Alain Desrosières 27 September 2002

Quality and statistics: origins of the movement and reflections on the criteria[1]

(to be published in a collective book by the University of Bergen)

The "quality" theme emerged in some European national statistical institutes (NSIs) and in Eurostat during the 1990s. It has been addressed in many documents and by a collective project undertaken in a "leadership group" initiated by Statistics Sweden in 1998—a project in which INSEE has taken part. But the basic concept of "quality" has older roots in the world of manufacturing. Our paper offers some thoughts about these origins, followed by a suggested interpretation of the six quality criteria proposed by Eurostat and officially adopted by the 2001 SPC: *relevance*, *accuracy*, *timeliness*, *accessibility*, *comparability*, and *coherence*.[2] A theoretical question underlies our analysis: how do statisticians perceive and manage the tension generated by the fact that their objects can be seen *both* as "real" (their existence precedes their measurement) and as "built on conventions" (they are, so to speak, "created" by conventions). This issue, philosophical in appearance, actually recurs in many debates about the circulation of statistics produced by the NSIs. We shall try to interpret the six quality criteria in terms of this distinction between two epistemologies, the "realist" and the "conventionalist" (also called "constructivist"). We show that statisticians committed to high professional standards must continually refer to these two seemingly contradictory epistemologies—a situation that, in a sense, provides the zest of their trade.

The way in which the "quality" theme has circulated since the 1930s, first in the manufacturing industry, then in other activities including statistical production, is a fine object of study for the historian of science and technology. Indeed, technical factors have always been inextricably intertwined with social concerns in this story: technical factors *qualify* and shape the ways in which people (engineers and their machines, enterprises and their employees, sellers and buyers, owners and contractors, etc.) view and organize their relationships and their reciprocal rights and obligations. Under a broader definition of "quality," economists have observed, since the 1970s, that the standard general-equilibrium model takes for granted a list of well-defined goods, whose "qualities" are not in dispute. "Conventionalist economists" have devoted special attention to the issue, seeking specifically to characterize the "quality conventions" theoretically implicit in a market economy (Akerlof 1970; Eymard-Duvernay 1989). But, under the "conventionalist" approach, "quality" remains a fairly abstract notion, as does the notion of "information" put forward by those same economists, precisely in order to characterize the "quality" of a good: the empirical study of the procedures for the concrete implementation of "quality" and "information" has yet to be undertaken. This is particularly true of what official statisticians in the Anglo-Saxon and Scandinavian countries elaborated in the 1990s by importing the tools of "total quality management" (TQM) from the manufacturing industry into their field.

The notion of quality was first discussed in France, in the 1920s, in a different context: mass manufacturing. The issue was "manufacturing control," for which a probabilistic sampling of items to inspect was introduced (Bayart 1998). An American statistician, Edwards Deming, then came up with the idea of transferring this same formal approach to industrial "quality control" (in the 1930s, Deming had taken part in organizing the first random-sample surveys on employment and unemployment [Didier 2000]). But the detection of irregularities and failures in series production systems led him to reexamine the organization of assembly lines, and, by extension, the entire system of workplace relationships. What appeared to be a purely technical inspection raised questions about all the social relationships between the members of a workforce. This context also gave rise to a fundamental distinction—which recurs throughout the rest of the story—between *product quality*, which users want, and *manufacturing-process quality*, which production organizers seek. The same distinction turned up

[1]The views and questions expressed in this paper are the authors' only. Their purpose is to generate discussion and criticism over the important issue of quality.

[2]Some are surprised that list of criteria contains no allusion to ethical issues such as data confidentiality, neutrality or independence from political pressure. The apparent reason is that the list was prepared from a *customer-oriented* standpoint focused on the *needs of statistics users*. The list therefore represents only a part of what, more generally, society is entitled to expect from its statistical system.

later in the system of “ISO standards” that firms demanded in order to be able to certify the good quality of their goods and services to their customers. But, in the United States of the 1950s, Deming’s proposals for reorganizing enterprises around the quality objective did not immediately encounter the success he was hoping for. As a result, he emigrated to Japan, whose manufacturing industry—then in the midst of reconstruction—adopted his language and his technical and social tools for achieving quality. This was followed by the development of “Japanese-style” management methods, most notably “quality circles,” which were supposed to bring employees together around a common objective: the improvement of production quality via the improvement of work organization. The definition of “quality” had thus shifted from a technicist definition, in terms of engineers, accuracy, and regularity of series production, to a social conception, formulated in the language of a human-relations manager. This wide gap, along with the contrast between “product quality” and “process quality,” has characterized the entire history of the quality issue. The trace of the divide would turn up again when official statisticians imported it in the 1990s (in so doing, incidentally, they often forgot the partly “statistical” origin of the vocabulary, linked to Deming’s career).

“Japanese-style” management effectively enjoyed its hour of glory in Europe in the 1980s, when “quality circles” were set up as a tool of “modern participative management” for harnessing the innovative initiatives of company employees (Boltanski and Chiapello 1999). Thus the quality issue, through successive metamorphoses, circled the globe via the United States and Japan. Meanwhile, corporate certification systems were also introduced, of which the most familiar are the ISO standards. The “quality movement” that appeared in some statistical institutes in the 1990s is the result of the encounter between (1) the above-mentioned management tools and vocabulary formulated in terms of “quality standards” and (2) two major changes in the management constraints of European governments: the trend toward contractualization—a consequence of the gradual deregulation of Europe’s economies—and the faster pace of European integration from the Single Act (1985) to the treaties establishing Economic and Monetary Union (1992) to the European Central Bank and the euro (1999).

