

ISSN 2281-4299



DIPARTIMENTO DI INGEGNERIA INFORMATICA
AUTOMATICA E GESTIONALE ANTONIO RUBERTI



SAPIENZA
UNIVERSITÀ DI ROMA

**A framework for the Assessment of
Research and its impacts**

Cinzia Daraio

Technical Report n. 4, 2017

A FRAMEWORK FOR THE ASSESSMENT OF RESEARCH AND ITS IMPACTS

CINZIA DARAIO

Department of Computer, Control and Management Engineering

A. Ruberti (DIAG), Sapienza University of Rome, Italy.

`daraio@dis.uniroma1.it`.

March 2017

Abstract

This paper proposes a framework for the development of *models* for the assessment of research activities and their impacts. It distinguishes three dimensions: theory, methodology and data, each of which is further characterized by three main building blocks: education, research and innovation (*theory*); efficiency, effectiveness and impact (*methodology*); and availability, interoperability and “unit free” property (*data*). The different dimensions and their nine constituent building blocks are attributes of an overarching concept, denoted as “quality”. Three additional quality attributes are identified as implementation factors (tailorability, transparency and openness) and three “enabling” conditions (convergence, mixed methods and knowledge infrastructures) complete the framework.

The paper illustrates the complexity of the evaluation describing the *generalized* “implementation problem” in research assessment, according to the proposed framework. A framework is required to develop models of metrics. Models of metrics are necessary to assess the meaning, validity and robustness of metrics. The proposed framework can be a useful reference for the development of the *ethics* of research evaluation. Three examples of application, as well as further directions for future research are provided.

Key Words: Evaluation of Research, Efficiency, Effectiveness, Impacts, Modelling, Implementation Problem, Responsible Metrics, Ethics of Research Evaluation.

Contents

1	Introduction and Main Contribution	1
2	The Framework	8
2.1	Theory	8
2.2	Methodology	11
2.3	Data	14
3	A Summary View and a Pragmatic Perspective	16
4	Model Selection, Implementation Factors and Enabling Conditions	19
5	Towards Responsible Metrics?	21
6	Three Examples of Application	25
6.1	A Doubly Conditional Performance Evaluation Model	25
6.2	Interpretation of university rankings	28
6.3	Setting Priorities of ANVUR	29
7	Conclusions and Further Research	30
8	Acknowledgements	32
A	Appendix: The Implementation Problem in Research Assessment	33

1 Introduction and Main Contribution

Recent trends in the policy of research and its development include, among others:

- the *explosion of the assessments* in the “evaluation” society (Gläser and Whitley, 2007; Dahler-Larsen, 2012).
- the *need of policy-makers* to have a comprehensive framework. We refer to the STAR metrics and UMETRICS in US (Lane, 2009; Lane et al. 2015)¹ and to the European Commission (2014) “Expert Group to support the development of tailor-made impact assessment methodologies for ERA (European Research Area)” in Europe.²
- the *criticisms* of the traditional assessment metrics. The traditional methods of research evaluation have been recently under attack in different contexts, in particular by the *San Francisco Declaration on Research Assessment* (DORA) and the Leiden Manifesto (Hicks et al., 2015) for the inherent problems of the evaluation of research, although some of the crucial limits and problems have already been known to the specialized community for decades; see e.g. Glänzel and Schoepflin (1994); Glänzel (1996) and Moed and Van Leeuwen (1996). A recent review on the role of metrics in research assessment and management (Wilsdon et al. 2015) has found that: “There is considerable scepticism among researchers, universities, representative bodies and learned societies about the broader use of metrics in research assessment and management” as one of the main findings of the study.
- the *crisis of science*. Benessia et al. (2016), identify the most heated points of discussion in: reproducibility (see Munafò et al. 2017), peer-review, publication metrics, scientific leadership, scientific integrity and the use of science for policy (see also “The end of the Cartesian dream” in Saltelli and Funtowicz, 2015).³
- the recent *debate on modelling* of research and innovation activities and on the use of qualitative or quantitative models for the analysis of science and innovation policies (Martin, 2016).

The advent of the *big data era* is another main recurring trend.

Recently, innovative data sources and tools offer new ways of studying science and technology and more data-driven knowledge discovery (Ding and Stirling, 2016). At the same time, these sources are casting some doubts on the extensive use of traditional data sources used by the scholars in the field (Feldman, Kenney and Lissoni, 2015). The results obtained are obviously linked to intrinsic potential or limitations

¹For more information see: <https://www.starmetrics.nih.gov/> and <https://www.btaa.org/projects/umetrics>.

²The first objective of the European Commission (2014) Expert Group was as follows: “Propose an analytical framework for identifying how the implementation of different ERA priorities and components observed at institutional level (i.e. research performing organisations) and national level (i.e. national policies and funding organisations policies) impact on the research system performance (at institutional and national level).”

³The transmission channel of this crisis from science to scientific advices is attributed to the collapse of the dual legitimacy system which was the basis of modernity, namely, the arrangement by which science provided legitimate facts, and policy, legitimate norms. The obsolescence of the classical opposition between scientific approach and dogmatic approach, generated by the problems of the empirical evidence (Saltelli and Funtowicz, 2015) may be a possible root of this crisis.

in the kind of data used in the analysis. This tendency, has led to the “computerization” of bibliometrics that has been linked to the development of altmetrics approaches (Moed, 2016). Is science really becoming increasingly data-driven? Are we moving towards a data-driven science (Kitchin, 2014), supporting “the end of theory” (Anderson, 2008), or will theory-driven scientific discoveries remain unavoidable (Frické, 2014)? There is little agreement in the literature. More balanced views emerging from a critical analysis of the current literature are also available (Ekbja et al, 2015; Debackere, 2016), leading the information systems community to further deeply analyse the critical challenges posed by the Big data development (Agarwal and Dhar, 2014). Data indeed “are not simply addenda or second-order artifacts; rather, they are the heart of much of the narrative literature, the protean stuff that allows for inference, interpretation, theory building, innovation, and invention” (Cronin, 2013, p. 435). Making data widely available is very important for scientific research as it relates to the responsibilities of the research community toward transparency, standardization, and data archiving. However, to make data available, researchers have to face the huge amount, complexity, and variety of the data that are being produced (Hanson, Sugden, Alberts, 2011). Moreover, the availability of data is not homogeneous for all disciplines and the cases of “Little data” and “No data” are not exceptions (Borgman, 2015).

These recent trends and the issues they underline, require a new framework for the analysis.

The theoretical framework (intended as a group of related ideas) that we propose in this paper is designed to be a reference for the development of models for the assessment of the research activities and their impacts. A framework is required to develop models of metrics. Models of metrics are necessary to assess the meaning, validity and robustness of metrics. We claim that our framework can support the development of the appropriate metrics for a given research assessment problem or for the understanding of existing metrics.

This is a very difficult question because, among other things, it refers to a complex phenomenon for which there is the lack of a *reference* or a *benchmark* to compare the metrics against. The purpose of our proposed framework is exactly to offer a *reference* to develop models of research assessment.

In this paper, *indicators* are *combinations* of data that produce figures, while *metrics* are considered as parameters or measures of quantitative assessment used for measurement, comparison or to track performance⁴. Hence, an indicator is a metric if it is used as a parameter in a research assessment. It is more difficult to develop metrics than indicators due to the “implementation” problem that we will discuss in the paper.

It is important to develop models for different reasons, including:

- learning*: to learn about the explicit consequences of assumptions, test the assumptions, highlight relevant relations;
- improving*, to better operate, document/verify the assumptions, decompose analysis and synthesis, systematize the problem and the evaluation/choice made, explicit the dependence of the choice to the scenario.

⁴Often, indicators and metrics are used as synonym. See also Wilsdon et al. 2015.

More specifically, a *model* is an abstract representation, which from some points of view and for some ends represents an object or real phenomenon. The representation of reality is achieved through the *analogy* established between aspects of reality and aspects of the model.

For quantitative models the analogy with the real world takes place in two steps:

1. *quantification* of objects, facts and phenomena in an appropriate way;
2. *identification* of the relationships existing between the previously identified objects, closest to the reality (that is the object of the model).

The practical use of a model depends on the different roles that the model can have and from the different steps of the decisional process in which the model can be used. A model can be considered as a tool for understanding the reality. The potentiality of models can be expressed for description, interpretation, forecasting and intervention. These different roles may be correlated or not, depending on the objective of the analysis and the way the model is built.

To be successful the modelling has to take into account the specificities of the processes and systems under investigation, and in particular consider that the behaviour is free and finalized to given aims; history and evolution matter as the behaviour of systems and process changes over time.

The modelling activity related to the assessment of research involves several methodological challenges.⁵

Evaluation is a complex activity that consists of at least three *levels of analysis*: outputs, processes and purposes.

The *finalization* of the analysis to the specific evaluation problem can help to specialize and simplify components, identifying those relevant for the purpose. The finalization may encourage a functional analysis of the systems. The external behaviour of the systems may be explained focusing the analysis to their aims and to their ways of interacting with the environment without entering into the details of the internal structures and organization (the organization becomes relevant only if it is a limit to pursuing the objectives of the system).

Some *pitfalls* of models are:

- *Theoretical* limits (limitation of the concepts and their relations considered relevant in the models);
- *Interpretative* and *forecasting* limits (uncertainty of the phenomena, necessity of exogenous assumptions, errors in the estimates, approximation between model and theory, deviations between theory and reality, evolution of behaviours);

⁵What is required is to develop models, able to characterize strongly connected or interdependent model components, dominated by their interactions, including complex model behaviour, emergent collective behaviour implies new and often unexpected model behaviour, counter intuitive behaviour and extreme events with less predictable outcomes, management based on setting rules for bottom up self-organization (Helbing and Carbone, 2012, p.15). This is very different from the traditional models, characterized by independent model components, based on simple model behaviour, where the sum of properties of individual components characterizes model behaviour, conventional wisdom works well and a well predictable and controllable top-down model seems to be inappropriate to capture the complexity and dynamics involved in the research assessment.

- Limits in the *decision context* (quantifiability of the objectives, multiplicity and variety of objectives, predictability of the external effects of the decisions, interdependencies with decisions of other subjects, computational complexity, and implementation of the decisions).

There are some *difficulties*, which arise in modeling:

1. Possibility that the targets are not quantifiable, or are multiple and conflicting; or that there are several decision makers with different interests;
2. Complexity, uncertainty and changeability of the environment in which the controlled system works and, after environmental stimuli, the difficulty of predicting the consequences of certain actions and relative responses;
3. The limits (in particular of an organizational nature) within which the controlled system adapts to the directives of the decision maker;
4. The intrinsic complexity of calculation of the objective of the analysis.

The ambition of our framework is to be a general basis able to frame the main dimensions (features) relevant for developing multidimensional and multilevel models for the evaluation of research and its impacts⁶.

We propose a framework, illustrated in Figure 1, based on three dimensions:

- *Theory*, broadly speaking, identifies the conceptual content of the analysis, answering the question “what” is the domain of interest, and delineating the perimeter of the investigation.
- *Methodology*, generally refers to “how” the investigation is handled, what are the kind of tools that can be applied to the domain of interest, tools which represent the means by which the analyses are carried out.
- *Data*, largely, and roughly, are instances coming from the domain of interest, and represent the means, by which the analyses are carried out.

We detail each dimension in three main building blocks and identify three operational factors for implementation purposes. The main building blocks of Theory are: 1. EDUCATION⁷, 2. RESEARCH⁸, 3. INNOVA-

⁶For a systematic presentation of unit and levels of analysis in research assessment, see e.g. Vinkler (2010).

⁷In general, education is the process of facilitating the acquisition or assignment of special knowledge or skills, values, beliefs and habits. The methods applied are varied and may include storytelling, discussion, teaching, training and direct research. It is often done under the guidance of teachers, but students can also learn by their self. It can take place in formal or informal settings and can embrace every experience that has a formative effect. Education is commonly organized into stages: preschool, primary school, secondary school and after that higher education level. See the International Standard Classification of Education (ISCED, 2011) for a more technical presentation.

⁸According to the OECD’s Frascati Manual (2002), Research and Development (R&D) is the “creative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of man, culture and society, and the use of this stock of knowledge to devise new applications”. The term R&D covers three activities: basic research, applied research and experimental development. Basic research is experimental or theoretical work undertaken primarily to acquire new knowledge of the underlying foundation of phenomena and observable facts, without any particular application or use in view. Applied research is also original investigation undertaken in order to acquire new knowledge. It is, however, directed primarily towards a specific practical aim or objective. Experimental development is systematic work, drawing on existing knowledge gained from research and/or practical experience, which is directed to producing new materials, products or devices, to installing new processes, systems and services, or to improving substantially those already produced or installed. R&D covers both formal R&D in R&D units and informal or occasional R&D in other units. See also the more recent Frascati Manual (2015).

TION⁹. The main building blocks of Methodology are: 1. EFFICIENCY, 2. EFFECTIVENESS, 3. IMPACT. The main building blocks of Data are: 1. AVAILABILITY, 2. INTEROPERABILITY, 3. *Unit-free* PROPERTY.

 FIGURE 1 HERE

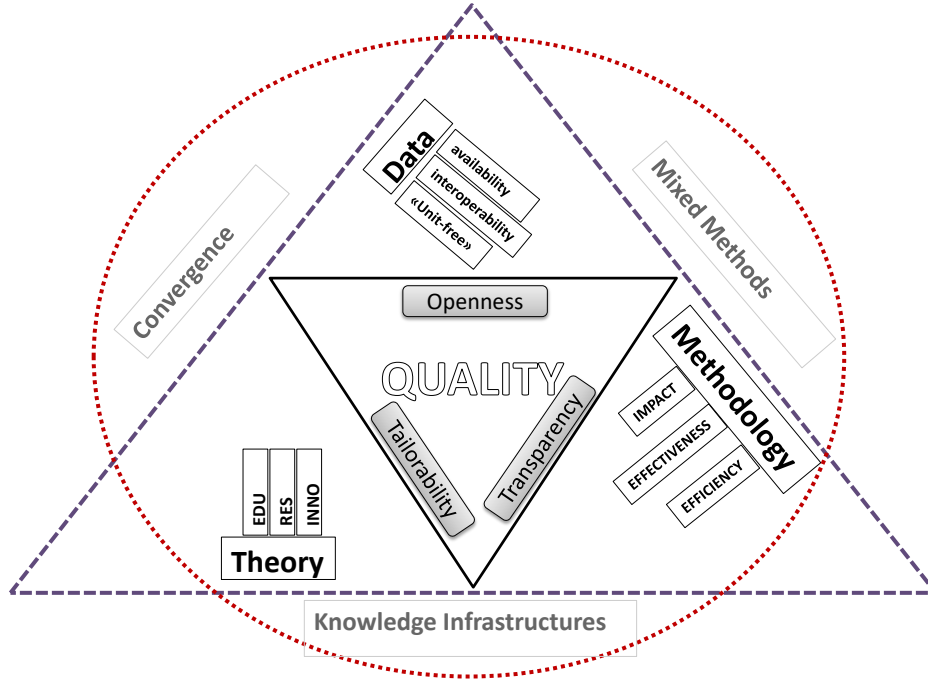


Figure 1: *An illustration of our framework including its three-implementation factors (tailorability, transparency and openness) and its three-enabling conditions: convergence, mixed methods and knowledge infrastructures.*

The problem of evaluation of the research activities, in our set-up, is framed in a systematic way, taking into account also education and innovation together with the other components of the Methodology and Data dimensions.

The main three implementation-factors (see Section 4) we propose are:

1. TAILORABILITY (broadly, the adaptability to the features of the problem at hand)
2. TRANSPARENCY (approximately, description of the choices made and underlying hypothesis masked in

⁹According to the Oslo Manual (2005), an innovation is “the implementation of a new or significantly improved product (good or service), or process, a new marketing method, or a new organizational method in business practices, workplace organization or external relations. The minimum requirement for an innovation is that the product, process, marketing method or organizational method must be new (or significantly improved) to the firm. Innovation activities are all scientific, technological, organizational, financial and commercial steps which actually, or are intended to, lead to the implementation of innovations. Innovation activities also include R&D that is not directly related to the development of a specific innovation.”

the proposed/selected theory/method/data combination)

3. OPENNESS (roughly, accessibility to the main elements of the modelling).

The more we are able to go to the deep, fine-grain of the most atomic level-unit of analysis (i.e. the higher the level of tailorability), the higher the level of openness and transparency, the better will be the conceptualization and formalization of quality within a model.

