

The Socio-Political Construction of a European Census of Higher Education Institutions: Design, Methodological and Comparability Issues

Benedetto Lepori · Andrea Bonaccorsi

Published online: 25 July 2013
© Springer Science+Business Media Dordrecht 2013

Abstract This paper reports on an experiment concerning the social construction of statistical definitions, where the first census of Higher Education Institutions in Europe has been developed. It conceptualizes the construction of indicators as a social process of definitions and boundaries' negotiation, involving value judgments, social and political opinions, as well as practical interests and power strategies of actors. The paper exemplifies this process on three issues, namely the social demand for establishing a census, the controversy concerning the definition of a perimeter as well as the selection of indicators, and the nature of comparability judgments. We first conclude that the socio-political dimension has to be explicitly taken into account when designing statistical systems; second, that social scientists involved in this process need to openly recognize the conflicts around the definition of indicators; third, that the objectified and taken for granted status of indicators makes them a powerful instrument to influence policy decisions and, that indicator designers need to make their own value judgments and interests fully transparent.

Keywords S&T indicators · Classification systems · Higher education institutions · Sociology of measurement · European Research Area

Introduction

The current study was commissioned by an authoritative statistical institution (EUROSTAT) and two Directorates General of the European Commission, an

B. Lepori (✉)
Faculty of Economics, University of Lugano, via Lambertenghi 10a, 6904 Lugano, Switzerland
e-mail: blepori@usi.ch

A. Bonaccorsi
University of Pisa, Pisa, Italy

unusual combination of forces for a feasibility study. Our task was practical: we had to support the creation of a census of all institutions of higher education (HEI) in Europe. For the larger census population, and for the research active sub-population, we had to ensure the delivery of a set of indicators and lay down the foundation for regular data collection by EUROSTAT. We approached the task with the kind of scientific and professional attitude needed in studies with a practical aim, i.e. to deliver an outcome according to the highest professional standard needed to reach a practical and usable result. In doing so, our attitude was technical and oriented towards a thorough definition of phenomena for accurate classification and measurement. Thus we anticipated a number of definitional, technical, and practical issues, for which we were confident a reasonable and technical solution could be found.

Over time, however, we first had to admit that questions of a non-technical nature were at stake and that classification and measurement issues are deeply embedded into political concerns, and more largely, into the way society represents and intervenes with higher education and research. Second, it also became evident that the perspective of an external, objective observer, who crafts a set of definitions according to rigorous methodological rules and is not involved in the conflict of interpretations, was not adequate. Any position supposed to be neutral and merely technical turned out, inescapably, to involve value judgments and we had to recognize that it would have been impossible to fulfill our task without entering into the interpretation and political game. In turn, this raises deep questions on the role of our profession and on the value of the indicator systems we are developing.

In this paper we propose a reinterpretation of this experience, focusing on how we addressed these socio-political issues and how we managed to find solutions which were acceptable both to the involved stakeholders as well as to our professional community; further, we inquire how the emerging representation of European higher education systems was influenced by these choices, as well as by the socio-political environment of the project.

Our account is an autobiographical one, as we managed the project, we wrote most reports and we participated actively in meetings, both inside the project and with the European Commission and EUROSTAT. Accordingly, while we provide reasonable arguments to support our interpretation, we cannot claim for objectivity or verifiability: not only were we too involved in the process for any scientific standard, but also most of our sources cannot be disclosed and verified by independent observers.

Rather, our aim is to reflectively interpret the grounds on which the involved actors negotiated the choices and constructed consensus, by highlighting the complex mix of interests and values that drove them. To this aim, *after the end of the project*, we began mobilizing a large set of literature covering sociology of science and S&T indicators to help us in this hermeneutic process—rather than as theoretical devices to develop hypotheses to be tested as in more conventional scholarly work. While scientists cannot claim to be free from values and interests when delivering political expertise, we claim that this commitment to reflectivity and to reasonableness should distinguish scientific expertise from other forms of competence in the political arena.

The paper is organized as follows. In section “[Mobilizing Theoretical Perspectives](#),” we make use of a number of theoretical perspectives on categorization, classification and measurement from different disciplines to identify some key issues to interpret the EUMIDA experiences. In the following sections, we organize the re-interpretation around three issues, namely the rationale for devising a HEI-level classification (“[From Mass Education to Differentiation: The New Social Demand for Measurement](#)”), the controversy on how to define the perimeter of higher education, (“[What are HEIs? The Controversy on Perimeter and Classification](#)”) and the operationalization of measures of HEI activities (“[The Conventional Nature of Statistics: Definitions and Measurement Issues](#)”). We conclude by contextualizing this experience in the broader framework of the relationships between science and politics.

Mobilizing Theoretical Perspectives

In general terms, our task was one of *categorization* (i.e. define how to partition observations, or the perimeter and boundaries of objects), of *classification* (i.e. define the rules by which an observation is assigned to one or another of categories) and of *measurement* (i.e. to define how the characteristics of organizations can be quantified and compared). These are clearly pre-requisite activities for any official statistics aimed at producing censuses. There is a rich body of literature that has discussed the cognitive, epistemological, sociological, and political dimensions of these activities.

Cognitive and Epistemological Perspectives on Classification

The act of placing objects into categories is one of the most fundamental activities in the life of individuals. Classical theories of categorization posit that categories are constructed by identifying the necessary and sufficient conditions for membership, consistent with philosophical objectivism, which assumes that it is possible to identify the “important physical properties” that allow for the division of the world into kinds (Bryant 2000).

A consistent literature in cognitive psychology has challenged this assessment, suggesting two alternative views. One is that the membership into categories is not binary, but fuzzy. Objects are assigned to categories on the basis of a similarity judgment, which is the sum of features shared by various objects weighted for their salience, or the frequency with which these appear in the category. The features then become abstract across instances, following a probabilistic approach, which is also called “family resemblance” (McCloskey and Glucksberg 1978). Another view suggests that categories are not constructed on the basis of features, but of mental models that link the features to other deep characteristics, typically of a causal or functional kind (Murphy and Medin 1985). While classical and probabilistic theories could be used to support the prediction that categorizations of the world converge, more recent theories strongly support a pluralistic view. There is no way

to develop a rule or procedure to eliminate the possible divergence in classifications. We need theories of higher order if we want to make sense of why certain classifications are adopted and others discarded.

As scientists should be better equipped to construct theories that allow convergence, one could hope that recourse to scientific expertise (as in the case of EUMIDA) helps to address this problem. Unfortunately, epistemological reflections have shown that this is not the case. Even in one of the apparently more robust taxonomic exercises in biology—the identification of species—there are several legitimate alternative classifications (Duprè 1993). On the basis of these and other contributions, a pluralistic view of scientific classification has been proposed, which admits the coexistence of alternative taxonomic systems (Bryant 2000).