The *contractualization trend* concerns (1) relations between government bodies and their “users” (now described as “customers”) and (2) relations between government bodies themselves, now required to bill their services and to try to balance their “budgets.” In Sweden, for example, government departments have been carved up into “agencies” (one of which is the “Central Statistical Office”) operating on this quasi-market principle, even though a large portion of their revenues is generated from contracts with other public agencies. Many operations are thus outsourced to other agencies, the private sector, or university research units. This requires two things. First, the specifications and “qualities” of their services must be spelled out in the invitations to tender and in specification sheets; second, the services provided should, as far as possible, be guaranteed and pre-certified along the lines of ISO standards for industrial enterprises. In the case of statistics, the reality status of the object procured through this procedure is heavily determined by the specifications legally written into such contractual documents. The pressure to objectivate “quality standards” is partly linked to this broader trend toward contractualization and billing of services between government entities. Some forms of these practices also appeared in France with the 1993 “Balladur circular,” one of whose more significant provisions was that government departments should bill each other for services performed.

The other factor that has stimulated the dissemination of the “quality movement” has been the *quicker pace of European integration*, and the far greater pressure to harmonize European statistics. Already, in the past, the demand for comparisons between national statistics has acted as a powerful incentive to compare—and, when appropriate, to challenge—tools that, in their strict national contexts, were sometimes previously encapsulated in “black boxes” for a long period. But since the end of the 1980s, the pressure to reopen the black boxes of statistical production has greatly intensified. National statistics—most notably the national accounts—now have direct consequences, spelled out in Community regulations: the size of national contributions to the EU budget, regional subsidies, and the enforcement of the clauses of the EMU treaties and the Stability Pact (debt and public-deficit ratio requirements). The “qualities” of the statistics provided are thus far more subject to scrutiny and potential challenge. It is precisely when controversies erupt over the use of reputedly harmonized statistics, for the implementation of the clauses of such Community regulations, that the “real” or “constructed” status of the statistical measurements used leaves the confines of academic debate between epistemologists and can become the focus of bitter dispute. Such issues indirectly inform the “quality” guidelines discussed below, despite the fact that their style is usually neutral and technical, which can discourage a naive reading. It will therefore be useful to address the six “quality criteria” for

statistics one by one, in an attempt to determine their impact and their implications in terms of semantic content (What do they say? To whom?) and pragmatic content (What do they do? For whom?). It is precisely with respect to these issues that the initial question of the tension between the "realist" and "conventionalist" points of view can be analyzed.

Now these issues find little explicit formulation in the official documents on quality, as the main function of the latter is to secure a minimum consensus on general work-organization principles. This "quality movement" is embedded in a broader program to mobilize all the European governments around the quality issue, with the aim of establishing Total Quality Management (TQM) tools and language in government departments. TQM is, in turn, promoted by the European Foundation for Quality Management (EFQM), set up by fourteen leading European corporations in 1988. We shall not dwell here on the analysis of this managerial program, which is not specifically statistics-oriented, although several NSIs have chosen to import it in order to structure and *qualify* their internal reorganization efforts. The EFQM context is important, however, as it enables us to understand the origin of the action categories emphasized in the quality documents. The chief category seems to be "customer orientation," although, paradoxically, the documents contain almost no analysis of the actual uses of official statistics.[3] This confirms the "all-purpose" character of the "quality approach."[4]

A sociological reading of the six quality criteria for official statistics

Before examining the significance of the six criteria set out in the official European documents on statistics, we should note that the arguments and thought categories that inform them differ from those that social scientists might imagine *a priori*. The widest gap concerns the notion of *relevance*. We are in another "socio-epistemological" world, where statisticians and their "clients" are separated and maintain service relationships with one another on a market or non-market basis. In fact, in most NSIs other than INSEE, "methodologist statisticians" and "subject-matter specialists" are two different professions. The former have studied statistics (mathematics), whereas the latter have degrees in economics, sociology, demography or geography.[5] This social and technical distinction explains the sharp divide between the *accuracy* and *relevance* criteria, which are under the responsibility of the first and second groups respectively. The predominance of a realist epistemology is strengthened by this social division of labor, since the first group "measures" objects whose reality status is guaranteed by an *external* entity: the second group.

Besides, in these documents, the main concern is not so much to discuss the origin and implications of the criteria. Rather, it is to attempt to *quantify their application level* (or, to put it differently, to quantify the qualities of the quantities)—even though, on several occasions, the quantification effort seems acrobatic to the documents' authors themselves. This point is relevant to the central question addressed by our paper: what is the purpose of the very act of quantification, and whom does it serve? Quantification serves to *provide tools for comparisons,* to coordinate them by standardizing them, and to control and stimulate players by ranking their performances on normative scales. In other words, the purpose is to construct test situations.[6] Quantifying quality criteria involves shaping an *equivalence*

[3]The only exception is an allusion to the crucial distinction between the uses of statistics as a "*public good*" (or "*universal service*"), i.e., free of charge and intended for "citizens of a democratic society," and as a "*private good*," billed to "customers" in the context of a market relationship. This distinction also occurs in the debates on the privatization (or transformation into "agencies") of certain public services.

[4]On the relationship between the two aspects of industrial quality—production and customers—see Duclos (1996).

[5]This analysis is less true for France than for other countries. The distinction is less clearcut in France, where the "economist-statisticians" graduating from the National School of Statistics and Economic Administration (École Nationale de la Statistique et de l'Administration Économique: ENSAE) and National School of Statistics and Information Analysis (École Nationale de la Statistique et de l'Analyse de l'Information: ENSAI) receive training that provides them—so it is assumed—with both competencies. In practice, they often work in both areas. French universities, as well, offer programs in mathematical statistics, but their graduates have few contacts with the official statistical system. Their counterparts in the Anglo-Saxon countries often work under contract for NSIs or Eurostat, but that is seldom the case in France.