In this paper, we assert that the ability of developing (and afterwards understanding and using effectively) models for the assessment of research is linked and depends, among other factors, on the degree or depth of the conceptualization¹⁰ and formalization¹¹, in an *unambiguous* way, of the underlying idea of quality. QUALITY, here, is intended as “fitness for use”.

The level of conceptualization and formalization of quality, however, is neither objective nor unique. It depends on the purposes and the *subject* or unit of the analysis (e.g. scholars, groups, institutions, up to meso or macro aggregated units, as regional or national entities) and it relates, in the end, to the specific evaluation problem under investigation.

We propose, finally, three *enabling conditions* that foster the connection of our framework with the empirical and policy worlds. The three enabling conditions are:

1. CONVERGENCE (as an evolution of the *transdisciplinary* approach, which allows for overcoming the traditional paradigms, increasing the dimensional space of thinking);
2. MIXED METHODS (as an intelligent combination of quantitative and qualitative approaches);
3. KNOWLEDGE INFRASTRUCTURES (as networks of people that interacts with artifacts, tools and data infrastructures).

We maintain that these three enabling conditions contribute to the conceptualization and formalization of the idea of quality that is related and foster the overlap of the different perspectives, namely modelling world, empirical world and policy world (see more in the following).

¹⁰Conceptualization here has to be intended as the formulation of the content of the general ideas and of the most important details.

¹¹Formalization here has to be intended as “to make it official” or refers to a defined structure.

Summing up, evaluating research and its impacts is a real complex task. Perhaps the key problem is that research performance is not fully quantifiable. Hence, research assessment has to deal with *non-fully quantifiable* concepts. There are several approaches to evaluating research. In order to adopt and use our framework, the following three postulates, intended as general validity conditions or principles, have to be accepted.

Postulate 1: Models of metrics

Each metric is based on at least one model. The model can be implicitly or explicitly defined and discussed.

This postulate is a proposition that we assume to be true because it is obvious. The implication of Postulate 1 is that if the model underlying the metric is not described, this does not mean that it is more robust to modelling choice. It simply means that you do not explicit and account for the underlying theoretical choices, methodological assumptions and data limits considerations. Put in other words, the metric can not be more robust than the model, and it is possible to assess the robustness of the model only if it is explicitly described.

Postulate 2: Conceptualization and formalization of “quality”

The accuracy, completeness and consistency of the research assessment depends on the level of conceptualization and formalization, in an unambiguous way, of the different layers and meanings of “quality” in our framework.

This is the cornerstone postulate of our framework. The accuracy, completeness and consistency of the research assessment depends upon and is limited by, among other factors, the *complexity of the research evaluation* which emerges from the description of the implementation problem (see Appendix A).

Postulate 3: “Responsible” Metrics

A metric developed according to a model that conceptualizes and formalizes in an unambiguous way the idea of quality in its different layers/meanings (according to our framework) is able to substantiate and give content to the concept of “responsible” metrics.

Postulate 3 should be considered as an open conjecture that needs to be further studied and demonstrated (see further discussion in Section 5).

The main contributions of the paper are:

- to introduce a simple framework that could be helpful in developing models for metrics of research assessment;
- to shed some light on the complexity of the research assessment process, by addressing the *implementation* problem in a generalized way, connected to the proposed framework;
- to propose a basis for the research of the ethics of research evaluation;
- to offer three examples of possible use of the framework;

- to outline directions for further research.

Our framework acts as a *common denominator* for different analytical levels and relevant aspects and is able to embrace many different and heterogeneous streams of literature along its three main dimensions. An outline is described in the next section.

2 The Framework

2.1 Theory

For Theory, we mean the set of general ideas or notions that defines and delineates the boundary of the investigation. In this paper, we are interested in the assessment of the research activity and its impact.

Research is an important driver for innovation, economic progress and social welfare (e.g. Adams, 1990; Griliches, 1998; Henderson et al. 1998; Mansfield, 1995; Rosenberg and Nelson, 1994). Scientific activities produce spillovers that have short and medium term effects on industrial innovation (Mansfield, 1991).

Salter and Martin (2001) review the works on the economic benefits of publicly funded basic research.¹²

Table 1 reports some streams of literature which have considered research and innovation, which are somewhat overlapping, as the main interplay of Science and Society together with education.

TABLE 1 HERE

¹²They classify three main methodological approaches that have been adopted: econometric studies, surveys and case studies. They detect six main categories of benefits of publicly funded basic research: increasing the stock of useful knowledge; training skilled graduates; creating new scientific instrumentation and methodologies; forming networks and stimulating social interaction; increasing the capacity for scientific and technological problem-solving; creating new firm. According to the authors, the relevance of these different forms of benefit differs along with the scientific field, technology and industrial sector. Hence, they conclude: “consequently, no simple model of the economic benefits from basic research is possible”.

Table 1: *A non-exhaustive overview of the literature on the Theory dimension.*

<i>Literature stream</i>	<i>References</i>
Economics of science ¹³ and technology	Audretsch et al. (2002) Stephan (2012a,b) Mirowski and Sent (2002) ¹⁴
Theories of Growth ¹⁵	Aghion, Howitt (2009) Solow (1957), Abramovitz (1956) Nelson, Phelps (1966), Romer (1986, 1994)
Quantitative science ¹⁶ and technology research	Moed, Glänzel, Schmoch (2004) Ding, Rousseau, Wolfram (2014) Egghe, Rousseau (1990), Egghe (2005) Cronin, Sugimoto (2014, 2015)
Innovation studies ¹⁷ Economics of innovation ¹⁸	Martin (2012, 2013) Hall, Rosenberg (2010)
Science of Science ¹⁹ Science and Society ²⁰	Fealing et al. (2011, p. 4) Gibbons et al. (1994) Etzkowitz, Leydesdorff (2000) Edquist (2001), Aghion et al. (2009) Helbing, Carbone (2012) Scharnhorst et al. (2012)
Economics of Knowledge	Antonelli and Link (2014)
Economics of Education	Blaugh (1966); Johnes, Johnes (2004) Checchi (2006) Hanushek et al. (2016)
Education and Society	Roper, Hirth (2005) Texeira et al. (2004, 2011)
Societal impact Science and Public Policy studies	Bornmann (2013) Ebrahim et al. (2014) Perkmann et al. (2013) Veugelers and Del Rey (2014) Hill (2016)

From the economics of education we know that education is an investment in human capital analogous

¹³It draws on the fields of economics, public policy, sociology and management; it includes also an *interdisciplinary* economics of science.

¹⁴Geuna (2003, p. 460) states that “The major attraction of this book lies in its ability to put together in an organized way essays that examine the ‘economics’ of science from different disciplinary perspectives: economics, philosophy and sociology”.

¹⁵Including old and new ones.

¹⁶It studies quantitative studies of science system, of technology system and of science-technology interface. The focus here is -though not exclusively- on scholarly publications and patents, it embraces bibliometrics, scientometrics and informetrics, more recently starting to consider also other non-scholarly and societal “altmetrics” dimensions.

¹⁷As the “economic, management organizational and policy studies of science, technology and innovation with a view to providing useful inputs for decision-makers concerned with policies for, and the management of science, technology and innovation” (Martin, 2012).

¹⁸It is at the core of several different economic fields Including macroeconomics, industrial organization (strategies and interactions of innovative firms), public finance, policies for encouraging private sector innovation, and economic development (innovation systems and technology transfer).

¹⁹Aiming “to develop the evidentiary basis for decision making by policy practitioners”. It is based on an interdisciplinary community including, non-exhaustively, economics, engineering, the history of science, operations research, physics, political science, psychology, and sociology (National Research Council, 2014a; Largent, Lane, 2012; Lane et al. 2015).

²⁰It include interplay models, including socio-economic as well as complex dynamic models.

to an investment in physical capital.

People represent the *link* between all these streams of literature. People in fact attend schools and higher education institutions, acquiring competences and skills. People are educated first and after that do research and carry out innovative activities during which they continue to learn, acquiring/extending their competences and skills and so on.²¹

The existing literature, summarized in Table 1, can be systematized around the *knowledge production activity*, defined in a broad way as “a complex of ideas, methods, norms, values, that are the cognitive and social norms which must be followed in the production, legitimation and diffusion of knowledge (Gibbons et al. 1994, p.2)” which is based on processes: sets of activities performed by agents, through time.

These knowledge activities include stock of inputs (including for instance cumulated results of previous research activities in relevant publications, and embodied in authors competences and potential); infrastructural assets; flows of inputs (such as the time devoted by a group to a current research project); time and resources devoted to teaching and service activities; joint effect of resources in teaching activities; competence of teachers; skills and the initial level of education of students; educational infrastructures, and other resources. Research and teaching institutions provide their environment infrastructural and knowledge assets. These act as resources in the assessment of the impact of those institutions on the innovation of the economic system. The transmission channels of the impact which emerge from previous literatures are, just to cite a few, mobility of researchers, career of alumni, applied research contracts and joint use of infrastructures. In this context, different theories and models of the system of knowledge production and allocation could be developed and tested.²²

²¹ Moreover, higher education systems are increasingly expanding their interplay with the society moving towards markets in higher education systems or going beyond. There are some science and public policy studies that have analysed the elements of societal impact, mostly rooting it into universities and public research characteristics (Bornmann, 2013) whilst others, mostly refer to approaches developed by practitioners (Ebrahim et al. 2014). An interesting survey on university-industry relations can be found in Perkmann et al. (2013). All these theoretical considerations related to the so-called third mission activities of higher education institutions and research centers (Veugelers and Del Rey, 2014) have to be considered in relation to the specific research and innovation activities carried out, including their interrelations with the educational activities conducted.

²² According to Gibbons et al. (1994), knowledge is produced by configuring human capital that is more malleable than physical capital. Indeed human capital can be configured in different ways to generate new forms of specialized knowledge. The new economics of production can be interpreted as a shift from search for economies of scale to economies of scope where the latter arise from the ability to reconfigure human resources and particularly knowledge in new ways (see Gibbons et al. 1994, p. 63). This offers a comparative advantage to more educated systems. This calls also for a new management style for Mode 2 of knowledge production which is characterized by transdisciplinary socially distributed knowledge production, emerging from the development of ICT and massification of education and research activities. According to Gibbons et al. (1994) the main characteristics of Mode 2 of Knowledge production are: the research is carried out in the context of application, that is in a specified and localized context (as opposed to research within the academic context of Mode 1); transdisciplinary research (as opposed to specialized or disciplinary); heterogeneous actors involved in the research activity (vs more homogeneous); heterarchical and transient organization (vs hierarchical one); external accountability and quality assessment (vs. internal self/peer-evaluation); more socially accountable and reflexive; number of participants wider, more heterogeneous and more temporary (vs fixed and closed participants). Traditional and new forms of knowledge creation (Mode 1 and Mode 2 according to Gibbons et al. (1994) definitions) co-exist and dynamically evolve. The dynamics of knowledge production, distribution, co-creation and evolution obviously matters for the assessment of research and its impact. Within this context, the communication become crucial, both for formal collaborative agreement and for informal networks. Gibbons et al. (1994) identified the need of “increasing permeability of boundaries” to facilitate efficient communication between heterogeneous groups, and the intervention of governments and agencies as *brokering* agents. For a discussion on the government role and the need of overcoming a view of government as fixer of failures towards a view of entrepreneurial state, see Mazzucato (2013). An interesting parallel can be done between smart and inclusive growth postulate (Mazzucato, 2013) and the feature of the balance of cooperation and

The assessment of research can not be addressed in isolation: without education and innovation. It requires the specification of variables and indicators consistent with a *systemic* view.

Results can widely differ at different levels of aggregation, for instance at the public research organization and higher education institution level or individual university/research center, or faculty or team down to individual scholar. At these different levels, the possible moderating variables or causes of different performances may change too. Examples of possible moderating variables are: the legislation and regulation, public funding, teaching fees and duties; geography, characteristics of the local economic and cultural system, effectiveness of research and recruiting strategy, budgeting, infrastructures (at the university or department level); intellectual ability of researchers, history and stability of the team, ability to recruit doctoral students, world-wide network of contacts (at the research group and individual level), and the like.

2.2 Methodology

Methodology, in the setting of our framework, identifies the range of methods, techniques and approaches that are relevant for the evaluation of research. Before entering into the more detailed description, a preamble is necessary here. The discussion on Methodology relates to two general interconnected questions which are “what to assess” and “how to assess”. These questions, in turn, are related to the organization of the assessment tasks and strategies (including priorities’ setting) and to the communication of the assessment results. We distinguish the “subject (the thing that is being considered)” of the assessment (what to assess), that we identify in: outputs, efficiency, effectiveness and impact, from the “means” of the assessment that can be qualitative (including *peer-review* and *case studies*), quantitative (including econometric approaches and tools from the physics of complex systems) and *combined* (quantitative-qualitative) approaches, including the so called *informed peer-review*²³. Evidently, the means should be identified in accordance with the subject of the assessment. The organization and the communication aspects of the evaluation however, fall within the sphere of policy and governance.

We propose three building blocks for methods: EFFICIENCY, EFFECTIVENESS and IMPACT, considering the outputs as a kind of *baseline* or step zero in the analysis, followed by the subsequent steps. See Table 2.

TABLE 2 HERE

competition in a chaotic system (Baranger, 2000).

²³A quite complete comparison of the main advantages and disadvantages of quantitative approaches, such as citation based indicators, vs qualitative approaches, such as peer-review, can be found, e.g. in Hemlin (1996). Specific “quali-quantitative” approaches may be requested for the assessment of interdisciplinary research, see also Bammer (2016).

Table 2: *Dimensions of Methodology: subject and means in our framework.*

<i>Dimension</i>	<i>Type/category</i>	<i>Content</i>
Subject (of the assessment)	Output (baseline)	result of a transformation process which uses inputs to produce products or services
	Productivity and Efficiency	partial or total factor productivity productivity with respect to a reference
	Effectiveness	considers inputs, outputs and account for the aims of the activity
	Impact	all contributions of research outside academia ²⁴
Means (of the assessment)	Quantitative approaches	
	Qualitative approaches	
	Quali-quantitative approaches	

A distinction between productivity and efficiency is in order. *Productivity* is the ratio of the outputs over the inputs. *Efficiency*, in the broad sense, is defined as the output/input with respect to an estimated reference frontier, or frontier of the best practices (Daraio and Simar, 2007, pag. 14). The econometrics of production functions is different from that of production frontiers as the main objective of their analysis differs: production functions look at average behaviour whilst production frontiers analyse the whole distribution, taking into account the best/worst behaviour (Bonaccorsi and Daraio, 2004). Obviously, assessing the impact on the average performance is different from assessing the impact on the best/worst performance. Accounting for inequality and diversity is much more natural in a model based on best/worst performance frontiers than in a standard (average, representative) behaviour. This is because in the former case the *whole distribution* is considered instead of only the central tendency. This distinction between “average” versus “frontier” is considered in recent theory of growth (Acemoglu, Aghion and Zilibotti, 2003, 2006; Vandenbussche, Aghion and Meghir, 2006) and in the managerial literature (Chen, Delmas and Lieberman, 2015). As far as quantitative methods are concerned, different approaches, both parametric (Galán et al., 2014) and nonparametric (Bădin, Daraio and Simar, 2012, 2014; Daraio and Simar, 2014) have been proposed, highlighting the changes required by the attempt to disentangle the *impact* of external-heterogeneity factors on the efficient frontier from that on the distribution of inefficiency. This trend witnesses the need to move from the assessment of efficiency towards the assessment of impacts. Some precursors of methodological challenges and changes within the frontier approach may be identified, without being complete, in:

- *Statistical approach* to nonparametric frontier estimation (Simar and Wilson, 2015; Daraio, Simar and Wilson, 2016): trend towards a data-driven modeling;

²⁴See also the Research Excellence Framework (REF, 2014) and Moed and Halevi (2015) definition as “the contribution of research outcomes to the advancement of scientific/scholarly knowledge (scholarly impact) and to the benefits for society, culture, the environment, or the economy (social impact)”.

- *Models averaging* in stochastic frontier estimation (Parmeter et al. 2016): trend towards robustness of modeling;
- Using *information* about *technologies*, markets and behaviour of institutions in productivity indices (O'Donnell, 2016); trend towards more comprehensive informational setup;
- From an implementation point of view, *interactive benchmarking* (Bogetoft, 2012); trend towards developing analytics for policy decision making support.

Moving from efficiency to effectiveness is an important step²⁵.

The methodological dimension should handle how to evaluate what, providing an appropriate account of reliability and robustness (see Glänzel, 2010; Glänzel and Moed, 2013), and uncertainty²⁶. These are all considerations which refer to the Quality-Methodology intersection.

