish a correct and significant representation of the observed phenomenon. Consequently, the more complex a phenomenon, the more difficult it will be to represent it. Furthermore, system monitoring, given its specific aims, carries an extra danger: the broader and more differentiated the reference system, the harder it will be to give a correct and significant representation of the phenomenon which remains valid within a systemic logic. This is also due to the difficulty in identifying the system's main internal mechanism, as well as the possible links with other external systems. In this respect, it is useful to recall the polysemy inherent in the concept of "system": a complex assemblage of diverse, more or less structured elements which are not necessarily interdependent but are at any rate interconnect and interact both with one another and with the external environment, within an organic and functionally unitary whole. From this point of view, the educational system represents a highly fertile terrain for inquiry, but, at the same time, it is also extremely complex and multi-faceted, because it presents us with a plurality of internal articulations, as well as strong connections with the main dimensions of the country's social life: it offers deeply interesting perspectives for inquiry, not at all limited to dimension of schooling in the strict sense of the term, but instead branching out to its interconnections with many other dimensions of society. The school is the institution which transmits to the young generations the fundamental culture and knowledge of a civilization, preparing them at the same time for their entry into the workplace. Its aim is, therefore, that of educating and training. But there is more to schooling than this. As educational agencies spread across the country at a grass roots level (Note 5), schools are one of the main spaces of organized socializing for young people: at school, the latter meet and engage with one another, they develop their identity and manifest their aspirations, with an eye also to a future life project. They express their personal life experience on the planes of learning, relationships, behavior, psychology, and emotion. Thus, schools also constitute important sociocultural subjects, as the space where important phenomena emerge and develop, in connection of course with the educational dimension of learning, but also with individual variables from outside of school, such as for example a young person's psychological and character traits as well as his or her family and cultural context. In this sense schooling is both a system and a site of highly relevant social phenomena, such as youth maladjustment and unrest, bullying, youth unemployment, child labor, NEET (Note 6), etc.

The Italian Education System currently includes a number of institutional subjects, both national and

regional, as well as agencies, other bodies and social partners. The administration and governance of the educational and formative offer are, as the case may be, either the exclusive prerogative of the State or Regions, or it takes place with the cooperation of the two different subjects (Note 7), excepting the autonomy of single schools. As is clear from this initial discussion, the Italian Education System features an intricate intertwinement of different levels of governance: it is founded on the principle of integration—both with the structures and agencies which constitute its internal articulation, and with the structures and agencies belonging to different systems.

With regard to its internal articulation, it presents an organization based on at least four institutional levels: administration at central level, administration at regional level, administration at local level and finally the administrations belonging to individual schools.

At central level, the Ministry of Education, Universities and Research (MIUR) (Note 8), which is in charge of determining policy and is responsible for general administration at national level, is organized into three central Departments (Note 9) headquartered in Rome and, at a decentralized level, operates through peripheral administrations called "Regional School Offices" (USR) (Note 10). The USR, which are branches of the central level located in every first principal cities of the Country, oversee "observance of general provisions for education and minimum performance requirements, the effectiveness of training action and observance of standards" (Italian Unit of Eurydice, 2014, p. 9). Every USR is equipped with the most "Local Territories". Furthermore, the MIUR has several bodies and agencies operating at national level (Note 11).

At regional level, "the Regions (Note 12) have a joint legislative role along with State on issues relates to education; conversely, they are solely responsible for planning, management and provision of vocational education and training through recognized institutions: through the State/Regions Conference, the Regions work closely with the Ministry of Education and the Ministry of Labour, which define the minimum national standards for the education system and the vocational education and training system. Education and training issues within the jurisdiction of the Regions are handled by specific education offices" (Note 13) (Italian Unit of Eurydice, 2014, p. 11).

At local level, the administration includes Provinces (Note 14), Metropolitan cities and Municipalities, which have responsibilities in different areas and at different levels of the education system: "Provinces are assigned specific functions for upper secondary education only. Municipalities, often representing small residential communities and restricted areas, are distributed through Italy and have their own or regionally or provincially delegated responsibilities for functions and services relating to pre-primary, primary and lower secondary schools" (Italian Unit of Eurydice, 2014, p. 11).

Finally, at institutional level, individual schools, following the devolution introduced into the system in 1999 (Note 15), retain a degree of didactical, organizational, and managerial autonomy: each school draws up its own educational offer plan, the basic document setting out its cultural and social identity.