[6]One of the incentives devised by the promoters of the TQM procedure is a series of awards to the best European NSIs for their successful compliance with these criteria (the awards are handed out in

space, and the *equivalence conventions* that structure it. The conventions are essential for comparing, coordinating, conducting competitive reviews, and measuring how successfully the players pass the proposed tests. Now the (partly distinct) functions of coordination, self-testing, and external testing are central to the social and political impact of the quantification approach (Porter 1994; Hopwood and Miller 1994; Thomas 1996). The expository mode and the content of the reference documents on the qualities expected from statistics seek to quantify these qualities. They thus offer a second-degree confirmation of this general hypothesis adopted by the studies on the sociology of quantification.

The ***relevance*** criterion offers the clearest example of the tension between (1) the "cognitive" (if not "scientific") vision that social scientists may have of the statistical approach in terms of *significance*, and (2) the implicit approach to statistics in the documents on quality criteria. One of the fullest Eurostat documents (and also one of the only ones initially drafted in French, because of the nationality of its author, Bernard Grais) characterizes *relevance* as follows:

"*A statistic is relevant when it meets user needs. Identifying users and their expectations is thus essential. To ensure that the produced statistics continue to meet user needs, Eurostat and the Member States should conduct a survey of main users at regular intervals, for example, every five years."* (Grais 1998, p. 32)

This description of relevance may seem remote from the semantic density that generations of economists or sociologists have assigned to the notion. Yet such a way of presenting relevance is less surprising than it may seem when a statistical function is autonomized at the social and technical level—a function offering its services to an extreme variety of users (and henceforth of "customers"), of whom researchers represent only a small proportion. In any event, it is inherent in a division of labor of this kind. Of course the suggested survey would be absurd if it aimed at a quantified result, i.e., the percentage of satisfied respondents; on the other hand, it could be useful if it provided an opportunity for investigating the social uses of official statistical products, about which we still know little.

Yet *relevance* issues are often debated in NSIs, even if the word is not always used. The public controversies over the definitions and measurements of "concepts" such as gross national product (GNP), unemployment, and inflation regularly resonate in the institutes. Admittedly, within the NSIs, such issues tend to be discussed by "subject-matter specialists" rather than by "methodologists," whereas the NSI representatives in European working groups on quality tend to belong to the second group. This may explain their downplaying of *relevance*: they are experts on *accuracy* issues, to which they devote greater attention in their documents. But the division of labor between *relevance* specialists and *accuracy* specialists is precisely in a close relationship with realist epistemology, which, in turn, is generated by a very widespread social demand. Inspired by natural science, realist epistemology is the only one that can endow statistics with the legitimacy it needs to be seen as external to the social debate, to which it provides an accepted common language and resting-points. This "exteriority" borrows both from the universality of *science* and from that of *law*, with a novel linkage between them that has received relatively little attention in its own right. Yet the model of law—with its equivalence classes embodied in legislation, and its coding procedures based on the assignment of individual cases to these classes—has much in common with the *upstream* part of the statistical production chain. (Admittedly, the upstream part is less familiar than the *downstream* part, which concerns mathematical methods for processing data pertaining to large numbers.) There is also a genuine affinity between the *social effects* of statistics and those of law: both seek to provide benchmarks to establish agreement between social players.

Quality promoters are more at ease with the ***accuracy*** criterion than with *relevance.* Here, they have an abundant technical reference literature. Historically, the three words *quality*, *reliability*, and *accuracy* have long been roughly synonymous for statisticians. Let us quote Bernard Grais again:

"Accuracy is defined as the closeness between the estimated value—in some cases calculated on a sample—and the (unknown) actual value for the total population. This aspect of quality has been studied in depth by many statistical offices and university research institutes. The assessment of accuracy rests mainly on an analysis of the total error associated with it, which is distributed between the sampling errors and non-sampling errors (errors in the sampling frame […], processing errors

---

very formal ceremonies). We find a similar use of quantification to compare and score high schools and hospitals in the ranking tables published by the French press since the 1990s.

[…],non-response errors […], measurement errors […], errors in the hypotheses of the model […]).” (Grais 1998, p. 32)

The very notion of “error” is significant. The distinction between “sampling errors” and “non-sampling errors” masks the fact that the first category is well defined and bounded thanks to probability theory, whereas the second is a blend of heterogeneous items. Using the single word “error” to describe both conceals essential issues concerning *relevance.* For example, the list of “causes of error” makes no reference to potential “uncertainties due to interactions between interviewers and respondents,” “enumeration imprecision,” or “coding conventions” (Thévenot 1983). The only notion mentioned is “error,” imported from the realist metrology of natural science (most notably astronomy: Armatte 1995). But our purpose here is not to reiterate the criticisms of the use of quantitative methods in the social sciences, often voiced since Cicourel’s famous book *Method and Measurement in Sociology* (1964). Rather, we ask the sociological question of why these criticisms have had so few consequences, even though they may have helped to increase the sophistication of some questionnaire-based survey methods. The only possible explanation lies in the existence of a social demand for “realist” statistics legitimated by an “unchallenged” institution. Now only the notion of “error” is consistent with this social demand; the other notions mentioned above, which involve *judgment conventions,* would entail a return to the *relevance* criterion*,* and more generally to the issue of the uses and the pragmatic sequences in which the statistical argument plays a part.

