Working **Paper**

2016 • 3

UMR 201 D&S Développement et Sociétés

From UNESCO's descriptive statistics to deductive Big Data: the role of human annotation in quantification processes

Roser Cussó

UNIVERSITÉ PARIS 1 PANTHÉON SORBONNE

IRD Institut de Recherche pour le Développement FRANCE

From UNESCO's descriptive statistics to deductive Big Data: the role of human annotation in quantification processes

Roser Cussó (roser.cusso@univ-paris1.fr)
IEDES, UMR 201, Université Paris 1 Panthéon-Sorbonne

Abstract: Analysis of the characteristics and activities of UNESCO's statistical personnel indicates: (i) the intertwined nature of practices involved in data-production processes and of the knowledge this requires; (ii) the importance of human-annotation to production and maintenance of databases; (iii) the shift from descriptive statistics to statistical inference, all in the context of structured data. These findings help to define four recent quantification trends. As massive unstructured data takes over: (i) there is a greater reliance on machine learning and modeling; (ii) the use of supervised learning implies increasingly complex and diverse data annotation–which may modify the role-played by the social sciences; (iii) in unsupervised learning, based on non-annotated data, the role of statistical models is enhanced; (iv) in both cases inductive and deductive approaches may be of use. These trends are taken here to be represented by the expression "deductive quantification".

Keywords: big data; quantification stages; deductive quantification; UNESCO data production; human annotation; modeling

Résumé: L'analyse des caractéristiques et des activités du personnel statistique de l'UNESCO signale: (i) la nature articulée des pratiques impliquées dans les processus de production des données ainsi que des connaissances que ces pratiques nécessitent; (Ii) l'importance de l'annotation humaine pour la production et l'entretien des bases de données; (Iii) le passage de la statistique descriptive à l'inférence statistique, le tout dans le contexte de données structurées. Ces résultats aident à définir quatre tendances récentes dans le domaine de la quantification. Au fur et à mesure que les données massives et non structurées se développent: (i) s'accroit l'usage du *machine learning* et de la modélisation; (ii) l'utilisation de l'apprentissage supervisé, qui implique des annotations de données de plus en plus complexes et diverses, peut modifier le rôle joué par les sciences sociales; (iii) dans l'apprentissage non supervisé, basé sur des données non annotées, le rôle des modèles statistiques est renforcé; (iv) dans les deux cas, les approches inductive et déductive peuvent être utiles. Ces tendances sont ici considérées sous l'expression "quantification déductive".

Mots-clés: données massives; étapes de quantification; quantification déductive; production statistique de l'UNESCO; annotation humaine; modélisation

1-How, Why and Who of Quantification: the Rise of the Human Factor

The founder of the sociology of quantification in France, Alain Desrosières (2008), considered that research endeavoring to contribute to this field cannot study separately the manner data are produced (how), its purpose (why), and the actors and institutions producing and using it (who). These three aspects are intertwined in quantification processes, generating particular articulations, and thus allowing to define a specific domain of knowledge with its own case-studies, concepts and theoretical discussions and background.

The how-dimension of quantification concerns the methodological dimensions: questionnaires or other new supports, definitions/categories attached to data, calculations, indicators, graphical representation of data, analysis and model creation (descriptive, inferential, interdisciplinary, algorithms…). It is the stage of the "mise en equivalence" and the establishment of conventions (Desrosières & Thévenot, 1988), the more technical description of those aspects being relatively under-studied in the literature of the sociology of quantification (Armatte, 2001; Cussó, & D'Amico, 2005). The why-dimension is connected to the use given to data. Being the most manifestly political facet of quantification, it is more deeply analyzed. Several contributions can be underlined for motivating new typologies and interpretations of instruments and indicators (Lascoumes & Le Galès, 2005; Salais, 2003; Bruno, 2008). Finally, the who-dimension is also quite well-studied; the examination of the institutions generating quantification is crucial to understand both data features and use (Cussó, 2007a).By contrast, and as in the case of the study of technical dimension of "how", analysis of detailed practices of "human annotation" of data (interpretation and labeling, experts' characteristics…) is less developed, especially its role in structuring databases and thereby shaping analysis. We can expect that this gap will be reduced. The rise of big data has, somewhat surprisingly, intensified the importance of data annotation through the development of supervised machine learning and modeling (Aroyo & Welty, 2015) while amplifying the interest in minimizing it through research on unsupervised learning[1].

I claim that focusing on the more technical aspects of the *who-how* interaction is crucial to understanding and defining the different stages in quantification evolution. Rather than concentrating on the particular use of data, or trying to grasp the general political logic of quantification, I maintain here that focusing on the articulations engaging human annotation and data analysis & models may be the key to identifying both past and still current quantification features (of UNESCO's or other structured traditional datasets) as well as new quantification trends (massive unstructured datasets). My thesis is three-fold. First, big-data produces an uncontrollably large diversity of data-interpretations and information ontologies to analyze data, this because such meta-data are essential to supervised machine learning activities. Second, it follows that social sciences have indirectly played and are going on to play a preeminent role in close relation to statistics, mathematics and IT science. Three, the quest for merging of datasets and machine learning may be tentatively taken here to be represented by the expression "deductive quantification". In parallel to globalization, what I call the "global and concurrent annotation" also raises questions of cultural dominance and harmonization; of unequal technical capacities and resources; of public control or dependency[2].

This paper is organized as follows. After a general introduction covering UNESCO's personnel to illustrate the interrelated nature of data activities (section 2), we study the pre-reform quantification processes of UNESCO's statistical services (section 3). The reform between 1996 and 2001 clearly identified the importance of data interpretation and analysis choices (section 4). Pressure was exerted on descriptive statistics-oriented activities in order make them evolve towards statistical inference analyses (section 5). This reform helps raise original questions on the underlying quantification nature of big data (section 6), especially as regards the role of human annotation, the social sciences and statistical-mathematical models. Finally, the conclusions (section 7) focus on a discussion on what I call "deductive quantification".