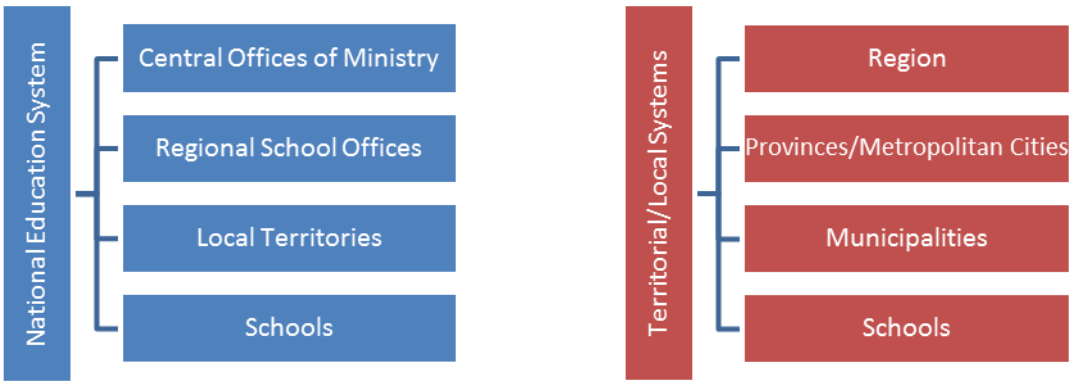

Figure 2. Administration and Governance of the Italian Education System

With regard instead to the connections with the "outside world", the Italian Educational System is called upon to interact with a number of different dimensions. These in turn change depending on which segment we consider: first or second cycle, tertiary education, adult education and training. Thus, we can speak of a particularly complex and structured system, which entails a number of forms of integration: between the different institutional levels of the mainframe State apparatus (the MIUR); with any other public administration apparatuses and agencies (such as other ministries, regions, other local administrations, welfare and insurance agencies, etc.); between the different systemic actors operating at a grass roots level (schools, enterprises, training agencies, universities, trade unions, representation associations, volunteering associations, tertiary sector associations, public services (health, transport, etc.); with entities specific to the various local contexts, each carrying its own unique training needs and capabilities for production, also in terms of local development policies. In this sense, then, an individual school located at the local level is configured simultaneously as both an element belonging to a broader system (the national educational system), and as an autonomous system in its own right. It is therefore called upon to engage with a multitude of different dimensions which exert an influence on the educational service that a school can make available. Therefore, a school will also be in dialogue with subjects belonging to different systems: the production and services system, the economic system, the political system, the cultural system, the arts and show-business systems, the media system, local administration systems, and so on.

Therefore, the individual school is both a stand-alone system, an attribute of the national education system and an attribute of the local systems; moreover, its aims views, it also interacts with other systems and areas of contemporary society.

Through its multidimensional perspective, this research intends to analyze the processes of change within the school system in relation to simultaneous processes of social and cultural change in the current historical period.

5. The Case Study: The Sector of Alternance Training

5.1 General Outline

The European Union has for a long time promoted the strengthening of the relationship between schooling and the workplace through a vast array of recommendations and documents aiming at broadening the possibilities young people have of coming into contact with professional, civic and cultural life of our society. Even just the past few years have seen a number of important documents deal with this question: *Key competences for a changing world* (2009), *Council Recommendation "Youth on the move"—promoting the learning mobility of young people* (2011) and the *Joint progress report of the Council and the Commission on the implementation of the "Education and Training 2010 work program"* (2010/C 117/01), the latter clearly asserting that "Learners should be given more and better opportunities to gain practical experience and insight into professional, civic and cultural life". Working upward from this argument, the *Education & Training 2020 program* (ET, 2020), furnishing the strategic framework for European cooperation in the Vocational Education and Training sectors, takes on the objective of promoting creativity, innovation, and entrepreneurship. It is in this perspective that the program itself valorizes partnerships between the entrepreneurial world and formation institutions, encouraging the creation of wider teaching communities, which may also include representatives of civil societies and other interested parties. According to the EU education policies, we ought to give more importance to workplace learning, apprenticeships, and to volunteering programs, not only in the official sphere of Vocational Education and Training and in adult education, but also in upper secondary schools and higher education.

In Italy, this perspective has rekindled a long-standing debate, which over time has dominated discussions within the country's political, educational, cultural and socioeconomic institutions, testifying to an often impetuous confrontation which has been taking place since the 1990s. At the time, the debate centered on what strategies to adopt to promote Technical and Vocational Education, on the integration between the fields of education and vocational training, on how to link secondary and tertiary education, on adult education, on how to promote a culture of integrating local systems which would be functional to improving the quality of the educational service and, at the same time, promote young people's access to the workplace. A new idea was taking shape: that of guaranteeing to young people and adults alike the opportunity of learning throughout their lives through a vast array of training channels alternative or complementary to the traditional educational system, adopted thanks to the spread of new scientific theories (Gardner, 1983; Kolb, 1984; Lave & Wenger, 1991). Recently, the law (Note 16) has drawn again on the main points of that movement, giving new life to the reflection on the possibility of overcoming the monads of schooling and entrepreneurial know-hows, and moving instead towards a more advantageous coming together of the two. This has also been made possible by the circulation of solutions at the international level such as the German dual system. Although the latter does not have a direct equivalent in the Italian Education System, such a model, which in Italy is

often associated with apprenticeship contracts, has contributed to the revitalization of the debate on the theme of linking school and workplace. In Italy, existing models of connection between school and workplace offer a wide range of opportunities, the most typical characteristic of which is to be found in variety, in difference understood as an added value which carries the possibility of enriching the public.