Let us take the example of unemployment. The “concept” of unemployed person is defined by the International Labor Office (ILO): a person is considered unemployed if he or she is “out of work, seeking work, and immediately available.” Thus spoke the “employment subject-matter specialists” before passing on the hot potato to the methodologists for them to measure it through labor-force surveys.[7] But at that point, myriad problems not predicted by the *subject-matter specialists* arise. They consist of the plethora of annoying borderline cases: makeshift jobs, internships, replacements, hospital stays, and early retirement. And there is more: in France, periodic visits to the National Employment Agency (ANPE) (to remain eligible for unemployment benefits) are regarded as “job-seeking” activity. Now Eurostat would like to stop these visits from being classified as such, a decision that could reduce the “number of unemployed” by about 10%. We have listed these issues here merely to suggest that they might all be viewed in one of two ways: (1) in terms of *relevance*, if we seek to interpret “ILO intent” as judges do for “lawmakers’ intent,” or (2) in terms of *accuracy,* by treating the cases mentioned as sources of “non-sampling errors,” which is what methodologists almost invariably do.

Two well-known methodologists, Richard Platek and Carl-Erik Sarndal, published a summary study in 2001 (accompanied by sixteen commentaries) on fifty years of research on accuracy and errors in sampling surveys. They deplore the fact that, while *sampling-error* theory has flourished for decades, *non-sampling error* theory remains underdeveloped and inconsistent. As a result, they argue, one cannot confidently provide a “confidence interval” for non-sampling error as robust as the interval that probability theory ensures for sampling error. This strange regret can be understood only with reference to the division of labor mentioned earlier and to the realist epistemology inevitably linked to it. The crucial issues concerning the *judgment* needed before the decision to classify a specific individual as unemployed cannot fit into the epistemological framework outlined by methodologists. But, here as well, the point is not to criticize methodologists; it is to understand why this epistemology is so “natural,” and to determine whether it is *socially* possible to speak another language. To sum up: how should one *jointly* conceptualize and discuss the cognitive and political issues raised by these problems, which are not only cognitive: employment policies, unemployment benefits, and minimum social benefits have effects that are all the more tangible as they are solidly built and embedded in institutions, among which the official statistical system is now one of the most important.

The responsibility of those who address these issues is thus inseparably scientific and political. The ethical principles championed both by statisticians and by those who rely on the work of statisticians to buttress their own arguments often refer to a separation between these two moments—that of science

[7]The division of labor is often more complex: methodologists are responsible for the theoretical design of the survey (sampling plan, sample selection, checks, adjustments, estimation of variables and their confidence intervals, etc.), but *data collection in the field* is, in some countries, often subcontracted to specialized private firms. This further increases the need for written *pre-objectivation* of “quality criteria.”

and that of its uses. A "great divide" of this kind can hardly dispense with a realist epistemology. Any allusions to the conventions presiding over the measurement could, in some cases, jeopardize such a division of labor, by raising doubts about the credibility and therefore on the legitimacy of a statistic. Thus scrupulous statisticians may sometimes be torn between two constraints: (1) a professional requirement to spell out their methods and conventions, by adopting *de facto* a quasi-constructivist position, and (2) an implicit realist social demand not to "burden the reader with these technical details," which may sow doubt or restrict the impact of the reported results. Controversies often center on the possibility of generalizing from "partial results," i.e., results disseminated with methodological caveats that disappear as the results are put to successive uses. We saw this in France, for example, in 2001 and 2002 with the debates over "poverty traps" and the "effects of cuts in employers' social contributions."[8] We shall return to these questions in our discussion below on the fourth quality criterion: "*accessibility and clarity*."

The third criterion, ***timeliness and punctuality***, has mobilized considerable energy in the reorganization of European statistical production channels since 1998 (Elissalt 2001). This is the result of the creation of the Economic and Monetary Union (EMU) and the euro. European monetary policy is now managed by a new institution established in 1998: the European Central Bank (ECB). To carry out its mission, the ECB requests "economic indicators" (indices for prices, labor costs, industrial production, and foreign trade) needed to manage the variables for which it is responsible—in particular, key lending rates. The ECB makes its requests to Eurostat, which forwards them to the NSIs. Now these indicators must cover the entire euro area, which thus constitutes an entirely new space of social and cognitive equivalence, for which the "equivalence conventions" must be at once redesigned and reconstructed. One of these conventions has turned out to be very difficult to implement: the *timeliness* of indicator *publication*. The *market* context in which the ECB operates generates a strong constraint: the speed of the bank's intervention. The combination of these two constraints—the construction of a new equivalence space and the standardization of the (shortened) publication times—has entailed a radical reorganization of some statistical production chains.

The model for setting common schedules is that of the U.S. Federal Reserve Bank. The standard is that the indicator should be available no later than one month (or 22 working days) after the end of the period concerned. This comparison of the performances of European and U.S. statistical channels has given rise to a large-scale *benchmarking* operation.[9] It did emerge, of course, that the setting of this standard implied a substantial shortening—in some cases, a halving—of the time allotted to computing the index. This could be achieved only at the expense of the second quality criterion, *accuracy,* in particular because the only solution to meet the timeliness requirement was often to "estimate" (i.e., to invent by resorting to econometric methods) the figures for the final days of a month or the final weeks of a quarter; at a later date, one could publish "revised" indices including the final periods estimated earlier. The emergence of "speed" as a vital quality criterion for statistics may be seen as characteristic of the growing market-oriented character of economics, symbolized by the functioning of the stock market and the financial markets. In the U.S., where market orientation is a well-established feature, there have long been strict timing rules for the publication of selected economic indicators *down to the second.* For example, the publication of the unemployment rate has instant effects on the Dow Jones: if unemployment rises, so does the stock market, and vice versa.[10]

The "realism" of time series differs from that of so-called "structural" statistics. The aspect of time series that is assumed to be "real" is the percentage *changes* (not the *levels*), since they trigger interpretation and action. The changes are derivatives with respect to time. But this is not only mathematical time; it is also the time that, in establishing links and creating equivalences or discontinuities between the past and the future, allows "projections" if not "forecasts," hence guidelines or justifications for action. "Realism" is not, in this case, very different from that of ordinary language: a

[8]These problems encountered by statisticians in their work, and the sometimes bitter controversies they elicit, are comparable to the problems faced by journalists, who are also often torn between conflicting requirements (Lemieux 2000).