In terms of our task, this literature hints to the fact that we cannot expect to rely on objective and clear-cut criteria to identify and classify HEIs, but that this will be a contestable exercise, where fuzzy and uncertain cases will emerge, while the criteria for classification will be subject to disagreement, related to diverging assumptions on the core characteristics of HEIs. Further, we expect that classification will be contingent and subject to controversy among the actors involved.

Classifications, Social Order and Controversies

Social sciences have insisted that classifications have a social nature and are by no means features of an objective reality (Espeland and Stevens 1998). Categories are important because they allow people to keep appropriate distance from others in social interactions, allowing secondary socialization involving anonymous persons (Berger and Luckmann 1966). Thus, classification systems have a tendency to hide the controversial nature of their establishment and to present themselves as naturally occurring systems, while the conflicts over their origin are removed (Bowker and Leigh Star 1999).

Several authors emphasize that this process is intrinsically related to the emergence of conflicts. In the theory of distinction, social conflict is not only generated by the fight for material goods which are subject to conditions of scarcity, but also by the need to exploit the *rent from distinction*, which results from the possession of symbolic capital (Bourdieu 1984). This requires the ability to stabilize the classifications that separate social groups, establishing the rules for access or exclusion. These classifications have the important property of transforming the continuous distribution of wealth into a discrete distribution of status, marked by binary rules of inclusion/exclusion. Titles of education granted by schools and universities are considered one of the most important tools to enforce distinction.

Consequently, controversies over classifications are a central element in the societal power struggle (Latour 1987) and their stability is not related to objectivity, but rather to a transient outcome from the resolution of controversies, where social actors reach an agreement on how to include individual cases into predefined categories (Thévenot 2001). The revision of classifications is thus central to the reversal of social power structures.

In terms of our task, we expect that disagreement over classifications of HEIs also reflects struggles between actors concerning their role and power in governing higher education, while the quest for a HEI-level classification and data collection is also likely to reflect broad changes in governance of higher education. Moreover, we expect that existing classification schemes are largely conventional and based on lasting domination structures rather than on objective choices and that they assume a taken-for-granted status which makes their revisions highly difficult.

Modern Bureaucracy and Measurement as a Technology of Distance

A third perspective relates the development of statistical systems to the emergence of the bureaucracy and constitution of the modern national State (Porter 1995; Desrosières 1993).

According to this view, classifications are a fundamental tool to govern societies without being obliged to address and solve individual cases. Through the use of classifications and their material embodiments (maps, lists, forms etc.) it is possible to “govern at a distance,” that is, to control actions that are displaced in time and space, beyond the boundaries of locality and community (Porter 1995). In this perspective, the measurement of social reality through statistical systems is necessary for accountability, as the modern State faces the challenge to offer valid arguments to citizens who are not satisfied by propaganda (Desrosières 2008). This argument also applies to the measurement of science and technology, one of the more recent branches of statistical activity (Godin 2005).

This literature shows that in most cases the rules for taxonomies and classification are subject to considerable controversy and that statistical classifications are not robust, since different experts end up in classifying individual cases in largely different ways in a remarkable proportion of cases (Desrosières and Thevenot 1988).

Thus, we foresee a deep connection between the request for a census and the emergence of a European policy in research and higher education. Further, this perspective points to the lasting institutionalization and conventional nature of existing statistical systems in the field and thus that revising their content will be deeply entrenched with political and institutional struggles at the European and national level.

Standardization and Sociology of Quality

There is a strong rationale for the creation of systems of classifications as instruments for the enhancement of efficiency in economic transactions. In order to conclude a contract among parties, there must be agreement on the content of the exchange, an issue which inevitably includes a procedure to determine the quality of goods, as well as a system able to determine the citizenship of traders, in case a dispute over the terms of the trade must be brought to the attention of the jurisdiction. As these systems are subject to indivisibilities—a census is complete,

or it has no value—from an economic point of view it would be highly inefficient if measurement costs were the burden of individual traders (Barzel 1982).

The economic analysis of measurement and standardization also offers a useful perspective on this issue (Brunsson and Jacobsson 2000). On the one hand, there are obvious gains in reaching universal systems of measurement because of measurement costs savings. Thus we should expect convergence around common standards of measurement. At the same time, due to the possibility that a plurality of agents engage in standardization activities and due to the importance of network externality, it may be that in equilibrium several competing standards survive. Thus, there is not a compelling economic argument for uniqueness of standards.

A further relevant perspective is the one of economic sociology and, more specifically, of quality markets, where quality is difficult to measure and more important than price in economic exchanges (White 1981). Typical examples are markets for customized professional services, high level expertise, or some areas of cultural industry (Karpik 2007). These goods are of uncertain quality, so that the perception of quality by customers is crucial for market exchanges. This perception does not depend on actions undertaken by producers, such as the classical devices examined by economics (reputation, signaling, guarantees etc.), but rather on external devices to which individuals delegate their trust in selecting producers of singular goods (Blank 2007). Two of them, denominations and rankings, are of direct interest here.

Denominations are used almost everywhere as a consequence of accreditation processes, typically managed by the State. Uncertainty is reduced by giving accredited institutions the monopoly of the use of the name, and by preventing others to use the denomination “university.” On the one hand, the coordination between demand and supply based on denominations prevents establishing *internal* differentiation within any given category. While uncertainty is reduced across categories, it remains significant within categories, leaving consumers with a great deal of autonomy in the final judgment. On the other hand, the coordination based on rankings simplifies a large number of dimensions of quality into a small manageable number. Rankings try to reduce the uncertainty *within* categories, at the expense of simplifying the criteria for choice. The prediction is one of incompatibility between denomination and ranking, or more generally, between mechanisms of relative and absolute qualification (Karpik 2007).

In terms of our task, this points to the close connection between classification and measurement issues on one side, the conception of higher education as a quality market on the other, as well as to the conceptual difference between the categorization of HEIs and educational activities and continuous measurement across a scale, providing the “customers” with more power in selecting the provider. Both are related to the lasting debate on steering of higher education, around notions like quasi-markets (Teixeira et al. 2004), competitive differentiation (Deiaco et al. 2010), as well as the emergence of “steering at distance” in New Public Management (Ferlie et al. 2008).

From Mass Education to Differentiation: The New Social Demand for Measurement

If statistical systems are complex devices aimed at governing societies at a distance, where is the demand for introducing a radical innovation such as a register? How did the main actors involved in the project—the European Commission, EURO-STAT and Member States, represented by the National Statistical Authorities (NSA)—respond to this challenge? And finally, what were the implications for the design of the census?

Traditionally, the two fields of statistics involved in EUMIDA—educational and R&D statistics—were largely content with the production of national-level aggregates and indicators. These allowed international organizations like UNESCO, OECD, and the European Commission to track the progress of individual countries based on indicators such as enrolment, graduation rates, and investments in R&D (Godin 2001). Interestingly, the data on students and R&D expenditures did not have to match, as they were considered to be two distinct sectors of economic activity.