In this setting, suggestions and interventions aiming at improving the quality of the educational offer and at favoring youth employment have gained a newly central position. A wider and well rooted educational model is being problematized, as it is perhaps outmoded because it has little to do with reality any more. The aim now is to locate schooling within social phenomena and changes. Among the many possibilities for an integration between school and work, the present research will concentrate on the higher secondary segment of education and especially on the theme of "Alternance training" (Balboni et al., 2003-2004). The latter is an innovative didactic methodology which, as the name suggests, is constituted by an alternation of instances of classroom learning with instances of on-the-job learning (which may take place also through business simulations).

Inspired by the concept of "learning by doing" (Dewey, 1916), Alternance training was introduced in Italy in 2003, as an acceptable method for following second level secondary educational programs (Note 17), and some years later was outlined in its fundamental principles (Note 18). Alternance training has found widespread application in the Italian Education System especially following legislation for the reorganization of the upper secondary schools happened in 2010 (Note 19): its value has been recently reaffirmed by the law (Note 20), in order to promote its consolidation and spread across the entirety of the Country, making it compulsive in all kinds of upper secondary schools (general, technical and vocational institutes). Its main aims are: reconnecting schooling with the workplace, promoting knowledge, broadening the processes, moments and environment and spaces of learning, so as to adjust them to the requirements of today's society. Individual schools, within the space of the autonomy allowed for by the educational system, have accrued a plurality of experiences, sometimes expressing a level of excellence, and expressing "from the grass roots" multifarious tendencies and instances often very different from each other. As is suggested by the proliferation of schools that have taken up its methodology, Alternance training has grown constantly in terms of the paths it has activated (especially in the North and Centre of Italy) and in terms of the number of students who have decided to follow these paths, thus showing their appreciation for this didactic methodology. However, despite this positive trend, phenomena such as dropping out, NEETs, youth unemployment or underage labor are taking on dimensions in Italy that experts do not hesitate to define as worrying. These phenomena invite our society to deeply rethink the meaning of schooling and the necessity of renewing our didactic methodologies, overcoming outdated models of how to transmit knowledge (Bruner, 1996). Thanks also to the spread of new technologies, over the past few years our society has undergone a number of radical transformations: it will be difficult for the schooling system to fulfill its mandate if it does not quicken its pace in drawing nearer to the educational needs of our

time (Bauman, 2012). Therefore, following pan-European orientation switch point towards promoting the connection between school and workplace, it is fitting to interrogate ourselves on what effects the method of Alternance training has produced. In this respect, there are a number of questions to which it is necessary to give an answer. What is meant by "competence" within the school, and what instead at the workplace? What competencies are held to be fundamental at school and at the workplace? What competencies does the educational system intend to build up in young people looking forward to their professional futures? What competencies are looked for at the workplace in a freshly graduated young person? What divergences and similarities can we find between the two perspectives? What is the best way of integrating competencies acquired in Alternance training within schools' current curricula? What is the cultural identity of the Alternance training educational paths within general, technical and vocational education in Italy? What is the impact of Alternance training methods on students in motivational terms? The experiences favor the emergence of their personal vocations? Are these teaching methodologies congenial to the learning processes necessary to developing the competencies required by each educational profile?

The complex nature of this theme is evident. The relationship between education and workplace is positioned in between two very different systems: the educational system and the production and services system, in which a number of different institutional subjects converge across a range of levels. Furthermore, the pulverization of sector-specific languages within different reference contexts has led to the creation of a terminological babel, whereby the same words are used with different meanings. All this risks hindering the articulation of a dialogue, which is indispensable, and halts the realization of Alternance training's full educational potential. Its weak point seems in fact to be an insufficient sharing of meanings. Much depends on this dialogue: the organizational and choices carried out jointly by schools and businesses, the efficacy of young people's experiences, the modalities of recognition for achieved outcomes, and so on—in short, the efficacy of the Alternance training method itself.

In this perspective, the world of scientific research can offer an important knowledge contribution in support of policy makers and governance agencies, by stimulating the debate within the scientific community and the development of forms of synergy and joint action with stakeholders, so as to overcome current critical points and to valorize schools' virtuous practices with respect to this didactic methodology—which has been after all imagined precisely to "systematically connect classroom formation with practical experience" and to "realize an organic connection of educational institutions with the workplace and with civil society, so as to allow for an active participation of these subjects […] within formative processes" (Note 21).

Therefore, the present case study will be dedicated to analyzing the effects of Alternance training educational school-workplace paths, especially with respect to youth employment and successful training. Through the information retrieved in the different institutional dataset we will try to take a photograph of the school, to focus on the strengths, limitations, complexity of the institution. This

photograph will be linked with other information available in open formats, by different contexts such as that professional, economic, cultural, in order to monitor how the Alternance's paths activated by the school fit into the cultural and social and labor dimension.

It will do this also with the aim of tracing the processes of development, innovation and improvement of the Italian schooling, in keeping with the general aims of the national educational policies: to promote a culture of integration between various local systems functional to improving the quality of the educational service, with the objective also of contrasting drop-outs; to promote the development of key competencies individuated at a European level; and to facilitate young people's transition from schooling to the workplace.

Our research will be conducted also through use of the rich knowledge base accrued by the National Institute of Documentation, Innovation and Educational Research (INDIRE), which has been charged by the MIUR with constructing and implementing an extraordinary array of informational instruments, including a vast number of databases and national level management platforms, analyzing and documenting the processes internal to the Italian educational system in support of innovation.

