ARTICLE

The future of research assessment in the humanities: bottom-up assessment procedures

Michael Ochsner[1,2], Sven Hug[1,3] and Ioana Galleron[4]

ABSTRACT Research assessment in the social sciences and humanities (SSH) is delicate. Assessment procedures meet strong criticisms from SSH scholars and bibliometric research shows that the methods that are usually applied are ill-adapted to SSH research. While until recently research on assessment in the SSH disciplines focused on the deficiencies of the current assessment methods, we present some European initiatives that take a bottom-up approach. They focus on research practices in SSH and reflect on how to assess SSH research with its own approaches instead of applying and adjusting the methods developed for and in the natural and life sciences. This is an important development because we can learn from previous evaluation exercises that whenever scholars felt that assessment procedures were imposed in a top-down manner without proper adjustments to SSH research, it resulted in boycotts or resistance. Applying adequate evaluation methods not only helps foster a better valorization of SSH research within the research community, among policymakers and colleagues from the natural sciences, but it will also help society to better understand SSH's contributions to solving major societal challenges. Therefore, taking the time to encourage bottom-up evaluation initiatives should result in being able to better confront the main challenges facing modern society. This article is published as part of a collection on the future of research assessment.

[1] ETH Zürich, Zürich, Switzerland [2] FORS, Lausanne, Switzerland [3] University of Zürich, Zürich, Switzerland [4] University Grenobles Alpes, Grenoble, France
Correspondence: (e-mail: ochsner@gess.ethz.ch)

Introduction

While there is more than a 100 years of scientific inquiry on research and dissemination practices in the natural and life sciences, until recently bibliometric and social studies on science and technology research neglected the SSH (Hemlin, 1996). Therefore, there are methods for research assessment in the natural and life sciences that relate to the practices in these fields and are accepted by the community (even though there are more and more critical voices, see for example, Lawrence, 2002; Molinié and Bodenhausen, 2010) and the measurement properties are tested by bibliometric research. In the meantime, knowledge on research and dissemination practices in the SSH is scarce, while research assessment did not stop at the gate of the SSH disciplines (Guillory, 2005; Burrows, 2012). The growing pressure of accountability, prevailing government practices based on New Public Management and the availability of quantitative data led to the implementation of (quantitative) research assessments also in the SSH during the last decades (Kekäle, 2002; Hammarfelt and de Rijcke, 2015; Hamann, 2016). The creation of the European Research Area (ERA) increased the importance of research evaluation: the initial communication "Towards a European Research Area" listed under the first theme of action the "mapping of European centres of excellence" and "Financing plan for centres of excellence on the basis of competition" (Commission of the European Communities, 2000); 15 years later, the ERA Roadmap listed the following as the first among the Roadmap's priorities: "Strengthening the evaluation of research and innovation policies and seeking complementarities between, and rationalization of, instruments at EU and national levels". (European Research Area and Innovation Committee, 2015: 5). The vast majority of research assessments, however, were implemented in a top-down manner by either governments or university administrators. In addition, research assessment procedures usually apply bibliometric and scientometric methods developed for the natural and life sciences that do not reflect SSH research and disseminations practices. Bibliometric research shows that these methods cannot readily be used for the SSH (Hicks, 2004; Lariviere et al., 2006; Nederhof, 2006). Therefore, research assessment procedures (and oftentimes research evaluation in general) meet strong opposition in the scholarly communities of the SSH.

In the last decade, a number of projects were initiated in Europe to explore research assessment procedures that adequately reflect SSH research practices. These projects did not arise from within the discipline in the sense of auto-regulation or the discontent with the quality or the standing of the discipline. Rather, they are the reaction on how research is assessed through procedures not linked to the functioning of the disciplines itself but to top-down decisions on how research is to be evaluated. Also, with the ERA Roadmap in place, the discussion could no longer be whether research should be subject to systematic research assessments but rather how to assess it. With a few exceptions, however, the bottom-up initiatives unfortunately do not get the attention of research evaluators and policymakers they deserve.