As regards the fieldwork and the methodology underlying this paper, the analysis of UNESCO is a case study based on my professional experience in the organization's statistical services from 1993 to 2001. I participated in all aspects of data production: collecting, entering, treating, extracting-calculating and presenting international data on education. I participated in international conferences and prepared reports including data analysis. My exploration of big data derives from informal discussions with data scientists, and specialized articles and documentation. In general, I use the word

[1] "Unsupervised learning studies how systems can learn to represent particular input patterns in a way that reflects the statistical structure of the overall collection of input patterns. By contrast with SUPERVISED LEARNING or REINFORCEMENT LEARNING, there are no explicit target outputs or environmental evaluations associated with each input; rather the unsupervised learner brings to bear prior biases as to what aspects of the structure of the input should be captured in the output" (Dayan, 1999).

"k-means" clustering is an example of an unsupervised learning method, popular for cluster analysis in datamining. k-means clusters by iteratively partitioning n observations into k (far fewer, k<<n) clusters in which each observation belongs to the cluster with the nearest mean, which thus serves as a prototype of the cluster. This results in a partitioning of the data space into cells cut out by bisectors between these means.

[2] Strong criticism arises among mathematicians and data experts as "current models" are seen by some of them to "exacerbate inequality and endanger democracy".

http://mathbabe.org/2014/01/13/im-writing-a-book-called-weapons-of-math-destruction/

See also: http://www.amazon.com/Models-Behaving-Badly-Confusing-Illusion-Reality-Disaster/dp/1439164991 “data” to refer to figures and digital information and the term “statistics” or “statistical” to refer to the analytical activity.

2- Some previous considerations on UNESCO’s “statisticians”: intertwined and interdisciplinary activities

The hierarchy of professional functions is in principle very clear in international organizations (IO), as in most public institutions. Personnel status is defined in job-descriptions that establish the category, grade and salary of positions according to assigned activities. But international personnel are subject to other distinctions, beyond formal categories: nationality, level of education, experience, specific competences, “real” responsibilities, diversity of generations and work-traditions... Taking also into account the specificities of international missions, especially as regards quantification production, these aspects are likely to develop a complex articulation and/or overlapping of functions and practices.

UNESCO’s staff is divided into three main categories: general services, GS, (seven grades); the professional personnel, P, (five grades) and the directors, D, (two grades). At the top of the hierarchy are found the Adjoint Director General (ADG) and the DG. The GS category includes a wide variety of activities, from IT to carpentry, while directors and professionals are, in principle, program specialists (or support for program specialists) committed to the implementation of UNESCO’s missions, and including the management of the personnel. This contrasts with the fact that many professionals do not supervise any personnel, while almost all professionals conduct some kind of secretarial work, particularly with the widespread use of computers and the organization of work by project. Few professionals have a personal secretary. Correspondence, which is largely performed digitally, is often managed by each individual independently, for example[3]. That is why the trend is to evaluate professional positions according to the degree of responsibility, and its program-related management dimension, rather than the actual content of the post. Thus complex skills, knowledge and intermediate decision may be needed for GS positions, avoiding any explicit official decision-making.

A job description for a “statistician” of professional status may thus contain a number of sentences and key words underlying their responsibilities[4]: “Contribute to the implementation of the UNESCO Institute for Statistics (UIS)[5] work program”; “taking over-all responsibility for…”, “supervising and reviewing the work of and proposals of other team members”, “Maintain liaison with advisers or experts in education in international organisms, regional and sub-regional agencies and associations”. As regards the more technical part, we find: “assisting in design and development of the UIS database (in collaboration with programming and IT staff)” or “the extraction, analysis and evaluation of data and the preparation of tables and other outputs of statistical information for inclusion in reports and publications”. Those last points are significant since, as for the division of tasks between P and GS, the division of tasks between statisticians, IT staff and “report-writers” was not completely possible to establish in practice.

The terms “statistician” and “assistant statistician” covered in fact different profiles and functions which all contributed to the quantification process. IT staff, professionals or GS, could not be reduced to support personnel to the statisticians. Inversely, those commonly called “statisticians” did not have activities limited to the practice of statistical sciences, which was, as we shall see, very little developed in the pre-reform period. They contributed also to IT management and development, and to the organization of activities. The mixed functioning and functions of personnel is not so surprising if we consider that quantification activities involve the whole process of “mise en chiffre”. Logically, the personnel needed to achieve them must have different intertwined skills and knowledge. From collecting the data to producing the final tables, the micro-activities were diverse, as analyzed in the next section.

---

[3] Another element that is both the cause and the effect of the blurring of UNESCO’s hierarchy is the reduction in the proportion of GS in relation to professionals. Having been about two thirds of all employees in the 1980s, GS are now only a half of the total. The staff numbered about 3,500 people in 1980 compared to 2200, thirty-five years later.

[4] From my own job description in the UIS in 2001.

[5] The former UNESCO’s Statistical Office became the Division of Statistics in 1991 and the Division officially became the UIS in 1999.In October 1996, there were 32 members of the personnel; 17 professionals and 15 GS. The existence of a growing number of non permanent staff is also to be noted. During the Division of statistics reform the number of temporary staff rose up to nine persons.

3- The pre-reform process of quantification at UNESCO: comparative data, descriptive statistics

Table 1 bellow presents a schema on the pre-reform data production process, i.e. from the 1950s to the end of the 1990s. Itomits the first step, the decision on what had to be quantified, i.e. the general UNESCO-objectives to be measured. The origin of international missions and their link with international data programs have been studied elsewhere (Cussó, 2010;2012). The international decision-making has been defined as three interacting spheres of power: intergovernmental, international and transnational. For UNESCO's personnel, their (political) power to interpret and/or reinterpret the missions was not explicit before the reform. While the main mission of the Organization was the right to education, indicators on school-enrolment by grade and their evolution were to be calculated and published. The intermediate decisions (e.g. why this indicator?) were not discussed.