5.2 Linked Open Data: A Basis for Study and Research for Educational Sector System Monitoring

To monitor the effects of educational paths about of youth employment and training success it is necessary to provide a pool of information on which to base the design of analysis. Through data, the research will try to rebuild the educational, social, economic requirements, which establish an Alternance training project, to study how the same path set with national educational policies. How the project is a part of on a broader operational program promoted of the individual school? How the school invests on aspects of alternation? How the teachers are prepared to deal with the issue? How the path promotes a culture of integration between school, work and contrasting drop-outs?

Methodologically speaking, attempting to offer a new informational view in the sector of Alternance training means experimenting an array of virtuous practices, which are needed in order to collect and publish data, to connect up that data later on. The pivotal point of these practices is to be found between two main prerequisites. As we saw in paragraph 5.1, the technical prerequisite consists in publishing the data obtained by the public administration in a standard format which will automatically make the data machine readable (Note 22)—i.e., producing data in a format which is accessible, legible, interpretable, and most importantly of all, usable by a machine. This means producing data which is self-descriptive, the meaning of which must be explicitly defined with a string of code constituted by words and markings (both standard and semantic metadata) (Tim & Mark, 2000). However, in the context of public institutions, publishing data according to open parameters, offering the technical possibility of accessing the data so as to process them, represents only the first technical step in achieving a better usability of information on the web. The second prerequisite, this time methodological, consists in the creation of instruments, which may allow to infer knowledge from the aggregation and correlation of different datasets (Linked Open Data) (Note 23). The possibility of

aggregating different available information sources on the basis of the semantic relations between types of sources, and that of crisscrossing them and combining them independently of their format and most importantly of their producers, appears to be of particular strategic importance in order to reconstruct the reticular structure of the educational process as it is investigated in its most complex and intricate aspects.

The methodology suggested by our research is articulated in five distinct phases: 1) identifying and selecting the datasets; 2) cleaning of the institutional datasets; 3) analysis and analysis modelling; 4) enrichment through the addition of metadata; 5) interlinking.

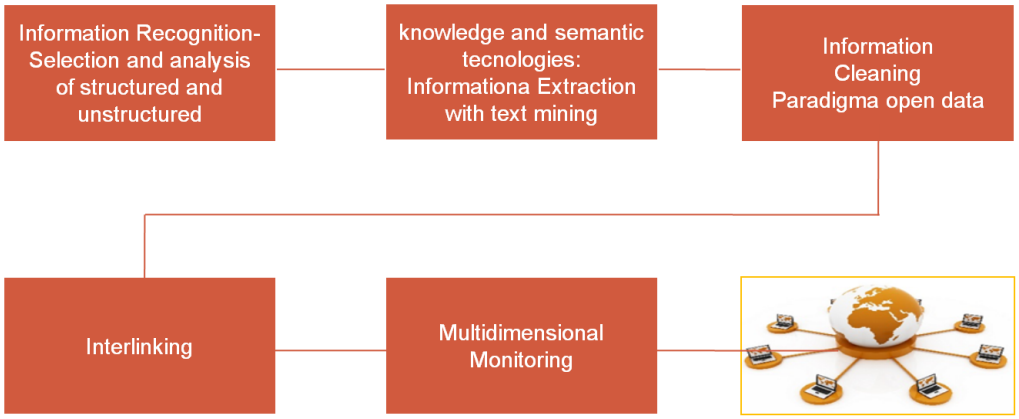

Figure 3. The Methodological Semantic Interoperability through LOD

5.2.1 Identifying the Datasets

If we consider the fact that the possibility of accessing a datum in and of itself represents the starting point for the development of a multidimensional analysis, we see how it is necessary to proceed first of all with a "data census", so as to identify the repository loci in which the data are stored. On the one hand, then, we have institutional datasets from available dominions, and on the other heterogeneous datasets built up by different local socio-economic agents. Some examples of Italian national-level virtuous practices with regard to opening up data, which are well consolidated in the educational sector, include *Open Coesione* (Note 24) initiatives and *Scuola in Chiaro* (Note 25). These entities may represent an informational and methodological starting point for our research. Not all the National public dataset are published according to criteria of interoperability and then will be necessary cleaning the relevant datasets for the research.

5.2.2 Cleaning the Datasets and Information Extraction

Cleaning the datasets mainly consists in technically transforming the format of already structured datasets (Note 26), that is to say, in verifying the publication criteria of the institutional datasets which are held to be of interest to a given study, and, if necessary, transforming the available information into an interoperable format and structuring it so as to make it possible to process it automatically. However,

part of the relevant information may remain underlying in individual documents, and is thus not recognizable and usable. It will then be necessary to automatically extract these information contents directly from the texts themselves, and then restructure them into an interoperable format (Lodi et al., 2014).