In this article, we give an overview of selected European initiatives that are genuinely reflecting the SSH research practices and were initiated or developed by scholars with an SSH background. Due to restrictions of space, we do not report how SSH research is assessed in unitary evaluation procedures, that is exercises that apply the same basic procedure for all disciplines (for sciences, technology, engineering and mathematics (STEM), as well as for SSH disciplines) and allow only for small adaptions to SSH research practices (for example, use of bibliometrics or not, types of eligible outputs). For this reason, we do not report how SSH research is evaluated in the RAE and REF in the United Kingdom[1] or the RQF and ERA in Australia as they are clearly top-down (see for example, Kwok, 2013), follow a unitary approach and the SSH are not having a major impact on the design of the exercise. Furthermore, the RAE/REF and RQF/ERA procedures are well-documented in the literature. For the SSH in the RAE/REF, see for example Arts and Humanities Research Council (2006, 2009); Butler and McAllister (2009); Hamann (2016); Johnston (2008); Norris and Oppenheim (2003); Oppenheim and Summers (2008). For the RAE/REF in general, see for example Barker (2007) and Hicks (2012). For SSH related matters in the Australian RQF/ERA, see for example Butler (2008), Butler and Visser (2006), Council for the Humanities, Arts and Social Sciences (2009), Genoni and Haddow (2009), Kwok (2013), Redden (2008). Because there is a wealth of such SSH initiatives in Europe, we also restrict our review to European initiatives and do not report other initiatives such as the Australian ERA and the Humanities Indicators project in the United States (www.humanitiesindicators.org).

In what follows, we first present the issues of research assessment in the SSH, such as the methodological issues and the SSH scholars' critique of the assessment procedures. We then move on to present several bottom-up initiatives taken up in (mainly continental) Europe by concerned SSH scholars. These initiatives set out at different levels and with different scope, from simply improving the situation of SSH data availability and accuracy to complex evaluation procedures involving a broad range of quality criteria and indicators. Some initiatives take place at a local level, others at a national level; and there are even European initiatives concerned with bottom-up research evaluation in the SSH. We conclude with some recommendations for future research evaluation in the humanities.

Research assessment in the SSH

To describe the current situation of research assessment in the SSH, we analyse them from two perspectives. First, we take the perspective of bibliometricians and scientometricians and focus on what they say regarding the adequacy of their methods for SSH research. Second, we analyse the critiques of the SSH scholars regarding those methods, which gives us hints at how to design adequate methods for research assessment in the SSH.

Bibliometrics and scientometrics in SSH research assessments

The application of bibliometric methods to the SSH proved to be problematic and yielded unsatisfying results, so that even bibliometricians caution from applying bibliometric methods to SSH disciplines (see for example, Nederhof et al., 1989; Glänzel, 1996; Lariviere et al., 2006). This is because of several reasons, that we summarize in two main issues: coverage issues and methodological issues.

Coverage issues arise for several reasons. First, in the SSH, chapters in books and monographs are more frequently used as publication channels and get cited more often than journal articles (Hicks, 2004; Nederhof, 2006). This leads to severe coverage issues in the most important databases for bibliometric analyses, which are mainly or exclusively based on scholarly journals (van Leeuwen, 2013). Furthermore, even internationally oriented European journals are not covered well in the relevant databases compared with American journals (Nederhof, 2006).

Second, some SSH disciplines are characterized by a more pronounced national and regional orientation (Nederhof, 2006). Nederhof states in his review of bibliometric monitoring in the SSH: "Societies differ, and therefore results from humanities or social science studies obtained in one country may not always be very useful to researchers in other countries" (Nederhof, 2006:

83). Thus, even though the topics might be internationally relevant, this kind of output is less visible, as often written in the national languages, seldom covered in the bibliometric databases (see for example, Chi, 2012), or even published in other publication channels that are not covered at all (example, reports and other publications directed to national or regional readership).

Third, SSH scholars write not only for the scholarly readers but also for the lay public (Hicks, 2004). This type of literature is usually not taken into consideration in evaluations and certainly not included in the databases used for bibliometric analyses. However, non-scholarly publications are an important part of SSH research and its societal impact.