[9]French government bodies use the English term (which may be rendered in French as *étalonnage* or *référence*) to denote this construction of a European equivalence space whose aim is to achieve convergence in NSI performances (most notably for publication timeliness).

[10]Traders see higher unemployment as a short-term signal of a decline in the "inflation risk"—a risk that would prompt the Fed to lift its key rates, thereby slowing economic activity. In the longer run, higher unemployment is seen as putting downward pressure on wages and hence driving up profits**.**

forecast is realistic if it is plausible, credible—in sum, "achievable." The interpretation of time series has given rise to highly specific statistical methods (calculation of moving averages, seasonal adjustment, stochastic linear models) designed to untangle realities of different orders. These range from short-term "random variations" to various types of cycles to "long-term, centuries-long trends." In the case of time series, the ties between realist rhetoric and conventionalist rhetoric are very distinctive and would deserve special study.[11]

The fourth criterion, ***accessibility and clarity****,* is, at least as much as *relevance,* the one that could characterize the forms of concrete interaction between statistical information and its users. Whereas *relevance* concerns the significance that users attach to a statistic, *accessibility* and *clarity* relate to guidelines for practical situations. One could envisage a research program on the material aspects of statistical usage practices (documentary research, database queries, construction of tables and charts, use of statistical software), using methods already field-tested by the "sociology of situated action" (Conein and Jacopin 1994). Such investigations are still uncommon. Yet this is the level of statistical practice covered by the fourth criterion, defined as follows:

"The value of statistical data is all the greater if they are easily available to users, are produced in a form that suits users, and are properly documented. Statistics suppliers must also provide all the assistance needed for their use and interpretation." (Grais 1998, p. 33).

This formulation implicitly assumes a "social demand". But the latter expression, while convenient, is simplistic. Realist social demand, discussed earlier, exists only at the end of a long process that has led to the emergence of "something" as an "issue" and as the possible object of an action, a condemnation, or a claim. The enactment of quantification procedures is one of the ways—along with political and labor-union struggles, law, economic theory, literature, or philosophy—that help to confer a social existence upon the "something" and to harden it. We can cite a number of historical, well-documented examples of such processes: poverty in England (1880s), unemployment or public opinion in the U.S. (1930s), the GNP growth rate (1950s), and child abuse (1980s). This does not mean that the object did not exist before its quantification; rather, it means that quantification has endowed the object with an entirely different form of social existence, in which two "measures" are closely intermingled: measures in the metrological sense, and measures in the sense of a government that "takes measures" and wants, in particular, to *evaluate* their effects. It is this semantic and pragmatic complex that confers reality upon the object now in circulation, which is the focus of a sometimes impatient "social demand," as we saw with the increase in *benchmarking* requests from the ECB. Likewise, *accessibility and clarity* can be assessed only with respect to the diversity of the modes of application of statistical arguments. The effort to enforce the constraints implied by this criterion may generate a "metadata paradox," due to the difficulty of yoking together a realist mode of expression and a conventionalist professional ethic.

Since the 1980s, documentation systems containing "metadata" have been developed, most notably as a result of the growing computerization of the production and storage of official statistics in databases. The systems allow the filing of, and on-demand access to, information on the methods used to construct the stored data. For example, survey metadata would include initial purpose, field of coverage, sampling plan, questionnaires, classifications, collection methods, response imputation, and so on. Statisticians regard the preservation and supply of such information as a basic professional requirement. Metadata are, at least officially, in strong demand from many users, in particular social scientists. Paradoxically, however, as with great monuments, exhibiting these "scaffoldings" may—whether one likes it or not—blur if not the beauty at least the argumentative efficiency of the factual evidence consisting in the publication of a plain, bare figure. Can too much metadata kill the data? This paradoxical idea may, of course, be rejected in the name of a rigorous, scrupulous approach to argumentation. But it would be better to envisage it in an empirical and non-normative fashion as well, by studying the diversity of the circumstances of metadata use or non-use. That would provide a field of investigation for a sociology of the uses of quantification—a subject to be invented—in relation to the sociology of argumentation.

[11]Some material on this may be found, for example, in the work of Michel Armatte (1992), Mary Morgan (1990), and Judy Klein (1997). Klein's book, *Statistical Visions in Time: A History of Time Series Analysis, 1662-1938*, offers an extraordinary survey of the evolution of these forms of rhetoric and of these tools for untangling realities from the flow of time.