More precisely, some data are produced in a structured format (i.e., kinds of schools and formational programs present on the national territory in a given academic year, number of students per school divided by gender, age, and class, and so on), whilst others (think for example of the administrative documentation needed to present a project proposal, or of the reports detailing local formational needs, or of competence certifications, etc.) must be extracted directly from the texts themselves, after processing the documents through an informatics aid (Note 27). The use of textual processing tools will been designed to detect and extracted, through a study of exploratory and comparative, information elements referrals from the Alternance domain of the relevant documents for this first phase of the reconstruction in question.

In particular, within the context of the project, NLP-based tools will be used with two main aims. First of all, we intend to promote the use of more refined and well consolidated techniques of automatic information extraction from textual documents as possible objects of quantitative and qualitative analysis to enhance the domain study, so as to construct a system monitoring model for the educational sector which will be able to take full advantage of the capabilities of current technology. Statistical analysis which has up until now been circumscribed to quantitative type structured data in fact limits the possibilities of gaining in-depth knowledge of studied phenomena. A lack of analyses conducted on semi-structured or non-structured textual information risks constituting a deficit of analysis then a severe limit both to the process of institutional decision making, and to self-reflection in the school sector. Secondly, we aim at creating a meta-dataset and automatic classification of the documentation produced from scholastic administration, so as to archive and to publish these data as well in interoperable formats and to widen the possibilities for multivariate analyses. The potential for inquiry afforded by these technologies can thus offer a key contribution to research activities.

5.2.3 Analysis Modelling for Developing a Data Warehouse and Business Intelligence System

This phase has two objectives: to develop an architecture which will allow to access the data through generalized data analytics tools, and to formalize the conceptual model to outline a coherent representation of the reference analysis frame. In order to interpret the data, it is in fact necessary to develop general relations between the data themselves in the shape of models built on the basis of dominion know-hows relative to the analyzed sector, the so-called “domain ontologies”. In this sense, it need to use technological infrastructure in which the concepts are uniquely identified and where software agents recognize these objects and implement associations and equivalences between them, by referring to ontologies, formal representations, shared and explicit specific knowledge domains. Ontologies are used to represent the entities through the description of their characteristics and the

identification of the relationships between them, and thus the semantics that binds these entities, mainly used to make categorizations and deductive reasoning (Guerrini & Possemato, 2013).

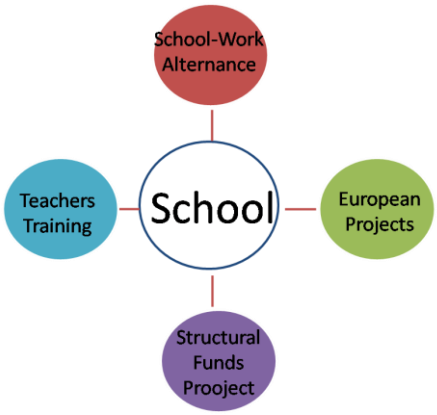

Figure 4. The Ontological Representation

This phase represents a logical and conceptual restructuring of the data: similarly to methods of software analysis and design, it is possible to say that a process of reengineering and refactoring of the informational base is activated.

5.2.4 Data Enrichment

In this phase, the previously cleaned and modelled data are enriched through by making explicit a range of background information which simplifies their reuse (metadating). This is done by adding an informative content which may be obtained also through the use of automatic information extraction or automatic reasoning. Just as with data, metadata are also expressed in RDF, a format which was born precisely as a standard for the semantic noting of webpages, and which then became the basic tool for the creation of Linked Data. Some of the most useful metadata in terms of data interoperability and reuse are constituted by information on data semantics (such as comments, tags, definitions, etc.) which serve to clarify the meaning of data and to facilitate their use, to describe the meaning of attributes, properties, and kinds of entities (Jain et al., 2011) (Note 28).

5.2.5 Interlinking

In this phase, local information content is connected to other information, which may be found in different datasets produced by the same administration, or by data collections available throughout the web. Information that is published in open standard can be compared and computed with other datasets, developed within the educational sector as well as in other ones (such as industries, universities, etc.), for instance: data set about the employment rate, the study sector, the main skills required by work's sector, the principal sector of business, or at any rate it can be publically reused and controlled. The information held in several datasets, both the institutional and open data released from other experiences, will be linked by a "domain ontology" (see 5.2.3).

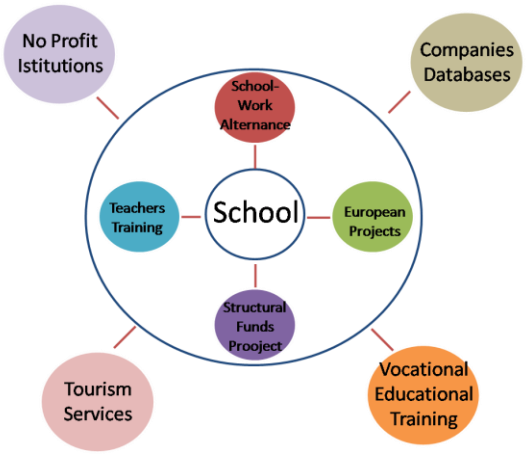

Figure 5. Interlinking Several Information

With this aim in mind it is necessary to develop common vocabularies with which to share the meaning of information that is related between individual repositories. In this way, Semantic interoperability is achieved, which represents the key instrument for the elaboration of information, including from external or secondary sources, without losing the real meaning of such information in the elaboration process. This holds a crucial role in the interchange of information between institutions, and as such constitutes the basis of any innovative e-government process (Bedini et al., 2013).