Table 1- UNESCO's pre-reform quantification process

Definition	Support	Collection	Treatment	Calculations	Extractions	Presentations	Analyses	Publication
GERand other indicators	Questionnaire	From countries already trained	Entering in DB	DB instruments	DB instruments	Excel, Word…	Global consistency	Figures and texts
Formula	Manuals	Correspondence	Data check	Fortran – SQL	Fortran - SQL	Graphs, tables	Understandable	Reports, Website
General interpretation	Workshops		Correction verifications	Data check	Data check	Indicators	International comparability	Political dimension

As already evoked, very little statistical analysis was done before the reform. The true core of the activity was to check data as regards international standards, to decide on their reliability & consistency, and to present them in an intelligible way. Only basic calculations were performed as part of these activities while indicators published could be considered to be "descriptive statistics" illustrating the education situation, and thus the policies, of Member States (MS)[6]. The Gross enrolment ratio (GER) is a percentage of enrolled children related to school-age population and the pupil-teacher-ratio is a mean of the number of children per teacher. The school-life expectancy is similar to the demographic life-expectancy. Nevertheless, no inferential statistics (based on probability theory) were applied at UNESCO's Division of statistics, i.e. we did not draw conclusions from data using correlations, central limit theorems, conditional probabilities, or Baye's rule[7]. UNESCO produced estimates on enrolment, but they were essentially used to fill in the gaps for regional means. With few exceptions, UNESCO's thousands of variables were not combined in order to find significant links and thus to generalize them to monitor MS planning. This planning was the sovereign realm of each MS as long as they accepted the general principle of the right to education[8].

This did not mean that the will to influence MS was absent. While UNESCO's most important publication, the Statistical Year-Book, presented the countries by region and in alphabetical order, and did not contain any text or graphical representations of data, comparisons of countries and its implicit competing effect were nevertheless present. It was possible to see that some countries enrolled all their children and that others did not; that some countries spent more money on education than others… Moreover, MS accepted an international entity publishing their data all together, in standardized form, as regards the definition of education and expenditure for instance, and this fact

[6] UNESCO's statistical services also produced data on other fields such as culture or sciences. They are not analyzed here.

[7] Statistics science can be divided into two main fields: descriptive statistics and statistical inference. In the former, the main emphasis is placed on arithmetic and graphical distillation of data to help describe the existing situation. Statistical inference involves much more sophisticated notions concerned with general decision-making and prediction. The idea is to find general laws from particular cases and observations based on probabilistic methods.

[8] A similar trend in statistical analysis is observed in the World Bank demographic production of quantified knowledge between 1945 and 1980 followed by a deep institutional reform in the early 1980s (Cussó, 2001). OECDs reform is well documented by Papadopoulos (1994).

was a sign of the rise of an international quantification of its own, i.e. a particular configuration of power and knowledge.

Back to table 1, it is worth spending some time to explain the detailed activity in the Division of statistics. Until 1997, once the questionnaires from MS had been received, the data were entered by GS in a non interactive database (Fortran). Printed A3 listings permitted to check data consistency by hand, doing some pencil/calculator controls. When the Division direction was informed about an imminent reform, some technical improvements were planned. The creation of a new database in 1997 was outsourced to Cap Gemini (CG). Several meetings between UNESCO's and CG's experts were needed and the whole process implied the presence of a CG's team at UNESCO offices for several months. The first meeting was the very strategic moment of definition of the database structure. ACG's senior member received UNESCO's personnel demands and translated them into the architecture of the future database. Once the basic functioning was agreed on, a long work of codification started. Each variable received a 10 digit number (identifier)including each pertinent dimension (country, year, grade, public/private, age, sex, depending on the variable).

Instead of the team of GS introducing data from separate offices, every individual's computer could be connected to the database. This was a relatively important change in work habits. My first job at the statistical services consisted in completing the data on secondary pupils by field of study in technical and vocational programs. I had the opportunity to use the A3 listings on which annotations by hand had to be written down in order to be entered in the old database. Exported to Excel I calculated percentage distributions of pupils by field of study and by sex. The objective was to illustrate the gender parity or disparity situation by country. No statistical inferences on the variables "explaining" the field-specialization by gender were presented in the report[9].When the new database was operative, thanks to a connected Oracle server, I could directly work on the database from entering to extracting data. The IT activity became both less visible and more intertwined since it was not possible for "statisticians" to get some degree of autonomy without developing some IT capacities. It was the mark of the increasing integration of both dimensions. More workload was thus transferred to program specialists in charge of data production, giving rise to a sort of ancestor of "data scientists".

The main person responsible for updating data calculations was both a statistician and an IT specialist professional seconded by an IT professional. He supervised the intermediate and simultaneous calculation of all indicators and the updating of population and economic series, once a year. This procedure meant that any data entered later on would not be included in the current Year-Book. Table 2 summarizes the YBK activities focusing on responsibilities and functions. The more the data treatment and related decisions advanced, the more "statisticians" and IT personnel worked on activities of management, especially those who were professionals.

Table 2- UNESCO's Statistical Year-Book

Reception of questionnaires	Entering process	Writing to MS if necessary	Pre-Year-Book calculation	Extractions	Presentations	Analyses	Publication
First internal data check	Second data check: hand-made calculations or by the entry seizure (automatic intermediate calculations)	Modifying or confirming data from questionnaires		Verifying process, data consistency study	Tables	Global consistency	
Subject or area specialists & assistants	…as the process advances, increasingly working…						Managers

The pre-reform quantifying model and some recommended changes cohabited during the period of restructuring. Only once the personnel were completely replaced, by the end of 2001, could a new culture of international quantification prosper, essentially linked to a new way to analyze data in an inferential and probabilistic logic.

9 Division of statistics. 1994. *Notes statistiques STE-17: Enseignement technique et professionnel du second degré, la participation féminine dans les différents domaines d'études 1980 et 1992*, UNESCO, Paris.