More detailed data was produced by national governments for the purpose of steering the system, but they were held by the State administration and not disclosed to the public. This system matched a mode of governance in most European countries where higher education was considered to be part of the public administration, regulated by bureaucratic means concerning budgets and personnel, with the relevant actors in the political game being the State and the academic communities (Clark 1983). The non-existence of HEIs as political actors also implied their non-existence as statistical units—as confirmed by the counterexample of the US where a national census has existed for many years (the Carnegie classification; McCormick and Zhao 2005).

This picture has evolved along with deep modifications to the governance of higher education systems (Paradeise et al. 2009). New public management narratives introduced notions like “steering at distance” and competitive differentiation (Ferlie et al. 2008), considering HEIs as strategic actors which should develop their specific profile (Bonaccorsi and Daraio 2008). Despite differences among countries (Bleiklie et al. 2011), the steering of higher education in Europe moved towards granting more autonomy and requesting accountability from HEIs, introducing evaluation systems (Whitley and Gläser 2007), as well as competition in the allocation of public funds (Jongbloed 2008). These narratives also spurred an increasing request for measurement by other audiences like students wishing to decide where to enroll, funding agencies and private companies in need of evaluating the quality of research. It is fair to say that we largely adhered to a vision of higher education as a competitive quasi-market where HEIs are a strategic actor (with all the nuances that educated scholars might attribute to these notions) and that this vision, shared with the European Commission, constituted the normative foundation of the choices made in constructing the census.

Responses of national statistical systems to these changes have been highly differentiated: institutional-level data on education (students, graduates) were thought to be the least problematic, and in many European countries, HEI-level data

has been made available by the NSA (Bonaccorsi et al. 2007). On the contrary, measures of research output were considered to be highly controversial, as they allowed customers to directly judge quality; hence their production by national authorities has been limited to a few countries who endorse the NPM rationale (e.g. UK with the Research Assessment Exercise). The social demand in this area has been covered by independent providers through different types of rankings; public controversies on their validity are unsurprising as there is no broad social agreement on which measures to use—rankings are social constructions, but they cannot claim to hold a validity endorsed by the State (Weingart and Maasen 2007).

The socio-political context was also shaped by the process of European integration and the increasing role of the European Union in research and higher education policies after the launch of the European Research Area (ERA); while international organizations in the past were largely content to compare countries and to advise national states, the ERA foresees a European space of research where individual actors compete across countries and thus request openness in their national systems and a transfer of competencies from the national to the European level (Kuhlmann 2001). Different scenarios concerning the configuration European research policy raise issues concerning the relationships between global and local and the extent to which a European governance system is preferable to a federal setting allowing adaptation to national specificities. As a matter of fact, the European statistical system in higher education is still largely based on the federal approach developed by the OECD, with EUROSTAT loosely coordinating the NSA's mastering their own data on the grounds of national interest. On the contrary, DG Research has taken a more proactive stance when producing analyses at the HEI level using non-statistical data (European Commission 2011).

Thus, the project was characterized by a complex constellation, where the cooperation of actors with different goals and attitudes was required to achieve a consensus. DG research had the role of promoter, as micro-data on HEIs were a priority in the monitoring of the European Research Area, but at the same time it did not own the resources and the legitimacy to establish long-term data collection. EUROSTAT was involved through its educational division, a logical choice due to the fact that data collection at universities is part of educational statistics (UOE 2006), but which made it difficult to address their specificities (in respect to schools); EUROSTAT was caught between the push of the Commission for the census and concern for the issues it would have raised in terms of relationships with national statistical authorities—as such they stated that their main role was to produce national aggregates, whereas micro-data should have been produced by other actors. In turn, the NSA had different attitudes depending on national cultures and policies; those in countries with a tradition of openness who were respectively adopting new public management practices, held no reservations against a European census but were concerned about two issues: the risk of an additional workload and the possibility that data would be integrated in the census, where no consensus existed at the political level, thus jeopardizing their political legitimacy. On the contrary, in countries with a stronger tradition of centralization and secrecy concerning data considered sensible for national policies the NSA had a negative attitude towards the census as such.

In this context, our role was not to propose technical alternatives for decision-makers, but rather to mediate between ideas and interests by devising solutions that would gather sufficient support to be accomplished (Barré 2004). In this respect, the project was highly successful: the census published by the European Commission only excluded France—whose authorities refused to make HEI level data public—as well as Denmark and the French-speaking part of Belgium for organizational reasons at the NSA. It is reasonable that, if the census becomes widely used, these countries will provide their data.

However, retrospectively this context had deep implications on how the census was constructed. Thus, issues and topics which held a risk of raising the opposition of many national states had to be avoided—such as the measurement of research output, data that was not generally available, or financial data that was considered confidential by some national laws.

A helpful technical device in managing this tension between coverage and completeness was the distinction between a full perimeter, for which we were requested to collect a core set of data, and a restricted perimeter of research-active HEIs, for which the tender requested an extended set of data characterizing research activities. We managed this distinction strategically by including in the core set of data only indicators available in all countries—a few organizational descriptors (like region, foundation years, etc.) and the number of students and staff. Completeness, non-confidentiality and the lack of controversies characterized this dataset and allowed the European Commission to publish it and to foresee a key role of EUROSTAT in a regular data collection. On the contrary, the extended dataset included indicators that would be considered as more interesting for analytical purposes, like data on revenues and expenditures, the share of international students and staff, R&D expenditures. Not only are there many unsolved measurement and comparability issues for this data, they are also unavailable for a substantial share of countries due to their unsystematic collection or for confidentiality reasons (like financial data in most new member states).

We further resorted to a strategy already adopted in the Frascati manual (OECD 2002), where the handbook proposes general guidelines, but leaves the NSA with the responsibility of deciding how to implement hiding national differences behind a shield of uniform definitions. This avoided pushing standardization too far and directly addressing national specificities related to political reasons, to differences in the organization of higher education or diverging practices of data collection, which would have made a consensus impossible and would have impacted national practices.

What are HEIs? The Controversy on Perimeter and Classification

If statistical systems rest on reaching agreements on how to solve controversies, the main controversy in EUMIDA concerned the core of the census, namely the definition of what HEIs are, as well as of operational criteria in order to decide on inclusion in the perimeter.

The conceptual roots of the definition were at stake, since educational statistics are constructed around the notion of educational programs and considerable effort was invested by the NSA to refine their classification by level, while national aggregates are computed by attributing students and teachers to educational levels. In this perspective, higher education meant the aggregation of all curricula classified under levels 5 and 6 in the International Standard Classification of Educational Degrees (ISCED; UOE 2006), usually labeled as tertiary education. From our perspective, this approach, conceptually lacked an understanding of HEIs as organizations. Furthermore, in countries like Germany and Switzerland, professional associations also offer vocational degrees as a supplemental activity, yet these can hardly be considered as *educational institutions*.