The fifth criterion is that of data ***comparability*** in time and space. It is totally bound up with the notion (seldom invoked by statisticians) of *equivalence convention,* which confers social and cognitive legitimacy upon comparisons, since the *convention* is a social procedure whereas *equivalence* is a logical category. Such a convention postulates the permanence in time, or the identity in space, of objects whose existence logically precedes measurement procedures. The construction of "long series" by quantitative historians implies conventions that other historians may challenge—as occurred, for example, with the series on "two centuries of labor in France" published by Olivier Marchand and Claude Thélot (1991). In such controversies, the objects' reality status and permanence is effectively at issue.[12]

The efforts to harmonize European statistics, which have greatly intensified since the 1980s, have highlighted the polarization between the two definitions of the issue, namely, the "harmonization of products" (or "of outputs"), and the "harmonization of methods" (or "of inputs"). These definitions, in turn, are informed by two approaches, one more realist, the other more conventionalist. *Product* harmonization implies that "subject-matter specialists" from the EU countries have met and agreed on the definition of a "concept," which they have then communicated to the methodologists of their respective NSIs. The task of the methodologists is to measure the concept as best they can, each acting individually with their own methods, and taking into account each country's specific characteristics and sources. This implicitly realist approach has historically been that of national accountants since the 1950s. It is partly justified by economic theories, which help to impart a social existence to objects (gross domestic product, investment, consumption) before they are measured (Vanoli 2002). But, since the 1990s, this approach has been sharply criticized because of the perceived significance of the cases in which the national specificity of measurement methods has had major effects on the measurements obtained. The harmonization of *methods*, offered as an alternative, is a sort of epistemological dream in which the complete process of data construction and collection would be standardized for the entire EU. Eurostat, of course, is pushing in this direction. We can see the link between this goal and that of the administrative unification of the continent. As some have already shown, political unification and statistical unification go hand in hand. This tie is implicit in the very etymology of the word *statistics*: the science of the State.[13] The issue of statistical harmonization shows the links between the cognitive and political dimensions of the tension between realism and conventionalism, and also, in another way, the affinity between statistical conventions and legal conventions.

The sixth criterion, ***coherence****,* raises a crucial issue in all NSIs: for many of the variables requested by users, the estimates based on different sources (censuses, surveys, administrative records) differ. Should they be "imputed," by imposing common fields of coverage, classifications, and all the factors contributing to data generation, so as to make the sources consistent—in an attempt to use this commonality as a means of "getting as close as possible to reality"? The documents on quality seem to recommend this approach, arguing that it meets users' presumed wishes, but they skirt the epistemological issues raised by coherence:

"When they derive from a single source, the statistics are coherent because it makes sense to combine the basic concepts into more complex sets. When they derive from different sources, and, in particular, from surveys conducted at different frequencies or using different administrative records, the statistics are coherent only to the extent that they are based on common definitions, classifications, and methodological standards. When that is indeed the case, the messages that these statistics convey will be clearly inter-related or at least will not contradict one another." (Grais 1998, p. 34)

Should the statistics provided to the public be made fully coherent? There is no obvious answer to this question. An examination of the arguments for the competing solutions offers another opportunity for an empirical study of the tension between the realist and constructivist attitudes. Here as well, national

[12]On this topic, see the debate organized by Florence Weber in *Genèses,* 9, October 1992, pp. 90-119. In this case, the epistemological compromise put forward by the challenged authors consists in falling back on the notion of "order of magnitude": the convention concerns only the accuracy of the measurement, but not an intrinsic reality, which is assumed to be optimally proxied by the conventional algorithms proposed. This type of argument should be studied systematically, by viewing it as one of a range of formulations available for managing the tension between realism and conventionalism.

[13] See, for example, Anderson (1988) for the United States and Patriarca (1996) for Italy.

accounting had historically opted for the comparison and imputation of sources, hence for a realist approach. Indeed, its very principle is to fill in—as best as possible given the available sources—a set of theoretically coherent and exhaustive tables describing a nation's total *macroeconomic* flows. By construction, these tables are doubly balanced, by agent and by transaction. The choice to target macroeconomic coherence is therefore the keystone of the approach. It has played a decisive role in developing and improving sources, by stimulating mutual criticism. It was consistent with the use of national accounts in Keynesian models, which had their moment of glory between the 1950s and 1970s. Since then, the uses of statistics have considerably diversified, and total consistency is no longer such an obvious requirement. Once these sources start to be used in different contexts, often for industry-specific or local issues, the potential advantage in undertaking the costly quest for consistency may be smaller than the loss due to subjecting basic sources to processing operations intended solely to obtain that overall coherence.

In fact, not all European NSIs address the coherence issue in the same way. An interesting borderline case is the Netherlands, which in the 1990s designed a sort of statistical utopia functioning as a thought experiment: this had the merit of highlighting the limits of a realist epistemological fiction (Bochove 1996; Desrosières 1999). The concept consists in imagining a theoretical table in which all "individuals" are recorded in the rows and all the "variables" are recorded in the columns! Two types of tables are planned, depending on whether the "individuals" are persons or enterprises. The tables would then be gradually filled by extrapolating and making "coherent," at the most microeconomic level, the sample surveys and available administrative records.[14] The justification provided by the project's authors is that it would satisfy the consistent desire of users to have "one figure for one variable." The project must not be judged on its "realism" in the ordinary sense of "feasibility." On that criterion, it is totally unrealistic. But it is "realistic" in the philosophical sense used in our article, since it postulates the existence, for each individual, of "variables" with a real single value, independent of all judgment and questioning procedures. In fact, it is not realistic in the first sense because it is too realistic in the second sense. NSIs in other countries have viewed the Dutch project with skepticism, not to say apprehension: it would evoke George Orwell's *1984*, if the Netherlands had not been a highly democratic nation for centuries. However, it has the merit of showing the limits of the desire for total coherence—a desire often expressed, with legitimate arguments, by many users of official statistics. The more prudent choice of not completely eliminating the inconsistencies between sources by cleaving to the initial data-construction process is actually inspired by a rather conventionalist epistemology. This is because there is a smaller risk of erasing the traces of the conventions in the process—which is what happens with the realist-oriented operations to achieve coherence in conformity with user wishes.