4- UNESCO's statistical services reform: more formalized annotations, more normative analysis

The reform of UNESCO's Division of statistics was spurred by the criticism from other already reformed IO and their programs. The World Bank proposed, in 1983, a fund for the improvement of education research. Later on, a study –prepared by Unicef and the World Bank for the meeting of the Board on International Comparative Studies in Education (BICSE) in 1993– was to strongly question the quality of UNESCO statistics (Puryear, 1995). In J. M. Puryears' opinion, the Organization was not able to adapt to the demand for new statistical information, due, to a great extent, to the influence of certain countries which were reticent to compare the efficiency of their educational systems (Puryear, 1995: p89). In response to this criticism, the Director-General invited, in late 1994, the BICSE to prepare a plan of action aiming to improve the quality of international education statistics. The BICSE published a report entitled *Worldwide Education Statistics: Enhancing UNESCO's Role*. Published in 1995, this report was financed by UNESCO and the World Bank, as well as by the National Center for Education Statistics and the National Science Foundation (United States) (Guthrie and Hansen, 1995).

The recommendations of the BICSE report related to several fields: the mission of the program, institutional structure, data production activities… As regards the latter, it could be noted the proposal of "documenting the underlying data base" and that of "carrying out analytical activities"[10]. In contrast, when the BICSE recommended the development of externally financed projects, that was, to some extent, because they already existed at the Division of Statistics (i.e. UNICEF's financing for specific studies), as were programs aimed at developing the statistical capacity-building of the Member States, notably the program on National Education Statistical Information Systems (NESIS), financed since 1991 by the Swedish International Development Agency (SIDA).

Following the "documenting" recommendation, presented as a response to the need for modernized data processing methods, it was suggested, for instance, that clerks' knowledge on national data be stored using a formal comment-recording system instead of being recorded on paper. This implied the development of software to automatically identify statistical inconsistencies, the systematic keeping of written records on metadata, and the formalization of data collection procedures. The enhancement of computer equipment and the reinforcement of technical staff were also discussed. It was thus thought that the creation of a "documentation data base" to store the staff's knowledge could render reproducible the interpretative/subjective dimension of the human annotation then being practiced. An unawareness of the intertwined nature of data production is evident there.

The recommended "carrying out of analytical activities" was rather loosely defined in the beginning. Most of the BICSE report was devoted to defining with precision the expected changes in the orientation of the statistical program (a new mission) and the organization of the activities (new management, autonomy). After remarking that the statistical services had not benefited from American or British expertise (Guthrie and Hansen, 1995: p19), the authors stressed the need for taking into account the evolution of requests for information in the international context, requests relating to, in particular, the performance of education systems: "[...] modern-day statistics users are interested in more and more accurate measures of student performance" (Guthrie and Hansen, 1995: p35). Governments were no longer in control of national economic policy, and countries competed with each other to attract the investments of international firms: "Modern technology, both its existence and pursuit of its development, has contributed to the formation of massive amounts of private-sector investment capital, often (or even usually) outside the immediate control of governments." In this context "[...] there is [...] an intense reliance on human capital formation to sustain a nation's global competitive status and internal civic structure". "International comparative data [...] display the capacity of the other nations that may be trade and investment competitors." (Guthrie & Hansen, 1995: p33, 30 and 32). One of the most important aims of the report was to encourage change of policy orientation and insight of international education statistics for the sake of "the altered human capital needs of

[10] "establishing common definitions and data standards; regularly collecting and disseminating a core set of education statistics and indicators; maintaining and documenting the underlying data base; planning and coordinating a strategic research and development effort; carrying out analytical activities; and playing the role of catalyst in spurring the development of statistical capacity and systems in member states [...]" (Guthrie and Hansen, 1995: p5).

member states, growth in internationalism among private-sector companies, and the emergence of major third-party agencies concerned with social infrastructure planning and development throughout the world" (Guthrie and Hansen, 1995: p47).

The UIS creation was decided by UNESCO's General Conference in 1997. To guide the new mission and functions, preliminary or parallel changes had also to be made in management as well as in the make-up and profile of personnel. In order to put in motion the transformation of the Division of Statistics into the new Institute, a Steering Committee was created during the meeting of the UNESCO Executive Board in May 1998. The paper *UNESCO International Institute for Statistics-Report*, was released in 1998. The document was known as the Thompson report from the name of the PricewaterhouseCoopers consultant who wrote it (Thompson, 1998).

Submitted in 1999, a second report defined the organizational chart and the major positions. The permanent staff was to be reduced to twenty-six members. They were supposed to be seconded by temporary personnel and discharged of certain activities by subcontracting. The report also recommended that the Institute should operate under its own rules, applying, for instance, a system of flexible recruitment in which salaries are not fixed in advance, a system at odds with UNESCO's rules and regulations. A large number of recommendations were adopted by UNESCO's Executive Board of May 1998. Emphasis was put on respecting UNESCO's regulations, and the proposed personnel policy was drawn up without mention of any precise decision concerning the future of the former Division's staff members. The Steering Committee recommended the new Institute's Director at the end of 1998, the position taking effect at the beginning of 1999, prior to the decision for the transfer of the Institute outside UNESCO's headquarters.

5- A new quantification? Statistical inference still attached to structured data

As already noted, the reform included two important technical recommendations: the quest for formality as regards the diverse and multiple activities attached to data interpretation and production, and the introduction of the goal of showing which evidence-based political reforms should be undertaken by MS –what they called data-pertinence. A bit too vague at first, the BICSE's recommendation of "carrying out analytical activities" was rapidly redefined more precisely. The suggested harmonization with OECD activities and methods indicated what data analysis was to be expected.

It was not sufficient to check and calculate standardized primary education GERs or repetition rates and show them in a table; it was necessary to identify explanatory variables for increasing or decreasing GERs. The door was open for inferential statistics to direct MS policies instead of for descriptive statistics to inspire them. The MS's (relatively) independent educational decision-making was to transform, and adopt common best practices and evidence-based policies.

The Fast Track Initiative (FTI) of the World Bank is a good example of this evolution. The FTI presented benchmarks on education and on expenditure to monitor the policies in indebted countries (Cussó, 2007).The principle is simple: the means of several education indicators (expenditure as percentage of PIB or pupil-teacher ratios) for the countries with high educational attainment (reference indicator) become predictors of educational success. If indebted countries matched the benchmarks it is probable that they reached the goal of education for all. OECD's Program of international school assessment (PISA) is an example of more sophisticated use of statistical inferences. School autonomy, for instance, is to be applicable to all participating countries as it is positively correlated to high PISA scores. The proliferation of studies on the investment return rate of education can be included in this probabilistic-prediction trend as well.

