# 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

**Cinzia Daraio** is an Associate Professor of Management Engineering at the University of Rome "La Sapienza" where she teaches Productivity and Efficiency Analysis, Quantitative Models for Economic Analysis and Management, and Economics and Business Organization. She has been invited keynote speaker, speaker and/or visiting researcher in several international research centers in Belgium, China, Denmark, France, Germany, Romania, Spain, Sweden, Switzerland, and US. She is specialized in science and technology indicators, higher education microdata, and methodological and empirical studies in productivity and efficiency analysis. Over the years, she participated in several international and national research projects and expert groups at the European Commission on these subjects. She is member of the Editorial Board of *Scientometrics*, *International Transactions in Operational Research*, *Journal of Productivity Analysis*, *Journal of Data and Information Science*, *Scholarly Metrics and Analytics* and others. She is referee for more than 35 journals. She was project evaluator for the Rustaveli National Science Foundation, external reviewer for the Portuguese Foundation for Science and Technology and member of the Commission of Experts for the Evaluation of the Third Mission Activities of the Italian National Agency for the Evaluation of Universities and Research. She authored more than 150 publications including a monograph, an edited book, many articles in international peer-reviewed journals, and several chapters and papers presented at conferences and in proceedings.

**Abstract:** This paper proposes a *holistic* framework for the development of *models* for the assessment of research activities and their impacts. It distinguishes three dimensions, including in an original way, *data* as a main dimension, together with *theory* and *methodology*. Each dimension of the framework 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

† Corresponding author: Cinzia Daraio (E-mail: cinzia.daraio@uniroma1.it).

7

**Perspective**

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. 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. It can act as a *common denominator* for different analytical levels and relevant aspects and is able to embrace many different and heterogeneous streams of literature. Directions for future research are provided.

**Keywords** Evaluation of research; Efficiency; Effectiveness; Impacts; Modeling; Responsible metrics; Ethics of research evaluation

## 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* (Dahler-Larsen, 2012; Whitley & Gläser, 2007).
- The *need of policy-makers* to have a comprehensive framework. We refer to the STAR METRICS[1] in the US (Largent & Lane, 2012) 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 recently been 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

---

[1] STAR METRICS is a data platform that is voluntarily and collaboratively developed by US federal science agencies and research institutions to describe investments in science and their results (Largent & Lane, 2012).

[2] The first objective of the European Commission (2014) Expert Group, in which the author of the present paper took part, was indeed to "propose an analytical framework for identifying how the implementation of different ERA priorities and components observed at institutional level (i.e. research performing organizations) and national level (i.e. national policies and funding organizations policies) impact the research system performance (at institutional and national level)."

societies about the broader use of metrics in research assessment and manage-
ment" 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 also Munafò et al. (2017)), peer review,
  publication metrics, scientific leadership, scientific integrity, and the use of
  science for policy (see also Saltelli & Funtowicz (2015) in *The End of the
  Cartesian Dream*). The transmission channel of this crisis from science to
  scientific advice 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, policy, and legitimate norms. The obsolescence of
  the classical opposition between scientific approach and dogmatic approach,
  generated by the problems of the empirical evidence (Saltelli & Funtowicz,
  2015) may be a possible root of this crisis.
- The recent debate on *modeling 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 & 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. The results obtained are obviously linked to intrinsic
potential or limitations 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 toward a
data-driven science (Kitchin, 2014), supporting "the end of theory" (Anderson,
2008), or will theory-driven scientific discoveries remain unavoidable (Frické,
2015)? There is little agreement in the literature. More balanced views emerging
from a critical analysis of the current literature are also available (Debackere, 2016;
Ekbia et al., 2015), leading the information systems community to further deeply
analyze the critical challenges posed by the big data development (Agarwal & Dhar,
2014).

Data sources 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

**Perspective**

have to face the huge amount, complexity, and variety of the data that is 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.

Often, indicators and metrics are used as synonyms (see also Wilsdon et al. (2015)). In this paper, indicators are combinations of data that produce *values*, while metrics are considered parameters or measures of quantitative assessment used for measurement, comparison, or to track performance. 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 (see Daraio (2017a) for further details).

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

- *Learning*, to learn about the explicit consequences of assumptions, test the assumptions, and highlight relevant relations;
- *Improving*, to better operate, document/verify the assumptions, decompose analysis and synthesis, systematize the problem and the evaluation/choice made, and state clearly and in detail the dependence of the choice to the scenario.

More specifically, a model is an abstract representation, which from some points of view and for some ends represents an object or real phenomenon[③]. The representation of reality is achieved through the *analogy* established between aspects of reality and aspects of the model.

---

[③]  Some interesting readings on modeling can be found in Morris (1967), Pollock (1976), Willemain (1994), Myung (2000), and Zucchini (2000).

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; and
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 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 modeling has to take into account the specificities of the processes and systems under investigation, and in particular consider that the behavior is free and finalized to given aims; history and evolution matter as the behavior of systems and processes changes over time (see e.g. Georgescu-Roegen (1971)).

Hence, the modeling activity related to the assessment of research involves several methodological challenges. What is required today is to develop models, able to characterize strongly connected or *interdependent* model components, dominated by their *interactions*, including complex model behavior, emergent collective behavior which implies new and often unexpected model behavior, counter intuitive behavior, and extreme events with less predictable outcomes, and management based on setting rules for *bottom up self-organization* (Helbing & Carbone, 2012, p. 15). This is very different from the traditional models, characterized by independent model components, based on simple model behavior, where the sum of properties of individual components characterizes model behavior, 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.

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 aspects for the purpose. The finalization may encourage a functional analysis of the systems involved in the assessment. The external behavior of the systems may be explained focusing the analysis on the aims and the ways of interacting with the environment without entering into the details of the internal structures and organization (the organization may become relevant only if it is a limit to pursuing the objectives of the system).

---

® In this paper evaluation and assessment are used as synonyms.

**Perspective**

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, and evolution of behaviors);
- 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:

- Possibility that the *targets* are *not quantifiable*, or are *multiple* and *conflicting*; or that there are several decision-makers with different interests;
- *Complexity*, *uncertainty*, and *changeability* of the *environment* in which the observed system works and, after environmental stimuli, the difficulty of predicting the consequences of certain actions and relative responses;
- The *limits* (in particular of an *organizational* nature) within which the analyzed system adapts to the *directives of the decision-maker*; and
- 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 to developing multidimensional and multilevel models for the evaluation of research and its impacts[⑤].

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

1) *Theory*, broadly speaking, identifies the conceptual content of the analysis, answering the question of "what" is the domain of interest, and delineating the perimeter of the investigation;
2) *Methodology*, generally refers to "how" the investigation is handled, what are the kind of tools that can be applied to the domain of interest, and tools which represent the means by which the analyses are carried out; and
3) *Data*, largely, and roughly, refers to instances coming from the domain of interest, and represents the means, on (or through) 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) education, 2) research, and 3) innovation. See Table 1 for their definition.