Methodological issues arise amongst others from the fact that citation behaviour is different in the SSH disciplines. The age of references is remarkably high. Glänzel noted for example in his analysis from 1996 that a 3-year citation window is too short. Given the distribution of the citations over time, almost a 10-year citation window would have to be applied, leading to an obsolete publication set for evaluation purposes (Glänzel, 1996). Furthermore, the citation culture is different (Hellqvist, 2010; Hammarfelt, 2012; Bunia, 2016). Hicks (2004) notes also that SSH journals are usually more transdisciplinary, which leads to methodological problems such as field normalization.

While this is not a comprehensive analysis of methodological issues of quantitative assessments, it shows that there are several problems with the application of bibliometric indicators in research assessments in the humanities. Importantly, it makes evident that today's bibliometric methods do not reflect SSH scholarship.

SSH scholars' critique of quantitative research assessments

If research assessment procedures are to be accepted and the tools and methods should help determining the quantity and quality of humanities research without significant delays, refusal or boycott by the scholarly community, the criticisms put forward by humanities scholars become an important issue. We have analysed SSH scholars' critique of (quantitative) research assessments elsewhere and summarized them into four main reservations (Hug and Ochsner, 2014). We will only briefly summarize our findings, as relevant for the purpose of this article.

The first reservation relates to the section above: the methods were developed for, and reflect the research practices in, the natural and life sciences (Vec, 2009). This means not only that the assessment practices do not account for SSH dissemination practices (monographs, diverse languages, local orientation, individual scholarship) as noted in the section above, but also that the assessment practices follow the natural sciences' linear understanding of progress while the SSH scholars share the notion of the "coexistence of competing ideas" (Lack, 2008: 14), that is, an ever-increasing knowledge base. This conception of knowledge that is diverse and not dying out is not reflected in most evaluation practices.

Second, SSH scholars have strong reservations about quantification. A joint letter by 24 international philosophers to the Australian government as a reaction to the journal ranking in the Excellence in Research for Australia (ERA) exercise points to this issue: "The problem is not that judgments of quality in research cannot currently be made, but rather that in disciplines like Philosophy, those standards cannot be given simple, mechanical, or quantitative expression" (Academics Australia, 2008). Other scholars argue that research does not produce products or goods in a free market, in which value can be defined according to the products' economic value or efficiency (Plumpe, 2010; Palumbo and Pennisi, 2015). Thus, many SSH scholars fear that the

intrinsic benefits of the arts and humanities will be neglected or even lost because of the focus on quantitative measures. The report for the Humanities and Social Sciences Federation of Canada says for example that "some efforts soar and others sink, but it is not the measurable success that matters, rather the effort" (Fisher *et al.*, 2000, "The Value of a Liberal Education", para. 18; see also the report for the RAND corporation McCarthy *et al.*, 2004).

The third reservation is the fear of negative steering effects of indicators. SSH scholars anticipate many dysfunctional effects such as mainstreaming or conservative effects of indicators, a loss of diversity of research topics or disciplines due to selection effects introduced by the use of indicators, or importance of spectacular research findings leading to unethical reporting of findings (Fisher *et al.*, 2000; Andersen *et al.*, 2009; Hose, 2009; Burrows, 2012). More and more such negative steering effects of indicators are observed also in the natural sciences (Butler, 2003, 2007; Mojon-Azzi *et al.*, 2003; Moonesinghe *et al.*, 2007; Unreliable research. Trouble at the lab, 2013). Such findings support the fear of negative steering effects in the SSH.

Fourth, the SSH are characterized by a heterogeneity of research topics, methods and paradigms. Finding shared quality criteria or standards for research assessments becomes an intricate task if there is no consensus on research questions, the suitability of the methods applied and even the definition of disciplines and sub-disciplines (Herbert and Kaube, 2008; van Gestel *et al.*, 2012; Hornung *et al.*, 2016). If criteria can be found, they are usually informal, refer to one (sub-)discipline and cannot easily be transferred to other sub-disciplines or evaluation situations (Herbert and Kaube, 2008).