Classical methods of impact assessment (see e.g. Bozeman and Melkers, 1993), including randomized evaluations, matching methods (such as propensity score matching), double-differences, instrumental variables, regression discontinuity, distributional impacts and structural and other modeling approaches (see Khandker et al., 2010, for an overview) are challenged by the “problem of evaluation [that] is that while the program’s impact can truly be assessed only by comparing actual and counterfactual outcomes, the counterfactual is not observed. [...] Finding an appropriate counterfactual constitutes the main challenge of an impact evaluation” (Khandker et al., 2010, p. 22). These classical methods appear inadequate to the checklist of *sensitivity auditing* (Saltelli and Guimarães Pereira et al. 2013; Saltelli and Funtowicz, 2014, 2015)²⁷.

We should move on from efficiency, to effectiveness, and then towards impact, shifting our current paradigm, including *quality indicators* to assess effectiveness instead of efficiency; considering the *quality* of the applied method and overall the *quality* of the model.

²⁵At this purpose, the inclusion of *managerial* and more qualitative aspects in the quantitative benchmarking models could be beneficial. Peter Drucker (1967) sustained that “Efficiency is doing the thing right. Effectiveness is doing the right thing.” Interestingly, similar to principle 6 of Saltelli and Funtowicz (2014, see the next footnote) another important managerial aspect to consider is that “There is nothing quite so useless, as doing with great efficiency, something that should not be done at all”.

²⁶Uncertainty analysis focuses on the quantification of the uncertainty in the model output. *Sensitivity analysis* instead analyses the relative importance of different input factors on the model output. *Global* sensitivity analysis (see Saltelli, Ratto et al. 2008) refers to the investigation of how the uncertainty of a model inputs is attributed to the uncertainty of the output of the model. It is based on the application of statistical tools for interpreting the output from mathematical or computational models. *Partial* sensitivity analysis, also called once-at-a-time sensitivity analysis, is based on the change of one variable or assumption at a time. *Sensitivity auditing* is an extension of sensitivity analysis to the entire evidence-generating process in a policy context.

²⁷The sensitivity auditing checklist, proposed by Saltelli and Funtowicz (2014), is based on the seven following principles: 1. (use models to clarify, not to obscure): models as useful tools to represent and clarify reality; 2. (adopt an assumption hunting attitude): listing the underlying assumptions of each approach; 3. Detect pseudoscience (uncertainty, spurious decisions, Garbage-In Garbage-Out): Make approximation by keeping into account data representativeness and role of variables; 4. (Find sensitive assumptions before they find you): find the critical points in the theoretical framework that deserve attentions; 5. Aim for transparency (increasing the diffusion of the used models basic ideas avoiding jargon); 6. Don’t do the sums right but do the right sums: concentrate the analysis on the most important components/aspects; 7. Focus the analysis (check sensitivity analysis not on one factor at a time but changing the different parameters together).

2.3 Data

The data dimension is characterized by a kind of “data paradox”. On the one hand, we are in a “big data” world, with open data and open repositories that are exponentially increasing. On the other hand, in a lot of empirical applications the “data constraints” look pretty much the same as those described in Griliches (1986, 1994, 1998). Data are a relevant dimension often neglected in modelling building. Data have a problematic definition because it depends on their use not on inherent characteristics of the data (Borgman, 2015, p. 74).²⁸

Their properties and their weaknesses affect both the modelling and the empirical results. The concepts of big data, little data, and even no data remains poorly understood in the current big data era. Efforts to promote better data management, sharing, credit, and attribution are well intentioned, but stakeholders disagree on the starting points, the end goals, and the path in between. Lacking agreement on what entities are data, it remains difficult to establish policies for sharing releasing, deposing, crediting, attributing, citing, and sustaining access that can accommodate the diversity of data scholarship across domains. Sustaining access to data is a difficult and expensive endeavour (Borgman, 2015, p.271): “Despite the overall lack of agreement, most scholars would like better means to manage whatever they do consider to be their data. Better management is likely to lead to more sustainable data and in turn to better means of discovering and sharing data. These, however, are expensive investments. Better access to data requires investments in knowledge infrastructures by research communities, funding agencies, universities, publishers, and other stakeholders” (Borgman, 2015, p.287).

The main building blocks we identify to characterize the Data dimension are: availability, interoperability, *unit-free* property. *Availability* refers to general alternatives and choices that affect the data which have to be used, for instance (without being complete): sampling vs census, freely available vs controlled or undisclosed ones, data as consumption vs participation (see Ekbja et al. 2015 for a critical discussion). Obviously, the minimal requirement for the elaboration of data refers to their availability in a *usable* way. This opens to the discussion on commercial versus publicly available (or open) data; institutional provided data, issues of privacy and confidentiality.

Interoperability is the way in which heterogeneous data systems are able to communicate and exchange information in a meaningful way (Parent and Spaccapietra, 2000). It is crucial for data integration of heterogeneous sources (see Daraio and Glänzel, 2016. See also the discussion on continuity vs innovation in Ekbja et al. 2015).

A great improvement in Data Integration for research assessment could come by the adoption of an Ontology-Based-Data-Management (OBDM) Approach (Calvanese et al. 2011; Lenzerini, 2011; Poggi et al., 2008). An OBDM approach is a form of information integration based on a three-level architecture: 1) the *ontology*, that is a conceptual, formal description of the domain of interest, expressed in terms of relevant

²⁸Alternative more positive definition of data include: facts and statistics collected together for reference or analysis; data as representations, reinterpretable representation of information in a formalized manner, suitable for communication, interpretation, or processing, up to data as “infrastructure” (Frischmann, 2012).

concepts, attributes of concepts, relationships between concepts, and logical assertions characterizing the knowledge domain); 2) the *sources*, that are the heterogeneous (maintained independently) repositories where the data concerning the domain are stored; 3) the *mappings*, that represents the correspondence between the data contained in the data sources and the elements of the ontology. The main advantages of an OBDM approach for integrating research and other scholarly data (Daraio, Lenzerini et al. 2016a) are: accessibility of the data through the elements of the ontology; explicit representation of the domain, facilitating the re-usability of the acquired knowledge; explicit specification of the relationships between the domain concepts and the data through the mappings, facilitating documentation and standardization; flexibility of the integrated system, that does not require the integration of all the data sources at once; extensibility of the system by means of incremental addition of new data sources or new concepts when they become available.²⁹

Unit-free property refers to the need of having consistent and coherent observations (instances of data) at different levels of analysis, to ensure a robust empirical evidence of a given phenomenon. The *Unit-free* property of data is somewhat interconnected to the possibility of *multiscale* modeling³⁰ of the problem at hand. It explicits the exigence of having data that are *independent* from the unit of analysis and hence can be used coherently in a multiscale model of the problem.

See Table 3 for an overview of the Data dimensions and their characterization in our framework.

TABLE 3 HERE

²⁹Halpern and Vardi (1991) propose the construction of semantic models to represent agents’ knowledge in a number of different contexts and discuss about model checking techniques from an artificial intelligence perspective. Their discussion about “Learning extensions” and use of heuristics, may be helpful also here, although, in our context here, it is even more difficult because of the general framework (discussed in Section 2) made by Theory-Method-Data dimensions.

³⁰The multiscale modeling is an interdisciplinary area of research (ranging from mathematics, to physics, engineering, bioinformatics and computer science) to explain problems which have significant characteristics at multiple scales (e.g. time and/or space). Its aim is “by considering simultaneously models at different scales, we hope to arrive at an approach that shares the efficiency of the macroscopic models as well as the accuracy of the microscopic models” (Weinan, 2011, p. viii). According to Horstemeyer (2009), the rapid growth of multiscale modeling is the result of the confluence of parallel computing power, experimental capabilities to characterise structure-property relations down to the atomic level, and theories that admit multiple length scales.

Table 3: *A characterization of the Data Dimension in our framework.*

<i>Dimension</i>	<i>Characterization</i>
Availability	usability sampling vs census freely, controlled or undisclosed consumption vs participation commercial vs publicly available open, institutional provided privacy/confidentiality (see Ekbia et al. 2015)
Interoperability	a very high level is obtained by an OBDM approach (see Daraio, Lenzerini et al. (2016b))
Unit-free property	independence of the data from the unit of analysis

A relevant connection, also for the following developments of modelling is the relationship between data and information. According to Floridi (2014), ICT (Information and Communication Technologies) have brought new opportunities as well as new challenges for human development and have led to a revolutionary shift in our understanding of humanity’s nature and its role in the universe, the “fourth revolution” according to which “we are now slowly accepting the idea that we might be informational organisms among many agents..., *inforgs* not so dramatically different from clever, engineered artefacts, but sharing with them a global environment that is ultimately made of information, the *infosphere* Floridi (2014). The information revolution is not about extending ourselves, but about “re-interpreting who we are” (Floridi 2008a).³¹ Within this context emerged the philosophy of information (Floridi 2010, 2012), in which the understanding of the ultimate nature of reality shifts from a materialist one to an informational one, in which all entities, both natural and artificial, are analysed as informational entities.

3 A Summary View and a Pragmatic Perspective

Our general framework is derived integrating relevant dimensions, grounded in existing approaches, according to three dimensions. The main building blocks of these dimensions are summarized in Figure 1. This framework could allow for combining the fine-grained results of case studies, with the ability to replicate and route them, taking them to a higher level, thanks to an integrated view, which maps the interfaces,

³¹An interesting and perhaps connected change, due to the developments introduced in information processing including novel algorithms, protocols, and properties of information bring to shift from the classical to the quantum computation paradigm and recently lead to derive Quantum Theory as a special theory of information (D’Ariano and Perinotti, 2016). An ontological analysis of the study of social reality (Lawson, 2012), consistent with modern interpretations of quantum field theory, argues that social science can be scientific in the sense of natural science; “All forms of established science have objects of study that are effectively (synchronically irreducible) emergent forms of organisations-in-process, and in this respect social science is no different. And just as with other sciences, social science is especially concerned with the (irreducible) causal properties of its domain of study (Lawson, 2012, p. 382)”.

interdependencies, complementarities among the three dimensions and allows for analysing the constraints on the three dimensions that may make analysis difficult.

In the field of education, much progress has been made. The *quality* of education has been demonstrated as relevant for research and innovation.³² Much more work is needed for research and innovation due to the inherent difficulties that arise for their specific content, context and complexity. The main object of a research evaluation is represented by the results of given research activities, which can be considered as the research effort (Hemlin, 1996). The outputs of a given research activity are the result of a complex set of interacting characteristics and activities that involve, but are not limited to: ability, talents, social aspects, luck, incentives, motivations, trade-offs, commitment, financial resource, efforts, infrastructure, education, personality skills, network, organization, curiosity, communication skills and contextual and institutional factors. These all interact dynamically, giving rise to complex processes. The evaluation of research is done in a context characterized by many more different factors that interact as well. Hemlin (1996, p.210) points out that “all evaluation of research quality must be based on an idea of the meaning of this concept. [...] The variety in meaning of scientific quality reflects the fact that research evaluations are being made in a context in which a number of different factors interact and where the interplay between these factors is essential to the concept of quality in science... not only the real interplay between factors is important, but also the evaluators conceptions of this interplay is crucial.”³³ The meaning of scientific quality and its difficulties in delimiting what is meant by it are related to the nature of research itself. The conception of what is good or bad research varies between different research areas and periods, constantly changing as the result of an interactive process between scientific development and events in the world outside the scientific community.

All these aspects show the complexity of the evaluation of research.

Issues of *uncertainty*, and, closely related, those of quality of information, are involved whenever policy-related research is utilized in the policy process (Funtowicz and Ravetz, 1990, p. 11). In assessing research, it is important also to consider the interactions of quality with uncertainty and policy, “in a situation where major decisions, on the most complex and uncertain issues, must frequently be made under conditions of urgency” (Funtowicz and Ravetz, 1990, p. 13).

From a methodological point of view, the inclusion of quality indicators in the analysis, may allow us to move from efficiency to effectiveness. Effectiveness can be captured then by using in the analysis “qualitative-adjusted” quantitative measures. It is the quality of education, research and innovation, which has an *impact*

³²Several contributions have analysed the impact of education quality on economic growth (e.g. Hanushek and Woessmann, 2007, Hanushek et al. 2008, Aghion, 2009).

³³Hemlin (1996) on the base of a review on the previous research identifies the following system of factors (context) influencing perspectives and important for the assessment of “research quality”:

- Research effort (the product of the research that is scientific knowledge in the form of a paper or a doctorate and so on);
- Researcher (competence and personality, motivational and emotional factors);
- Research environment and research resources (colleagues, students, premises, and the supply of research resources such as economic means, apparatus, indirect effect on the motivation and interests of the researcher);
- Research effects (two main effects: intra-scientific effects on the current state of scientific knowledge and extra-scientific effects on society in a wide sense);
- Research financing, organization and policy.

on the development of the society.

Finally, it is on the data dimension that the quality issues are of primary importance in all the three main building blocks proposed (availability, interoperability, unit free property).³⁴

Quality of available data is crucial; in data quality there have been relevant advances, going from data quality to information quality (Batini and Scannapieco, 2016) and developing a philosophy of data information quality (Floridi et al. 2014). The quality of the interoperability is important in the integration of heterogeneous datasets which are useful for research and innovation studies. Finally, the “unit-free property” of data, in terms of data quality aims at reaching a kind of “objectivity”, for empirical purpose and for data reuse³⁵.

Quality as acceptability (suitability) for application (*fitness for purpose*) is the overarching concept, which keeps together the building blocks of the three dimensions. It is a characteristic in all the three dimensions. The nine building blocks, from 1 to 3, are attributes of Quality. The quality of Theory, as dimension, is related to the problem of *boundaries* and philosophical representation of the reality. The degree of implementation of the assessment of quality is related to the level and intensity of the resolution of the underlying “valuative problem”. It is linked to the implementation factor *tailorability*. The quality of Method, as dimension, refers to the *transparency* and suitability in the context of application (again tailorability). Quality of Data is related to the quality of information and plays a crucial role at the implementation level. It is also linked to the degree of *openness* of data and information.

From the description so far, it emerges that the assessment of the research activity is indeed a complex task. In addition, in Appendix A we describe the “generalized” implementation problem in research assessment and show how it adds complexity to the assessment of research.

Now, our finding could be interpreted in two ways:

- *Impossibility option*: given that it is so difficult, we must abandon it and conclude that it is not possible to assess the research; or,
- *Pragmatic option*: use our knowledge on the difficulty of the assessment of research and use our proposed framework with a pragmatic purpose, which is, to develop possibly *meaningful* models of research assessment.

The latter is exactly what we pursue here.

³⁴Data quality according to the OECD (2011) Quality Framework is defined with respect to user needs, and it has seven dimensions: relevance (“degree to which data serves to address their purposes”); accuracy (“how the data correctly describes the features they are designed to measure”); credibility (“confidence of users in the data products and trust in the objectivity of the data”); timeliness (“length of time between their availability and the phenomenon they describe”); accessibility (“how readily the data can be located and accessed”); interpretability (“the ease with which the user may understand and properly use and analyse the data”); coherence (“the degree to which they are logically connected and mutually consistent”).

³⁵The Provenance initiative (Moreau et al. 2008) is a clear example of describing at better data for different purposes, including also the opening or sharing of data.

4 Model Selection, Implementation Factors and Enabling Conditions

Due to the complexity of the evaluation of research described so far, it is more appropriate to talk about model *development* rather than of model *selection*, as the selection is very difficult to handle. What can be done, according to the pragmatic perspective pursued in this paper, is monitoring the model development and its evolutions, including the characterization of the *quality*, according to our framework dimensions.

In our framework we identify three implementation factors and three enabling conditions that may be helpful to monitor the model development.

We highlight that our framework is able to act as a *common denominator* of many different strands of literature, collecting them under the same conceptual scheme. In the following we report just a few examples, leaving a systematic analysis of the related literature for future research.

In Theory, Tailorability refers to flexibility of the model for problem solving and its related learning: taking into account *absorptive capacity* and innovation processes (Cohen and Levinthal, 1990). In Methods we should account for a multimethodology approach (Mingers, 2006). In a Data perspective, tailorability is linked to the usability and end-users personalization of platforms. Transparency and Openness are two implementation factors that can be detailed along the main building blocks of our framework and have a self-evident importance.

For Theory we have open education (see e.g. DeMillo and Young, 2015) which refers to the transformation of higher education towards new ways of disseminating knowledge at lower cost, as MOOCS (Massive Open Online Courses), thanks to technology fuelled innovations, and research on learning processes.