Table 3 below summarizes the new quantification arrangements. As for World Bank and OECD data production and analysis, UNESCO's quantification process is hybrid. The former data production, which lacked inferential analysis, persisted since it was still necessary to have a consistent structured data framework to refer to. In parallel, following PISA, a survey-based analysis was developed. Other surveys such as SACMEQ or PASEC were created for non-PISA countries (Cussó, 2007b).

Table 3- UNESCO's quantification from descriptive to inference statistics

UNESCO's former data process	+ New surveys on the "quality" of education
Weak normativity	Strong normativity
Strong data consistency and comparability	Weaker data consistency and comparability
How many children are enrolled?	+ How to enroll more children expending the less? Which are the determinants of school scores? To what high female GER are correlated?
Traditional DB Descriptive data and analysis	+ SPSS or SAS Inferential analysis Econometric models
From YBK	… to EFA Reports

If the former UNESCOs program had a particular political orientation (a "why" dimension attached to both the development doctrine and some kind of MS sovereignty), the ensuing international quantification, relying on both inferential statistics and evidence-based prescriptions, seeks to become a globalizing reform. This reform is more or less integrated in the indicators and models underlying the data collection. Since school autonomy and possible related independent variables are measured, it is logical to eventually find correlations between them. As discussed in the following section, inference from statistics applied to structured data may identify relationships as previously defined but without explaining why.

This evolution is similar to that observed in state-based data and statistics. Descriptive and inferential analytics developed in parallel, the latter being progressively combined with the former in the sciences and in the public sphere (Graunt, 1662), especially through insurance (Lengwiler, 2009) and econometrics (Armatte & Desrosières), and culminating in the neo-Keynesian,1940-1970s, period (Bourdieu & Passeron, 1964). From the 1980s, when econometrics and microeconomics are unified, their logics and methods begin to take over public decision-making. Inferential statistics, as used in the management of public reform, need different types of data for different policy changes. In both of these periods, structured data remained the basis of both the descriptive and inferential statistical approaches, the latter presenting the problems of verification already cited[11].

Replacement of the personnel, still needing descriptive data

Before the reform, UNESCO's statistical services prepared data to be presented to the international community. It was an internationally standardized form of data, allowing general comparisons. MS were deemed to favor universal and/or increasing schooling of girls and boys in what MS agreed to call pre-primary, primary, secondary and higher education. Comparison attested to the existence of a global community sharing a number of rights and principles. The reform recommended the use of statistical inference as it was meant to establish and verify hypotheses on how the general shared goals should be reached as regards expenses or school contents, for instance. It was an intrusion in MS policies, that was difficult to accept for UNESCO's personnel. The latter had harmonized data to loosely indicate the direction towards school modernization while new recommended methods were clearly oriented to austerity and neoliberal policies. UNESCO's personnel then tried two strategies: to respond to quality criticism (Cussó, 2006) and to alert MS about the political consequences of the reform.

The move of the statistical services to Montreal in 2001 was key to their ultimate transformation. It entailed the almost complete replacement of its personnel. The capacity of harmonizing and deciding data has since been undertaken by the new staff. The production of basic data by the new Institute is still at the core of its activity while the production of inferential statistics seems to have been kept in a

[11] Bourdieu & Passeron's work was based on administrative and survey data, analyzed to show a correlation between the socio-professional category of pupils' parents and pupils' school-level attainment. The link between low school attainment and low socio-professional category was considered to be the effect of social reproduction instead of seeing it as a slowness of social mobility, which can only be fully tested by long time series and with a considerable number of variables.

quite stationary stage[12]. Nevertheless, the EFA Report created in 2002 has introduced several other sources and analyses of data. The report is now produced within UNESCO's HQ offices but in practice outsourced. A rapid overview of the EFA reports shows the inclusion of statistical-inference based analysis as for example in 2010 EFA report, chapter 2 (Figure 2.14: Poor and rural children have much less chance of going to school in Burkina Faso and Ethiopia; Figure 2.16: The relationship between enrolment and gender parity varies across countries).Only the internationally comparative data remains UNESCO's as is shown in the YBK-style annexes.

6- When Big Data emphasizes diversity in human annotation and statistical models

Somewhat surprisingly, recent discussions on supervised machine learning not only highlight subjectivity and diversity of human data annotations and treatment but also bring it to light as never before. By presenting a proposal for better human annotation directed at supervised learning, Aroyo and Welty (2015) give an interesting overview of the increasing importance of both data interpreters, and interpretation criteria, to making big-data analysis both reliable and useful. As in UNESCO's statistical services, where "statisticians" checked figures and indicators or added footnotes (meta-data) if further explanations were judged necessary, -- big-data managers need human annotations to label sentences, images or figures, and this in huge amounts of "raw" unstructured data. This labeling is not to be incorporated to the data for good. It may be continuously revised and several statistical models may coexist.

Source: Janowicz *et al.*, 2015.

Aroyo & Welty focus on annotations involving causality identification in medical sentences. These annotations can be done by experts (as in UNESCO) or thru crowd-sourcing techniques. According to the authors, if the goal is to get the "correct labeling" of each sentence, the exercise will always be unsatisfying and incomplete, especially for ambiguous sentences. We are thus forced to redefine our idea of truth to be probabilistic; a percentage of (or given number of) interpreters' answers. Such "crowd truths" can be integrated into algorithms and thus, enable machine learning, to analyze millions of documents. Table 4 below includes some examples of sentences analyzed in the article. Causality is not uniformly identified in some of the sentences, depending on the interpreters. Aroyo & Welty also point out the need for revising the established probability attached to annotations. The sentence on bin-Laden given in table 4 could be interpreted as a free of reference to "terrorism" during the war in Afghanistan against the Soviet Union, and as a "terrorism-related" sentence afterwards.

Table 4– Sentences interpretation

Causality	"Terrorism"
[GADOLINIUM AGENTS] used for patients with severe renal failure show signs of [NEPHROGENIC SYSTEMIC FIBROSIS].	[Osama bin Laden] used money from his own construction company to support the [Muhajadeen (*sic)*] in Afghanistan against Soviet forces.
[Antibiotics] are the first line treatment for indications of [TYPHUS].	