⑤  Vinkler (2010) presents a systematic view of units and levels of analysis in research assessment.

Table 1.   Definitions of education, research, and innovation.

| Term | Definition |
| --- | --- |
| Education | In 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 themselves. 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. |
| Research | According 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 toward 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 (OECD, 2015b). |
| Innovation | According to the OECD (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 main building blocks of *methodology* are: 1) efficiency, 2) effectiveness, and 3) impact. The main building blocks of *data* are: 1) availability, 2) interoperability, and 3) unit-free property.

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

The three main *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 the proposed/selected *theory*/*methodology*/*data* combination); and
3) *Openness* (roughly, accessibility to the main elements of the modeling).

The more we are able to go to the deep, fine-grain of the most atomic level-unit of analysis, the higher the level of tailorability, the higher the level of transparency
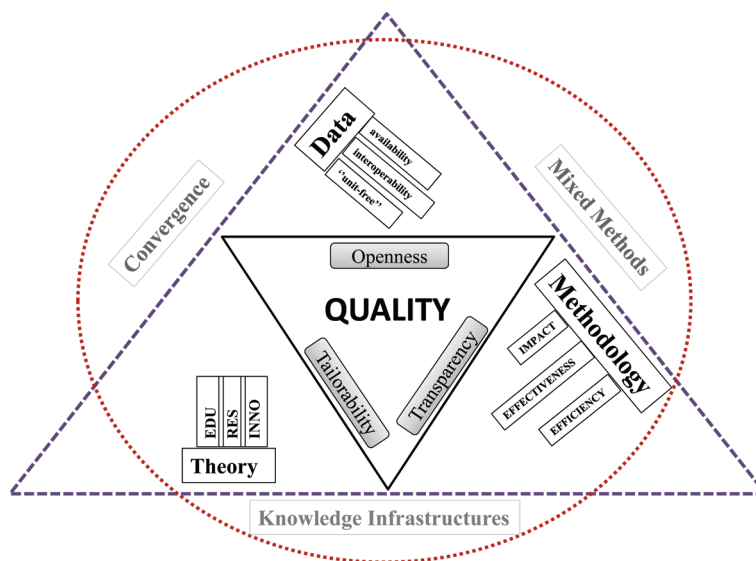
Figure 1.    An illustration of our framework including its three implementation factors (tailorability, transparency, and openness) and its three enabling conditions: convergence, mixed methods, and knowledge infrastructures.

and openness may be, and the better will be the conceptualization and formalization of quality within a model.

In this paper, we assert that the ability of developing (and afterward understanding and effectively using) models for the assessment of research is linked and depends, among other factors, on the degree or depth of the *conceptualization* (intended here as the formulation of the content of the general ideas and of the most important details) and *formalization* (intended here as "to make it official" or explicit), 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 and increasing the dimensional space of thinking);

2) *Mixed methods* (as an intelligent combination of quantitative and qualitative approaches); and

3) *Knowledge infrastructures* (as networks of people that interact 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 fosters the overlap of the different perspectives, namely modeling world, empirical world, and policy world (see Section 4 and Figure 2 in Section 5).

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 modeling choice. It simply means that you do not state clearly and in detail and account for the underlying theoretical choices, methodological assumptions, and data limits considerations. Put in other words, the metric cannot 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 (degree) of conceptualization and formalization, in an unambiguous way, of the "Quality" and its different layers and meanings.*

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. A further discussion on this issue can be found in Daraio (2017a).

**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 and meanings is able to substantiate and give content to the concept of "responsible" metrics.

Postulate 3 should be considered to be 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 (e.g. a kind of *checklist* when practitioners plan an assessment);
- To propose a basis for the research of the *ethics* of research evaluation; and
- 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. 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, Jaffe, & Trajtenberg, 1998; Mansfield, 1995; Rosenberg & 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. 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 firms.

Table 2 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.

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

Moreover, higher education systems are increasingly expanding their interplay with society moving toward markets in higher education systems or going beyond. There are some science and public policy studies that have analyzed the elements

Table 2.   A non-exhaustive overview of the literature on the *theory* dimension.

| Literature stream | References |
| --- | --- |
| *Economics of science and technology* | Audretsch et al. (2002), Stephan (2012), Mirowski & Sent (2002) |
| *Theories of growth* | Solow (1957), Abramovitz (1956), Nelson & Phelps (1966), Romer (1986; 1994), Aghion & Howitt (2009) |
| *Quantitative science and technology research* | Egghe & Rousseau (1990), Egghe (2005), Moed, Glänzel, & Schmoch (2004), Ding, Rousseau, & Wolfram (2014), Cronin & Sugimoto (2014; 2015), Glänzel et al. (in press) |
| *Economics of innovation, innovation studies* | Fagerberg, Martin, & Andersen (2013), Hall & Rosenberg (2010) |
| *Science of science, science and society* | Fealing et al. (2011, p. 4), Gibbons et al. (1994), Etzkowitz & Leydesdorff (2000), Edquist (2001), Aghion, David, & Foray (2009), Helbing & Carbone (2012), Scharnhorst, Borner, & van den Besselaar (2012), Nowotny, Scott, & Gibbons (2001), Barré (2004) |
| *Economics of knowledge* | Antonelli & Link (2014) |
| *Economics of education* | Blaug (1966), Johnes & Johnes (2004), Checchi (2006), Hanushek, Machin, & Woessmann (2016) |
| *Education and society* | Roper & Hirth (2005), Teixeira et al. (2004), Teixeira & Dill (2011) |
| *Societal impact, science and public policy studies* | Bornmann (2013), Ebrahim & Rangan (2014), Perkmann et al. (2013), Veugelers & Del Rey (2014), Hill (2016) |
| *History of science* | Godin (2002; 2004), Gingras (2016) |

of societal impact, mostly rooting it into universities and public research characteristics (Bornmann, 2013) whilst others, mostly refer to approaches developed by practitioners (Ebrahim & Rangan, 2014)[©]. All these theoretical considerations related to the so-called *third mission activities* of higher education institutions and research centers (Veugelers & 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.

The existing literature, summarized in Table 2, 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 on teaching activities; competence of teachers; skills and the initial level of education of students; educational infrastructures, and other resources.

[©]  An interesting survey on university-industry relations can be found in Perkmann et al. (2013).

**Perspective**

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 literature 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. 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 re-configure human resources and particularly knowledge in new ways (Gibbons et al., 1994, p. 63). 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 matter for the assessment of research and its impact.

The assessment of research cannot 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, and infrastructures; intellectual ability of researchers, historical paths, ability to recruit doctoral students, world-wide network of contacts, and the like.