Bottom-up procedures for research assessment in the humanities

Despite these critiques of both bibliometricians and scientometricians on the one hand and SSH scholars on the other hand, more and more research assessments in the SSH are implemented. Usually, the procedures for research assessments are implemented in a top-down manner, not taking the situation at the coal face of research into account. However, there are several initiatives that reflect the characteristics of SSH research. In the following, we focus on initiatives that come from within the SSH research communities or are at least developed by scholars from SSH disciplines, genuinely taking into account SSH research practices in their approaches[2]. All of them address at least one of the issues mentioned in the previous section. While these bottom-up initiatives are more likely to be accepted by SSH scholars, some of them still face strong opposition or are boycotted.

Improving the databases

Considering that typical SSH publications (for example, books, proceedings, publications in local languages) are badly represented in current databases[3], efforts have been made in several countries to improve coverage, especially in the countries with a performance-based funding model, like Spain, Norway, Denmark, Belgium (Flanders) and Finland (Giménez-Toledo *et al.*, 2016). There was also an attempt to create a full-coverage bibliographic/bibliometric database for Europe, but it did not result in an implementation of a European-wide database or standard (Martin *et al.*, 2010). In parallel, the ERIH project intended to create a European journal list for the SSH to overcome the problems of under-representation of (European) SSH journals in the main bibliometric databases; however, the project faced strong opposition (Andersen *et al.*, 2009), had to be remodelled (see Lauer, 2016) and was relaunched under the name ERIH Plus[4].

Attempts to create publication databases suitable for the humanities have sometimes also been organized at the level of disciplines. The EERQI project included such a database for the educational sciences on the European level; it also investigated methods for using the data in research evaluations in a meaningful way (Gogolin et al., 2014; Gogolin, 2016). The database allows scholars to search for publications using keywords in one language, while retrieving results in all four languages covered in the database. Therefore, beyond evaluative purposes, centralized and systematic coverage of SSH production appears as an endeavour with multiple potential benefits, such as improving information retrieval for scholars and widening access to publications in multiple languages.

In all cases, consciousness is raising about the need to compile complete and interoperable databases of SSH scholarly and non-scholarly outputs, so as to gain accurate knowledge about productivity and publication behaviour in these very diverse disciplines. At the same time, the creation of such databases should go hand in hand with the development of standards regarding their use, including standards on how not to use them.

An SSH approach towards bibliometrics and scientometrics

Bibliometric analyses face many problems when applied to SSH disciplines (Nederhof et al., 1989; Archambault et al., 2006; Nederhof, 2006; van Leeuwen, 2013). However, Hammarfelt (2016: 115) observes a shift from investigating coverage issues towards studying the characteristics of SSH publication practices and developing bibliometric approaches sensitive to the organization of SSH research fields. This includes, but is not limited to, extending bibliometric analyses to non-source items (Butler and Visser, 2006; Chi, 2014) or the relatively new Book Citation Index (Gorraiz et al., 2013), using other databases like Google Scholar (Kousha and Thelwall, 2009) or data from social media services, the so-called altmetrics (Holmberg and Thelwall, 2014; Mohammadi and Thelwall, 2014; Zuccala et al., 2015; Zuccala and Cornacchia, 2016), analysing the inclusion in library catalogues (White et al., 2009), exploring national databases with full coverage (Giménez-Toledo et al., 2016), extending data to references in research grant proposals (Hammarfelt, 2013) or to book reviews (Zuccala and van Leeuwen, 2011; Zuccala et al., 2015), exploring collaboration (Ossenblok and Engels, 2015) and publication patterns (Chi, 2012; Ossenblok et al., 2012; Verleysen and Weeren, 2016). From a more pragmatic point of view, attempts are made to "weigh" the various outputs, such as journals or books in the SSH, similar to the journal impact factor, commonly used in the sciences (Giménez-Toledo, 2016).