Unsurprisingly, issues of legitimacy compounded this controversy: our expertise in constructing statistics was questioned by our statistical partners—as it was considered that legitimacy is rooted in State power and the professional tradition of statisticians. It was pointed out that while in scholarly research data is debatable by definition, statistical data claims a more objective status. Obviously, the distinction is not an epistemological one, but political and functional, statistics cannot fulfill its functions without granting it an objective status and excluding the possibility of revisions.

To address this issue, we adopted different strategies: first, we showed that the debate was related to two different conceptions of higher education, one centered around educational programs, the other one around HEIs as (multifunctional) organizations—a view which matched the goals of DG research, whose main focus was on characterizing the European-level research universities. Second, at the technical level, we proved our competence in educational statistics by carefully referring to the Educational Statistics manual (UOE 2006), but also by questioning its objectivity by displaying inconsistencies in some of its definitions. Third, we strategically used the Swiss and German cases to display that the NSA is not willing to collect institutional-level data in such a broad perimeter for data availability and confidentiality reasons. The outcome of this debate allowed for the construction of the census around an organizational definition of HEIs, which did not yet exist in official statistics, thus transferring some insights of higher education studies into it.

However, functional criteria alone are not sufficient to identify a perimeter, as categories and distinctions between organizational populations are cognitive constructs related to representations of the world that are specific to one audience (Ruef 2000). In our case, the definition of higher education institutions is a political decision, endorsed in most countries through a legal act, with practical implications in terms of status, right to award diplomas and access to public resources. This socially and politically constructed perimeter is continuously evolving, with institutions striving to be included (Kyvik 2004). Unfortunately, there was no single audience to which we could refer, as decisions on inclusion are largely national and similar institutions might be included or not depending on the country.

The EUMIDA handbook defines “*higher education institutions* as entities recognizable as distinct organizations and whose main activity is providing education at the tertiary level, i.e. at the levels 5 and/or 6 of ISCED. Recognizable means that their boundaries can be identified rather unambiguously, they have an

internal organizational structure and, at least in principle, their own budget.” Further, the handbook provides examples and delimitation criteria, including the stipulation that education should be the main activity, thus excluding public research organizations.

Interestingly, it turns out that this definition can hardly be operationalized. For most public organizations, the boundaries criterion is ambiguous—and, if taken seriously, it would have excluded many universities from the perimeter as not having their own budget or even not hiring their own personnel (like in France). This reveals that its function was not operational, but symbolic, to refer to the literature on public-sector organizations which considers the definition of a boundary and control over a budget as central components of their transformation towards corporate actors (Brunsson and Sahlin-Andersson 2000). At the same time, decisions on the inclusion in the perimeter were taken at the national level based on political or legal considerations. This definition, exactly because it cannot be seriously operationalized, allowed *decoupling* the conceptual framework (which needs to be consistent at the European level) from the definition of the perimeter by NSA (where national specificities need to be taken into account).

Nevertheless, a few local controversies emerged. Some categories of institutions, like military schools or arts schools, are only included in some countries; these differences are unavoidable, e.g. if military schools are considered as part of higher education in one country, whereas as internal training activities of the army in another.

Second, while in most countries private institutions accredited by the State were included in the perimeter, controversies emerged in two cases. In Spain, the Statistical Authority refused to include private institutions in the register, claiming they were not part of the public educational system; as a consequence, the perimeter includes only the 47 public universities. In a few countries in Eastern Europe, Statistical Authorities included private institutions in the census, but warned that it would be impossible to obtain financial data. In practice, even though private institutions are accredited by the State and funded, they obtain a treatment under which their financial accounts are not publicly available. We discovered that considerable political controversies were in place in these countries, with a direct impact on the availability of official statistics.

Third, a controversy concerning public research organizations (PRO) emerged, since some PROs oriented towards academic science training of PhD students, even if few of them have the right to award the PhD degree. The criterion that education should be a *major* activity excludes them from the perimeter, which we considered as a fair choice to avoid inclusion of most of public-sector research. However, in Bulgaria, the Academy of Sciences accounts for 1/6 of all PhD degrees and it would have been politically not acceptable to exclude it—a situation reflected in R&D statistics, where the Academy of Sciences and the French CNRS are included in the higher education sector.

In the end, the EUMIDA perimeter includes 91% of the students at the tertiary level in educational statistics (source: EUROSTAT), displaying that the EUMIDA perimeter is not much smaller than the whole tertiary education. Among the 2,400 HEIs in the perimeter, only about 900 are doctorate-awarding institutions. Thus, the

census extends well beyond research universities: not only does it include the non-university sector in binary countries, but also small-scale educational providers like military academies, arts schools, etc. However, 78% of ISCED 5 students are enrolled in the doctorate-awarding institutions and half of the HEIs in the sample include 95% of the ISCED 5 students (the concentration of research activities is expected to be even stronger); accordingly, analyses based on the volume of activity are not very sensitive to the choice of the perimeter, but institutional-level averages (e.g. number of students per HEI) are. In cognitive and political terms, this perimeter represents an unprecedented act of labeling as HEI institutions in the remote periphery of the system, whose long-term implications are difficult to evaluate and which does not fully match our own representation of higher education.

Identifying Research-Active Institutions

The European Commission also requested the identification within the perimeter of a smaller set of so-called *research-active institutions*, for which additional data would have been collected concerning their research activities.

The normative value of a research mandate, with many HEIs in Europe striving to acquire this status (Kyvik and Lepori 2010), makes this definition politically sensible; statistical systems are in this respect powerful instruments of institutionalization and thus actors are likely to look at them in terms of status and political implications. A controversy emerged also in the EUMIDA team itself, as related to different scholarly backgrounds, but also national situations. Those focusing on (academic) research and from countries characterized by unitary systems considered the PhD awarding status as the main criterion, while those with a focus on comparative higher education pointed to the fact that in binary countries non-PhD awarding institutions have a large volume of (applied) research activity (Kyvik and Lepori 2010). Thresholds in terms of volume are also problematic as there is no natural cut-off point, as beyond the top international universities the distribution of research activities is more gradual, with a long tail of institutions performing some (and different types of) research.

Again, the final choice was to adopt an extensive criterion of *institutionalization of research activities*, which can be justified, as a research mission is likely to drive strategic choices, as well as representations of other actors and thus can have practical implications. This definition could also be operationalized in terms of verifiable criteria, like the existence of an official research mandate, presence of research units, inclusion in R&D statistics, awarding of PhDs and regular funding for research projects. This was supported by DG research, which emphasized that *research active* was a broader concept than *research intensive* and thus that the restricted perimeter should not be limited to universities appearing in international rankings.

Among the 2,420 units with full information in the census, 1,405 have been characterized as research-active, including 555 institutions that do not grant the PhD degree. In eight countries, all HEIs in the perimeter are considered as research-

active; most of them are binary systems where non-university institutions have a research mission, like Austria, Switzerland, Finland, and the Netherlands; however, in Norway colleges are not included, even if they have a research mandate (Kyvik and Skodvin 2003). In other countries, specific categories of institutions have been excluded, among them colleges, art schools, and business schools. Other countries made choices based on thresholds (for example, number of PhD students), while very few countries split categories of institutions, like Germany where two-thirds of the Fachhochschulen are considered research-active. These differences are largely related to the organization of national higher education systems, as well as to representations of the importance of the research mandate; clearly, national considerations were more important than using comparable criteria across Europe.