According to OECD (2015), open science refers to “efforts by researchers, governments, research funding agencies or the scientific community itself to make the primary outputs of publicly funded research results publications and the research data publicly accessible in digital format with no or minimal restriction as a means for accelerating research; these efforts are in the interest of enhancing transparency and collaboration, and fostering innovation.”³⁶

West et al. (2014) in reviewing the open innovation literature since Chesbrough (2003) identify three main directions of research: better measurement, resolving the role of appropriability and linking open innovation to the management and economics literature.³⁷ The exponential increase and development of information

³⁶ “[...] Three main aspects of open science are: open access, open research data, and open collaboration enabled through ICT. Other aspects of open science post-publication peer review, open research notebooks, open access to research materials, open source software, citizen science, and research crowdfunding are also part of the architecture of an open science system” (OECD, 2015, p. 7). Nielsen (2012) develops the concept of open research a bit further, talking about “data driven intelligence” controlled by human intelligence which amplifies collective intelligence: “To amplify cognitive intelligence, we should scale up collaborations, increasing cognitive diversity and the range of available expertise as much as possible. Ideally, the collaboration will achieve *designed serendipity*...”. According to Nielsen (2012) this could be achieved by conversational critical mass and collaboration which becomes self-stimulating with online tools, which may establish architecture of attention that directs each participant where it is best suited. This collaboration may follow the patterns of open source software: commitment to working in modular way; encouraging small contributions; allowing easy reuse of earlier work; using signalling mechanisms (e.g., scores) to help people to decide where to direct attention.

³⁷ See also Chesbrough, Vanhaverbeke, and West (2006, 2014) that describe in more details the main building blocks of the open innovation literature and the new challenges the field is facing.

availability and the development of the information society is leading us towards an open innovation society (see e.g. Chesbrough, 2006) based on a Quadruple Helix Model (Leydesdorff, 2012) of bottom up interactive policy framework.³⁸ Although the Quadruple Helix model gives emphasis to the broad idea of cooperation in innovation, it is not a very well established and much used concept in research and innovation studies, because of its conceptual and practical elusiveness. We argue here that our framework could be a valid support for the conceptualization and the implementation of a Quadruple Helix model.

Mixed methods is the first enabling condition and relates to the combination of quali-quantitative analysis. It offers strengths that offset the weaknesses of both quantitative and qualitative research (e.g. Creswell and Plano Clark, 2011). Quantitative methods are weak in understanding the context, qualitative methods (on the other hand) are weak because of personal interpretation and difficulty in generalizing. A bridge across adversarial divide, between quantitative and qualitative; encourages the use of multiple paradigms (beliefs and values), is *practical* to solve problems, combine inductive and deductive thinking.

The formalization of concepts and measurements is necessary, as it offers the flexibility of qualitative research and allows for accountability, intended and unintended consequences and monitoring mechanisms.

The second enabling condition refers to *convergence*³⁹ intended as “the coming together of insights and approaches from originally distinct fields”, “provides power to think beyond usual paradigms and to approach issues informed by many perspectives instead of few” (National Research Council, 2014b).

The third enabling condition refers to the *knowledge* infrastructure intended as “robust networks of people, artifacts, and institutions that generate, share, and maintain specific knowledge about the human and natural worlds” (Edwards et al., 2013).⁴⁰

In the next section, Figure 2 illustrates the connections of our modelling framework with the empirical, policy and real world. The enabling conditions foster these connections.

³⁸In this model, government, industry, academia and civil participants work together to co-create the future and drive structural changes far beyond the scope of what any one, organization or individual, could do alone. This model encompasses also user-oriented innovation to take full advantage of ideas’ cross-fertilization, leading to experimentation and prototyping in real world setting. Different forms and levels of co-production with consumers, customers and citizens challenge public authorities and the realization of public services. These new forms, comprised in the fourth helix of the Quadruple Helix model, allow overcoming the traditional linear top-down approach, expert-driven, to the development/realization of production and services. Carayannis and Campbell (2009) show the connection of the Quadruple Helix model with a mode 3 innovation system based on innovation network and knowledge clusters. They show that the Quadruple Helix model facilitate the “democratization” of knowledge (von Hippel, 2005), that is the co-development and co-evolution of different paradigms of knowledge creation, diffusion and use. von Hippel (2016) extends the analysis of the democratization of innovation, based on user-centered innovation systems, to a “free innovation” paradigm in which there are no transactions but a peer-to-peer free interaction and diffusion.

³⁹According to the OECD definitions, in multidisciplinary research the subject under study is approached from different angles, using different disciplinary perspectives, but integration is not accomplished. Interdisciplinary research leads to the creation of a theoretical, conceptual and methodological identity, hence more coherent and integrated results are obtained. Transdisciplinary research goes one-step further and refers to a process in which convergence among disciplines is attained.

⁴⁰Within some research projects funded by Sapienza in 2013 and 2015 we did an experiment of a knowledge infrastructure, a case of an “open science of science” exercise, around Sapiientia: The Ontology of Multi-Dimensional Research Assessment (Daraio, Lenzerini et al. 2016a,b). Sapiientia represents an effort of going towards a common platform which can show which data have to be collected; by offering the opportunity of making analysis under different perspectives, testing different models, but sharing the same common conceptual characterization.

5 Towards Responsible Metrics?

The discussion so far seems incomplete: what is missing? Perhaps much, but we identify two things at least: the connection to the real world and a “reference” against which monitor the development of the model of research evaluation. We try to illustrate the contribution of our framework with respect to the different “representations” of the real world involved in research evaluation processes. Figure 2 shows the interconnections between the different views of the real world, made by the policy world, the modelling world and the empirical world. The illustration of the different representations as concentric ellipses denotes the fact that each world is perceived differently from other worlds.

FIGURE 2 HERE

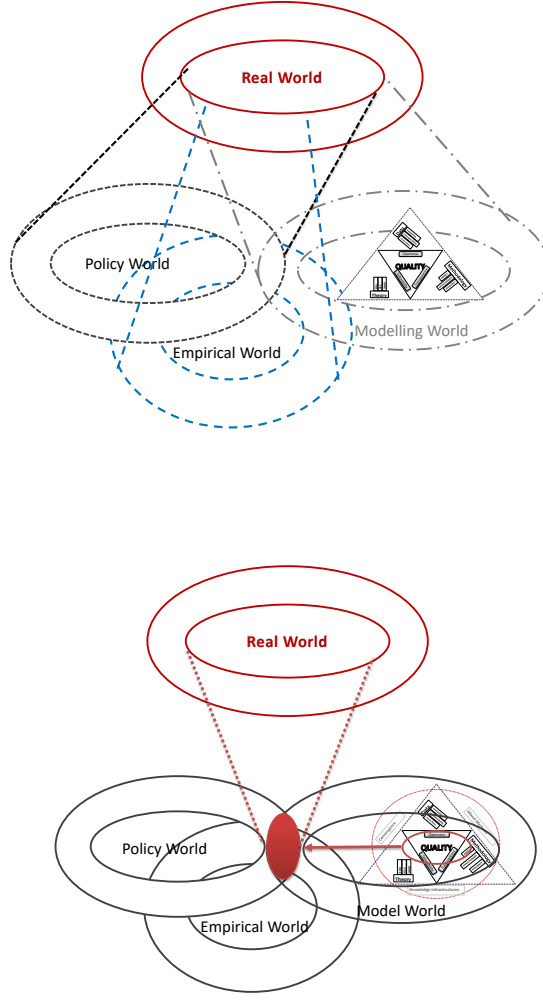


Figure 2: *An illustration of the relationship between modelling world, empirical world and policy world: they are all somewhat overlapping visions or projections of the real worlds.*

Figure 2 shows the role of our modelling framework in its interplay with the empirical and policy world for the understanding of the real world. We claim that the more the quality is conceptually and formally specified, the more the overlapping area among modelling, policy and empirical worlds is, and closer to the real world the model is.

This statement is basically the postulate 2 of our framework (see Section 1). It is linked to the second missing item introduced before, namely the need to have a “reference” for checking the development of the model. It calls also for the introduction of the third postulate which is, the monitoring of the developments and the evolutions of the modelling activity can be carried out in relation to the “responsibility” of the metrics proposed and involved.

But what does be a “responsible metric” mean in an evaluation process? According to the Cambridge dictionary, to be responsible could be defined as “be responsible for something or someone” that means

“to have control and authority over someone or something and the duty of taking care of it”; or as “be responsible to something or someone” that means “to be controlled by someone or something”.

Does “responsible” relate to metric itself or to its use, or both?

Wilsdon et al. (2015, p. x) propose the notion of responsible metrics as “a way of framing appropriate uses of quantitative indicators in the governance, management and assessment of research [...]”⁴¹ Interestingly, also Benessia et al. (2016) propose *responsible metrics* at the end of their discussion on the crisis of science.

After the publication of the Independent Review of the Role of Metrics in Research Assessment and Management whose report, “The Metric Tide”, published in July 2015 (see Wilsdon et al., 2015) a website for *responsible metrics* has been established.⁴²

Coming back to our framework, we identify some connections of its enabling conditions with the oeuvre of Alasdair MacIntyre.⁴³ They are illustrated in Table 4.

TABLE 4 HERE

⁴¹ “Responsible metrics can be understood in terms of the following dimensions: Robustness: basing metrics on the best possible data in terms of accuracy and scope; Humility: recognising that quantitative evaluation should support but not supplant qualitative, expert assessment; Transparency: keeping data collection and analytical processes open and transparent, so that those being evaluated can test and verify the results; Diversity: accounting for variation by field, and using a range of indicators to reflect and support a plurality of research and researcher career paths across the system; Reflexivity: recognising and anticipating the systemic and potential effects of indicators, and updating them in response(Wilsdon et al.” (2015, p. x).

⁴²From the website <https://responsiblemetrics.org/> (last accessed on 10th February 2017) its main aim is: “The metric tide is rising. But we have the opportunity and a growing body of evidence to influence how it washes through higher education and research. This site is intended to broaden debate and encourage action to ensure that metrics and indicators develop in more positive and responsible ways.” On responsible research and innovation see also Owen et al. (2012) and Stilgoe et al. (2013).

⁴³See for instance see Lutz (2017) that describes an overview on MacIntyre’s oeuvre, reporting a rich bibliography on his works.

Table 4: *Towards an Ethics of Research Assessment? Some connections of our framework with MacIntyre’s oeuvre.*

<i>Enabling Condition</i>	<i>Potential connection</i>	<i>MacIntyre work</i>
Convergence	Invitation to overcome the fragmentation of knowledge, and excessive specialization.	<i>The end of education</i> , MacIntyre (2006).
Mixed Methods	The need to go beyond a pure quantitative approach (abstract representation of the reality) and include qualitative cases (narratives, story telling). ⁴⁴	<i>After Virtue</i> , MacIntyre (2007), <i>Whose justice, which rationality?</i> MacIntyre (1988).
Knowledge Infrastructure	Retrieve the values of tradition in communities of practice that regulate themselves by defining their own standards.	<i>After Virtue</i> , MacIntyre (2007)

The third postulate of our framework, reported in the Introduction, gives the ability to give content to the concept of “responsible metrics” to the grade (level) of conceptualization and formalization, in an unambiguous way, of the different layers/ meanings of “quality”.⁴⁵ These activities of conceptualization and formalization of Quality are strictly linked to the production, use and effects of “standards”. It is useful to recall here a precursor paper on the need for standards in bibliometrics. We refer to the work of Wolfgang Glänzel (1996), still relevant today, more than twenty years after its publication.

As clearly illustrated by Brunsson and Jacobsson (2002a), standardization may be a valid alternative to market forces and to organizational forms as an institutional arrangement for coordinating and controlling complex exchanges. Brunsson and Jacobsson (2002b) summarize the arguments in favour of standardization in “more effective use of information, better coordination of activities, simplification, and the advantages of large-scale production” (Brunsson and Jacobsson (2002b, pag.170). On the other hand, they summarize the arguments against standardization in those similar to the objections against rules and regulation in general, lack of trust in the expertise and goodwill of those who set the rules, critics of those that prefer markets to standards, or of those that want, on the other hand, a stronger formal coordination way (such as directives) (see Brunsson and Jacobsson, 2002b, 171-172). In concluding their essay and the entire book, Brunsson and Jacobsson (2002b, p. 172) state that “Standardization deserves to be paid a good deal more attention than it has received up to now”, and “... we may have something to learn from the old Greek myths. In a way, standardizing is the art of constructing a procrustean bed. Procrustes was a legendary bandit in Greek mythology, a bandit who placed his victims on a specially constructed bed. The bed was a

⁴⁴Recent research shows that there are new machine learning models (Graves et al., 2016) that are able to reason from entirely non-symbolic systems that learn through experiences.

⁴⁵This could permit to give content to the somewhat “vague” idea of “excellence” (Moore et al., 2017).

pattern and a yardstick intended to create conformity... (p. 173)”. We share their conclusions, and believe that their reference to the procrustean heritage could be an interesting starting point to further explore and develop the connections of our framework with MacIntyre’s oeuvre (see Table 4). Further research on the connections with MacIntyre’s oeuvre could help to fill an existing gap providing new tools to assess efficiency together with equity (Hinrichs-Krapels, 2016). It would be very interesting to investigate whether and how to extend, specify and apply MacIntyre’s philosophy to develop an *Ethics* of evaluation with our framework as a background. This is out of the scope of the present paper and is left to future research.

6 Three Examples of Application

6.1 A Doubly Conditional Performance Evaluation Model

In this section we introduce a *Doubly Conditional* Performance Evaluation Model which is one possible model for the assessment of Research coherent with our framework. “Doubly conditional” means that the evaluation is conditioned two times: on the information we have and on that we do not have.

The performance evaluation model unfolds mainly along the Methodological dimension, while Theory and Data dimensions are specific to the problem content and application context.

Performance measurement in public management is a challenged subject (Johnsen, 2005). Woelert (2015) identifies a proliferation of performance indicators in a kind of technical and quantitative escalation. Lewis (2015) highlights the almost neglected consideration of the consequences of performance measurement in evaluation of public policies. We have seen that the output of the research activity has some features that include complexity, uncertainty and indeterminacy. Among the challenges of their assessment, there are: 1) bring about communication and debate about assumptions, choices and uncertainties, and about the limits of scientific knowledge; 2) to allow for articulation of different types of (scientific, local, indigenous, political, moral and institutional) knowledge; 3) to provide room for a transparent negotiation among standpoints (participatory processes)(van den Hove, 2007, see p. 815 for more normative requirements for science-policy interface).

Our framework can be helpful in combining the advantages of the partial convergence indicators approach (Martin and Irvine, 1983; Martin, 1996) mainly measurability and possibility of calculating the indices of performance, with a need for a multidimensional approach to the assessment of research and its impact (Moed and Halevi, 2015).

Martin and Irvine (1983) introduce the idea of converging partial indicators approach for assessing scientific performance, based on the multidimensional nature of research and its outputs: “All quantitative measures of research are, at best, only partial indicators influenced partly by the magnitude of the contribution to scientific progress and partly by other factors. Nevertheless, selective and careful use of such indicators is surely better than none at all. Furthermore, the most fruitful approach is likely to involve the combined use of multiple indicators. However, because each is influenced by a number of other factors, one

needs to try and control for those by matching the groups to be compared and assessed as closely as one can” (Martin, 1996, p. 351). Hence, given the partial character of indicators, it is only possible to draw reliable conclusions in those cases in which the indicators provide convergent results, keeping the influence of non-relevant factors low. Moed and Halevi (2015) developed further the notion of multidimensionality of research, extending the work of AUBR (2010). Rather than focusing on a single output dimension and underlining the need to obtain convergence among a set of different indicators in order to produce valid and useful outcomes, they aimed at proposing “a consolidated multidimensional methodological approach addressing the various user needs, interests and purposes, based on the notion that “indicators designed to meet a particular objective or inform one target group may not be adequate for other purposes or target groups.” Diverse institutional missions, and different policy environments and objectives, require different assessment processes and indicators. They focus on the purpose, objectives, and policy context of research assessments, and demonstrates how these characteristics determine the methodology and metrics to be applied. For instance, publication counts are useful instruments to help discriminating between those staff members who are research active and those who are not, but are of little value if research active scientists are to be compared one with another according to their research performance. Moed and Halevi (2015) introduce the concept of a *meta-analysis* of the units under assessment in which metrics are not used as tools to evaluate individual units, but to reach policy decisions regarding the overall objective and general setup of an assessment process. Their base underlying assumption, that we share here, is that “the future of research assessment exercises lies in the intelligent combination of metrics and peer review. A necessary condition is a thorough awareness of the potentialities and limitations of each method.” We develop further this approach here, building on the contribution of van den Hove (2007) on the science-policy interfaces in which science and policy are considered as intersecting domains of the human activity which are in co-evolution; “science-policy interfaces are defined as social processes which encompass relations between scientists and other actors in the policy process, and which allow for exchanges, co-evolution and joint construction of knowledge with the aim of enriching decision-making” (van den Hove, 2007).⁴⁶

Figure 3 illustrates the main component of our proposed doubly conditional performance evaluation model which is based on our combination and extension of Johnsen (2005); van den Hove (2007) and Lewis (2015).