## 2.2   Methodology

*Methodology*, in the setting of our framework, identifies the range of methods, techniques, and approaches that are relevant to the evaluation of research. A preamble is necessary here before entering into the more detailed description. 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) from the "means" or the tools of the assessment (how to assess). We identify the *subject* in: outputs, efficiency, effectiveness, and impact and the *means* in *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.

A 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 e.g. Bammer (2016).

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 *methodology*: efficiency, effectiveness, and impact, considering the outputs as a kind of baseline or step zero in the analysis, followed by the subsequent steps (Table 3).

Table 3.    Dimensions of *methodology*: subject and means in our framework.

| Dimension | Type/category | Content |
|---|---|---|
| *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 with respect to a reference |
| | Effectiveness | Considering inputs and outputs, and accounting for the aims of the activities |
| | Impact | All contributions of research outside academia |
| *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 relation with respect to an estimated reference frontier, or frontier of the best practices (Daraio & Simar, 2007, p. 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 behavior whilst production frontiers analyze the whole distribution, taking into account the best/worst behavior (Bonaccorsi & 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 or representative

behavior model. This is because in the former case the whole distribution is considered instead of only the central tendency. This distinction between "average" *vs* "frontier" is considered in recent theory of growth (Acemoglu, Aghion, & Zilibotti, 2003; 2006; Vandenbussche, Aghion, & Meghir, 2006) and in the managerial literature (Chen, Delmas, & Lieberman, 2015) as well.

As far as quantitative methods are concerned, different approaches, both parametric (Galán, Veiga, & Wiper, 2014) and non-parametric (Bădin, Daraio, & Simar, 2012; 2014; Daraio & 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 toward 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 non-parametric frontier estimation (Daraio, Simar, & Wilson, 2016; Simar & Wilson, 2015): trend toward a data-driven modeling;
- Models averaging in stochastic frontier estimation (Parmeter, Wan, & Zhang, 2016): trend toward robustness of modeling;
- Using information about technologies, markets, and behavior of institutions in productivity indices (O'Donnell, 2016); trend toward more comprehensive informational setup; and
- From an implementation point of view, interactive benchmarking (e.g. Bogetoft, Fried, & Eeckaut 2007); trend toward developing analytics for policy decision making support.

Moving from efficiency to effectiveness is an important step. At this purpose, the inclusion of managerial and more qualitative aspects in the quantitative benchmarking models could be beneficial. According to Drucker (1967), *effectiveness* is "doing the right thing" while *efficiency* is "doing the thing right." Effectiveness is similar to principle 6 of Saltelli and Funtowicz (2014, see below).

The methodological dimension should handle how to evaluate what, providing an appropriate account of reliability and robustness (Glänzel, 2010; Glänzel & Moed, 2013), and uncertainty.

An interesting distinction exists between uncertainty and sensitivity analysis. *Uncertainty* analysis focuses on the quantification of the uncertainty in the model output. *Sensitivity analysis* instead analyzes the relative importance of different input factors on the model output. Global sensitivity analysis (Saltelli et al., 2008) refers to the investigation of how the uncertainty of the inputs of the model 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.

These are all considerations which refer to the *quality-methodology* intersection. Classical methods of impact assessment (see e.g. Bozeman & 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, Koolwal, & Samad (2010) for an overview) are challenged by the "problem of evaluation [which] 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, Koolwal, & Samad, 2010, p. 22). These classical methods appear inadequate to the checklist of *sensitivity auditing* (Saltelli et al., 2013; Saltelli & Funtowicz, 2014; 2015), proposed by Saltelli and Funtowicz (2014, p. xxx), which 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 attention;
5) Aim for *transparency* (increasing the diffusion of the used models basic ideas avoiding jargon);
6) Do not 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).

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

**Perspective**

### 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* is a relevant dimension often neglected in modeling building. According to Frischmann (2012), data includes facts and statistics collected together for reference or analysis; data is representation, reinterpretable representation of information in a formalized manner, suitable for communication, interpretation, or processing, up to data as "infrastructure."

Besides this positive view on data, data has a *problematic* definition because it depends on its use not on inherent characteristics of the data (Borgman, 2015, p. 74): "Their properties and their weaknesses affect both the modeling 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 endeavor." "Despite the overall lack of agreement, most scholars would like better means to manage whatever they do consider to be their data" (Borgman, 2015, p. 271). 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, and unit-free property.

*Availability* refers to general alternatives and choices that affect the data to be used, for instance (without being complete): sampling *vs* census, freely available *vs* controlled or undisclosed ones, data as consumption *vs* participation (see Ekbia et al. (2015) for a critical discussion). Obviously, the minimal requirement for the elaboration of data refers to its availability in a usable way. This opens to the discussion on commercial *vs* publicly available (or open) data; institutional provided data, and 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 & Spaccapietra, 2000). It is crucial for data integration of heterogeneous sources (see Daraio &

Glänzel (2016). See also the discussion on continuity *vs* innovation in Ekbia 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 (Lenzerini, 2011). An OBDM approach may be considered to be a kind of *information integration* based on: a) a conceptual description of the knowledge domain of interest, called the *ontology*; b) the different databases where the data of the domain is kept, called the *sources*; c) the correspondences between the data contained in the *sources* and the concepts of the ontology, called the *mappings*. The main advantages of an OBDM approach for integrating research and other scholarly data (Daraio 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.

*Unit-free property* refers to the need to have consistent and coherent observations (instances of data) at different levels of analysis, to ensure robust empirical evidence of a given phenomenon. The unit-free property of data is somewhat interconnected to the possibility of *multiscale modeling* of the problem at hand. The 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 characterize structure-property relations down to the atomic level, and theories that admit multiple length scales. This kind of modeling makes clear the need of having data that is independent from the unit of analysis and hence can be used coherently in a multiscale model of the problem. See Table 4 for an overview of the *data* dimensions and its characterization in our framework.

A relevant connection, also for the following developments of modeling is the relationship between data and information. According to Floridi (2014), information and communication technologies (ICT) 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

Table 4.    A characterization of the *data* dimension in our framework.

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

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, 2008).

An interesting and perhaps connected change, due to the developments introduced in information processing including novel algorithms, protocols, and properties of information brings to shift from the classical to the quantum computation paradigm and recently leads to derive quantum theory as a special theory of information (D'Ariano & Perinotti, 2016. For an introduction, see Nielsen & Chuang (2010).).

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 analyzed 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 the three main dimensions illustrated 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, interdependencies, complementarities among the three dimensions and allows for analyzing the constraints on the three dimensions that may make analysis difficult.

Concerning Quality, in the field of education, much progress has been made. The quality of education has been demonstrated as relevant for research and innovation.

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

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 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."

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 are closely related to those of quality of information. Problems of quality of information are involved whenever policy related research is utilized in the policy process (Funtowicz & 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 & 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. In the end, maybe, although difficult to assess, it is the quality of education, research, and innovation, which has an impact 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, and unit-free property).