Retrospectively, this issue was far less controversial than expected *from the perspective of political authorities and statistical institutions* as it quickly became clear that it was not realistic to expect the integration of additional data for the research-active institutions in official statistics. As indicated by social conflict theories, controversies are only raised when decisions on classifications bear direct implications in terms of power and resources.

The Conventional Nature of Statistics: Definitions and Measurement Issues

A further task was to select a set of indicators and collect them for a single year, in order to demonstrate the feasibility of the census and to estimate its costs. This phase also displayed the relevance of operational and economic considerations: we had to demonstrate that comparable data could be collected for the whole census and for all countries and to design a system which would not put an excessive burden on the NSA, meaning in practice that no additional data collection could be envisaged.

On the Shoulders of Official Statistics

When looking at indicators, a paradox could be noticed. On the one side, EUMIDA reports declare that their selection is based on a conceptual model of universities as multi-functional organizations and on scholarly work on how to measure these dimensions (Van Vught 2009; Daraio et al. 2011); on the other side, the list of indicators is composed by a few descriptors (like region, year of foundation, degrees awarded) and by the usual indicators from educational statistics, including the number of students, staff, and budgetary data. When looking at some relevant dimensions, *research activities* were measured only through PhD students and *internationalization* through the share of foreign students, while no indicator is provided for *knowledge exchange* and *regional engagement* activities.

Departures from the “ideal” list of indicators are explained by missing data, data quality problems, or by a lack of agreement concerning definitions and methodology for data collection. Following this argument, the EUMIDA team declares to have pragmatically made the best of available data sources.

From the perspective of literature on the social construction of statistical systems, the interpretation is more ambiguous: for example, there is no lack of research output measures—the fact that university rankings exist being a clear witness—what is lacking is a consensus among statistical authorities and national authorities on these measures. Moreover, for some dimensions data exists outside of official statistics, but the project defined the perimeter of available data sources as those collected by the NSA, the official reason being that statistical data are of better quality than those from other sources.

The case of measurement of R&D expenditures in HEIs casts some doubts on this explanation. Namely, that data is regularly collected by the NSA based on a procedure defined in the Frascati manual (OECD 2000), but only about half of the countries regularly perform a survey, while in many countries the calculation is based on national coefficients derived from older surveys or even as “rules of thumb” adopted by the NSA. As a result, in EUMIDA, only about 500 institutions (38% of the research-active institutions) provided data on R&D expenditures and their usability for European-level analysis is questionable. There seems to be a limited willingness to invest in this area because of methodological problems, the high effort required for data collection, and limited political interest.

In this respect, the political context of the census mattered. Namely, we considered that DG research decided to cooperate with EUROSTAT for practical reasons, but for political ones as well, since official statistics are the only source of data endorsed by member states; in this respect, the EUMIDA census can also be seen as a political response to the emergence of international rankings. While in general terms, our quest for a broader set of measures covering HEIs and their activities was reasonable, in the framework of this politico-institutional game, it became a legitimizing argument to support the superiority of educational statistics on other sources. Retrospectively, our behavior largely corresponded to theories of rational behavior which foresees that humans tend to search for alternatives only in the frame of reference they are given and to choose satisficing solutions given cognitive and institutional limitations (Simon 1991), while trying to legitimize them on more objective grounds.

This hints to the fact that scholars involved in indicators design and production should pay attention to their forms of institutionalization because of their lasting impact on the produced indicators (Lepori et al. 2008). The footprint of educational statistics was in fact so strong that even two usual practices in business registers managed by EUROSTAT itself—namely notations for multi-site organizations and for demographic changes like mergers—were not retained and implemented in the data collection.

A less visible but substantial influence of educational statistics concerns definitions. As it was clear that the NSA would reuse available data, we had to take as given the existing definitions from the UOE manual even if these do not always fit the requirements of HEI-level statistics (UOE 2006). One case concerns capital expenditures, which according to the UOE manual should be accounted for on a cash accrual basis, a suitable procedure for public budgets, but questionable for HEIs since they have increasingly been granted financial autonomy.

However, a few adaptations were made. For example, the UOE manual foresees a procedure to clean double counting of degrees in countries where the bachelor is mostly an intermediary diploma, which works reliably only for national-level data; in EUMIDA, it was preferred to single out the long degrees without an intermediary qualification in order to improve comparability. In a few other cases, ongoing improvements of statistical definitions were integrated from the onset, like using the Bologna structure to classify degrees, since this organization of studies is now widespread throughout Europe.

A domain where EUMIDA had more leeway to shape definitions concerned descriptors, as these were considered mostly of scholarly interest, but unlikely to be used for political and funding decisions, and they are non-statistical variables and were collected by EUMIDA experts directly and not by the NSA. Despite their apparent simplicity, some of them led to complex discussions in the EUMIDA team itself. For instance, the year of foundation might be a well-defined concept for Bologna University, but it is more problematic for institutions with a complex history. Hence, the choice was made to include a *foundation year*—the year of foundation of the first traceable ancestor—and a *current status year*—when the institution got the present legal status and configuration. These choices are likely to have a profound impact on the analysis of the demography of HEIs in Europe.

This discussion reveals also a specificity of the process of institutionalization of S&T indicators coherent with sociological theory, namely that the fact that the transition from experimental indicators to official ones, legitimized and objectivized by official statistics, is first a conventional one, and second, when indicators move to a taken-for-granted status existing controversies and methodological issues are removed. While we considered our definitions of HEI descriptors as largely experimental, following a scholarly culture where refinement is considered a normal process, it might well be that they become integrated in an official census without further questioning.

The Ghost of Rankings: Avoiding Controversial Relative Qualifications

The debate on the measures of research output—and its final outcome that no measures were included in the census—is also revealing of the difference between absolute and relative qualification and its deep impact in practice.

Namely, European universities have traditionally benefited, in the European tradition, from absolute qualification, i.e. that the only distinction was between universities and other educational institutions, but no relative qualification in terms of different levels of quality was legitimate.

In the relation with their governments, associations of universities and conferences of rectors have systematically refused the notion that it could be possible to rank national institutions. However, rankings are published and influence the decisions of students, families, and other interested audiences, even if they lack any official recognition. While the increasing mobility of students and international competition on skilled researchers makes it impossible for national governments to control these processes, they remain free to delegitimize rankings pointing to their

methodological limitations, without directly addressing the issue of relative qualification.

In EUMIDA, two statistical offices were against even a simple classification including existing basic indicators such as the number of students and staff. As the officials stated, their governments were actively engaged against the practice of ranking universities; once a census is produced, and basic data is officially published, then nobody can prevent people from calculating their own indicators (such as staff per student, or even cost per student) and draw rankings.