⁴⁶See Table 2 p. 815 which describes the normative requirements for the interfaces. van den Hove (2007, p. 824) identifies the following methodological issues to account for in the design, implementation and assessment of the science-policy interfaces: “(i) the reinforcement and enlargement of scientific quality and validation processes; (ii) the development of transdisciplinary research methodologies; (iii) transparency, participation and dynamism of interfaces, in particular the role of other stakeholders and the public; (iv) accountability of the different actors; (v) translation of scientific knowledge into policy-relevant knowledge and of policy knowledge into science-relevant knowledge; (vi) the inclusion of a diversity of knowledges and intelligences; (vii) the development of dialogical dissemination channels for scientific knowledge which specifically target the various potential user groups; and (viii) the institutionalisation of science-policy interfaces in a democratic context”. See also Kowarsch et al. (2016).

FIGURE 3 HERE

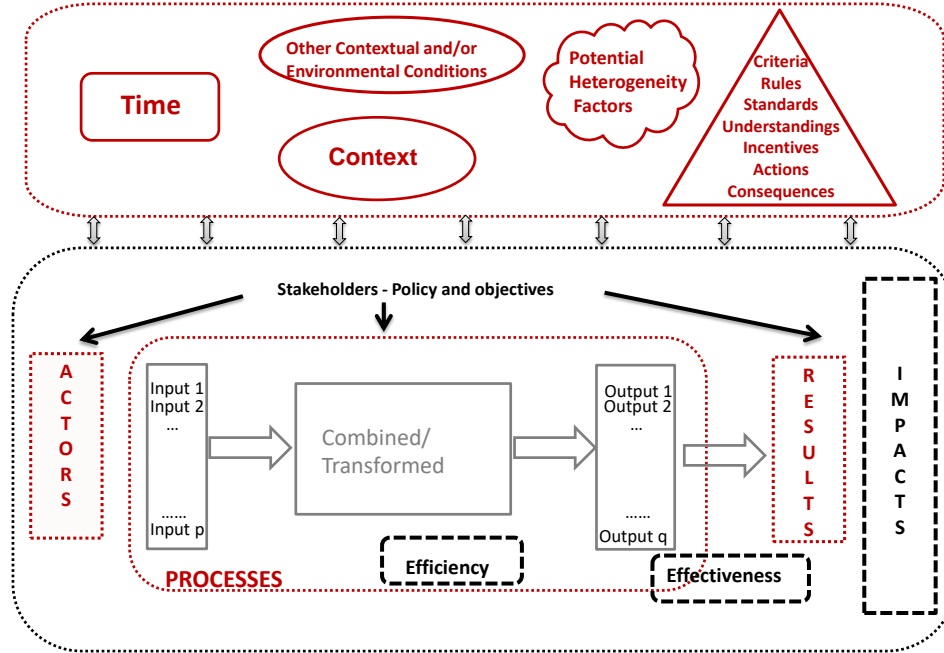


Figure 3: A Doubly Conditional Performance Evaluation Model.

The red elements represent the main items on which the conditioning could be done.

We distinguish two kind of conditioning. Conditioning on the items reported in the bottom of the Figure 3 (actors, processes and results) means to compare comparable entities, setting appropriate reference sets. We call this as *internal* conditioning or *normalization*. On the other hand, conditioning on the items reported in the top of the Figure 3 equals to account for *heterogeneity* factors that we call *external* conditioning or *contextualization*. According to our model of performance evaluation, it's all a matter of appropriate normalization and contextualization.

The main contribution of our approach to research assessment estimation relays on the:

1. Identification of the components of the analysis (in terms of our theory-method-data characterization) that are excluded (what remains outside) in the specific context of the evaluation;
2. Interpretative value of the measure (or metrics or indicators) of research assessment calculated, that has to be considered as a *residual*, what remains after the consideration of the dimensions we pursued, that is due to other factors/components not accounted for;

3. A step toward the democratization of the evaluation practice, able to balance the opposite views of external accountability and internal improvement (Ewell, 2009), composing contrasting trends towards competition and cooperation through cohesion.

Our performance evaluation model might be helpful to identify *constitutive effects* of indicators (Dahler-Larsen, 2014) and perhaps also their “unintended consequences”. Our framework is useful for the interpretation of the results coming out from our model, and to identify discrepancies, what is the *residual*, our “ignorance”. It is also helpful to identify gap and which variables may be added to explain a part of these discrepancies. It is a contemporary revisiting and revalidation of the Leibnsteins x-inefficiency concept (Leibenstein 1966, 1975, 1978a,b; see also Leibenstein and Maital, 1992).

The indicators empirically calculated are interpreted as the residual or our ignorance on the phenomenon and it is possible to identify the neglected aspects of the analysis carried out. The neglected component can be useful for suggesting alternative or additional dimensions of research assessment, of interest for the subjects of the assessment that are scholars, institutions and so on.

Another underlying idea of our doubly conditional performance evaluation model is that for each subject under assessment it can be found a *dimension* of performance along which the evaluated entity can outperform or do better than the others. The identification of the best performing dimension of each entity subjected to the evaluation is important for developing strategy for identifying and establishing sustainable and durable value creation, going *beyond competitive advantages* (Zenger, 2016), exploiting the existing *information* (Porter and Miller, 1985), and finding out their own specialty-role in the knowledge production system through profiling of the activities, sharing evaluation models, improving knowledge and by learning. These are relevant factors among others for the development of a *participatory learning* role of the subjects involved in the research assessment exercise. This approach could be a first step towards a formative *democratic approach* to evaluation in which indicators are used as learning tools instead of target of policy.

6.2 Interpretation of university rankings

In this section we briefly describe how our framework could be applied for the interpretation of university rankings.

Let us consider the interpretation of the university rankings debate in Nordic countries. Piro and Sivertsen (2016) try to explain the differences in the scores of universities on the basis of the different assumptions made, while, Elken, Hovdhaugen, and Stensaker (2016) found that “rankings have a relatively modest impact on decision-making and strategic actions in the Nordic universities studied, and that there are few signs of rankings challenging the existing identities of the universities in this region”. The latter seems to support an alternative role of rankings in the national context (see also Douglass, 2016). On the other hand, Daraio, Bonaccorsi and Simar (2015a)⁴⁷ found that Nordic universities, taking into account their multidimensional

⁴⁷Daraio, Bonaccorsi and Simar (2015a) describe the main criticisms of rankings that their proposed approach (based on conditional directional distance functions) addressed, namely: 1. Monodimensionality (universities perform several missions:

activities, perform in line with the European reference standard. What can the present framework tell us about this rankings comparison? We can frame the comparison of rankings results according to our three dimensions (theory, methodology and data) and their constituent building blocks. It would appear clearly then where the differences are. The differences observed can be interpreted as a measure of our ignorance (or the *residual*) that is due to the non-considered components, such as third mission activities, Higher Education Institutions non included for missing data or other non-European countries in the comparison and so on.⁴⁸

6.3 Setting Priorities of ANVUR

In period of budgetary restrictions, policy makers need timely and inexpensive answers to their questions. While our framework highlights the need to invest in knowledge infrastructures as an investment for the development of a new generation of assessment models, it offers, at the same time, a pragmatic scheme to identify priorities in policy actions. Coherent with our framework is the view of priority-setting as a problem in system design, which is “best understood as a systemic process, with outcomes determined by the incentives and inter-relationships of choice rather than by ex-ante calculation” (Stewart, 1995). To show an example, we use our representation of the relevant dimensions for the assessment of Research to frame the activities carried out so far from the Italian national agency for the evaluation of universities and research centers (ANVUR).⁴⁹ In the Top Panel of Figure 4 we illustrate the main activities carried out by ANVUR so far, where AVA (Autovalutazione, Valutazione periodica, Accreditamento) stands for the evaluation of teaching, VQR (Valutazione della Qualità della Ricerca) for the evaluation of the quality of the research activity and Third Mission stands for the assessment of third mission outputs of universities and research centers. ANPRES is an acronym that indicates the Registry (Anagrafe) of the Italian scholars, created by the current legislation but not yet implemented. For that reason in Figure 4 it is reported as a wide circle. ORCID indicates the measure introduced by law to mandatory ask to each academic staff to ask and obtain an ORCID code, a successful measure for the standardization and calculation of bibliometric indicators of the assessment exercise. Although rough and approximated, the picture of the Top Panel of Figure 4 immediately allows us to identify the priorities in the next planning of the activities, that are (see Figure 4 Bottom Panel) first along the data dimension, on data availability (accessibility) then along the method dimension moving from the output towards impact, and finally along the theory dimension.

teaching, research, third mission) 2. Statistical robustness (multiple rankings, probabilistic rankings, robustness of rankings) 3. Dependence on university size and subject mix (rankings favour old and large universities where scientific, technical and medical disciplines prevail) 4. Lack of consideration of the input-output structure (ignore the amount of resources used). They propose a conditional directional distance approach as a *fair comparison*: a novel two-stage approach of conditional efficiency scores (nonparametric location scale model) allowed them to estimate the *managerial* efficiency scores as the *residuals* (what remains after having eliminated size and disciplinary specialization effects). They propose a bootstrap based approach for estimating error bounds for managerial efficiency scores and hence provide some statistical robustness.

⁴⁸Daraio and Bonaccorsi (2016) suggest that it is possible to go beyond university rankings through the intelligent integration of existing data that may lead to an open-linked data platform that permits the construction of new indicators without designing the indicators on a custom basis.

⁴⁹On the reform of the Italian system, see e.g. Moscati (2008), while for some comparative analyses of the Italian system with France and UK respectively, see Boffo, Dubois and Moscati (2008), Rebora and Turri (2013).

 FIGURE 4 HERE

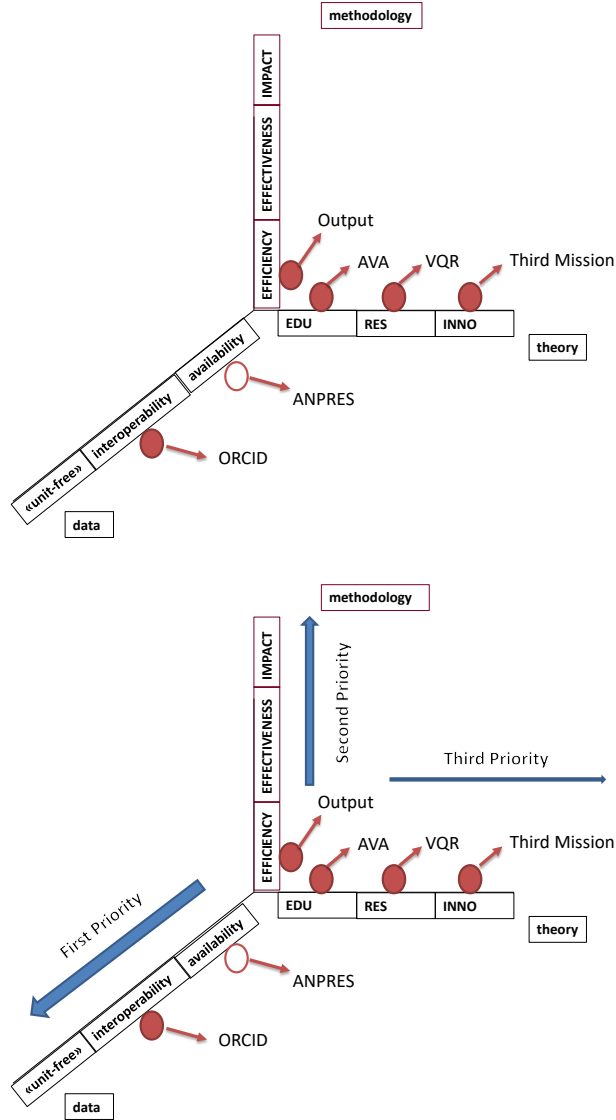


Figure 4: *An application of the framework to the ANVUR's activities.*

7 Conclusions and Further Research

The main objective of this paper is to provide a comprehensive framework able to serve as a basis for the development of models for the assessment of research and its impacts that be “quality- aware” (in the broad meaning discussed in the paper) i.e. fitness for use. We show that with our framework, composed of 3

dimensions (theory, methodology and data) of 3 building blocks each (education, research and innovation; efficiency, effectiveness and impact; availability, interoperability and unit-free property), 3 implementing factors (tailorability, transparency and openness) and 3 enabling conditions (convergence, mixed-methods and knowledge infrastructure), all joined together around the overarching idea of quality, we are able to embrace many different and heterogeneous streams of literature. We introduced a doubly conditional performance evaluation model, coherent with the proposed framework that may pave the way to the democratization of evaluation, and show the interpretative usefulness of the proposed framework by discussing university rankings and priority setting for the Italian national agency for the evaluation of research (ANVUR).

Our framework may be particularly useful to develop models of research assessment, to frame the traditional problems of evaluation in a wider perspective and to facilitate the introduction of new methods for the assessment of research relevant to support their governance. The framework introduced has the ambition of being general and valid for different units and layers of analysis. For this reason it needs to be corroborated, tested and extended to different specific evaluation cases.

The paper may open the way to many extensions and further research:

- Testing the proposed framework for developing effective checklists for designing and implementing policy-monitoring mechanisms on the assessment of research activities.
- Running additional research for providing a systematic analysis and classification of the existing literature, having our framework as a *common denominator*.
- Corroborating the framework facing the problem of the *democratization* of the evaluation (Acemoglu and Robinson, 2006).
- Extending the proposed framework to the characterization of different governance systems (Capano et al., 2015) for analysing their systemic connection with their performance.
- Applying the framework, *mutatis mutandis*, to the evaluation of *education* (Daraio, 2015) and *innovation* (Borrás and Edquist, 2013) activities.
- Investigating the Ethics of evaluation by exploring the connections between our framework and MacIntyre's oeuvre.
- Corroborating the framework for the *regulation* of the evaluation of research.

Finally, our framework may pave the way for new *revolutionary* models⁵⁰ of research assessment, closer to the represented reality, even if to do so much additional research is needed.

⁵⁰Here revolutionary refers to the Kuhn (1962)'s idea of change of the representations of the investigated reality ("A scientific theory is usually felt to be better than its predecessors not only in the sense that it is a better instrument for discovering and solving puzzles but also because it is somehow a better representation of what nature is really like" (Kuhn (1969) Postscript, p. 206).

8 Acknowledgements

This work originated from Daraio (2015a) which pointed out the unavailability of a best evidence on the “efficiency, effectiveness and impact of research and innovation” due to the lack of a suitable framework for a comprehensive analysis. See also Daraio (2015b). Preliminary versions of this paper were presented at the Workshop “Efficiency, Effectiveness and Impact of Research and Innovation, 19 February 2015 Sapienza University of Rome, LeccEWEPA June 2015, ISSI 2015 Conference, Istanbul, June 2015; Anvur Seminar, Rome July 2015; CHER Conference Lisbon, September 2015, at a seminar at IFQ Berlin September 2015, NAPW Quebec City, June 2016 and RSA AIIG Bergamo, October 2016.

I am indebted to many people for useful discussions on this work. In particular, I thank warmly Henk Moed for his accurate comments on different versions of the work, Andrea Bonaccorsi, Giliberto Capano, Daniele Checchi, Alessandro Daraio, Marta Degl’innocenti, Kristof De Witte, Daniel Deybe, John Douglass, Hal Fried, Clara Eugenia Garcia, Massimiliano Guerini, Sybille Hinze, Stefan Hornbostel, Hugo Horta, Michele Meoli, Alessio Moneta, Roberto Moscati, Stefano Paleari, Monica Scannapieco, Robin Sickles, Lèopold Simar, Pedro Teixeira for comments/discussions during the presentations in seminars and workshops, and finally, my Sapienza’s colleagues, with which I have worked on several projects related to this work Claudio Leporelli, Maurizio Lenzerini, Giancarlo Ruocco, Paolo Naggar, Giuseppe Catalano, Tiziana Catarci and Alberto Nastasi.

This work would not have been accomplished without the support and encouragement of my husband Ado to which it is dedicated.