Data 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 it is designed to measure"); credibility ("confidence of users in the data products and trust in the objectivity of the data"); timeliness ("length of time between its availability and the phenomenon it describes"); accessibility ("how readily the data can be located and accessed"); interpretability ("the ease with which the user may understand and properly use and analyze the data"); coherence ("the degree to which it is logically connected and mutually consistent").

Quality of available data is crucial; in data quality there have been relevant advances, going from data quality to *information quality* (Batini & Scannapieco, 2016) and developing a *philosophy of data information quality* (Floridi & Illari, 2014). The quality of the interoperability is important in the integration of heterogeneous data sets 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 reuse. The provenance initiative (Moreau et al., 2008) is a clear example of describing better data for different purposes, including also the opening or sharing of data.

Quality as acceptability (suitability) for application (fitness for purpose) is the overarching concept, which keeps together the building blocks of the three dimensions of our framework. It is a characteristic in all the three dimensions.

The nine building blocks are attributes of Quality. The quality of *theory*, as dimension, is related to the problem of boundaries and philosophical representation of reality. The degree of implementation of the assessment of quality is related to the level and intensity of the resolution of the underlying "problem of evaluation." It is linked to the implementation factor tailorability. The quality of *methodology*, 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. 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 research and its impacts; 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.

## 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 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 evolution, 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 again 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 review and 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 *à la* Cohen and Levinthal (1990). In *methodology* we should account for a *multimethodology* approach (Mingers, 2006), according to which, instead of a single method, a combination of methods, both *hard* and *soft*, is used.

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 & Young (2015)) which refers to the transformation of higher education toward new ways of disseminating knowledge at lower cost, such as Massive Open Online Courses (MOOCS), thanks to technology fuelled innovations, and research on learning processes.

According to OECD (2015a), 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."[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

---

[⑦] "[…] 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, 2015a, p. 7).

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 a modular way; encouraging small contributions; allowing easy reuse of earlier work; using signaling mechanisms (e.g. scores) to help people to decide where to direct attention.

The exponential increase and development of information availability and the development of the information society is leading us toward an open innovation society (see e.g. Chesbrough (2012))® based on a Quadruple Helix model (Leydesdorff, 2012) of bottom up interactive policy framework.

In 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 facilitates the "democratization" of knowledge (von Hippel, 2005), which 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.

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.

The first enabling condition is *mixed methods* which relates to the combination of quali-quantitative analysis. It offers strengths that offset the weaknesses of both quantitative and qualitative research (e.g. Creswell & Clark, 2011). Quantitative

---

® West et al. (2014) in reviewing the open innovation literature identify three main directions of research: better measurement, resolving the role of appropriability, and linking open innovation to the management and economics literature.

methods are weak in understanding the context, and qualitative methods (on the other hand) are weak because of personal interpretation and difficulty in generalizing. A bridge across the adversarial divide, between quantitative and qualitative, encourages the use of multiple paradigms (beliefs and values), and is practical to solve problems and 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 monitored mechanisms.

The second enabling condition refers to *convergence* 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, 2014).

The third enabling condition refers to the *knowledge infrastructures* 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)⑨.

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

## 5   Toward 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 to 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 modeling 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 shows the role of our modeling 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

---

⑨ Within some research projects funded by Sapienza University in 2013 and 2015 we did an experiment of a knowledge infrastructure, a case of an "open science of science" exercise, around Sapientia: The Ontology of Multi-Dimensional Research Assessment (Daraio et al., 2016a; 2016b). Sapientia represents an effort of going toward a common platform which can show which data has to be collected; by offering the opportunity of making analysis under different perspectives, testing different models, but sharing the same common conceptual characterization.
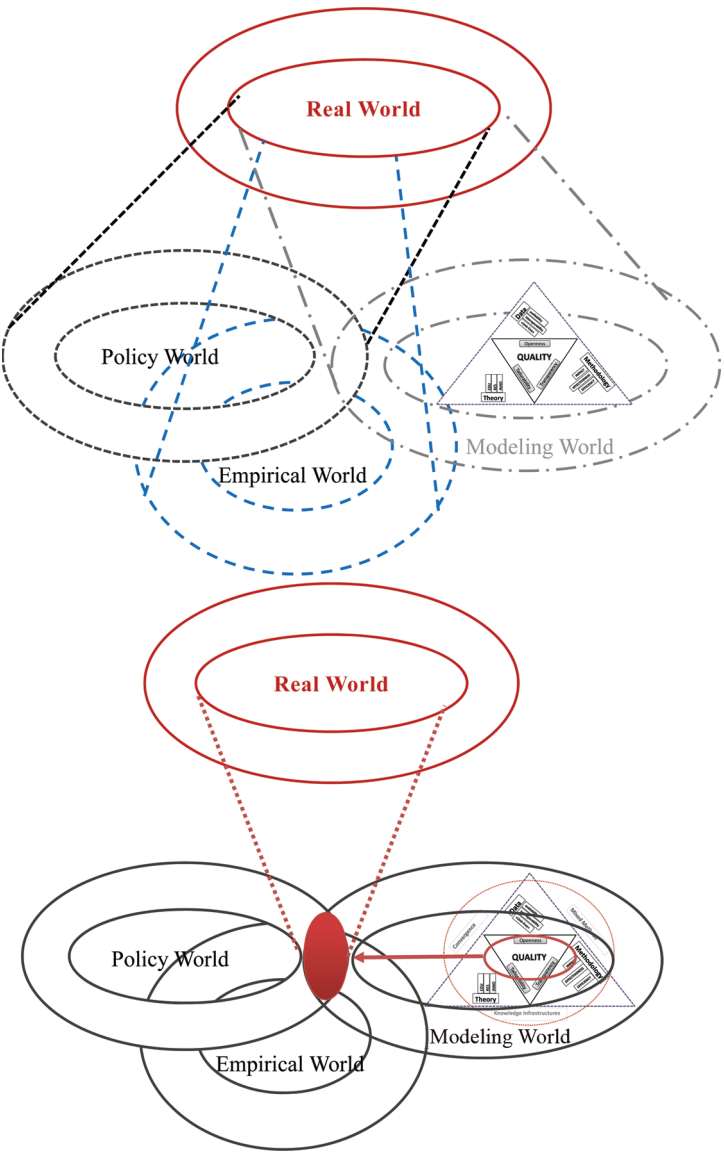
Figure 2.    An illustration of the relationship between modeling world, empirical world, and policy world: they are all somewhat overlapping visions or projections of the real worlds.

area among modeling, policy, and empirical worlds is, and closer to the real world the model is.

This statement is basically 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 also calls for the introduction of the third postulate which is the monitoring of the developments and the evolutions of the modeling activity can be carried out in relation to the "responsibility" of the metrics proposed and involved.