12 http://www.uis.unesco.org/Pages/default.aspx

With [Antibiotics] in short supply, DDT was used during World War II to control the insect vectors of [TYPHUS].	

Crowd-sourcing appears to be both a more "democratic" way of labeling and a relativization of truth even in the realm of the logical thinking, as in the case of causality. This need for human annotation in big data (images, gas consumption, transportation, twitter or map interpretations) and the change in labeling as questions, physical models, periods or cultures vary, opens up an unprecedented diversity of unstructured data treatments and subsequent analyses. Putting aside unsupervised learning, big data is the largest exercise of data codification ever undertaken. The potential role of social scientists in the analysis of this coding activity seems evident, as they have long worked on closely related issues in sociology, philosophy and history of sciences (Kuhn, 1962; Bachelard, 1975; Latour, 1979) as well as on the sociology of quantification (Desrosières, 2008).

Furthermore, one wonders whether this diversity in data interpretation is close to what one observes in social sciences as a discipline, when combined with the issue of diversity of information ontologies. There is no such thing as one social reality, but unlimited (scientific and founded) social interpretations and theories, and this regardless of whether the underlying vision is positivist or relativist. Data-production and data-analysis are thus linked in big data as they are in social knowledge.

"There is no such thing as raw data. Data is always created for a certain purpose, following work-flows and observation procedures, depends on the used sensors and technologies, comes with an intrinsic uncertainty, reflects the theories and viewpoints of the people that recorded the data, and so forth. To give a concrete example, the body position at which a blood pressure measure was taken matters for the interpretation of the results" (Janowicz et al., 2015). For the Semantic Web[13], for instance, Janowicz *et al.* consider that agreement on information ontologies is not required."Consequently, data publishers can define their own local ontologies, reuse existing ontologies, or combine multiple ontologies. Strictly speaking, the semantic web stack does not impose any restrictions on the use and quality of ontologies. For instance, one can describe data using only selected predicates from different ontologies and add additional, even inconsistent, axioms. While this would not be considered good practice and will prevent the usage of certain (reasoning) capabilities of semantic web technologies, basic queries will still be possible" (Janowicz *et al.*, 2015)[14].

The World Bank's document *Big Data for action in development* contains several examples of big data analysis, results and limits[15]. Let's focus on the "Tracking FoodPrice Inflationusing Twitter Data" project, which is part of the UN Global Pulse program[16]. Datasets used are diverse: unstructured and structured. The Global Pulse lab in charge of the study used over 100,000 Tweets generated between March 2011 and April 2013 in Jakarta. This unstructured dataset was "temporally referenced, spatially referenced by region and identifiable by Twitter account, and at times, by person". This data was complemented by structured public datasets regarding food and fuel prices including datasets typically generated through questionnaires and surveys. Soybean inflation data from the USA was also collected from the World Bank.

The objective was to "investigate the possibility of utilizing social media data to give an indication of social and/or economic conditions", and this by examining "the relationship between food and fuel prices, Twitter posts and the corresponding changes in official price index measures". Following the

13 The Semantic Web is an extension of the Web through standards which promote common data formats and exchange protocols on the Web. It provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries.

14 See as well Vashist (2015) for a discussion on cloud computing infrastructure and use.

15 http://live.worldbank.org/sites/default/files/Big%20Data%20for%20Development%20Report_final%20version.pdf

16 UN Global Pulse. (2014). Mining Indonesian Tweets to Understand Food Price Crises. http://www.unglobalpulse.org/sites/default/files/Global-Pulse-Mining-Indonesian-Tweets-Food-Price-Crises%20copy.pdf "Global Pulse is a flagship innovation initiative of the United Nations Secretary-General on big data. Its vision is a future in which big data is harnessed safely and responsibly as a public good. Its mission is to accelerate discovery, development and scaled adoption of big data innovation for sustainable development and humanitarian action. The initiative was established based on a recognition that digital data offers the opportunity to gain a better understanding of changes in human well-being, and to get real-time feedback on how well policy responses are working". http://www.unglobalpulse.org/about-new

World Bank's summary, by "conducting the research in the Indonesian context, the research benefited from a large user base--the city of Jakarta has the largest Twitter presence in the world with 20 million user accounts".

The Global Pulse lab "secured" the tweets through the use of the Crimson ForSight Hexagon software, including a classification algorithm, "which can analyze strings of text and sort them into categories of interest". For the study, "data was categorized through an initial filter" to identify those tweets which dealt with the price increases. A "researcher manually classified" randomly selected tweets based on sentiment as "positive", "negative", "confused/wondering", or "realized price high/high no emotion. "This manual selection by the researcher essentially "trains" the algorithm, which can, in turn, automatically classify the remaining tweets". As regards the statistical methods, the categorized tweets, forming part of a dataset, "could be analyzed using simple statistical regression techniques. Using such techniques, correlation coefficients could be estimated to analyze the relationship between twitter conversations and official food price inflation data, among other questions".

The results were mitigated. There was, in principle, "a relationship between official food inflation statistics and the number of tweets about food price increases". In the World Bank's opinion, "this initial effort to analyze social media data indicated the potential to utilize social media data to analyze public sentiment as well as objective economic conditions". Nevertheless, in view of the "abundance of false positive relationships, i.e. large changes in twitter data with no corresponding changes in actual inflation measures", "more research is certainly needed to improve the classification process as well as the process of geolocation".

Two things are to be noted here: first, official structured data are to some extent considered the reference to test Twitter data and the algorithm, which is to say that official data accuracy remains central –and still based on regular data production, close to that defined by UNESCO. It is thus suggested that a "higher granularity of official statistics are needed in order to more effectively compare it to the correspondingly spatially and temporally specific twitter data". Second, the improvement of the classification process is indirectly seen as based on the centrality and diversity of (changing) human annotation, as it is noted by the World Bank: "To the extent that classification algorithms are strengthened (by data interpreters?), and more fine grained economic data with which to train algorithms are made available (by official regular statistics producers?), the potential to implement ongoing real-time analysis of such data appears to be closely within reach". In other words, supervised machine learning and analysis depend on classification and official data quality, so on human annotation and coding criteria.