The majority of the NSA endorsed a moderate position, accepting to participate in the creation of a census and considering relative qualification as unproblematic when relating to volumes of activities, without any direct implication in terms of measurement of quality. Most of them were probably aware that the census was likely to be used also for relative qualifications, but their issue was not to make this comparison impossible by keeping the basic data, but rather to avoid being involved in the controversy by introducing these measures directly in the census.

In our view, this controversy conceals profound differences in the underlying conception of HEI steering, on the one side, and of the function of S&T indicators, on the other. In a more bureaucratic and centralizing tradition, the State is the only actor deputed to rule higher education; accordingly, relevant information for managing the system should be available internally, while publicly available indicators should be used instrumentally to justify decisions taken on different grounds (Godin 2005). On the contrary, in a more decentralized conception, where the State is only one among several social actors involved in higher education, openness and availability of information is relevant both for construction of consensus (Barré 2004) and for allowing customers to make informed decisions in quasi-market settings. Accordingly, the design of national higher education policies influences the availability of micro-data as well—in some countries not even the number of employees of universities are publicly available, whereas in others detailed information on research performance is made public.

The opposition, even to a census including indicators with no reference to quality or performance is in our view revealing the profound changes that the bare existence of a census is likely to have. Namely, a great deal of data at the level of individual HEIs is currently available from different sources, like university reports, international databases, and independent S&T indicator producers (Lepori et al. 2008). On the symbolic side, an official census would reinforce the legitimacy of undertaking comparisons at the level of individual HEIs (even if with data from other sources). On the practical side, its availability could help to address some of the criticisms these sources are subject to, like the inability to correctly identify the perimeter—because of the lack of an *official* list of HEIs—or of assessing the coverage of samples derived from bibliometric databases (for example, as share of students or academic staff in the whole perimeter). Obviously, data on students and staff would also allow computing some interesting indicators like productivity per staff at least for those HEIs included in bibliometric databases—and thus strongly enhance the value of other sources.

Against our normative position, that openness and informed choice by customers is relevant for a well-functioning higher education system and should be weighted

more than power interests of ministries or national prestige, our strategy was a rational one; namely, we considered that establishing the principle of a census (and its feasibility) was more important than the choice of indicators to include; in this perspective, accepting that the census did not include many relevant indicators for measuring HEI output was a reasonable decision. Our course of action was more driven by political than by scholarly considerations.

Comparability as a Conventional Notion

Comparability obviously plays a central role in establishing a census: it requires the creation of definitions and categories, the specification of boundary cases, the allocation of individual cases to categories, and finally the comparison among members of the same category. In all these activities there is room for conflicts. If we consider PhD students, a number of common definitional elements are beyond discussion, but there is considerable difference between those systems in which the doctoral degree is considered a complement to undergraduate education to get an employment and those in which it is largely limited to a career in research. Simply comparing numbers of PhD students in international settings does not do justice to reality. Another example is the definition of academic staff: depending on institutional or normative idiosyncratic aspects, what comes to be defined as academic staff is subject to considerable differences. Given the large diversity of higher education institutions in Europe, proper contextualization becomes a central aspect in using and interpreting the data (Almklov 2008).

In doing this, a commonly used practice can be defined as damage avoidance: two things are considered comparable if comparison does not generate harm, distortions, or unacceptable consequences, for (almost) all actors involved. Two relevant issues in this respect are the notion of measurement error and the definition of professional standards concerning data validation.

Sociology of measurement demonstrated that the notion of *measurement error* plays a critical role when bridging local realities with measures (Mallard 1998; Derksen 2000); for example, measurement errors originate from a lack of data, the use of proxies instead of real measures, and imprecision in the measurement techniques—so defining them as errors supports the validity of the measure, but it also specifies the conditions of validity—for example, by allowing control when observed differences between units are statistically meaningful. EUMIDA was, in this respect, not different from other S&T indicators, where a look at methodological information reveals an impressive number of departures from definitions and data-collection procedures which are potential sources of errors (Godin 2005): for example, some countries provided staff data in headcount instead of full time equivalents, different definitions of foreign students were adopted, budgetary data for some countries include only third-party-funds.

As well-educated scholars in quantitative research, we assumed the position that metadata should be considered as an integral part of a dataset and carefully looked at; accordingly, the census is accompanied by a detailed document illustrating departures from definitions. Retrospectively, we recognized that in official statistics

comparability problems are not dealt with through the dissemination of metadata, but through limitation and control of the usage; indicators are produced on request by professional statisticians which are entitled to refuse—based on their expertise—to produce them if not sufficiently robust. This is coherent with their taken-for-granted status and their role as control instrument for the State and enhances their usability and reach.

EUMIDA raises new issues in this respect. First, HEI-level indicators lack the effect of aggregation, where measurement errors at the level of individual units are expected to compensate for national-level averages. Second, it provides data meant to be combined by different audiences with other sources to produce meaningful and relevant indicators and thus means that the NSA is likely to lose their control on usage. Third, while national aggregates are statistical constructions, which can hardly be verified externally, institutional-level data (like numbers of students for individual HEIs) can be checked by the actors themselves.

The conventional nature of comparability does, however, not mean to subscribe to a view of social entities as infinitely malleable. Conventionality is not arbitrariness. When actors come to an agreement, they also define standards and establish procedures to ascertain the conformity of reality to standards. Despite its status as an experimental project, EUMIDA was subject to this request and thus we devised and implemented procedures for data quality and checking, including largely symbolic acts such as the nomination of two quality experts, but also controls of international coherency, statistical checks of outliers, and comparisons with expected figures, as well as control of consistency of national totals with EUROSTAT national aggregates.

It is fair to say that a number of mistakes and departures from definitions were revealed by these procedures, but many more have emerged by using the dataset in a later stage; for instance, data on foreign PhD students in a large European country did not match those in national official publications and, most likely, they were mistaken. This process would have required a much larger effort and thorough testing of the data, while it was difficult to ascertain comparability without detailed knowledge of the underlying reality and control over the data collection processes. Both remarks underline the difficult situation of EUROSTAT when it comes to certifying the validity of HEI-level data in a context where the NSA is explicitly allowed to adapt the guidelines, while the documentation on data collection practices is at best scarce.

Retrospectively, this discussion highlights that the current setting of the European statistical system is tailored to the production of aggregated indicators from micro-data that are not publicly accessible. National aggregates achieve their taken-for-granted status largely because they cannot be compared with other sources and thus the implications of the choices made (as well as possible distortions due to national specificities and data quality problems) are not visible outside the realm of official statistics.

While adhering in EUMIDA to the federal approach borne by the OECD was a convenient political choice for a pilot project, we consider that, should this data be integrated in the European statistical system, stronger standardization and a reduction of national specificities will be required, as well as a more systematic

checking of data provided by the NSA. Otherwise, comparison by users with information from other sources, like university reports, might shed doubt on the reliability of this data, and potentially affect the legitimacy of all educational statistics.