But what does being a "responsible metric" mean in an evaluation process? According to 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 [. . .]":

> "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: recognizing 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: recognizing and anticipating the systemic and potential effects of indicators, and updating them in response" (Wilsdon et al., 2015, p. x).

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*," was published in July 2015 (Wilsdon et al., 2015) a website for responsible metrics has been established[®].

Coming back to our framework, we identify some connections of its enabling conditions with the oeuvre of Alasdair MacIntyre[®]. They are illustrated in Table 5.

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)

---

[®] From the website https://responsiblemetrics.org/ (last accessed on February 10, 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 Owen, Macnaghten, and Stilgoe (2012) and Stilgoe, Owen, and Macnaghten (2013).

[®] See for instance Lutz (2017) and Ballard (2000) that describes an overview on MacIntyre's oeuvre, reporting a rich bibliography on his works.

**Perspective**

Table 5.   Toward an ethics of research assessment? Some connections of our framework with MacIntyre' oeuvre.

| Enabling condition | Potential connection | MacIntyre work |
|---|---|---|
| *Convergence* | Invitation to overcome the fragmentation of knowledge and excessive specialization. | *The end of education* MacIntyre (2006) |
| *Mixed methods* | The need to go beyond a pure quantitative approach (abstract representation of the reality) and include qualitative cases (narratives and storytelling). | *After virtue*, MacIntyre (2007), & *Whose justice, which rationality?* MacIntyre (1988) |
| *Knowledge infrastructures* | Retrieve the values of tradition in communities of practice that regulate themselves by defining their own standards. | *After virtue* MacIntyre (2007) |

of conceptualization and formalization, in an unambiguous way, of the different layers/meanings of "Quality."

This could permit to give content to the somewhat "vague" idea of "excellence" (Moore et al., 2017) as well. 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 Glänzel (1996), still relevant today, more than 20 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 favor of standardization in "more effective use of information, better coordination of activities, simplification, and the advantages of large-scale production" (Brunsson & Jacobsson (2002b, p. 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) (Brunsson & Jacobsson, 2002b, pp. 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 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 (Table 5). 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 & Grant, 2016) and *sustainability* in a consistent way. It would be very interesting to investigate whether and how to extend, specify, and apply MacIntyre's philosophy to develop an *ethics* of evaluation<sup>⊛</sup> with our framework as a background. This is out of the scope of the present paper and is left to future research.

## 6    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 could be "quality-aware" (in the broad meaning discussed in the paper) i.e. fitness for use. We show that with our framework, composed of three dimensions (*theory*, *methodology*, and *data*) of three building blocks each (education, research, and innovation; efficiency, effectiveness, and impact; availability, interoperability, and unit-free property), three implementing factors (tailorability, transparency, and openness) and three enabling conditions (convergence, mixed-methods, and knowledge infrastructures), all joined together around the overarching idea of Quality, we are able to embrace many different and heterogeneous streams of literature.

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.

This 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 along the lines of Daraio (in press);
- Running additional research for providing a systematic review, 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 (Daraio, 2017b);
- Extending the proposed framework to the characterization of different governance systems (Capano, Howlett, & Ramesh, 2015) for analyzing their systemic connection with their performance;

---

<sup>⊛</sup> More reading can be found in Furner (2014).

**Perspective**

- Applying the framework, *mutatis mutandis*, to the evaluation of education (Daraio, 2015) and innovation (Borrás & Edquist, 2013) activities;
- Investigating the *ethics* of evaluation by exploring the connections between our framework and MacIntyre's oeuvre; and
- Corroborating the framework for the regulation of the evaluation of research.

Finally, our framework may pave the way for new *revolutionary* models of research assessment, which include *data* as a relevant conceptual dimension, and which are closer to the represented reality. Here "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). However, to do so much additional research is needed.

## Acknowledgements

## References

Abramovitz, M. (1956). Resource and output trends in the United States since 1870. The American Economic Review, 46(2), 5–23.

Acemoglu, D., Aghion, P., & Zilibotti, F. (2003). Vertical integration and distance to frontier. Journal of the European Economic Association, 1(2–3), 630–638.

Acemoglu, D., Aghion, P., & Zilibotti, F. (2006). Distance to frontier, selection, and economic growth. Journal of the European Economic Association, 4(1), 37–74.

Adams, J.D. (1990). Fundamental stocks of knowledge and productivity growth. Journal of Political Economy, 98(4), 673–702.

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.

Aghion, P. (2009). Growth and education. Commission on growth and development working paper, no. 56. Washington, DC: World Bank Publications.

Aghion, P., & Howitt, P. (2009). The economics of growth. Cambridge, MA: The MIT Press.

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.

Anderson C. (2008). The end of theory: The data deluge makes the scientific method obsolete. Wired Magazine. Retrieved on July 30, 2017, from https://www.wired.com/2008/06/pb-theory/.

Antonelli, C., & Link, A.N. (2014). Routledge handbook of the economics of knowledge. London: Routledge.

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.

Bǎdin, L., Daraio, C., & Simar, L. (2012). How to measure the impact of environmental factors in a nonparametric production model. European Journal of Operational Research, 223(3), 818–833.

Bǎdin, L., Daraio, C., & Simar, L. (2014). Explaining inefficiency in nonparametric production models: The state of the art. Annals of Operations Research, 214(1), 5–30.

Ballard, B.W. (2000). Understanding MacIntyre. Lanham, MD: University Press of America.

Bammer, G. (2016). What constitutes appropriate peer review for interdisciplinary research? Palgrave Communications, 2 (palcomms201617). Retrieved on July 30, 2017, from http://www.nature.com/articles/palcomms201617.

Barré, R. (2004). S&T indicators for policy making in a changing science–society relationship. In H. Moed, W. Glänzel, & U. Schmoch (Eds.), Handbook of Quantitative Science and Technology Research (pp. 115–132). Dordrecht: Springer Netherlands.

Batini, C., & Scannapieco, M. (2016). Data and information quality. Cham, Switzerland: Springer International Publishing.

Benessia, A., Funtowicz, S., Giampietro, M., Pereira, Â.G., Ravetz, J., Saltelli, A., . . . & van der Sluijs, J.P. (2016). Science on the verge. Tempe, AZ: Consortium for Science, Policy, & Outcomes at Arizona State University.

Blaug, M. (1966). Economics of education; a selected annotated bibliography (No. 370.193 B5). Retrieved on July 30, 2017, from http://www.sciencedirect.com/science/article/pii/B9780080206271500025.

Bogetoft, P., Fried, H.O., & Eeckaut, P.V. (2007). The university benchmarker: An interactive computer approach. In A. Bonaccorsi, & C. Daraio (Eds.), Universities and Strategic Knowledge Creation: Specialization and Performance in Europe (pp. 443–462). Cheltenham: Edward Elgar Publishing.