While most of this research is done by bibliometricians and scientometricians, there are more and more SSH scholars still focusing on their SSH career and at the same time investigating research practices in their disciplines, such as citation practices (Drabek et al., 2015; Bunia, 2016), the influence of databases (Lauer, 2016), the relation of bibliometric indicators to research practices (Gogolin, 2016) or career building and dissemination (Williams and Galleron, 2016). Also, more methodological analyses are conducted by SSH scholars, such as the investigation of the inter-rater reliability of research assessment procedures (Riordan et al., 2011; Plag, 2016) or the correlation of bibliometric and expert-based procedures (Ferrara and Bonaccorsi, 2016). While Hammarfelt requests to build a "bibliometrics for the humanities" (Hammarfelt, 2016: 115), Zuccala (2016: 149) goes further and demands that bibliometricians find ways to teach bibliometrics to humanities students so that a "new breed of humanistic bibliometrician can emerge successfully".

Bunia (2016), a German literature scholar, argues that the problem of applicability of citation analyses might, besides coverage and technical issues, as well be intrinsic to the field of literary studies: literature scholars seem not to read the work of their colleagues in the same field or at least they do not use or cite them in their own publications. He advocates using bibliometric analyses to study the citation behaviour of literary scholars since this is also important knowledge for the scholarly community in the field. The use of bibliometric methods in research assessment will not be possible until light is shed on this issue.

Summarising the situation of bibliometrics and scientometrics in the SSH, bibliometric methods cannot be readily used for research assessment in the SSH. But bibliometrics adapted to the SSH can help to study research practices, publication and citation practices as well as other practices important for knowledge production in the SSH. A thorough look at citation habits can also broach some delicate issues in research practice. Applied with some care, some quantitative indicators can also be used to complement peer review if they are defined bottom-up, that is, from within the disciplines.

Funding SSH research grants

Third-party funding becomes more and more important because, first, a higher share of the research budget in most countries is competitively distributed through funding organizations (van den Akker, 2016), second, because the amount of third-party funding is used in most assessment procedures at least as an information criterion (Ochsner et al., 2012). Especially for the careers of young scholars, grant allocation gains importance: on the one hand, job opportunities of young researchers are more and more characterized by short-term contracts based on external funding (van Arensbergen et al., 2014b); on the other hand, allocated grants serve as a prove of excellence in talent selection decisions (van Arensbergen et al., 2014a).

Third-party funding implies ex-ante research assessment, that is, research is assessed before it has been conducted. While most ex ante assessments are based on peer review, many of them use bibliometric data to inform the peers. Certainly, these processes have been already in place for some time, mainly unnoticed by most SSH scholars because research grants are less important for them as they do not need expensive infrastructure to do their research (Krull and Tepperwien, 2016). The growing importance of grants in science policy at the national and international level, however, has drawn the attention of SSH scholars to the processes of distributing research grants because there are huge differences in the distribution of grants between the STEM and SSH disciplines (Krull and Tepperwien, 2016), not to mention the differences of amounts.

The lower chances and the lower amount of acquired third-party funding have their roots in the epistemic differences of research practices between the STEM and SSH, as well as in a different disciplinary organization and divergent practices of research evaluation. Only a minority of SSH scholars needs expensive instruments to conduct experiments, as opposed to the basic needs SSH scholars usually express, which are a computer, access to archives, travel expenses and research time (Krull and Tepperwien, 2016). Therefore, third-party funding did not play a role for a long time in most SSH disciplines and grants are usually of a comparably low amount.

Second, the way SSH scholars appreciate research output of colleagues is quite different from how STEM researchers do. SSH scholars are much more critical. They criticize even work they value as excellent. A bit-by-bit examination is considered a proof of love. In interdisciplinary panels, STEM researchers do not agree on funding research that is heavily criticized. Because SSH scholars always do criticize the work of their colleagues, irrespective of the quality of the research, SSH scholars are often

discriminated in interdisciplinary granting schemes (Krull and Tepperwien, 2016) even though this practice of criticizing works fine within SSH disciplines (König, 2016; Krull and Tepperwien, 2016).

Third, in the STEM disciplines, paradigmatic issues are usually disputed internally while at the outside there is coherence. The SSH disciplines, however, do not resolve such issues but allow for diversity within their fields (van den Akker, 2016). Of course, this is rooted in a different understanding of scholarly work—linear progress in the STEM disciplines versus increase of the knowledge base in the SSH disciplines (Lack, 2008)—but it is also the result of a lack of organization. This leads to further marginalization as the SSH disciplines do not stand together to criticize univocally the short-sighted focus on the linear progress of science (van den Akker, 2016) and to demand funding schemes adequate for SSH research with a powerful united voice.