The financial support of the Sapienza Awards 2015 no. 6H15XNFS is gratefully acknowledged.

A Appendix: The Implementation Problem in Research Assessment

The operationalization of the general framework proposed in this paper needs the discussion of the implementation problem which is typical of quantitative methods (mostly mathematical models) in operations research and management science (Schultz and Slevin, 1975; Hildebrandt, 1977; Bryant, 1989; Keys, 1991; Mingers and Gill, 1997). It consists in applying methods developed as “basic research” to concrete organizations and contexts for specific problem-solving. A few general considerations are in order and primarily relate to the necessity of extending the implementation to the three dimensions of our framework. For this reason, we refer to the “generalized” problem, considering the interaction of method development with its useful application; that the implementation changes the unit of assessment; that knowledge and technological innovation may be at place. Needless to say, in our framework the identification of the *right* problem and the development of an *appropriate* model are crucial determinants of success.

Along Methodology, we observe an evolution of the implementation theory, thanks to the introduction of the *multi-methodology* approach⁵¹ (Mingers and Gill, 1997; Mingers, 2006).

In Figure 5, on the top-left side there is an illustration of three systems, which constitute the context of the intervention. According to Mingers (2006) who extended the Checkland (1981)’s two systems (problem solving and problem content systems), the three systems are the intervention system (agents undertaking the intervention), the problem content system (real-world situation of concern) and the intellectual resources system (available theories and methodologies).

A fruitful connection is the work on cognitive biases and their impact on decision-making introduced by Tversky and Kahneman (1974) which present three heuristics⁵² that are employed in taking decisions under uncertainty:

- *representativeness*, related to the use of categories to evaluate the probability that a given event belongs to a given class or a given process;
- *availability of instances or scenarios*, related to the evaluation the frequency of a given class or the likelihood of a particular progress; and

⁵¹Within the discipline of operations research and management science, many methods and techniques have been developed. Initially, these were generally based on mathematical or computer models. However, it was found in practice, particularly with complex, *wicked* problems, that many aspects of the situation, especially those concerning peoples’ viewpoints and values, could not be represented mathematically. This led to the development of a range of non-quantitative, *soft* methods, which were, nevertheless, rigorous and systemic. Examples are soft systems methodology (SSM), cognitive mapping, strategic choice analysis, and strategic assumption surfacing and testing (SAST). The question then became, which method should be used and when. However, rather than using just a single method, theory and practice have demonstrated that most complex problems are better tackled using a combination of methods, both *hard* and *soft*. This approach has been called *multimethodology* (Mingers, 2006).

⁵²Heuristics are considered here as simple rules, derived by evolutionary processes, proposed to explain how people make decisions and solve problems when facing complex problems with incomplete information. Metaheuristics are higher level heuristic-based algorithms and techniques of stochastic optimization to find optimal solutions to hard inverse problems (the so called *I know it when I see it* problems) on the base of very few or weaker assumptions with respect to classical optimization techniques. A compact presentation of the subject can be found in Luke (2015).

- *adjustment from an anchor*, related to numerical prediction, when a pertinent value is available.

These heuristics that are generally effective, lead to regular and expected biases.⁵³

Returning to the context of intervention, intervention systems are made by agents (managers, decision-makers...) who build models of problem content system (including inputs, organization, goals, criteria etc.) with the purpose of problem solving and obtaining outputs (activities, projects, solutions) which generate effects that interact and influence the previous systems. The analysis of the problem and context of intervention can be very difficult to address globally. This is due to the consideration of the two additional dimensions of Data and Theory. Here the approach of the Levels of Abstraction (LoA) proposed by Floridi (2002) could be helpful for our implementation problem to move from Data to information. Floridi's (2011) method of abstraction (Floridi 2002) relies on a method borrowed from a branch of theoretical computer science called Formal Methods, in which discrete mathematics is used to specify and analyse the behaviour of information systems.⁵⁴

The idea behind the method of Levels of Abstraction (Floridi 2008b) is quite simple and straightforward: reality can be viewed from different levels. Perhaps the most crucial feature of the method of LoA is that the identification relation between two observables is never absolute. Rather, the identification is always contextual and the context is a function of the level of abstraction chosen for the required analysis (Floridi and Sanders 2004). A LoA can be defined as a finite non-empty set of observables (which are interpreted typed variables, typed variables together with a statement of what features of the system under consideration they stand for) which are expected to be the components of a "theory"⁵⁵. This in our set-up, is given by the "theory-method-data" paradigm. A *gradient of abstraction* is made by a collection of LoAs and represents a kind of interface. It is used to analyse some systems from varying points of view or at different LoA. Hence, the method permits the analysis of systems by means of models developed at specific gradients of abstraction (interfaces). Figure 5 (Top Panel) reports an illustration of the method adapted to our set-up. LoAs can be nested, disjointed and overlapping and do not need to be hierarchically related or ordered (as atomic

⁵³The study of heuristics in decision-making introduced by Tversky and Kahneman (1974) extended Simon's research on human bounded rationality in problem solving (see e.g. Simon, 1969, 1982, 2000) which lead to the "satisficing" situation where people seek solutions or accept choices or judgments that are "good enough" for their purposes. The discussion on heuristics in human decision making and of their inherent biases is extended in Kahneman (2011) that, building on earlier contributions, describe two different ways of thinking, a "fast system", characterized by fast, automatic, frequent, emotional, stereotypic, subconscious and a "slow system", characterized by slow, effortful, infrequent, logical, calculating, conscious. On the base of heuristics, Kahneman (2011) asserts that the fast system involves the association of new information with existing patterns instead of building new patterns for each new event. These recent developments in behavioural economics decision making could be further explored in combination with recently developed statistical and machine learning approaches (see e.g. Mezard and Montanari, 2009; Barber, 2012). Indeed, *machine learning* techniques, lying at the intersection of computer science and statistics, are at the core of artificial intelligence and data science, and are showing increasing potentialities (Jordan and Mitchell, 2015). Kahneman, Lovallo, and Sibony (2011), building on previous works, propose a checklist of decision quality control that has several connections with our implementation set up. The checklist is made by the following items: -self-interested biases; -affect heuristic; -group-think; -saliency bias (analogy to a memorable case); -confirmatory bias; -availability bias; -anchoring bias; -Halo effect; -sunk-cost fallacy, endowment effect; -overconfidence, planning fallacy; optimistic biases, competition neglect; -disaster neglect; -loss aversion. For a detailed discussion, see Kahneman, Lovallo, and Sibony (2011).

⁵⁴This method, according to Floridi (2010) paves the way for defining a new macro-ethical theory, i.e., Information Ethics. By using this method, Floridi claims that the moral value of human actions is not different in kind from the moral evaluation of other informational objects.

⁵⁵Note that it differs from the "Theory" dimension of our framework as it refers to the modelling activity.

components in a molecular). This method makes the ontological commitment of a “theory” (in our case of a configuration of “theory-method-data”) explicit as follows: A configuration of “theory-method-data” (“theory” in the original Floridi scheme) comprises at least one LoA and one model. The LoA allows the “theory-method-data” configuration to analyse the system under investigation and to elaborate a model that identifies some properties of the system (in our set up “context of intervention systems”) at the given LoA. The ontological commitment of a theory can be recovered by distinguishing between a committing and a committed component, within the scheme (see Figure 5, Top Panel). By making the ontological commitment of a “theory-method-data” configuration explicit, the method of abstraction plays a crucial role in any context in which we acquire and process information. The resulting model, then, consists of informational objects and processes.

FIGURE 5 HERE

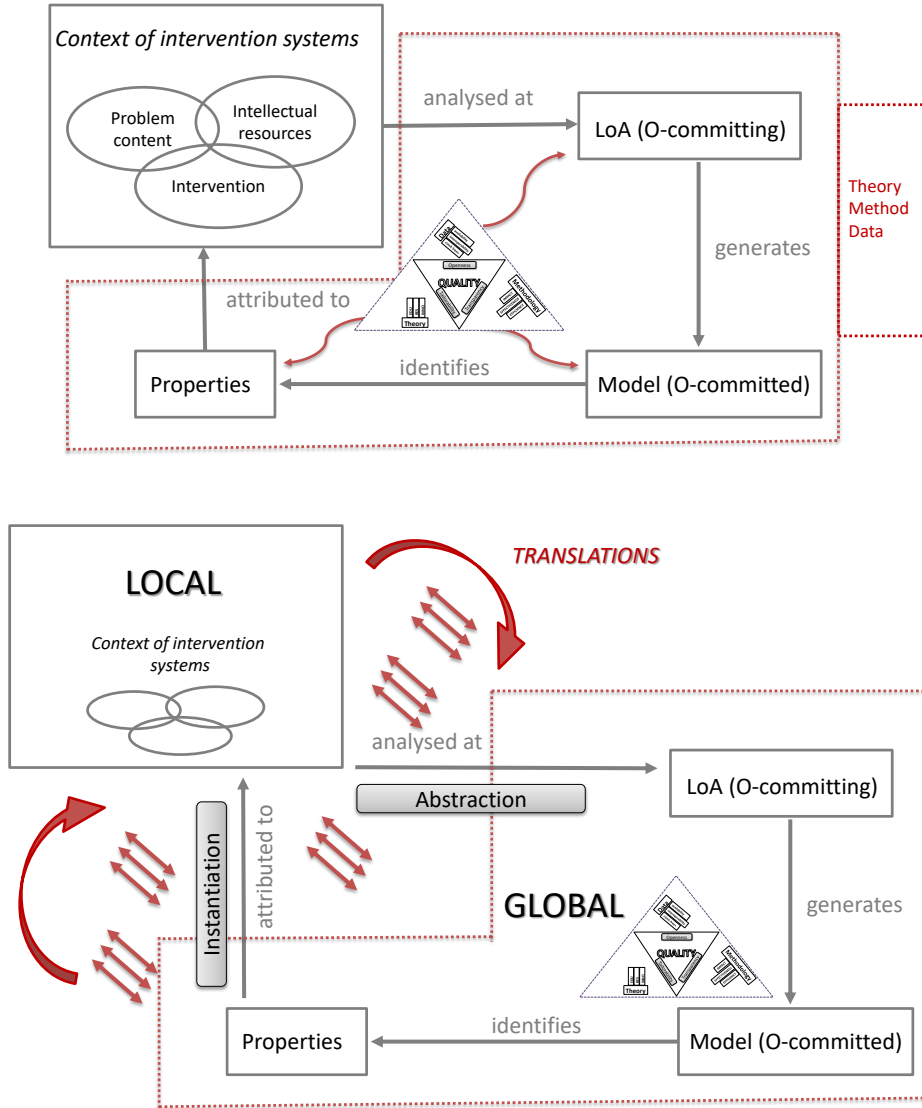


Figure 5: *An illustration of the generalized implementation problem. Context of Intervention and level of analysis (Top Panel) with inclusion of translations (Bottom Panel).*

We have seen the implementation problem from a methodological perspective⁵⁶ and under an informational

⁵⁶It has to be noted that in our framework, multimethodology refers also to empirical model building in statistics (Box and

point of view, as combined in Top Panel of Figure 5. In our framework, the Theory dimension refers to people (and organizations of people) which are involved in the research activities. How “translations” occur in such a way that processes of abstraction (from the local to the global) and of instantiation (from the global to the local) *modify* the actors involved in the translation process (Callon and Latour, 1981; Callon, 1986a,b; Latour, 1993). This is what we learn from sociology, to complete our “generalized” implementation problem, whose main references are summarized in Table 5. See Figure 5 Bottom Panel for a complete overview. It highlights the “translation” and the relative configurations and reconfigurations of mediations originated by the movements of the instantiation and abstraction that transform the actors involved in the process. This is why it is so difficult to “trace the social” (Latour, 2005).

TABLE 5 HERE

Table 5: *The generalized implementation problem in Research assessment.*

<i>Framework Dimension</i>	<i>Approach (Content)</i>	<i>Reference</i>
Methodology	Multimethodology (OR/MS, Statistics and Econometrics and Information theory...)	Mingers (2006) (Box and Draper (1987) Golan (2008) Judge and Mittelhammer(2011))
Data	Levels of Abstraction (Philosophy of Information)	Floridi (2008)
Theory	Translation process (Sociology of Science)	Callon and Latour (1981) Callon (1986a,b), Latour (1993, 2005)

Draper, 1987), applied econometrics (see e.g. the ten commandments of Kennedy (2008, p. 362) and its link with information theory.

References

- [1] Abramovitz, M. (1956). Resource and Output Trends in the United States Since 1870. *The American Economic Review*, 46(2), 5-23.
- [2] Acemoglu, D., Robinson, J. A. (2006). *Economic origins of dictatorship and democracy*. Cambridge University Press.
- [3] Acemoglu, D., Aghion, P., Zilibotti, F. (2003). Vertical integration and distance to frontier. *Journal of the European Economic Association*, 1(2-3), 630-638.
- [4] Acemoglu, D., Aghion, P., Zilibotti, F. (2006). Distance to frontier, selection, and economic growth. *Journal of the European Economic association*, 4(1), 37-74.
- [5] Ackoff, R. L. (1974a). *Redesigning the future: A systems approach to societal problems*. New York, 2.
- [6] Ackoff, R. L. (1974b). The systems revolution. *Long Range Planning*, 7(6), 2-20.
- [7] Adams, J. D. (1990). Fundamental stocks of knowledge and productivity growth. *Journal of Political Economy*, 673-702.
- [8] Agarwal, R., Dhar, V. (2014). Editorial Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research. *Information Systems Research*, 25(3), 443-448.
- [9] Aghion, P. (2009). *Growth and education*. World Bank.
- [10] Aghion, P., Howitt, P. (2009). *The Economics of Growth*. MIT Press Books, 1.
- [11] Aghion, P., David, P. A., Foray, D. (2009). Science, technology and innovation for economic growth: linking policy research and practice in STIG Systems'. *Research Policy*, 38(4), 681-693.
- [12] Anderson C. (2008). The end of theory: the data deluge makes the scientific method obsolete. *Wired Magazine*, 16 July.
- [13] Antonelli, C., Link, A. N. (2014). *Routledge Handbook of the Economics of Knowledge*. Routledge.
- [14] AUBR (2010). Expert Group on the Assessment of University-Based Research. Assessing Europe's University-Based Research. European Commission-DG Research. EUR 24187 EN.
- [15] Audretsch, D. B., Bozeman, B., Combs, K. L., Feldman, M., Link, A. N., Siegel, D. S., ... Wessner, C. (2002). The economics of science and technology. *The Journal of Technology Transfer*, 27(2), 155-203.
- [16] Bădin, L., Daraio, C. and L. Simar (2012), How to measure the impact of environmental factors in a nonparametric production model, *European Journal of Operational Research*, 223 (3), 818–833.
- [17] Bădin, L., Daraio, C. and L. Simar (2014), Explaining inefficiency in nonparametric production models: the state of the art, *Annals of Operations Research*, 214 (1), 5–30.
- [18] Bammer, G. (2016). What constitutes appropriate peer review for interdisciplinary research?. *Palgrave Communications*, 2 (palcomms201617).
- [19] Baranger, M. (2000). *Chaos, complexity, and entropy*. New England Complex Systems Institute, Cambridge.