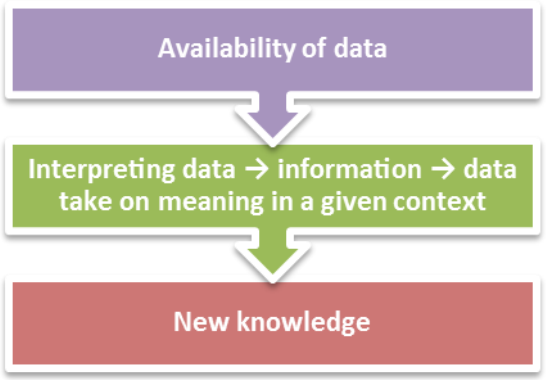

Figure 3. The Underlying Methodology of Alternance Training Monitoring

6. Conclusion

Having been born at a time when the paradigm of quantitative research seemed to be able to provide the answers needed to know any phenomenon, monitoring has found in the educational system a context of application not just with regard to observing its transformations, but also as an instrument to support and promote an improvement of the services offered by schools. The collection and management of a large quantity of data pertaining to public administrations, both State and local, has

given monitoring activities the possibility of utilizing a rich and well-articulated informational base. Over the years, however, the autonomy of individual administrations has brought them to build up many separate and heterogeneous data collection systems, weakening any systemic view of relevant phenomena. This has also happened in the educational sector, the complexity of which sees the presence of a number of diverse subjects, both public and private, pertaining to institutions or to the business or service worlds, and carrying heterogeneous competencies, but nonetheless working together to the building up of the Italian educational and formational system. The effects of government sponsored reforms and of the transformations which are taking place in the educational world can thus only be understood if we consider on the one hand the many factors which characterize a single phenomenon and on the other the relations which exist between many different phenomena. This entails the necessity of keeping together within a single analysis a large amount of data coming from different social entities. Integrating different information sources thus takes on great importance in getting to know the system. This knowledge can become broader and deeper and taken on a leading role in the decision making process relative to national educational. Given that schooling is a key part of an integral part of a country's social system, those elements which contribute to create certain effects and which are linked especially to the educational sector do not at any rate only concern specific political reforms, but also affect—and are affected by—other social contexts, such as economics, finance, and culture. The use of technologies for the management of structured and non-structured data configures itself as an expert support for the management and elaboration of the information contained in the several databases pertaining to various social contexts. The use of these technological paradigms is thus functional to the institutional necessity of having an accurate informative picture, which takes into account multiple factors even though they may seem mutually distant, so as to take decisions that are more aware better respondent to the needs of the context under consideration. On the other hand, with the aim of favoring the improvement of their own educational service, over the years, schools themselves have developed a heightened awareness of their need to, as it were, know themselves, so as to better reflect on their own work by seeing it within the broader social, economic and cultural context within which they act.

The research we have presented here aims to experiment, in the field of school-and-workplace Alternance training, with a methodology which may be able to enrich traditional quantitative and statistical analyses with a qualitative perspective. This methodology is centered on integrating data coming from different datasets and databases, both institutional and non-institutional, thanks to the possibilities afforded by semantic interoperability, by technological instruments finalized to relating structured data, and by specialist tools for the extraction of non-structured information from free-floating texts.

The information held in several datasets, both the institutional and open data released from other experiences, will be linked by an "ontology" and return a picture of the school in order to understand

the potential and limits of the organization with respect to the objectives of the Alternance training. Ultimately, with this research we intend to try out a new form of system monitoring which, thanks to the articulated informational load of potential analyses, can stand in as expert support for both the institutional decision making process, and for the self-reflection carried out by individual schools and oriented to improving their educational service.

References

Bahereini, K., & Elci, A. (2008). A New Software Architecture for J2EE enterprise environments via semantic access to web mining by distributed intelligent software agents. *IEEE international software and Applications conference*.

Balboni, C. A. (2003-2004). L'Alternanza scuola-lavoro. *Annali della Pubblica Istruzione, 5-6-2003, 1-2004*.

Barabási, A. L. (2002). *Linked: The New Science of Networks*. Cambridge: Perseus.

Bauman, Z. (2012). *Conversazioni sull'Educazione*. Trento, Italia: Erikson.

Bedini, I., Elser, B., & Velegrakis, Y. (2012-2013). The Trento Big Data Platform for Public Administration and Large Companies: Use cases and Opportunities. *Proceedings of the VLDB Endowment*, 1166-1167.

Berners-Lee, T., & Fischetti, M. (2000). *Weaving the web: The original design and ultimate destiny of the world wide web by its inventor*. New York: Harper Collins.

Bruner, J. (1996). *The culture of Education*. Cambridge, Massachusetts, London, England: Harvard University Press.

Caracelli, V. J., Green, J. C., & Graham, W. F. (1989). Toward a Conceptual Framework for Mixed Methods Evaluation Designs. *Educational Evaluation and Policy Analysis, 11*(3), 255-274.