This is well understood by Rob Munro(2013) who tries to take into account language, cultural and social criteria to interpret feeling: "Despite the differences to English in the US and elsewhere, it is relatively easy for machines to learn subtleties, e.g. that a "bit disappointed" means "annoyed" after the algorithms have seen a few examples. The crucial point here is to know that one must apply 'British English' criteria. This is the most important next step in sentiment analysis: automatically knowing what kind of analysis (context) to apply, depending on the genre, language or source of the utterance. By knowing something about the social and cultural context of the utterance, we can make smarter assumptions about the assumed the knowledge of speaker and more accurately tailor the sentiment predictions to specific types of communication. As for how we should do this in an automated system? We'll leave that for a future post".

Source: Munro, 2013.

The importance of data-interpretation in data analysis has raised the question of the need for automization of human annotation activities, and this well beyond BICSE's recommendation of a "documentation data base", as a result of developments in application of algorithms.

6- Conclusions: statistical models and deductive quantification

After having considered the shift from descriptive to inferential statistical analysis for structured data, I conclude by touching on the question of the evolution from supervised to unsupervised learning, what I describe here as thee merging statistical-mathematical deductive vision/project of quantified knowledge. As studied by Leo Breiman (2001), there are two cultures "in the use of statistical modeling to reach conclusions from data": "One assumes that the data are generated by a given stochastic data model. The other uses algorithmic models and treats the data mechanism as unknown"[17]. The algorithmic modeling culture concerns supervised and unsupervised learning.

Box 1- Two cultures of statistical modeling (Breiman, 2001)

1- The Data Modeling Culture: the values of the parameters are estimated from the data and the model then used for information and/or prediction. Thus the black box is filled in like this:

Y	linear regression; logistic regression; Cox model	X

Model validation. Yes–no using goodness-of-fit tests and residual examination.

2- The Algorithmic Modeling Culture: the inside of the box is considered complex and unknown. The approach is to find a function f(x) —an algorithm that operates on x to predict the responses y. The black box becomes:

Y	Unknown	X
	decision trees; neural nets	

Model validation. Measured by predictive accuracy.

Developed in the political realm to help structured data-based planning or, recently, to help deregulate the social state, inferential statistical methods are also used to sustain machine learning modeling, following the first culture, where human annotation is still central to work on unstructured datasets. This opens up the discussion on the role of inferential statistics in statistical-mathematical algorithms to do away with human annotation and divide the second statistical model culture into the supervised and unsupervised learning methodologies. Efforts to attain this still utopian stage are seen by Janowicz *et al.* (2015) as the articulation between inductive and deductive approaches:

> "In principle, the dedicated pursuit of combining inductive and deductive techniques for dealing with data may indeed have significant potential that remains to be unlocked, and the key would be in a best-of-both-worlds combination. Indeed, deductive methods are extremely powerful if used in special-purpose applications such as expert systems, with a well-defined use case, limited scenario, and a domain that is understood well enough so that expert knowledge can be captured in the form of crisp logical axioms. Inductive methods, on the other hand, excel if data is noisy, expert knowledge is not readily mapped, and the input-output relationship matches the search space, that is, can be captured by learning with the chosen method(s). Yet, deductive methods usually break down under noisy data, while inductive methods may solve a problem but may not help to understand the solution or to verify it".

Despite the benefit of "a more systematic study of combinations of deductive and inductive methods", those studies are mostly restricted to a few special problems (e.g. low level vision) or as preprocessing utilities in larger projects "for example, systems that first learn higher-level features, expressed as logical axioms, from data and then use these higher-level features for nontrivial deductive inferences".

[17] A statistical model is a mathematical model which is modified or trained by the input of data points. Statistical models are often but not always probabilistic. There are "mathematical models" which specify a (deterministic) relation among variables; probabilistic models which specify a probability distribution over possible values of random variables; trained models which use some training/learning algorithm to take as input a collection of possible models and a collection of data points and select the best model. Often this is in the form of choosing the values of parameters through a process of statistical inference (Norvig, 2012).

Another perspective would be that of “viewing a deduction problem as an information retrieval or binary classification problem, and then applying non deductive techniques to solve this task.”

Table 5 below summarizes the three perspectives here studied. Dominique Boullier distinguishes the principles of exhaustivity and representativity of his two first levels and talks about censuses and surveys as principles of validation. His third generation would be the “traceability” quantification[18]. The first level concerns "Explanations”; the second “Descriptive and later on predictive correlations”, the third “Predictive correlations”. I prefer to emphasize the structured or non-structured nature of data and the focus on inductive or deductive processes. The latter seems to be destined to dominate. As expressed by a data scientist in an informal discussion: “managers have not understood it yet, it is data that commands”.

Table 5- Three quantifications as regards statistical models

<table>
<tr><td colspan="2">Descriptive Quantification</td><td colspan="2">Inferential Quantification</td><td colspan="2">Deductive Quantification</td></tr>
<tr><td colspan="2">Structured data
Previously standardized data, subsequently treated to create international or state-based DB</td><td colspan="2">Structured data
Pre-established questionnaire
(+ traditional DB)</td><td colspan="2">Unstructured data: infinite
(to be annotated or not depending on the model)</td></tr>
<tr><td colspan="2">Example: UNESCO’s collection of data from MS school censuses</td><td colspan="2">Example: PISA survey –linked to school census data but independent in design and results</td><td colspan="2">Example: Indonesia
(food and fuel prices from Twitter data)</td></tr>
<tr><td colspan="2">Descriptive statistics: comparative indicators presented in alphabetical order</td><td colspan="2">Inferential statistical model
Responses fit to the model
Limited conclusions</td><td colspan="2">Supervised algorithms including
Inferential statistical model
Unlimited conclusions</td></tr>
<tr><td colspan="2"></td><td colspan="2"></td><td colspan="2">Unstructured non annotated data analyzed with unsupervised algorithms which “choose” the statistical-mathematical model</td></tr>
<tr><td colspan="2">PROBABLE-DESCRIPT. TRUTH</td><td colspan="2">PROBABLE TRUTH</td><td colspan="2">“CLOSE TO NATURE”TRUTH</td></tr>
<tr><td colspan="6">Political-economic periods</td></tr>
<tr><td colspan="3">Liberalism – Keynesianism 1880-1930; 1945-1980
Macroeconomics: structural analysis, econometrics
Census, administrative data→ Surveys
Statistical science– Econometrics-planning from the 1940</td><td colspan="3"></td></tr>
<tr><td></td><td colspan="3">Neoliberalism 1980-
Microeconomics: cost-effectiveness analysis, benchmarking and results-evaluation oriented tools
Surveys -- census, administrative data
Statistical science – Modeling/theoretical-planning</td><td>Big Data supervised learning
Statistical-mathematical</td><td>Big Data unsupervised learning?
Statistical-mathematical</td></tr>
</table>