- [20] Barber, D. (2012). *Bayesian reasoning and machine learning*. Cambridge University Press.
- [21] Batini, C., Scannapieco, M. (2016). *Data and Information Quality*. Springer International Publishing.
- [22] Benessia, A., Funtowicz, S., Giampietro, M., Pereira, Â. G., Ravetz, J., Saltelli, A., ... van der Sluijs, J. P. (2016). *Science on the Verge*. Amazon Book, in the Series The Rightful Place of Science Consortium for Science, Policy Outcomes Tempe, AZ and Washington, DC.
- [23] Blaug, M. (1966). *Economics of education*; a selected annotated bibliography (No. 370.193 B5).
- [24] Boffo S., Dubois P., Moscati R. (2008) Changes in University Governance in France and in Italy, *Tertiary education and Management*, 14 (1),2008,pp.13-26.
- [25] Bonaccorsi A., Daraio C. (2004), Econometric approaches to the analysis of productivity of RD systems. Production functions and production frontiers, in H.F. Moed, W. Glänzel and U. Schmoch (edited by), *Handbook of Quantitative Science and Technology Research*, Kluwer Academic Publishers, 51-74.
- [26] Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world*. Mit Press.
- [27] Bornmann L. (2013), What Is Societal Impact of Research and How Can It Be Assessed? A Literature Survey, *Journal of the American Society for Information Science and Technology*, 64(2), 217233.
- [28] Borrás, S., Edquist, C. (2013). The choice of innovation policy instruments. *Technological forecasting and social change*, 80(8), 1513-1522.
- [29] Box, G. E., Draper, N. R. (1987). *Empirical model-building and response surfaces* (Vol. 424). New York: Wiley.
- [30] Bozeman, B., Melkers, J. (1993), (eds). *Evaluating R&D Impacts: Methods and Practice*, Springer, New York.
- [31] Bryant, J. (1989). *Problem management: a guide for producers and players*. John Wiley & Sons Inc.
- [32] Brunsson, N., Jacobsson, B. and associates (Eds.)(2002a). *A world of standards*. Oxford/New York: Oxford University Press.
- [33] Brunsson, N., Jacobsson, B. (2002b). The Pros and Cons of Standardization - An Epilogue, In *A world of standards*, 169-173. Oxford/New York: Oxford University Press.
- [34] Callon M. and Latour B. (1981). Unscrewing the Big Leviathan: how actors macro-structure reality and how sociologists help them to do so. in K.Knorr-Cetina and A. Cicourel, *Advances in Social Theory*.
- [35] Callon M. (1986a). Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieux Bay. In *Power, action and belief: a new sociology of knowledge?*, ed. J. Law, London, Routledge, pp.196-223.
- [36] Callon, M. (1986b). The sociology of an actor-network: The case of the electric vehicle. In M. Callon, J. Law and A. Rip (Eds.), *Mapping the dynamics of science and technology: Sociology of science in the real world*. London, Macmillan, pp. 19-34.
- [37] Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Poggi, A., Rodriguez-Muro, M., et al. (2011). The Mastro system for ontology-based data access. *Semantic Web*, 2(1), 4353.

- [38] Capano, G., M. Howlett, M. Ramesh (2015), Eds., *Varieties of Governance*. Palgrave Macmillan UK.
- [39] Carayannis, E. G., Campbell, D. F. (2009). 'Mode 3' and 'Quadruple Helix': toward a 21st century fractal innovation ecosystem. *International Journal of Technology Management*, 46(3-4), 201-234.
- [40] Chatfield, C. (1995). *Problem solving: a statistician's guide*. CRC Press.
- [41] Checchi, D. (2006). *The economics of education: Human capital, family background and inequality*. Cambridge University Press.
- [42] Checkland, P. (1981). *Systems thinking, systems practice*. John Wiley & Sons, Chichester.
- [43] Chen, C. M., Delmas, M. A., Lieberman, M. B. (2015). Production frontier methodologies and efficiency as a performance measure in strategic management research. *Strategic Management Journal*, 36(1), 19-36.
- [44] Chesbrough, H. (2003). *Open innovation: The new imperative for creating and profiting from technology*. 2003. Cambridge, Harvard Business School Press.
- [45] Chesbrough, H. W. (2006). *Open innovation: The new imperative for creating and profiting from technology*. Harvard Business Press.
- [46] Chesbrough, H., Vanhaverbeke, W., West, J. (2006). *Open innovation: Researching a new paradigm*. Oxford University Press.
- [47] Chesbrough, H., Vanhaverbeke, W., West, J. (Eds.). (2014). *New frontiers in open innovation*. OUP Oxford.
- [48] Cohen, W. M., Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 128-152.
- [49] Creswell, J. W., Clark, V. L. P. (Eds.). (2011). *Designing and Conducting Mixed Methods Research*. SAGE Publications.
- [50] Cronin, B. (2013). Thinking about data. *Journal of the American Society for Information Science and Technology*, 64(3), 435-436.
- [51] Cronin, B., Sugimoto, C. R. (2014). *Beyond bibliometrics: harnessing multidimensional indicators of scholarly impact*. MIT Press.
- [52] Cronin, B., Sugimoto, C. R. (Eds.). (2015). *Scholarly metrics under the microscope: from citation analysis to academic auditing*. Information Today, Medford USA.
- [53] D'Ariano, G. M., Perinotti, P. (2016). Quantum Theory is an Information Theory. *Foundations of Physics*, 46(3), 269-281.
- [54] Dahler-Larsen, P. (2012). *The evaluation society*. Stanford University Press, California.
- [55] Dahler-Larsen, P. (2014). Constitutive effects of performance indicators: getting beyond unintended consequences. *Public Management Review*, 16(7), 969-986.
- [56] Daraio C. (2015a), What do we know about Efficiency, Effectiveness and Impact of Research and Innovation? Proceedings of Workshop, 20th February DIAG Sapienza University of Rome, edited by C. Daraio, Efesto Edizioni, Rome, pag. 13-25.

- [57] Daraio C. (2015b), Efficiency, Effectiveness and Impact of Research and Innovation: a framework for the analysis, in Salah, A.A., Y. Tonta, A.A. Akdag Salah, C. Sugimoto, U. Al (Eds.), Proceedings of ISSI 2015 Istanbul: 15th International Society of Scientometrics and Informetrics Conference, Istanbul, Turkey, 29 June to 3 July, 2015, Bogazii University Printhouse, pp. 1226-1227. ISBN: 978-975518381-7.
- [58] Daraio C. Simar L. (2014), - Directional Distances and their Robust versions. Computational and Testing Issues, *European Journal of Operational Research*, 237, 358-369.
- [59] Daraio, C. and L. Simar (2007), *Advanced Robust and Nonparametric Methods in Efficiency Analysis. Methodology and applications*, Springer, New York.
- [60] Daraio C., Simar L., P.W. Wilson (2016), Nonparametric Estimation of Efficiency in the Presence of Environmental Variables, Technical Report DIAG (Sapienza University of Rome) n. 2/2016.
- [61] Daraio, C., Bonaccorsi, A. (2016). Beyond university rankings? Generating new indicators on universities by linking data in open platforms. *Journal of the Association for Information Science and Technology*. Version of Record online: 28 MAR 2016, DOI: 10.1002/asi.23679
- [62] Daraio, C., Glänzel, W. (2016). Grand challenges in data integrationstate of the art and future perspectives: an introduction. *Scientometrics*, 108 (1), 391-400.
- [63] Daraio, C., Bonaccorsi, A., Simar, L. (2015a). Rankings and university performance: A conditional multidimensional approach. *European Journal of Operational Research*, 244(3), 918-930.
- [64] Daraio, C., Lenzerini, M., Leporelli, C., Moed, H. F., Naggar, P., Bonaccorsi, A., Bartolucci, A. (2016a). Data integration for research and innovation policy: an Ontology-Based Data Management approach. *Scientometrics*, 106(2), 857-871.
- [65] Daraio, C., Lenzerini, M., Leporelli, C., Naggar, P., Bonaccorsi, A., Bartolucci, A. (2016b). The advantages of an Ontology-Based Data Management approach: openness, interoperability and data quality. *Scientometrics*, 1-15.
- [66] Debackere, K. (2016). Let the Data Speak for Themselves: Opportunities and Caveats. *Journal of Data and Information Science*, 1(1), 3-5.
- [67] DeMillo, R. A., Young, A. J. (2015). *Revolution in Higher Education: How a Small Band of Innovators Will Make College Accessible and Affordable*. MIT Press.
- [68] Ding, Y., Rousseau, R., Wolfram, D. (2014). *Measuring Scholarly Impact*. Springer.
- [69] Ding, Y., Stirling, K. (2016). Data-driven Discovery: A New Era of Exploiting the Literature and Data. *Journal of Data and Information Science*, 1(4), 1-9.
- [70] Doornik, J. A., Hendry, D. F. (2015). Statistical model selection with ‘Big Data’. *Cogent Economics & Finance*, 3(1), 1045216.
- [71] Douglass, J. A. (Ed.). (2016). *The New Flagship University: Changing the Paradigm from Global Ranking to National Relevancy*. Springer.
- [72] Drucker, P. F. (1967). *The effective executive*. New York: Harper and Row.

- [73] Ebrahim, A., Rangan, V. K. (2014). What Impact?. *California Management Review*, 56(3), 118-141.
- [74] Edquist, C. (2001). The Systems of Innovation Approach and Innovation Policy: An account of the state of the art. In DRUID Conference, Aalborg (pp. 12-15).
- [75] Edwards, P. N., Jackson, S. J., Chalmers, M. K., Bowker, G. C., Borgman, C. L., Ribes, D., ... Calvert, S. (2013). *Knowledge Infrastructures: Intellectual Frameworks and Research Challenges* (p. 40). Ann Arbor, MI: University of Michigan. <http://deepblue.lib.umich.edu/handle/2027.42/97552>.
- [76] Egghe, L., Rousseau, R. (1990). *Introduction to informetrics: Quantitative methods in library, documentation and information science*.
- [77] Egghe, L. (2005). *Power laws in the information production process: Lotkaian informetrics*. Elsevier.
- [78] Ekbja, H., Mattioli, M., Kouper, I., Arave, G., Ghazinejad, A., Bowman, T., ... Sugimoto, C. R. (2015). Big data, bigger dilemmas: A critical review. *Journal of the Association for Information Science and Technology*
- [79] Elken, M., Hovdhaugen, E., Stensaker, B. (2016). Global rankings in the Nordic region: challenging the identity of research-intensive universities?. *Higher Education*, 1-15.
- [80] Etzkowitz, H., Leydesdorff, L. (2000). The dynamics of innovation: from National Systems and Mode 2 to a Triple Helix of universityindustrygovernment relations. *Research Policy*, 29(2), 109-123.
- [81] European Commission (2014), Expert Group to support the development of tailor-made impact assessment methodologies for ERA (European Research Area), Brussels, Belgium.
- [82] Ewell, P. (2009). Assessment, Accountability and Improvement: Revisiting the Tension. National Institute for Learning Outcomes Assessment, [http : //www.learningoutcomesassessment.org/documents/PeterEwell05.pdf](http://www.learningoutcomesassessment.org/documents/PeterEwell05.pdf), last accessed, 28thSeptember2016.
- [83] Fagerberg, J., Martin, B. R., Andersen, E. S. (Eds.). (2013). *Innovation studies: evolution and future challenges*. OUP Oxford.
- [84] Fealing K. H., Lane J. I., Marburger J. H. JIII, Shipp S. S. (Eds.) (2011), *The Science of Science Policy, A Handbook*. Stanford, USA, Stanford University Press.
- [85] Feldman, M., Kenney, M., Lissoni, F. (2015). The New Data Frontier. *Research Policy*, 44(9), 1629-1632.
- [86] Floridi, L. (2002). What is the Philosophy of Information?. *Metaphilosophy*, 33(1-2), 123-145.
- [87] Floridi, L. (2008). The method of levels of abstraction. *Minds and machines*, 18(3), 303-329.
- [88] Floridi, L. (2011). *The philosophy of information*. Oxford University Press.
- [89] Floridi, L. (2012). The road to the philosophy of information. In Luciano Floridi's Philosophy of Technology (pp. 245-271). Springer Netherlands.
- [90] Floridi, L. (2014). *The fourth revolution: How the infosphere is reshaping human reality*. OUP Oxford.
- [91] Floridi, L. (Ed.). (2008a). *The Blackwell guide to the philosophy of computing and information*. John Wiley & Sons.

- [92] Floridi, L. (Ed.). (2010). *The Cambridge handbook of information and computer ethics*. Cambridge University Press.
- [93] Floridi, L., Illari, P. (Eds.). (2014). *The philosophy of information quality* (Vol. 358). Springer.
- [94] Floridi, L., Sanders, J. W. (2004). The method of abstraction. Yearbook of the artificial. Nature, culture and technology. *Models in contemporary sciences*, 177-220.
- [95] Frascati Manual (2015), Guidelines for Collecting and Reporting Data on Research and Experimental Development, ISB9789264239012, available at: [http : //www.oecd – ilibrary.org/science – and – technology/frascati – manual – 2015_9789264239012 – en](http://www.oecd-ilibrary.org/science-and-technology/frascati-manual-2015_9789264239012-en) last accessed on 20thDecember2016.
- [96] Funtowicz, S. O., Ravetz, J. R. (1990). Science for Policy: Uncertainty and Quality. In *Uncertainty and quality in science for policy* (pp. 7-16). Springer Netherlands.
- [97] Frascati Manual (OECD, 2002), Proposed Standard Practice for Surveys on Research and Experimental Development, 6th. ed.
- [98] Frischmann, B. M. (2012). *Infrastructure: The social value of shared resources*. Oxford University Press.
- [99] Galán, J. E., Veiga, H., Wiper, M. P. (2014). Bayesian estimation of inefficiency heterogeneity in stochastic frontier models. *Journal of Productivity Analysis*, 42(1), 85-101.
- [100] Geuna, A. (2003). The Economics of Science: Good Old Wine in a New Bottle. *Social Studies of Science*, 33(3), 458-461.
- [101] Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., Trow, M. (1994). *The new production of knowledge: The dynamics of science and research in contemporary societies*. Sage.
- [102] Glänzel, W. (1996). The need for standards in bibliometric research and technology. *Scientometrics*, 35(2), 167-176.
- [103] Glänzel, W., Schoepflin, U. (1994). Little scientometrics, big scientometrics... and beyond?. *Scientometrics*, 30(2-3), 375-384.
- [104] Glänzel, W. (2010). On reliability and robustness of scientometrics indicators based on stochastic models. An evidence-based opinion paper. *Journal of Informetrics*, 4(3), 313-319.
- [105] Glänzel, W., Moed, H. F. (2013). Opinion paper: Thoughts and facts on bibliometric indicators. *Scientometrics*, 96(1), 381-394.
- [106] Gläser, J., Whitley, R. (Eds.). (2007). *The changing governance of the sciences: The advent of research evaluation systems*. Springer.
- [107] Golan, A. (2008). Information and Entropy Econometrics. A Review and Synthesis. *Foundations and Trends (R) in Econometrics*, 2(12), 1-145.
- [108] Graves, A., Wayne, G., Reynolds, M., Harley, T., Danihelka, I., Grabska-Barwinska, A., ... Badia, A. P. (2016). Hybrid computing using a neural network with dynamic external memory. *Nature*, 538(7626), 471-476.

- [109] Griliches, Z. (1986). Economic data issues. *Handbook of econometrics*, Volume III, edited by Z. Griliches and M.D. Intriligator, Elsevier Science Publishers, p. 1465-1514.
- [110] Griliches, Z. (1994). Productivity, R&D, and the Data Constraint. *American Economic Review* 84 (1), 1-23.
- [111] Griliches, Z. (1998). *R&D and productivity: The Econometric Evidence*. University of Chicago Press, National Bureau of Economic Research.
- [112] Hall, B. H., Rosenberg, N. (Eds.). (2010). *Handbook of the Economics of Innovation*. Elsevier, Amsterdam.
- [113] Halpern, J. Y., Vardi, M. Y. (1991). Model checking vs. theorem proving: a manifesto. *Artificial intelligence and mathematical theory of computation*, 212, 151-176.
- [114] Hanson, B., Sugden, A., Alberts, B. (2011). Making data maximally available. *Science*, 331(6018), 649-649.
- [115] Hanushek, E. A., Woessmann, L. (2007). *The role of education quality for economic growth*. World Bank Policy Research Working Paper, (4122).
- [116] Hanushek, E. A., Machin, S. J., Woessmann, L. (Eds.). (2016). *Handbook of the Economics of Education*. Elsevier.
- [117] Hanushek, E. A., Woessmann, L., Jamison, E. A., Jamison, D. T. (2008). Education and economic growth. *Education Next*, 8(2).
- [118] Helbing, D., Carbone, A. F. (Eds.) (2012). Participatory Science and Computing for Our Complex World. *The European Physical Journal*. Special Topics, 214, II.
- [119] Hemlin, S. (1996). Research on research evaluation. *Social Epistemology*, 10(2), 209-250.
- [120] Henderson, R., Jaffe, A.B., Trajtenberg, M., (1998). Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting, 1965-88, *Review of Economics and Statistics*, 119127
- [121] Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., Rafols, I. (2015). Bibliometrics: the Leiden Manifesto for research metrics. *Nature*, 520, 429-431.
- [122] Hildebrandt, S. (1977). Implementation of the operations research/management science process. *European Journal of Operational Research*, 1(5), 289-294.
- [123] Hill, S. (2016). Assessing (for) Impact: Future Assessment of the Societal Impact of Research. *Palgrave Communications*.
- [124] Hinrichs-Krapels, S., Grant, J. (2016), Exploring the effectiveness, efficiency and equity (3es) of research and research impact assessment. *Palgrave Communications*.
- [125] Horstemeyer, M. F. (2009). Multiscale modeling: a review. In Leszczynski, Jerzy; Shukla, Manoj K., Practical aspects of computational chemistry (pp. 87-135). Springer Netherlands.