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, & U. Schmoch (Eds.), Handbook of Quantitative Science and Technology Research (pp. 51–74). Dordrecht: Springer Netherlands.

Borgman, C.L. (2015). Big data, little data, no data: Scholarship in the networked world. Cambridge, MA: The MIT Press.

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), 217–233.

Borrás, S., & Edquist, C. (2013). The choice of innovation policy instruments. Technological Forecasting and Social Change, 80(8), 1513–1522.

Bozeman, B., & Melkers, J. (Eds.) (1993). Evaluating R&D impacts: Methods and practice. New York: Springer.

Brunsson, N., & Jacobsson, B. (Eds.). (2002a). A world of standards. Oxford: Oxford University Press.

Brunsson, N., & Jacobsson, B. (2002b). The pros and cons of standardization—An epilogue. In A World of Standards (pp. 169–173). Oxford: Oxford University Press.

Capano, G., Howlett, M., & Ramesh, M. (Eds.). (2015). Varieties of governance. Hampshire, UK: Palgrave Macmillan.

Carayannis, E.G., & Campbell, D.F. (2009). 'Mode 3' and 'Quadruple Helix': Toward a 21$^{st}$ century fractal innovation ecosystem. International Journal of Technology Management, 46(3–4), 201–234.

Checchi, D. (2006). The economics of education: Human capital, family background and inequality. Cambridge: Cambridge University Press.

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.

Chesbrough, H. (2012). Open innovation: Where we've been and where we're going. Research-Technology Management, 55(4), 20–27.

Cohen, W.M., & Levinthal, D.A. (1990). Absorptive capacity: A new perspective on learning and innovation. Administrative Science Quarterly, 35(1), 128–152.

Creswell, J.W., & Clark, V.L.P. (Eds.). (2011). 2$^{nd}$ Ed. Designing and conducting mixed methods research. Thousand Oaks, CA: SAGE Publications.

Cronin, B. (2013). Thinking about data. Journal of the American Society for Information Science and Technology, 64(3), 435–436.

Cronin, B., & Sugimoto, C.R. (Eds.). (2014). Beyond bibliometrics: Harnessing multidimensional indicators of scholarly impact. Cambridge, MA: The MIT Press.

Cronin, B., & Sugimoto, C.R. (Eds.). (2015). Scholarly metrics under the microscope: From citation analysis to academic auditing. Medford, NJ: Information Today.

D'Ariano, G.M., & Perinotti, P. (2016). Quantum theory is an information theory. Foundations of Physics, 46(3), 269–281.

Dahler-Larsen, P. (2012). The evaluation society. California: Stanford University Press.

Daraio, C. (2015). Assessing the efficiency, effectiveness and impact of education in the age of big data: Challenges and a way forward. Keynote presentation at Leuven LEER Workshop 'Efficiency in Education and the Use of Big Data', November 19–20, 2015, Leuven (Belgium).

Daraio, C. (2017a). Assessing research and its impacts: The generalized implementation problem and a doubly-conditional performance evaluation model, paper presented at the ISSI 2017 Conference, October 2017, Wuhan (China).

Daraio, C. (2017b). A doubly conditional performance evaluation model, the democratization of evaluation and altmetrics, paper presented at the STI 2017 Conference, September 2017, Paris.

Daraio, C. (in press). Econometric approaches to the measurement of research productivity. In W. Glänzel, H.F. Moed, H. Schmoch, & M. Thelwall (Eds.), Springer Handbook of Science and Technology Indicators.

Daraio, C., & Simar, L. (2014). Directional distances and their robust versions: Computational and testing issues. European Journal of Operational Research, 237(1), 358–369.

Daraio, C., & Simar, L. (2007). Advanced robust and nonparametric methods in efficiency analysis. Methodology and applications. New York: Springer.

Daraio, C., Simar, L., & Wilson, P.W. (2017). Central limit theorems for conditional efficiency measures and tests of the "separability" condition in nonparametric two-stage models of production. The Econometrics Journal. Retrieved on July 30, 2017, from https://doi.org/10.1111/ectj.12103.

Daraio, C., & Glänzel, W. (2016). Grand challenges in data integration—state of the art and future perspectives: An introduction. Scientometrics, 108 (1), 391–400.

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.

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, 108 (1), 441–455.

Debackere, K. (2016). Let the data speak for themselves: Opportunities and caveats. Journal of Data and Information Science, 1(1), 3–5.

DeMillo, R.A., & Young, A.J. (2015). Revolution in higher education: How a small band of innovators will make college accessible and affordable. Cambridge, MA: The MIT Press.

Ding, Y., Rousseau, R., & Wolfram, D. (Eds). (2014). Measuring scholarly impact. Cham, Switzerland: Springer International Publishing.

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.

Drucker, P.F. (1967). The effective executive. New York: Harper and Row.

Ebrahim, A., & Rangan, V.K. (2014). What impact? California Management Review, 56(3), 118–141.

Edquist, C. (2001). The systems of innovation approach and innovation policy: An account of the state of the art. In Druid Nelson and Winter Conference 2001 (pp. 12–15). Denmark: Aalborg University.

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. Retrieved on July 30, 2017, from http://deepblue.lib.umich.edu/handle/2027.42/97552.

Egghe, L., & Rousseau, R. (1990). Introduction to informetrics: Quantitative methods in library, documentation and information science. Amsterdam: Elsevier.

Egghe, L. (2005). Power laws in the information production process: Lotkaian informetrics. Amsterdam: Elsevier.

Ekbia, 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, 66(8), 1523–1545

Etzkowitz, H., & Leydesdorff, L. (2000). The dynamics of innovation: From national systems and mode 2 to a triple helix of university-industry-government relations. Research Policy, 29(2), 109–123.

European Commission (2014). Expert Group to support the development of tailor-made impact assessment methodologies for ERA (European Research Area), Brussels, Belgium.

Fagerberg, J., Martin, B.R., & Andersen, E.S. (Eds.). (2013). Innovation studies: Evolution and future challenges. Oxford: Oxford University Press.

Fealing, K.H., Lane, J.I., Marburger, J.H. JIII, & Shipp, S.S. (Eds.) (2011). The science of science policy, a handbook. Stanford: Stanford University Press.

Floridi, L. (Ed.). (2008). The Blackwell guide to the philosophy of computing and information. Hoboken, NJ: John Wiley & Sons.

Floridi, L. (Ed.). (2010). The Cambridge handbook of information and computer ethics. Cambridge: Cambridge University Press.

Floridi, L. (2012). The road to the philosophy of information. In H. Demir (Eds.), Luciano Floridi's Philosophy of Technology (pp. 245–271). Dordrecht: Springer Netherlands.

**Perspective**

Floridi, L. (2014). The fourth revolution: How the infosphere is reshaping human reality. Oxford: Oxford University Press.