In turn, this would require a deep change in the structure of the European statistical system, from the current federal setting to a more unitary structure where EUROSTAT can edict and enforce more precise guidelines on the type of data to be collected and how this should be done and is allowed to intervene in the case of non-compliance from the NSA. The current institutional setting of European statistics, where in principle, the European Union sets guidelines, but in practice the NSA is under the jurisdiction of national states, largely explains the difficulties in integrating the census in the European statistical system.

Discussion and Conclusions

In this paper we contribute to the sociological analysis of higher education, on the one hand, and measurement and statistics, on the other. By elaborating reflexively on a project whose initial goal was mainly technical, we come to a number of interesting conclusions. First, the process of defining categories for statistical analysis is bound with conceptual controversies. There are very different conceptions of what higher education is, of the legitimacy to deliver it, of the relation between higher education and research. Any categorical definition creates, by its very nature, boundaries that are contested. Second, while controversies are in most cases rooted in concrete clashes of interests, actors almost never formulate their positions in open terms. Power relations are vested into technical discussions and not expressed openly. There is clearly room for critical analysis. Third, social scientists should not advocate the role of knights of objectivity, although this is what policymakers typically ask of them. Rather, the appropriate standing is one of professionalism. Social scientists recognize the complexity of conflicts, support actors in the elaboration of arguments, and contribute to the elucidation of issues at stake. Thus, they would be better off assuming a pragmatic stance, without themselves claiming detachment from the complexity of interests and visions at stake. Rather, they should make their options transparent, asking actors to take a position.

It is relevant to reinterpret this experience in the frame of the studies dealing with the governance of science and, more precisely, with the relationships between experts, on the one side, and policymakers and administrators, on the other (Jasanoff 2004). A growing body of literature has demonstrated that this relationship cannot be represented as a linear process, where either experts produce solutions to political problems or alternatives selected by policymakers based on normative judgments, but rather as an interactive and recursive process, where both sides are jointly involved in the definition of the issues, in the production of scientific expertise, in value judgments and in the implementation of the chosen solution (Weingart 1999).

This helps to reveal the underlying normative and political interests that drove our course of action, beyond the legitimizing position of being purely technical

experts. While broader socio-political developments have created a demand for micro-data, as well as a favorable political environment for their production, the antecedents of EUMIDA have to be sought in a former project where we started HEI-level data collection: we strategically disseminated these results not only in scholarly publications, but also in expert reports and to the European Commission. Thus, we were not ingenuous scholars discovering the political dimension of S&T indicators in the course of action, but rather political entrepreneurs building on our scholarly status to influence public policy.

Second, we were also motivated by normative assumptions on how science policy should be made: in previous scholarly publications, we pointed to the lack of selectivity and performance orientation of European research policies as a major cause of the gap between Europe and the US. As economists, we strongly believe in the virtue of competition and on the value of information to support public and private choices; as citizens we think that in a pluralistic society information should be disclosed and support the public debate, rather than kept by administrators—we endorse an *agora model* of S&T indicators (Barré 2004) and thus we prefer models of S&T indicator-production where different sources and institutions compete to a monopoly by National Statistical Authorities (Lepori et al. 2008). As in the good tradition of scientific expertise, these assumptions are supported by arguments and empirical evidence—they are *reasonable*, but have a political rather than a scientific status.

Third, our interests were at stake, too. Promoting the use of S&T indicators for policymaking means displacing other types of scientific expertise with the one of S&T indicators, designers, and producers; this translates to stronger influence in the political arena, but in the scholarly community as well—admittedly, the field of S&T indicators is mostly financed by these projects and thus we were not just scientists illuminating political processes, we were also fighting for the (political, scholarly, and economic) development of our field. The project gave us access to a unique dataset on European higher education, on which we are publishing intensively; the alliance with policy and statistical offices allows for control over a critical resource for scholarly work, namely data. Of course, sociology of science demonstrated that power and interests are essential components of the scientific enterprise (Latour and Woolgar 1979); what is at stake, however, is transparency concerning the complex nature of these processes and the fact that, when dealing with boundary work with politics, interests and power relationships cannot be separated from ideal reasons and scholarly competence. What our account makes clear is that over the course of this project, we assumed the role of good citizens trying to improve how our society is governed, the one of scholars driven by our scientific interests and insights and of entrepreneurs constructing the material and political basis of our business.

Interestingly, this leads us to reinterpret somewhat differently the paradox of indicator production and circulation, namely the fact that once produced, indicators become social objects themselves. They turn out to be independent not only from the official source that has produced them, but even more so from social scientists that have contributed to the definition and from social actors whose interests and values have clashed. As scholars we were always concerned by the fact that, in this

process, the complexity and nuances of the construction of indicators is largely lost and we considered it dangerous that indicators are used without all the required methodological cautions. Yet, exactly this taken-for-granted status largely accounts for their role in the decision-making process and explains why their construction is so highly controversial—and why scholars and policymakers are willing to invest so much in this process: very much like other technical and organizational artifacts, once established, S&T indicators embody in a durable manner, value choices and power relationships and thus have a lasting influence on future decisions. By shaping indicators, we are shaping the future governance of science systems as well, in a way that we cannot, however, fully control.

Acknowledgments The authors would like to thank the European Commission which funded the EUMIDA project, as well as EUROSTAT for their continuing discussion and support. The views expressed in this article are those of the authors and do not necessarily reflect the opinion or position of the European Commission. We would like to thank the other members of the EUMIDA core team, namely Tasso Brandt, Daniela De Filippo, Andreas Niederl, Ulrich Schmoch, Stig Slipersaeter as well as the national experts who contributed to the data collection. Finally, we would like to thank two anonymous reviewers for useful comments.