At the same time, some funders are frustrated that their schemes do not attract more proposals from SSH disciplines (König, 2016), maybe because SSH scholars do not take the risk of writing a proposal when past experiences seem to make it likely that it will be turned down. Therefore, the Fritz Thyssen Stiftung and the VolkswagenStiftung have created a funding programme adapted to the needs of humanities scholars entitled "Focus on Humanities" that includes the grant *Opus Magnum* that could bridge the gap between the humanist way of doing research and at the same time adding a competitive component. In addition, the VolkswagenStiftung (2014) has established bottom-up guidelines regarding how to recognize intellectual quality in the humanities collected in a workshop with renowned scholars and young scholars.

SSH research practices and criteria for research quality

To assess research performance, there should be an explicit understanding of what "good" research is, since any assessment points out "high quality" research or tries to judge which research is "better" (Butler, 2007). However, not much is known what actually research quality means (see e.g. Kekäle, 2002), especially so in the SSH. The literature on research assessment actively avoids this topic, while existing tools and procedures of research assessment do not include an explicit understanding of research quality (Glänzel et al., 2016). Rather, authors revert to "impact", which is easier to measure but not congruent with "quality" (Gumpenberger et al., 2016)[5]. Therefore, if SSH research is to be assessed appropriately, there must be knowledge on what actually research quality means in these disciplines and assessment procedures must relate to the conceptions of research quality of the assessed scholars. To get a grasp on what guides judgement on what is good or bad research, we need empirical knowledge on research practices and the notions of quality that humanities scholars use to interpret, structure and evaluate the events and entities in their research activities.

During the last hundred years, scholars analysed research practices of the STEM disciplines, especially the natural sciences, in detail; however, the newly emerging field of social studies of science neglected its own (SSH) disciplines until recently (Hemlin, 1996: 53; Hammarfeldt, 2012: 164). The literature so far describes the characteristics of SSH research in the following way: a) SSH research is interpretative, that is, humanities research is mainly text- and theory-driven and social sciences are more concept-driven, while the natural sciences set up their studies to answer specific questions and are progress-driven (MacDonald, 1994; Guetzkow et al., 2004; Lamont, 2009); b) it is reflective and introduces new perspectives in academia, by fostering discursive controversy and competing visions (Fisher et al., 2000; Hellqvist, 2010). With regard to the society, they bring a decisive

contribution to the training of critical thinking as a prerequisite for democracy (Nussbaum, 2010) or to the critical examination of modern trends, such as technologisation (Luckmann, 2004); c) it is mainly individual (Finkenstaedt, 1990; Weingart et al., 1991), few publications are co-authored (Hemlin, 1996; Hellqvist, 2010) and research is often connected to the person conducting it (Hemlin and Gustafsson, 1996; Guetzkow et al., 2004); d) productivity is not that important for research performance in the SSH (Hemlin, 1993; Fisher et al., 2000; Hug et al., 2013); e) societal orientation is important, i.e. research is meant to influence society, direct interaction with society is part of SSH research (Weingart et al., 1991; Hellqvist, 2010; Hug et al., 2013); but f) the influence of society or other stakeholders outside of academia, such as external funding, on SSH research is evaluated negatively (Hemlin, 1993; Hug et al., 2013; Ochsner et al., 2013).