Floridi, L., & Illari, P. (Eds.). (2014). The philosophy of information quality (Vol. 358). Cham, Switzerland: Springer International Publishing.

Frické, M. (2015). Big data and its epistemology. Journal of the Association for Information Science and Technology, 66 (4), 651–661.

Frischmann, B.M. (2012). Infrastructure: The social value of shared resources. Oxford: Oxford University Press.

Funtowicz, S.O., & Ravetz, J.R. (1990). Science for policy: Uncertainty and quality. In Uncertainty and Quality in Science for Policy (pp. 7–16). Dordrecht: Springer Netherlands.

Furner, J. (2014). The ethics of evaluative bibliometrics. In B. Cronin, & C. Sugimoto (Eds.), Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact (pp. 85–107). Cambridge, MA: MIT Press.

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.

Georgescu-Roegen, N. (1971). The entropy law and the economic process. Cambridge, MA: Harvard University Press.

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. London: Sage.

Gingras, Y. (2016). Bibliometrics and research evaluation: Uses and abuses. Cambridge, MA: The MIT Press.

Glänzel, W. (1996). The need for standards in bibliometric research and technology. Scientometrics, 35(2), 167–176.

Glänzel, W., & Schoepflin, U. (1994). Little scientometrics, big scientometrics… and beyond? Scientometrics, 30(2–3), 375–384.

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.

Glänzel, W., & Moed, H.F. (2013). Opinion paper: Thoughts and facts on bibliometric indicators. Scientometrics, 96(1), 381–394.

Glänzel, W., Moed, H.F., Schmoch, H., & Thelwall, M. (Eds). (in press). Springer Handbook of Science and Technology Indicators.

Godin, B. (2002). Outline for a history of science measurement. Science, Technology, & Human Values, 27(1), 3–27.

Godin, B. (2004). Measurement and statistics on science and technology: 1920 to the present. London: Routledge.

Griliches, Z. (1986). Economic data issues. In Z. Griliches, & M.D. Intriligator (Eds.), Handbook of Econometrics Volume III (pp. 1465–1514). Amsterdam: Elsevier.

Griliches, Z. (1994). Productivity, R&D, and the data constraint. American Economic Review, 84(1), 1–23.

Griliches, Z. (1998). R&D and productivity: The econometric evidence. Chicago: University of Chicago Press.

Hall, B.H., & Rosenberg, N. (Eds.). (2010). Handbook of the economics of innovation. Amsterdam: Elsevier.

Hanson, B., Sugden, A., & Alberts, B. (2011). Making data maximally available. Science, 331(6018), 649–649.

Hanushek, E.A., & Woessmann, L. (2007). The role of education quality for economic growth. World Bank Policy Research Working Paper, No. 4122. Retrieved on July 30, 2017, from https://ssrn.com/abstract=960379.

Hanushek, E.A., Machin, S.J., & Woessmann, L. (Eds.). (2016). Handbook of the economics of education. Amsterdam: Elsevier.

Hanushek, E.A., Woessmann, L., Jamison, E.A., & Jamison, D.T. (2008). Education and economic growth. Education Next, 8(2). Retrieved on July 31, 2017, from http://www.freepatentsonline.com/article/Education-Next/177556508.html.

Helbing, D., & Carbone, A.F. (Eds.). (2012). Participatory science and computing for our complex world. The European Physical Journal. Special Topics, Vol. 214. Retrieved on July 31, 2017, from https://epjst.epj.org/index.php?option=com_toc&url=/articles/epjst/abs/2012/14/contents/contents.html.

Hemlin, S. (1996). Research on research evaluation. Social Epistemology, 10(2), 209–250.

Henderson, R., Jaffe, A.B., & Trajtenberg, M. (1998). Universities as a source of commercial technology: A detailed analysis of university patenting, 196588. Review of Economics and Statistics, 80(1), 119–127.

Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. Nature, 520, 429–431.

Hill, S. (2016). Assessing (for) impact: Future assessment of the societal impact of research. Palgrave Communications. Retrieved on July 31, 2017, from http://www.nature.com/articles/palcomms201673.

Hinrichs-Krapels, S., & Grant, J. (2016). Exploring the effectiveness, efficiency and equity (3es) of research and research impact assessment. Palgrave Communications. Retrieved on July 31, 2017, from https://www.nature.com/articles/palcomms201690#t2.

Horstemeyer, M.F. (2009). Multiscale modeling: A review. In J. Leszczynski, & M. Shukla (Eds.), Practical Aspects of Computational Chemistry (pp. 87–135). Dordrecht: Springer Netherlands.

ISCED. (2011), International Standard Classification of Education, UNESCO Montreal, Canada. Retrieved on December 20, 2016, from http://www.uis.unesco.org/Education/Documents/isced-2011-en.pdf.

Johnes, G., & Johnes, J. (Eds.). (2004). International handbook on the economics of education. Cheltenham: Edward Elgar.

Khandker, S.R., Koolwal, G.B., & Samad, H.A. (2010). Handbook on impact evaluation: Quantitative methods and practices. Washington, DC: World Bank Publications.

Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. Big Data & Society, 1(1), 1–12.

Kuhn, T.S. (1962). The structure of scientific revolutions. Chicago: University of Chicago Press.

Kuhn, T.S. (1969). The structure of scientific revolutions. Retrieved on July 31, 2017, from http://projektintegracija.pravo.hr/_download/repository/Kuhn_Structure_of_Scientific_Revolutions.pdf/.

Largent, M.A., & Lane, J.I. (2012). STAR METRICS and the science of science policy. Review of Policy Research, 29(3), 431–438.

Lenzerini, M. (2011). Ontology-based data management. CIKM, Proceedings of the 20th ACM International Conference on Information and Knowledge Management (pp. 5–6). New York: ACM.

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.

Lutz, C.S. (2017). Alasdair Chalmers MacIntyre. In The Internet Encyclopedia of Philosophy. Retrieved on January 7, 2017, from http://www.iep.utm.edu/mac-over/.

MacIntyre, A. (2006). The end of education: The fragmentation of the American university. Commonweal, 133(18), 10–14.

MacIntyre, A. (2007). After virtue: A study in moral theory. 3rd ed. Notre Dame, Indiana: University of Notre Dame Press.

MacIntyre, A. (1988). Whose justice? Which rationality? Notre Dame, Indiana: University of Notre Dame Press.

Mansfield, E. (1991). Academic research and industrial innovation. Research Policy, 20(1), 1–12.

Mansfield, E. (1995). Academic research underlying industrial innovations: Sources, characteristics, and financing. The Review of Economics and Statistics, 77(1), 55–65.

Martin, B.R. (2016). Introduction to discussion paper on the sciences are different and the differences are important. Research Policy, 9(45), 1691.

Mingers, J. (2006). Realising systems thinking: Knowledge and action in management science. Boston, MA: Springer.