References

- Almklov, P.G. 2008. Standardized data and singular situations. *Social Studies of Science* 38: 873.
- Barré, Rémi. 2004. S&T indicators for policy making in a changing science-society relationship. In *Handbook of quantitative science and technology research*, eds. H.F. Moed, W. Glänzel, and U. Schmoch, 115–132. Dordrecht: Kluwer.
- Barzel, Yoram. 1982. Measurement cost and the organization of markets. *Journal of Law and Economics* 25(1): 27–48.
- Berger, Peter L., and Thomas Luckmann. 1966. *The social construction of reality*. 1967: Garden City, NY: Doubleday Anchor.
- Blank, Grant. 2007. *Critics, ratings and society. The sociology of reviews*. Plymouth: Rowman & Littlefield.
- Bleiklie, Ivar, Juergen Enders, Benedetto Lepori, and Christine Musselin. 2011. New public management, network governance and the university as changing professional organization. In *Ashgate research companion to new public management*, eds. T. Christensen and P. Lægreid. Aldershot: Ashgate.
- Bonaccorsi, Andrea, and Cinzia Daraio. 2008. The differentiation of strategic profile of higher education institutions. New positioning indicators based on microdata. *Scientometrics* 74(1): 3–25.
- Bonaccorsi, Andrea, Cinzia Daraio, Benedetto Lepori, and Stig Slipersaeter. 2007. Indicators on individual higher education institutions: Addressing data problems and comparability issues. *Research Evaluation* 16: 66–78.
- Bourdieu, Pierre. 1984. *Distinction: A social critique of the judgment of taste*. London: Routledge.
- Bowker, G. S., and S. Leigh Star. 1999. *Sorting things out. Classification and its consequences*. Cambridge, Mass.: The MIT Press.
- Brunsson, N., and B. Jacobsson. 2000. *A world of standards*. Oxford: Oxford University Press.
- Brunsson, N., and K. Sahlin-Andersson. 2000. Constructing organizations: The example of the Public Sector Reform. *Organization Studies* 21(4): 721–746.
- Bryant, R. 2000. *Discovery and decision. Exploring the metaphysics and epistemology of scientific classification*. London: Associated University Presses.
- Clark, Burton R. 1983. *The higher education system. Academic organization in cross-national perspective*. Berkeley: University of California Press.
- Daraio, Cinzia, Andrea Bonaccorsi, Aldo Geuna, B. Lepori, et al. 2011. The European university landscape: A micro characterization based on evidence from the Aquameth project. *Research Policy* 40(1): 148–164.

- Deiaco, E., M. Holmén, and M. McKelvey. 2010. What does it mean conceptually that universities compete? In *Learning to compete in European Universities*, eds. M. McKelvey, and M. Holmén, 300–328. Cheltenham: Edward Elgar.
- Derksen, L. 2000. Towards a sociology of measurement: The meaning of measurement error in the case of DNA profiling. *Social Studies of Science* 30: 803.
- Desrosières, Alain. 2008. *Gouverner par les nombres*. Paris: Presses de l'école des mines.
- Desrosières, Alain. 1993. *La politique des grands nombres. Histoire de la raison statistique*. Paris: La découverte.
- Desrosières, Alain, and Laurent Thevenot. 1988. *Les catégories socio-professionnelles*. Paris: La Découverte.
- Duprè, John. 1993. *The disorder of things*. Cambridge, Mass.: Harvard University Press.
- Espeland, W., and M. Stevens. 1998. Commensuration as a social process. *Annual Review of Sociology* 24: 313–343.
- European Commission. 2011. *Innovation union competitiveness report*. Brussels: European Commission.
- Ferlie, Ewan, Christine Musselin, and Gianluca Andresani. 2008. The steering of higher education systems: a public management perspective. *Higher Education* 56: 325–348.
- Godin, Benoit. 2005. *Measurement and statistics on science and technology*. London: Routledge.
- Godin, Benoit. 2001. The emergence of science and technology indicators: Why did governments supplement statistics with indicators? *Research Policy* 32(4): 691.
- Jasanoff, Sheila. 2004. *States of knowledge. The co-production of science and social order*. London: Routledge.
- Jongbloed, Ben. 2008. Creating public-private dynamics in higher education funding. A discussion of three options. In *Public-private dynamics in higher education funding*, eds. J. Enders and B. Jongbloed, pp. 113–138. Bielefeld: Transcript.
- Karpik, Lucien. 2007. *L'économie des singularités*. Paris: Gallimard.
- Kuhlmann, Stefan. 2001. Future governance of innovation policy in Europe—three scenarios. *Research Policy* 30: 953–976.
- Kyvik, Svein, and Benedetto Lepori. 2010. *Research in the non-university higher education sector in Europe*. Dordrecht: Springer.
- Kyvik, Svein. 2004. Structural changes in higher education systems in Western Europe. *Higher Education in Europe* 29(3): 393–409.
- Kyvik, Svein, and Ole-Jacob Skodvin. 2003. Research in the non-university higher education sector—tensions and dilemmas. *Higher Education* 45: 203–222.
- Latour, Bruno. 1987. *Science in action: How to follow engineers and scientists through society*. Milton Keynes: Open University Press.
- Latour, Bruno, and Steve Woolgar. 1979. *Laboratory life. The construction of scientific facts*. New Jersey: Princeton University Press.
- Lepori, Benedetto, Rémi Barré, and Ghislaine Filliatreau. 2008. New perspectives and challenges for the design and production of S&T indicators. *Research Evaluation* 17: 33–44.
- Mallard, A. 1998. Compare, standardize and settle agreements: On some usual metrological problems. *Social Studies of Science* 28: 571.
- McCloskey, Michael, and Sam Glucksberg. 1978. *Natural categories: Well defined or fuzzy sets?* New York: Springer.
- McCormick, A.C., and C.M. Zhao. 2005. Rethinking and reframing the Carnegie classification. *Change* 37(5): 50–57.
- Murphy, Gregory, and Douglas Medin. 1985. The role of theories in conceptual coherence. *Psychological Review* 92(3): 290–316.
- OECD. 2002. *Frascati manual. Proposed standard practice for surveys on research and experimental development*. Paris: OECD.
- OECD. 2000. *Measuring R&D in the higher education sector. Methods in use in the OECD/EU member countries*. Paris: OECD.
- Paradeise, Catherine, Emanuela Reale, Ivar Bleiklie, and Ewan Ferlie. 2009. *University governance. Western European comparative perspectives*. Dordrecht: Springer.
- Porter, T. 1995. *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton: Princeton University Press.
- Ruef, Martin. 2000. The emergence of organizational forms: A community ecology approach. *The American Journal of Sociology* 106(3): 658–714.

- Simon, Herbert A. 1991. Bounded rationality and organizational learning. *Organization Science* 2(1): 125–134.
- Teixeira, Pedro, Ben Jongbloed, David Dill, and Alberto Amaral. 2004. *Markets in higher education. Rhetoric or reality?* Dordrecht: Kluwer.
- Thévenot, Laurent. 2001. Organized complexity. Conventions of coordination and the composition of economic arrangements. *European Journal of Social Theory* 4: 405–425.
- UOE. 2006. *UOE data collection on education systems. Manual. Concepts, definitions, classifications.* Montreal, Paris, Luxembourg: UNESCO, OECD, EUROSTAT.
- Van Vught, Frans. 2009. *Mapping the higher education landscape. Towards a European classification of higher education.* Milton Keynes, UK: Springer.
- Weingart, Peter. 1999. Scientific expertise and political accountability: Paradoxes of science in politics. *Science and Public Policy* 26(3): 151–161.
- Weingart, Peter, and Sabine Maasen. 2007. Elite through rankings—the emergence of enterprising university. In *The changing governance of the sciences*, eds. R. Whitley, and J. Gläser, 75–1011. Dordrecht: Springer.
- White, Harrison C. 1981. Where do markets come from? *American Journal of Sociology* 87: 517–547.
- Whitley, Richard, and Jochen Gläser (eds.). 2007. *The changing governance of the sciences. The advent of research evaluation systems.* Dordrecht: Springer.