These characteristics must be considered when assessing SSH research. Therefore, there are several bottom-up projects by SSH scholars that analyse how quality is perceived in the SSH disciplines. The European Educational Research Qualitiy Indicators (EERQI) project (Gogolin et al., 2014) started from the discontent with the current assessment practices applied to educational research (Gogolin, 2016: 105–106). The project lasted from 2008 to 2011 and aimed at the development of a set of tools (as opposed to a ranking or rating or a single indicator) to detect research quality (for a summary of the project and its tools, see Gogolin, 2016). The project differentiates between extrinsic quality indicators, that is, quality indicators that are not inherent to the text (such as number of citations, webometrics, authorships), and intrinsic quality indicators, that is, indicators that are inherent to the text (such as rigour, stringency). Part of this set of tools was a peer review questionnaire that included five intrinsic quality criteria for educational research: rigour, originality, significance, style and integrity. The criteria were developed in collaboration with experts in the field, mainly organized within national associations (Gogolin and Stumm, 2014). The project included also an exploratory natural language processing system to highlight the most important sentences in an article. The idea behind the tool was to help reviewers judge an article's quality by guiding their attention to the most important parts of an article (Sandor and Vorndran, 2014a). The tests with the tool showed that while texts in STEM disciplines follow a clear structure and reveal a high potential for automated highlighting, articles in SSH disciplines do not follow such a standard structure. Using keywords and different categories of sentences (for example, problem, summary), the authors argue that highlighting might considerably reduce the time needed for reviewing an article. However, highlighting did not cover two criteria appropriately, that is, integrity and rigour, thus, reviewers using highlighted versions of the article did not always rate those criteria. Furthermore, accuracy of the highlighting differs between (sub-) disciplines and the agreement between automated summaries and reviewers' summary differed between languages (Sandor and Vorndran, 2014a: 50–52). While the authors argue that automatic highlighting seems to work to a certain degree and that a highlighting tool is a promising help to ease peer review workload, the results suggest also that there are severe limits to its usefulness for the assessment of SSH manuscript, especially with regard to the quality criteria. Two out of five criteria tend to be overseen (i.e. integrity and rigour) and language and (sub-) discipline impact the results: summaries by English experts are closer to the sentences highlighted by the tool than the summaries of the French, while the error rate of the highlighting tool is higher for psychological articles than for sociological or historical. However, the authors used this tool also in the multilingual search engine for the EERQI-database and found that it can enhance the search results (Sandor and Vorndran, 2014b).

Also for educational research, Oancea and Furlong (2007) developed criteria for research performance. They define educational research as practice-based and state that such research is not confined to scientificity (that is, discoveries of universal findings or even laws), impact or economic efficiency but also encompasses, amongst others, methodological and theoretical rigour, dialogue, deliberation, participation, ethics and personal growth. They argue that the evaluation of practice-based research has to cope with the entanglement of research and practice, which means that evaluation still has to reflect reasoning and knowledge but it has also to open up for more experimental modes of knowledge coming from within a context of concrete situations and first-person action. While they do not aim at setting standards of good research practice, they conclude that research assessment needs to re-integrate a cultural and philosophical dimension that had been lost in the current discourse of research assessment (Oancea and Furlong, 2007).

A more descriptive approach was chosen by Guetzkow, Lamont and Mallard (2004). They analysed interviews with peer review panellists from multidisciplinary fellowship competitions and found that originality was the most frequently mentioned criterion for judging applications. They thus focused on analysing originality and found that originality is defined differently across different disciplines: Humanists referred often to originality of data and approach whereas social scientists emphasized originality of methods. Besides originality, however, there were also other important criteria, for example, clarity, social relevance, interdisciplinarity, feasibility, importance. Note that those criteria are not necessarily criteria for judging research quality but proposals for a fellowship. Because the authors focused on originality for a more thorough analysis, we do not learn whether there were also disciplinary differences in the salience of those other criteria and in the meaning that was given to the criteria. Given the results regarding originality, however, it is likely that such differences do exist.