The unsupervised level of big-data machine learning could conceivably attain a almost-complete representation of nature/society. To realize this deductive knowledge, a diversity of information ontologies will have to be available and in continuous revision, combining different statistical-mathematical models. Would it be a significantly enhanced quantification?

References

Amatte, Michel. 2001. “Le statut changeant de la corrélation en économétrie (1910-1944)”. *Revue économique*, vol. 52, nº3, May 2001: 617-631.

Bachelard, Gaston. 1975. *Le nouvel esprit scientifique*, Paris: PUF.

Breiman, Leo. 2001. “Statistical Modeling: The Two Cultures”. *Statistical Science*, Vol. 16, No. 3: 199-231.

Bruno, Isabelle. 2008. *A vos marques, prêts… cherchez! La stratégie européenne de Lisbonne, vers un marché de la recherche*, Paris: Editions du croquant.

Bourdieu, Pierre& Passeron, Jean-Claude. 1964.*Les héritiers , Les étudiants et la culture*, Paris: Minuit.

18 shs3g.hypothèses.org/17

Cussó, R. 2012. "L'activité statistique de l'Organisation économique et financière de la Société des Nations: un nouveau lien entre pouvoir et quantification".*Histoire & Mesure*, XXVII-2: 109-138.

Cussó, R. 2010. "La quantification internationale à la lumière de la SSP et des Congrès internationaux de statistique: continuités et ruptures".*Journ@l Électronique d'Histoire des Probabilités et de la Statistique*, Vol. 6, n°2, December. http://www.emis.de/journals/JEHPS/Decembre2010/Cusso.pdf

Cussó, R. 2007a. "La Fast Track Initiative de la Banque mondiale : financement et transformation politique" in Vinokur, A. (ed) *Pouvoirs et financement en éducation. Qui paye décide?*, Paris: L'Harmattan: 69-87.

Cussó, R. 2007b. "SACMEQ, PASEC, PISA… les enquêtes sur les acquis des élèves et les politiques de mondialisation", Les évaluations au niveau international : impacts, contradictions et incertitudes, CIEP, Sèvres: 79-89.

Cussó, R. 2006. "Restructuring UNESCO's statistical services: The "sad story" of UNESCO's education statistics four years later". *International Journal of Educational Development*, Vol. 26, Issue 5, September: 532-544. http://dx.doi.org/10.1016/j.ijedudev.2006.01.001

Cussó, R. & D'Amico, S. 2005. "From development comparatism to globalization comparativism: towards more normative international education statistics".*Comparative Education*, vol. 41, n° 2, May: 199-216.

Dayan, Peter. 1999. "Unsupervised Learning" in Wilson, RA & Keil, F. (eds). *The MIT Encyclopedia of the Cognitive Sciences*.

Desrosières, Alain. 2008. *Pour une sociologie historique de la quantification*, Paris: Presses des Mines.

Desrosières, Alain & Thévenot, Laurent.1988. *Les Catégories socio-professionnelles*, La Découverte: Paris.

Graunt, John. 1662.*Natural and Political Observations Made upon the Bills of Mortality*, London.

Janowicz, Krzysztof, Frank van Harmelen, James A. Hendler & Pascal Hitzler. 2015. "Why the Data Train Needs Semantic Rails". *Association for the Advancement of Artificial Intelligence*.

Kuhn, Thomas S. 1962. *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

Salais, Robert. 2003. "Les politiques des indicateurs. Du taux de chômage au taux d'emploi dans la stratégie européenne pour l'emploi (SEE)" in Zimmermann, B. (dir.) *Action publique et sciences sociales*, Paris: Editions de la Maison des Sciences de l'Homme.

Munro, Rob. 2013. "Why is sentiment analysis hard?", September 6th, Idibon. http://idibon.com/why-is-sentiment-analysis-hard/

Lascoumes, Pierre & Le Galès, Patrick. 2005. *Gouverner par les instruments*, Paris: Presses de Sciences Po.

Latour, Bruno & Woolgar, Steve. 1979. *Laboratory life: the construction of scientific facts*. Los Angeles:Sage Publications.

Lengwiler, Martin. 2009. "Double standards: the history of standardizing humans in modern life insurance" in Lampland, Martha; Leigh Star, Susan (Hg.)*Standards and their stories: How quantifying, classifying, and formalizing practices shape everyday life*, Cornell University Press, 95-122.

Norvig, Peter. 2012. On Chomsky and the Two Cultures of Statistical Learning. At *the Brains, Minds, and Machines symposium* held during MIT's 150th birthday party, Technology Review reports that Prof. Noam Chomsky. http://daselab.cs.wright.edu/nesy/NeSy13/norvig.pdf

OCDE. 2001. *Connaissances et compétences: des atouts pour la vie. Premiers résultats du Programme international de l'OCDE pour le suivi des acquis des élèves (PISA) 2000*, Paris: OCDE.

Papadopoulos, G. S. 1994. L'OCDE face à l'éducation, 1960-1990, Paris: OCDE.

Vashist, Renu. 2015. "Cloud Computing Infrastructure for Massive Data: A Gigantic Task Ahead". *Big Data in Complex Systems. Studies in Big Data*, Volume 9: 1-28.