Double consciousness and compromise solutions

We have tried to identify the ways in which statisticians and users of statistics experience the daily tension between the realist and conventionalist approaches. We have not sought to resolve the difficulty through epistemological formulas, which would give the "right theoretical answers" to this eternal problem. The drawback of these right answers is that they make no allowance for the practical situations in which these issues arise, or for the ways in which the players muddle through, for better or worse. We have tried to construct a sociology of statistical argument, rather than a normative epistemology, to which many studies have already been devoted. The choice between the two attitudes, realist and conventionalist, is not an existential choice to which people must commit for good. It is determined by the constraints of the situations that these players face, depending on the moment, and depending on the types of interactions with objects or with people they encounter. For example, it is often in situations of controversy and challenges to the routines of the social world that the earlier realism comes under fire from those who emphasize the conventionality of statistics. By contrast, reality is equated with "things that hold," in the triple sense of "being robust," "being tied together," and "holding people together by coordinating them." We could reinterpret the six quality criteria as prerequisites for this cohesion.

In their everyday practice, statisticians are plunged into a world of conventions, which they record or shape by themselves. The fact that the measurement results from this sequencing of conventional decisions is therefore self-evident to them. But later, they change hats without realizing it, and speak a

[14]The Dutch and Scandinavian NSIs have long chosen to develop their official statistics from administrative records, reducing the share of direct surveys.

realist language when addressing the outside world. There is no ambiguity here: in both cases, statisticians have done their job conscientiously. The attentive user and the researcher, who consult the metadata, are faced with the same difficulty as well. It is not specific to the statistician. How do the players, in practice, manage this tension—which no theoretical epistemology can solve for them? This question is reminiscent of the "double consciousness" phenomena encountered by anthropologists. Jeanne Favret-Saada (1977), in her study of witchcraft in rural Normandy, tries to understand how people can *at the same time* believe and not believe in spells. The typical phrase here is: "*I'm well aware* (that spells don't exist), *yet* (bizarre events happen)." How do people keep both ideas in their heads and in their lives? The context is, of course, different for statisticians, since their rationalism is not in doubt in either moment, realist or conventionalist. Yet there is a form of incompatibility between the two attitudes, which statisticians have to cope with somehow: "*I'm well aware* (that my statistics result from conventions), *yet* (I believe in a reality that I am asked to quantify)."

Several compromise solutions are available to ease this tension. They rely on the principle of separating the signifier (the measurement) and the signified (the object to be measured), while linking them by means of varied correspondence terms. For instance, *indicators* and *indices* do not pretend to measure something directly, as a physicist or astronomer would.[15] Rather, depending on the situation, they are summaries (as is an average), representatives, or spokespersons for silent, complex, and unattainable entities. In a certain sense, they are useful fictions. Here, we can speak of "near-conventionalism," since the fact that the measurement strongly depends on computational conventions is asserted and accepted, at least by those who produce it. But the indicator tends to become, for users, the thing itself. In this sense, we can say that statistics create reality. This has happened with objects such as the *poverty line* used since Charles Booth in the late nineteenth century (Hacking 2000), the *intellectual quotient* measured by psychometric tests (Martin 1997), and *public opinion* measured by Gallup polls (Blondiaux 1998). One should catalogue these compromise solutions, by relating them both to situations (sociology of argumentation and of situated action) and to the persons involved (sociology of actors). Even if it does not offer a clearcut epistemological choice concerning "reality,"[16] such an inquiry would help us to understand the often obscure debates and controversies triggered by statistical arguments, both in the public arena and in the social sciences.

REFERENCES

AKERLOF, G., 1970: "The market for '*lemons*': quality uncertainty and the market mechanism," *Quarterly Journal of Economics*, 84, pp. 488-500.

ANDERSON, M., 1988: *The American Census: A Social History*, New Haven: Yale University Press.

ARMATTE, M., 1992: "Conjonctions, conjoncture et conjecture. Les baromètres économiques (1885-1930)," *Histoire et mesure*, VII-1/2, pp. 99-149.

ARMATTE, M., 1995: *Histoire du modèle linéaire. Formes et usages en statistique et économétrie jusqu'en 1945*, doctoral dissertation, EHESS, Paris.

BAYART, D., 1998: "Statistique mathématique et gestion de la qualité. Recherches sur l'histoire de cognition dans l'organisation industrielle," *Cahier du CRG* (École Polytechnique), no. 14, May 1998.

BLONDIAUX, L.,1998: *La fabrique de l'opinion. Une histoire sociale des sondages*, Paris: Seuil.

BOCHOVE, C. van, 1996: "From Assembly Line to Electronic Highway Junction: A Twin-Track Transformation of the Statistical Process," *Netherlands Official Statistics*, vol. 11, Summer 1996, pp. 5-36.

BOLTANSKI, L. and CHIAPELLO, E.,1999: *Le nouvel esprit du capitalisme*, Paris: Gallimard.

CICOUREL, A., 1964: *Method and Measurement in Sociology*, New York: The Free Press of Glencoe.

CONEIN, B., JACOPIN, E., 1994: "Action située et cognition. Le savoir en place," *Sociologie du travail*, XXXVI, 4, pp. 475-500.