Mirowski, P., & Sent, E.M. (2002). Science bought and sold: Essays in the economics of science. Chicago: University of Chicago Press.

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.

Moed, H.F., & van Leeuwen, T.N. (1996). Impact factors can mislead. Nature, 381 (6579), 186.

Moed, H.F., Glänzel, W., & Schmoch, U. (2004). Handbook of quantitative science and technology research. Dordrecht: Springer Netherlands.

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. Retrieved on July 31, 2017, from https://www.nature.com/articles/palcomms2016105.

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). Berlin: Springer.

Morris, W.T. (1967). On the art of modeling. Management Science, 13(12), B707 – B717.

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.

Myung, I.J. (2000). The importance of complexity in model selection. Journal of mathematical psychology, 44(1), 190–204.

National Research Council. (2014). Science of science and innovation policy: Principal investigators' conference summary. Washington, DC: The National Academies Press.

Nelson, R.R., & Phelps, E.S. (1966). Investment in humans, technological diffusion, and economic growth. The American Economic Review, 56(1/2), 69–75.

Nielsen, M. (2012). Reinventing discovery: The new era of networked science. Princeton: Princeton University Press.

Nielsen, M.A., & Chuang, I.L. (2010). Quantum computation and quantum information. 10th Anniversary Edition. Cambridge: Cambridge University Press.

Nowotny, H., Scott, P., & Gibbons, M. (2001). Re-thinking science: Knowledge and the public in an age of uncertainty. Cambridge: Polity.

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.

OECD. (2002). Frascati Manual: Proposed standard practice for surveys on research and experimental development. Retrieved on July 31, 2017, from http://www.tubitak.gov.tr/tubitak_content_files/BTYPD/kilavuzlar/Frascati.pdf.

OECD. (2005). Oslo Manual: Guidelines for collecting and interpreting innovation data. 3rd edition. Paris: OECD Publishing.

OECD. (2011). Quality framework and guidelines for OECD statistical activities. Paris: OECD Publishing.

OECD (2015a). Making open science a reality. OECD Science, Technology and Industry Policy Papers No. 25. Paris: OECD Publishing.

OECD. (2015b). Frascati Manual 2015: Guidelines for collecting and reporting data on research and experimental development. Retrieved on July 30, 2017, from http://www.oecd.org/science/frascati-manual-2015-9789264239012-en.htm.

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.

Parent, C., & Spaccapietra, S. (2000) Database integration: The key to data interoperability. In M.P. Papazoglou, & Z. Zari (Eds.), Advances in Object-Oriented Data Modeling (pp. 221–253). Cambridge, MA: The MIT Press.

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.

Perkmann, M., Tartari, V., McKelvey, M., Autio, E., Brostrom, A., D'Este, P., . . . & Krabel, S. (2013). Academic engagement and commercialisation: A review of the literature on university-industry relations. Research Policy, 42(2), 423–442.

Pollock, S.M. (1976). Mathematical modeling: Applying the principles of the art studio. Engineering Education, 67(2), 167–171.

Romer, P.M. (1986). Increasing returns and long-run growth. The Journal of Political Economy, 94(5), 1002–1037.

Romer, P.M. (1994). The origins of endogenous growth. The Journal of Economic Perspectives, 8(1), 3–22.

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.

Rosenberg, N., & Nelson, R.R., (1994). American universities and technical advance in industry. Research Policy, 23(3), 323–348.

Saltelli, A., & Funtowicz, S. (2014). When all models are wrong. Issues in Science and Technology, 30(2), 79–85.

Saltelli, A., & Funtowicz, S. (2015). Evidence-based policy at the end of the Cartesian dream: The case of mathematical modelling. In G. Pereira, & S. Funtowicz (Eds.), Science, Philosophy and Sustainability: The End of the Cartesian Dream. Beyond the Techno–Scientific Worldview. Routledge's Series: Explorations in Sustainability and Governance (pp. 147–162). London: Routledge.

**Perspective**

Saltelli, A., Guimaraes Pereira, A., 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.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., . . . & Tarantola, S. (2008). Global sensitivity analysis: The primer. Chichester, UK: John Wiley & Sons.

Salter, A.J., & Martin, B.R. (2001). The economic benefits of publicly funded basic research: A critical review. Research Policy, 30(3), 509–532.

Scharnhorst, A., Borner, K., van den Besselaar, P. (Eds.). (2012). Models of science dynamics: Encounters between complexity theory and information sciences. Berlin: Springer.

Simar, L., & Wilson, P.W. (2015). Statistical approaches for nonparametric frontier models: A guided tour. International Statistical Review, 83(1), 77–110.

Solow, R.M. (1957). Technical change and the aggregate production function. The Review of Economics and Statistics, 39(3): 554–562.

Stephan, P.E. (2012). How economics shapes science. Cambridge, MA: Harvard University Press.

Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. Research Policy, 42(9), 1568–1580.

Teixeira, P.N., & Dill, D.D. (Eds.). (2011). Public vices, private virtues? Assessing the effects of marketization in higher education (Vol. 2). Rotterdam: Sense Publishers.

Teixeira, P., Jongbloed, B., Dill, D., & Amaral, A. (2004). Markets in higher education: Rhetoric or reality? Dordrecht: Kluwer.

Vandenbussche, J., Aghion, P., & Meghir, C. (2006). Growth, distance to frontier and composition of human capital. Journal of Economic Growth, 11(2), 97–127.

Veugelers, R., & Del Rey, E. (2014). The contribution of universities to innovation, (regional) growth and employment. EENEE Analytical Report. Munich, Germany: EENEE. Retrieved on July 31, 2017, from http://www.voced.edu.au/node/82516.

Vinkler, P. (2010). The evaluation of research by scientometric indicators. Cambridge: Chandos Publishing.

von Hippel, E. (2005). Democratizing innovation. Cambridge, MA: The MIT Press.

von Hippel, E. (2016). Free innovation. Cambridge, MA: The MIT Press.

Weinan, E. (2011). Principles of multiscale modeling. Cambridge: Cambridge University Press.

West, J., Salter, A., Vanhaverbeke, W., & Chesbrough, H. (2014). Open innovation: The next decade. Research Policy, 43(5), 805–811.

Whitley, R., & Gläser, J. (Eds.). (2007). The changing governance of the sciences: The advent of research evaluation systems. Dordrecht: Springer Netherlands.

Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Jones, R., . . . & Johnson, B. (2015). The metric tide: Report of the independent review of the role of metrics in research assessment and management. Retrieved on July 31, 2017, from http://dera.ioe.ac.uk/23424/.

Willemain, T.R. (1994). Insights on modeling from a dozen experts. Operations Research, 42(2), 213–222.

Zucchini, W. (2000). An introduction to model selection. Journal of Mathematical Psychology, 44(1), 41–61.