Commission Européenne (CE), F. s. (1999). *évaluer le programmes socio-economiques* (Collection MEANS, ed.). Luxembourg: Office des publications officielles des Communautés européennes.

Creswell, J. W. (2011). *Controversies in Mixed Methods Research*. Thousand Oaks, CA: Sage Publications.

Dell'Orletta, F., & Montemagni, S. (2011). Towards an NLP-based approach for measuring syntactic complexity: Preliminary experiments with Italian texts from different registers. *Workshop on Cross-linguistic and language-internal variation in text and speech: Focus on the joint analysis of multiple characteristics, Freiburg Institute for Advanced Studies (FRIAS), University of Freiburg*.

Dewey, J. (1916). *Democracy and Education: An Introduction to the Philosophy of Education*. New York: The MacMillan Company.

Driscoll, D. L., Appiah-Yeboah, A., Salib, P., & Rupert, D. J. (2007). Merging Qualitative and Quantitative Data in Mixed Methods Research: How To and Why Not. Ecological and Environmental Anthropology, 3(1), 19-28.

EU. (2000). Indicators for Monitoring and Evaluation: An indicative methodology.

EU. (2006). *Indicative Guidelines on Evaluation Methods: Minitoring and Evaluation Indicators*.

EU. (2014). *Guidance Document on Monitoring and Evaluation—European Regional Development Fund and Cohesion Fund*.

Eurydice, I. U. (2014, October). The Italian Education system. *I Quaderni di Eurydice, 30*.

Gardner, H. (1983). *Frames of Mind: The Theory of Multiple Intelligences*. New York: Basic Books.

Guerrini, M., & Possemato, T. (2013). Linked data: A new alphabet for the semantic web. *Italian Journal of Library, Archive and Information Science, 4*(1), 23-32.

Iannacci, F., Cornford, T., Cordella, A., & Grillo, F. (2009). Evaluating Monitoring Systems in the European Social Fund Context: A Sociotechnical Approach. *Evaluation Review, 33*(5), 419-445.

Jain, A., Khan, I., & Verma, B. (2011). Secure and Intelligent Decision Making in Semantic Web Mining. *International Journal of Computer Applications, 15*(7), 14-18.

Kalampokis, E., Hausenblas, M., & Tarabanis, K. (2011). Combining Social and Government Open Data for Participatory Decision-Making. In E. Tambouris, A. Macintosh, & H. de Bruijn (Eds.), *Electronic participation* (Vol. 6847, pp. 36-47).

Kusek, J. Z., & Rist, R. C. (2004). *Ten Steps to a Results-based Monitoring and Evaluation System: A handbook for development pratictioners*. Washington DC: World Bank.

Lave, J., & Wenger, E. (1991). *Situated Learning. Legitimate peripheral Participation*. Cambridge: Cambridge University Press.

Lodi, G., Maccioni, A., Scannapieco, M., Scanu, M., & Tosco, L. (2014). *Publishing Official Classifications in Linked Open Data*. 2nd International Workshop on Semantic Statistics (SemStats 2014) in conjunction with ISWC 2014. Riva del Garda, Italy.

Martelli, C., & Bellini, E. (2013). Using Value Network Analysis to Support Data Driven Decision Making in Urban Planning. *2013 International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)* (pp. 998-1003).

Masoni, V. (1996). *Monitoraggio e valutazione dei progetti nelle organizzazioni pubbliche e private*. Milano, Italia: Franco Angeli.

Masoni, V. (2008). Monitoraggio: Valutazione di processo. Lezioni dall'esperienza. *Rassegna italiana di valutazione, XII*(40).

Mazzeo Rinaldi, F. (2012). *Il monitoraggio per la valutazione*. Milano: Franco Angeli.

Ministero dello Sviluppo Economico. (2007). *Quadro Strategico Nazionale (QSN) per la politica regionale di sviluppo 2007-2013*.

Patton, M. Q. (1997). The Paradigm Debate and Utilitarism Synthesis. In M. Q. Patton (Ed.), *Utilization Focused Evaluatio. The New Century Text* (3rd ed.). London: Sage Publications.

Presutti, V. et al. (2014). Uncovering the Semantics of Wikipedia Pagelinks. *Knowledge Engineering and Knowledge Management—19th International Conference, EKAW 2014* (pp. 413-428).

Linkoping, Sweden, Proceedins.

Provost, F., & Fawcett, T. (2013). Data Science and its Relationship ti Big Data and Data-Driven DEcision Making. *Big Data*, *1*(1), 51-59.

Sivarajaha, U. et al. (2016). The role of e-participation and open data in evidence-based policy decision making. *Local government, Journal of Organizational Computing and Electronic Commerce*, *26*(1-2), 64-79.

Song, C. X., & Ma, K. (2008). Applications of Data Mining in the Education resource based on XML. *International Conference on Advanced Computer Theory and Engineering*.

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. Thousand Oaks, CA: Sage.