DESROSIÈRES, A., 1999: "La statistique aux Pays-Bas. Informatisation et intégration, un projet futuriste," *Courrier des statistiques*, no. 91-92, December 1999, INSEE, Paris, pp. 51-59. http://www.insee.fr/fr/ffc/docs_ffc/cs91g.pdf

DESROSIÈRES, A., 2000: *La politique des grands nombres. Histoire de la raison statistique* (reprint), Paris: La Découverte/Poche (English translation: *The Politics of Large Numbers: A History of Statistical Reasoning*, Cambridge, Mass.: Harvard University Press, 1998. Note: does not contain the afterword referred to in n. 16).

[15]The notion of *order of magnitude* is ambiguous, being midway between the *imprecise measurement,* which pertains to the field of classical metrology, and the *indicator*, which separates the measurement and the thing measured. It does not make a clear choice between the two alternatives.

[16]We shall not return here to the previously mentioned hypothesis that a statistic is all the more *real* as it is more solidly *built* and anchored in stable institutions. This "compromise solution" helps to understand many historical situations and to study them empirically. On some of the debates sparked by this hypothesis, see the afterword to Desrosières (2000).

DESROSIÈRES, A., 2001: “Entre réalisme métrologique et conventions d’équivalence: les ambiguïtés de la sociologie quantitative,” *Genèses*, 43, June 2001, Paris: Belin, pp. 112-127. (A slightly different version is available at: http://www.upmf-grenoble.fr/adest/seminaires/desros/)
DIDIER, E., 2000: *De l’échantillon à la population. Sociologie de la généralisation par sondage aux États-Unis avant la seconde guerre mondiale*, dissertation, Centre de Sociologie de l’Innovation, École Nationale Supérieure des Mines de Paris.
DUCLOS, L., 1996 : “L’exigence de qualité suffit-elle pour porter la parole du client?”, *Les cahiers de recherche*, GIP Mutations industrielles, no. 70, February 29, 1996, pp. 27-36.
ELISSALT, F., 2001: “La statistique communautaire au tournant du XXIème siècle. Nouveaux enjeux, nouvelles contraintes,” *Courrier des statistiques*, no. 100, December 2001, Paris: INSEE, pp. 41-51. http://www.insee.fr/fr/ffc/docs_ffc/cs100i.pdf (in English: “European statistics at the dawn of the twenty-first century,” *Courrier des statistiques*, English series, no. 8, 2002 [forthcoming], Paris: INSEE).
EYMARD-DUVERNAY, F., 1989: “Conventions de qualité et formes de coordination,” *Revue Économique*, vol. 40, no. 2, March 1989, pp. 329-59.
FAVRET-SAADA, J., 1977: *Les mots, la mort, les sorts*, Paris: Gallimard.
GRAIS, B., 1998: “Harmonisation statistique et qualité: le cas des statistiques sociales,” paper at Eurostat Seminar in Mondorf on “The Future of European Social Statistics” (4th session, March 26-27, 1998).
HACKING, I., 2000: “Façonner les gens: le seuil de pauvreté,” in: J.P. Beaud and J.G. Prévost (eds.), *L’ére du chiffre. Systèmes statistiques et traditions nationales*, Montreal: Presse de l’Université du Québec, pp. 17-36.
HOPWOOD, A., and MILLER, P. (eds.), 1994: *Accounting as social and institutional practice,* Cambridge (U.K.): Cambridge University Press.
KLEIN, J., 1997: *Statistical Visions in Time: A History of Time Series Analysis, 1662-1938*, Cambridge (U.K.): Cambridge University Press.
LEMIEUX, C., 2000: *Mauvaise presse. Une sociologie compréhensive du travail journalistique et de ses critiques*, Paris: Métailié.
MARCHAND, O. and THÉLOT, C. 1991: *Deux siècles de travail en France*, Paris: INSEE.
MARTIN, O., 1997: *La mesure de l’esprit. Origines et développements de la psychométrie, 1900-1950*, Paris: L’Harmattan.
MORGAN, M., 1990: *The History of Econometric Ideas*, Cambridge (U.K.): Cambridge University Press.
PATRIARCA, S., 1996: *Numbers and Nationhood. Writing Statistics in Nineteenth-Century Italy*, Cambridge (U.K.): Cambridge University Press.
PLATEK R. and SARNDAL, C.E., 2001: “Can a Statistician Deliver?”, *Journal of Official Statistics*, vol. 17, no. 1, March, Stockholm: Statistics Sweden, pp. 1-20. http://www.jos.nu/Contents/issue.asp?vol=17&no=1
PORTER, T., 1994: “Making things quantitative,” in: Power, M. (ed.), *Accounting and Science*, Cambridge (U.K.): Cambridge University Press, pp. 36-56.
RIVIÈRE, P., 1999: “Qualité et statistique,” *Courrier des statistiques*, Paris: INSEE, pp. 47-58. http://www.insee.fr/fr/ffc/docs_ffc/cs90k.pdf (in English: “Quality and Statistics,” *Courrier des statistiques*, English series, no. 7, 2001, Paris: INSEE, pp. 20-30).
THEVENOT, L., 1983: “L’économie du codage social,” *Critiques de l’économie politique*, no. 23/24, April-Sept. 1983; *Théorie économique et pratiques sociales*, La Découverte/Maspero, pp. 188-222.
THOMAS, R., 1996: “Statistics as Organizational Products,” *Sociological Research Online*, vol. 1, no. 3, http://www.socresonline.org.uk/socresonline/1/3/5.html
VANOLI, A., 2002: *Une histoire de la comptabilité nationale*, Paris: La Découverte.
WEBER, F. et al., 1992: “Histoire et statistique. Questions sur l’anachronisme des séries longues,” *Genèses*, 9, Paris: Belin, pp. 90-119.