- [126] ISCED (2011), International Standard Classification of Education, UNESCO Montreal, Canada, ISBN 978-92-9189-123-8. Available at: <http://www.uis.unesco.org/Education/Documents/isced-2011-en.pdf>, last accessed, on 20th December 2016.
- [127] Johnes, G., Johnes, J. (2004). *International Handbook on the Economics of Education*, Edward Elgar, Cheltenham, UK.
- [128] Johnsen, Å. (2005). What does 25 years of experience tell us about the state of performance measurement in public policy and management?. *Public Money and Management*, 25(1), 9-17.
- [129] Jordan, M. I., Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260.
- [130] Judge, G. G., Mittelhammer, R. C. (2011). *An information theoretic approach to econometrics*. Cambridge University Press.
- [131] Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux, New York
- [132] Kahneman, D., Lovallo, D., Sibony, O. (2011). Before you make that big decision. *Harvard Business Review*, 89(6), 50-60.
- [133] Kennedy, P. (2008). *A guide to econometrics*. Sixth Edition Wiley, Blackwell.
- [134] Keys P. (1991) *Operational research and systems: The systemic nature of operational research*. Springer, New York.
- [135] Khandker, S. R., Koolwal, G. B., Samad, H. A. (2010). *Handbook on impact evaluation: quantitative methods and practices*. World Bank Publications.
- [136] Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1), 1-12.
- [137] Kowarsch, M., Garard, J., Rioussset, P., Lenzi, D., Dorsch, M. J., Knopf, B., ... Edenhofer, O. (2016) Scientific assessments to facilitate deliberative policy learning, *Palgrave Communications* 2,16092(2016).
- [138] Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. Chicago (University of Chicago Press).
- [139] Lane, J. (2009). Assessing the impact of science funding. *Science*, 324(5932), 1273-1275.
- [140] Lane, J. I., Owen-Smith, J., Rosen, R. F., Weinberg, B. A. (2015). New linked data on research investments: Scientific workforce, productivity, and public value. *Research Policy*, 44(9), 1659-1671.
- [141] Largent, M. A., Lane, J. I. (2012). STAR METRICS and the Science of Science Policy. *Review of Policy Research*, 29(3), 431-438.
- [142] Latour B. (1993). *We Have Never Been Modern*. Harvard University Press, Cambridge, MA.
- [143] Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford University Press.
- [144] Lawson, T. (2012). Ontology and the study of social reality: emergence, organisation, community, power, social relations, corporations, artefacts and money. *Cambridge Journal of Economics*, 36(2), 345-385.

- [145] Leibenstein, H. (1966). Allocative efficiency vs. "X-efficiency". *The American Economic Review*, 56(3), 392-415.
- [146] Leibenstein, H. (1975). Aspects of the X-Efficiency Theory of the Firm. *The Bell Journal of Economics*, 580-606.
- [147] Leibenstein, H. (1978). On the basic proposition of X-efficiency theory. *The American Economic Review*, 68(2), 328-332.
- [148] Leibenstein, H. (1978). *General X-efficiency theory and economic development*. Oxford University Press.
- [149] Leibenstein, H., Maital, S. (1992). Empirical estimation and partitioning of X-inefficiency: a data-envelopment approach. *The American Economic Review*, 82(2), 428-433.
- [150] Lenzerini, M. (2011). Ontology-based data management. CIKM, 2011, 56.
- [151] Lewis, J. M. (2015). The politics and consequences of performance measurement. *Policy and Society*, 34(1), 1-12.
- [152] Leydesdorff, L. (2012). The Triple Helix, Quadruple Helix, and an N-tuple of helices: Explanatory models for analysing the knowledge-based economy? *Journal of the Knowledge Economy*, 3(1), 25-35.
- [153] Luke S. (2015), *Essentials of Metaheuristics*, Lulu, second edition, available at <http://cs.gmu.edu/~sean/book/metaheuristics/>, ISBN:978-1-300-54962-8.
- [154] Lutz C.S. (2017) Alasdair Chalmers MacIntyre, in The Internet Encyclopedia of Philosophy, ISSN 2161-0002, <http://www.iep.utm.edu/mac-over/> last accessed 7 January 2017.
- [155] MacIntyre, A. (2006). The end of education: The fragmentation of the American university. *Commonweal*, 133(18), 10-14.
- [156] MacIntyre, A. (2007). *After Virtue: A Study in Moral Theory*, University of Notre Dame Press, Third Edition.
- [157] MacIntyre, A. (1988). *Whose justice? Which rationality?*, University of Notre Dame Press.
- [158] Mansfield, E. (1991). Academic research and industrial innovation. *Research Policy*, 20, 1-12.
- [159] Mansfield, E. (1995). Academic research underlying industrial innovations: Sources, characteristics, and financing. *The Review of Economics and Statistics*, 77(1), 55-65.
- [160] Martin, B. R. (1996). The use of multiple indicators in the assessment of basic research. *Scientometrics*, 36(3), 343-362.
- [161] Martin, B. R. (2012). The evolution of science policy and innovation studies. *Research Policy*, 41(7), 1219-1239.
- [162] Martin, B. R. (2013), Innovation studies: An Emerging Agenda, in Fagerberg, J., Martin, B. R., Andersen, E. S. (Eds.). *Innovation studies: evolution and future challenges*. OUP Oxford, pp. 168-186.
- [163] Martin, B. R. (2016). Introduction to discussion paper on The sciences are different and the differences are important. *Research Policy*, 9(45), 1691.

- [164] Martin, B. R., Irvine, J. (1983). Assessing basic research: some partial indicators of scientific progress in radio astronomy. *Research Policy*, 12(2), 61-90.
- [165] Mazzucato, M. (2013). *The entrepreneurial state: debunking the public vs. private myth in risk and innovation*. Anthem, London.
- [166] Mezard, M., Montanari, A. (2009). *Information, physics, and computation*. Oxford University Press.
- [167] Mingers, J. (2006). *Realising systems thinking: knowledge and action in management science*. Springer Science & Business Media.
- [168] Mingers, J., Gill, A. (1997). *Multimethodology: theory and practice of combining management science methodologies*. Wiley.
- [169] Mirowski, P., Sent, E. M. (2002). *Science bought and sold: Essays in the economics of science*. University of Chicago Press.
- [170] Moed, H. F. (2016). Altmetrics as traces of the computerization of the research process. In C. R. Sugimoto (Ed.), *Theories of informetrics and scholarly communication*. A Festschrift in honor of Blaise Cronin (pp. 360-371). Berlin: De Gruyter.
- [171] Moed, H. F., Halevi, G. (2015). Multidimensional assessment of scholarly research impact. *Journal of the Association for Information Science and Technology*, 66(10), 1988-2002.
- [172] Moed, H. F., Van Leeuwen, T. N. (1996). Impact factors can mislead. *Nature*, 381 (6579), 186-186.
- [173] Moed, H. F., Glänzel, W., Schmoch, U. (2004). *Handbook of Quantitative Science and Technology Research*, Springer Netherlands.
- [174] Moore, S., Neylon, C., Eve, M. P., O'Donnell, D., Pattinson, D. (2017). Excellence R Us: University Research and the Fetishisation of Excellence. *Palgrave Communications*.
- [175] Moreau, L., Freire, J., Futrelle, J., McGrath, R. E., Myers, J., Paulson, P. (2008). The open provenance model: An overview. In International Provenance and Annotation Workshop (pp. 323-326). Springer Berlin Heidelberg.
- [176] Moscati R. (2008), Transforming a Centralised System of Higher Education: Reform and Academic Resistance in Italy, in A.Amaral, I.Bleiklie, C.Musselin (edited by), From Governance to Identity. A Festschrift for Mary Henkel, Dordrecht, Springer.
- [177] Moscati R. (2010), The Implementation of the Bologna Process in Italy, in A.Amaral, G.Neave, C.Musselin, P.Maassen (edited by), European Integration and the Governance of Higher Education and Research, Dordrecht, Springer, pp.207-226.
- [178] Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., du Sert, N. P., ... Ioannidis, J. P. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1, 0021.
- [179] National Research Council. (2014a). Science of Science and Innovation Policy, K. Huskisson Fealing, A.S. Beatty, and C.F. Citro, Rapporteurs. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

- [180] National Research Council (2014b). *Convergence: Facilitating Transdisciplinary Integration of Life Sciences, Physical Sciences, Engineering, and Beyond*. Washington, DC: The National Academies Press.
- [181] Nelson, R. R., Phelps, E. S. (1966). Investment in humans, technological diffusion, and economic growth. *The American Economic Review*, 56(1/2), 69-75.
- [182] Nielsen, M. (2012). *Reinventing discovery: the new era of networked science*. Princeton University Press.
- [183] O'Donnell, C. J. (2016). Using information about technologies, markets and firm behaviour to decompose a proper productivity index. *Journal of Econometrics*, 190(2), 328-340.
- [184] OECD (2011). *Quality Framework and Guidelines for OECD Statistical Activities*. OECD Publishing, Paris.
- [185] OECD (2015). *Making Open Science a Reality*. OECD Science, Technology and Industry Policy Papers No. 25, OECD Publishing, Paris.
- [186] Oslo Manual (2005) *Guidelines for Collecting and Interpreting Innovation Data*, 3rd edition, ISBN 978-92-64-01308-3, available at: [http : //www.oecd – ilibrary.org/science – and – technology/oslo – manual_9789264013100 – en](http://www.oecd-ilibrary.org/science-and-technology/oslo-manual_9789264013100-en) , last accessed on 20thDecember2016.
- [187] Owen, R., Macnaghten, P., Stilgoe, J. (2012). Responsible research and innovation: From science in society to science for society, with society. *Science and public policy*, 39(6), 751-760.
- [188] Parent, C., Spaccapietra S. (2000) Database integration: the key to data interoperability. In *Advances in Object-Oriented Data Modeling*, Papazoglou M.P., Zari Z. Eds. (2000), The MIT press, 221-253.
- [189] Parmeter, C. F., Wan, A. T., Zhang, X. (2016). A Model Averaging Stochastic Frontier Estimator, paper presented at the NAPW 2016 Quebec City, Canada June 2016.
- [190] Perkmann, M., Tartari, V., McKelvey, M., Autio, E., Broström, A., DEste, P., ... Krabel, S. (2013). Academic engagement and commercialisation: A review of the literature on university-industry relations. *Research Policy*, 42(2), 423-442.
- [191] Piro, F. N., Sivertsen, G. (2016). How can differences in international university rankings be explained?. *Scientometrics*, 1-16.
- [192] Poggi, A., Lembo, D., Calvanese, D., De Giacomo, G., Lenzerini, M., Rosati, R. (2008). Linking data to ontologies. In S. Spaccapietra (Ed.), *Journal on Data Semantics X, Lecture Notes in Computer Science* (Vol. 4900, pp. 133173). Berlin: Springer.
- [193] Porter, M. E., Millar, V. E. (1985). How information gives you competitive advantage.
- [194] Rebora, G., Turri, M. (2013). The UK and Italian research assessment exercises face to face. *Research Policy*, 42(9), 1657-1666.
- [195] Research Excellence Framework (REF, 2014), [http : //www.ref.ac.uk/about/](http://www.ref.ac.uk/about/), last accessed on December 21st2016.
- [196] Romer, P. M. (1986). Increasing returns and long-run growth. *The Journal of Political Economy*, 1002-1037.

- [197] Romer, P. M. (1994). The origins of endogenous growth. *The Journal of Economic Perspectives*, 8(1), 3-22.
- [198] Roper, C. D., Hirth, M. A. (2005). A history of change in the third mission of higher education: The evolution of one-way service to interactive engagement. *Journal of Higher Education Outreach and Engagement*, 10(3), 3-21.
- [199] Rosenberg, N., Nelson, R.R., (1994). American Universities and Technical Advance in Industry. *Research Policy* 23, 323-348.
- [200] Saltelli, A., Funtowicz, S. (2014). When all models are wrong. *Issues in Science and Technology*, 30(2), 79-85.
- [201] Saltelli, A., Annoni, P. (2010). How to avoid a perfunctory sensitivity analysis. *Environmental Modelling & Software*, 25(12), 1508-1517.
- [202] Saltelli, A., Funtowicz, S. (2015). Evidence-based policy at the end of the Cartesian dream: The case of mathematical modelling. In: *Science, philosophy and sustainability: the end of the Cartesian dream*, Routledge Explorations in Sustainability and Governance. Routledge, New York, 147-162.
- [203] Saltelli, A., Guimarães Pereira, Â., Van der Sluijs, J. P., Funtowicz, S. (2013). What do I make of your latinorum? Sensitivity auditing of mathematical modelling. *International Journal of Foresight and Innovation Policy*, 9(2-3-4), 213-234.
- [204] Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., ... Tarantola, S. (2008). *Global sensitivity analysis: the primer*. John Wiley & Sons.
- [205] Salter, A. J., Martin, B. R. (2001). The economic benefits of publicly funded basic research: a critical review. *Research Policy*, 30(3), 509-532.
- [206] Scharnhorst, A., Börner, K., van den Besselaar, P. (Eds.). (2012). *Models of science dynamics: encounters between complexity theory and information sciences*. Springer Science & Business Media.
- [207] Schultz, R. L., Slevin, D. P. (1975). *Implementing operations research/management science*. American Elsevier Pub. Co.
- [208] Simar, L., Wilson, P. W. (2015). Statistical Approaches for Nonparametric Frontier Models: A Guided Tour. *International Statistical Review*, 83(1), 77-110.
- [209] Simon, H. A. (1969). *The sciences of the artificial*. Cambridge, MA.
- [210] Simon, H. A. (1982). *Models of bounded rationality: Empirically grounded economic reason* (Vol. 3). MIT press.
- [211] Simon, H. A. (2000). Bounded rationality in social science: Today and tomorrow. *Mind & Society*, 1(1), 25-39.
- [212] Solow, R. M. (1957). Technical change and the aggregate production function. *The Review of Economics and Statistics*, 312-320.
- [213] Stephan, P. (2012b). Research efficiency: Perverse incentives. *Nature*, 484(7392), 29-31.

- [214] Stephan, P. E. (2012a). *How economics shapes science*. Cambridge, MA: Harvard University Press.
- [215] Stewart, J. (1995). Models of priority-setting for public sector research. *Research Policy*, 24(1), 115-126.
- [216] Stilgoe, J., Owen, R., Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, 42(9), 1568-1580.
- [217] Teixeira, P. N., Dill, D. D. (Eds.). (2011). *Public vices, private virtues?: Assessing the effects of marketization in higher education* (Vol. 2). Springer Science & Business Media.
- [218] Teixeira, P., Jongbloed, B., Dill, D., Amaral, A. (2004). *Markets in higher education: Rhetoric or reality?*. Dordrecht: Kluwer.
- [219] Tversky, A., Kahneman, D. (1974). *Judgment under Uncertainty: Heuristics and Biases*. Science (New York, NY), 185(4157), 1124-1131.
- [220] Van den Hove, S. (2007). A rationale for science-policy interfaces. *Futures*, 39(7), 807-826.
- [221] Vandenbussche, J., Aghion, P., Meghir, C. (2006). Growth, distance to frontier and composition of human capital. *Journal of Economic Growth*, 11(2), 97-127.
- [222] Veugelers, R., Del Rey, E. (2014). The contribution of universities to innovation,(regional) growth and employment. EENEE Analytical Report 18thJan.2014, [http : //www.eenee.de/dms/EENEE/AnalyticalReports/EENEE_AR18.pdf](http://www.eenee.de/dms/EENEE/AnalyticalReports/EENEE_AR18.pdf), last accessed, 28thSeptember2016.
- [223] Vinkler, P. (2010). *The evaluation of research by scientometric indicators*. Chandos Publishing, Cambridge, UK.
- [224] von Hippel (2005), *Democratizing Innovation*, MIT Press, Cambridge, MA.
- [225] von Hippel, Eric (2016) *Free Innovation*. MIT Press, Cambridge, MA forthcoming.
- [226] Weinan, E. (2011). *Principles of multiscale modeling*. Cambridge University Press.
- [227] West, J., Salter, A., Vanhaverbeke, W., Chesbrough, H. (2014). Open innovation: The next decade. *Research Policy*, 43(5), 805-811.
- [228] Wilsdon, J., et al. (2015). The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management. DOI: 10.13140/RG.2.1.4929.1363
- [229] Woelert, P. (2015). The logic of escalation in performance measurement: An analysis of the dynamics of a research evaluation system. *Policy and Society*, 34(1), 75-85.
- [230] Zenger, T. (2016). *Beyond Competitive Advantage*. Harvard Business Review Press, Boston.