Notes

Note 1. Retrieved from http://www.ec.europa.eu/idabc/en/document/2319/5644.html

Note 2. The data published using the Linked declination, i.e., using the RDF standard, promoted and kept up by W3C, for the publication of content that was up until now typically put forward in CSV and XML, are: highly "re-usable" by third parties, thus drastically reducing the costs of re-use activity; highly "inter-operable" with other datasets because they allow for the acquisition of added information published in other contexts by other agents; extremely "comprehensible" thanks to ontologies that are published alongside them and which make them both machine readable and intelligible to humans without mediation.

Note 3. The term "open data" is defined (according to the definition given by the Open Knowledge Foundation in the Open Definition project) "a content or datum is defined as open if anyone is able to use it, reuse it, and distribute it, subject only to a potential requirement of attribution and sharing in the same way".

Note 4. As required by article 68 of the C.A.D. and as advocated by the Italian National Digital Agency, the data format is defined as "open" when its semantics, syntax, operational context mode of use are made public through an exhaustive documentation. Such documentation, together with a format user guide, must be present in one or more issues by the producer of the standard. For example, a Microsoft Excel table (XLS format) is not open in the technical sense of the word, but is machine-readable, and in practical terms is infinitely more useful than a PDF, which is an open format from a certain point of view, but from which it can be extremely difficult to extract well-organized data. Some examples of excellent machine-readable (and open) formats include XML and CSV (Comma Separated Value).

Note 5. Italian Constitution lays down one basic principle: the State is obliged to provide a State-school system accessible to all young people. To study is a right-duty.

Note 6. "Not (engaged) in Education, Employment or Training".

Note 7. The State sets down general norms and determines the general principles of education; the

Regions and autonomous Provinces have legislative power with regard to job-oriented training; education is thus the object of concurrent legislation on the part of the State and Regions.

Note 8. Ministero dell'Istruzione, dell'Università e della Ricerca.

Note 9. Education, University, Research.

Note 10. Uffici Scolastici Regionali.

Note 11. Suchas: The High Council of Public Education (*Consiglio Superiore della Pubblica Istruzione, CSPI*), the National Institute of Documentation, Innovation and Educational Research (*Istituto Nazionale di Documentazione e Innovazione della Ricerca Educativa*, INDIRE), the National Institute for the Evaluation of the Education System (*Istituto Nazionale per la Valutazione del Sistema di Istruzione*, INVALSI), the Institute for the Development of Professional Training for Workers (*Istituto per lo Sviluppo della Formazione Professionale dei Lavoratori*, ISFOL), the National Council of University Students (*Consiglio Nazionale degli Studenti Universitari*, CNSU), the National Agency for the Evaluation of the University and Research System (*Agenzia Nazionale per la Valutazione del Sistema Universitario e della Ricerca*, ANVUR), the Conference of Rectors of Italian Universities (*Conferenza dei Rettori delle Università Italiane*, CRUI).

Note 12. The Regions are one of the constituent elements of the Italian Republic: each one is a local authority with its own statutes, powers and functions according to the principles established by the Constitution of the Italian Republic. The Italian regions are twenty, including five with a special statute of autonomy (one of these, Trentino-Alto Adige, consists of only two "autonomous Provinces", i.e., with legislative powers similar to those of regions). Italian regions are: Abruzzo, Basilicata, Calabria, Campania, Emilia-Romagna, Friuli Venezia Giulia, Lazio, Liguria, Lombardy, Marche, Molise, Piedmont, Puglia, Sardinia, Sicily, Tuscany, Trentino-Alto Adige, Umbria, Aosta Valley, Veneto.

Note 13. *Assessorati*.

Note 14. Provincial organizations have been revised; thus, the system may change in the future.

Note 15. National Law No. 59—March 15, 1999.

Note 16. National Law No. 107—Jul. 13, 2015.

Note 17. National Law No. 53—March 28, 2003.

Note 18. Legislative Decree No. 77—Apr. 15, 2005.

Note 19. Presidential Decrees Nos. 87 and 88 and 89—March 15, 2010.

Note 20. National Law No. 107—Jul. 13, 2015.

Note 21. Legislative Decree No. 77—Apr. 15, 2005.

Note 22. For example, a Microsoft Excel table (XLS format) is not open in the technical sense of the term, but it is machine-readable, and, in a practical sense, is infinitely more useful than a PDF, which is an open format from a certain point of view, but from which it can be extremely difficult to extract well-organized data. Some excellent examples of machine-readable (and open) formats are XML and CSV (Comma Separated Value).

Note 23. The Italian Digital Agency "National Guidelines for the Promotion of the Information Heritage Public" (p. 24), 2014.

Note 24. Retrieved from http://www.opencoesione.gov.it/

Note 25. Retrieved from http://www.archivio.pubblica.istruzione.it/scuola_in_chiaro/open_data/index.html

Note 26. The structure of a datum is the framework in which data can be stored and organized, so as to make it possible to process them automatically.

Note 27. To access such text-based content it is necessary to use automatic text information extraction tools (Natural Language Processing) (Montemagni, 2011).

Note 28. Some of the main examples of semantic resources: SKOS Simple Knowledge Organization System (Retrieved from http://www.w3.org/2004/02/skos/); FOAF-Friend of a Friend an experimental linked information system (Retrieved from http://www.foaf-project.org/).