The project "Developing and Testing Quality Criteria for Research in the Humanities" (Ochsner et al., 2016) applied a strict bottom-up approach and developed a framework for the exploration and development for quality criteria for SSH research (Hug and Ochsner, 2014) that consists of four pillars: *adopting an inside-out approach* (adequate representation of the scholarly community, also of young scholars, in the development process; discipline specific criteria), *applying a sound measurement approach* (linking indicators to quality criteria derived from the scholars' notions of quality), *making the notions of quality explicit* (apply methods that can elicit criteria from the scholars' tacit knowing of research quality to draw a comprehensive picture of what research quality is in a given discipline; make transparent which quality aspects are measured or included in the assessment and which are not), and *striving for consensus* (methods and especially criteria to be applied in research assessment have to be accepted by the community). This framework was applied to three humanities disciplines, known to be difficult to assess with scientometric methods: German literature studies, English literature studies and art history. In a first step, the scholars' implicit knowing about research activities was investigated, made explicit and summarized into different conceptions of research using Repertory Grid interviews (Ochsner et al., 2013). The results showed that two conceptions of research exist, specifically a modern and a traditional one. This differentiation is not connected to quality: both the modern as well as the traditional research can be of excellent or bad quality. Remarkably, the results also reveal that many commonly used indicators for research assessment, such as interdisciplinarity, internationality, cooperation and social impact, are, in fact, indicators for the modern conception of research and are not related to quality

(Ochsner et al., 2013). Besides the observations about scholars' conceptions of research, quality criteria were extracted from the scholars' notions of quality. In a second step, these quality criteria were completed and rated by all scholars in the three disciplines at the Swiss and LERU universities (League of European Research Universities), thus identifying consensual quality criteria for research using the Delphi method (Hug et al., 2013). According the measurement approach, indicators were identified for the consensual quality criteria (Ochsner et al., 2012) and also rated by the scholars. The results of the project indicate that there are a lot of quality criteria for research in the humanities to consider in research assessments. Many criteria are common to all three disciplines but there are also some discipline specific criteria. Furthermore, there is a mismatch between the humanities scholars' quality criteria and the criteria applied in evaluation procedures (Hug et al., 2013). Importantly, only about 50% of the relevant quality criteria can be measured with quantitative indicators. Therefore, humanities scholars will be critical of research assessments by means of *indicators*. Concerning a research assessment by means of *quality criteria* the studies show that a broad range of quality criteria must be applied and disciplinary differences have to be taken into account. With a certain amount of care, research indicators linked to the relevant criteria can be used to support the experts in research assessments (informed peer review). The project shows that humanities scholars are ready to take part in the development of quality criteria for research assessment if a strict bottom-approach is followed and transparency is assured (Ochsner et al., 2014).

In the context of a broad examination of research assessment in law studies, Lienhard et al. (2016) present quality criteria for research in law studies drawing from the first findings of the project described above (Hug et al., 2013) and complementing them with discipline specific criteria from the law studies. Being a discipline closely connected to a profession, the authors also included professionals (lawyers) into their analysis and find differences in the preferences for quality criteria between professors and lawyers, such as originality, reflexivity and theoretical soundness being emphasized much more by professors than lawyers, while clear language and correctness was more important to lawyers. Besides differentiating evaluations by different stakeholders, for example professors, lawyers or funders, they also differentiate between different assessment situations, for example, research evaluation, assessment of dissertations and habilitations or assessment of scholarly journals (Lienhard et al., 2016: 177).

In France, the Maison des Sciences de l'Homme en Bretagne (MSHB) supported two bottom-up projects related to research assessment in the humanities (for an overview see Williams and Galleron, 2016). The first project, IMPRESHS, was destined to investigate the dissemination practices and impact paths of research conducted by Breton scholars from various SSH disciplines (see https://www.mshb.fr/projets_mshb/impreshs/2314/). Through focus group interviews and a thorough analysis of CVs, the project tried to identify publications with potential impact outside academia, as well as non-academic stakeholders of SSH researchers. The goal of the project was to understand what kind of relations SSH scholars build with these stakeholders, and to what extent one finds practices of co-creation of knowledge in France, such as described within the European project SIAMPI (http://www.siampi.eu). One of the major outcomes of the project is to have uncovered that many SSH scholars exercise a form of auto-censorship when it comes to declaring forms of research or outputs destined to a broader or non-scholar readership, these not being included in institutional forms of reporting or in CVs. This finding draw the attention of the project team upon the problems French scholars face when they come to declaring their

work, since available fields in templates from AERES (the national agency for evaluation of higher education and research), or metadata structure in national repositories (such as HAL—Hyper Articles en Ligne) do not do justice to the large variety of outputs SSH research produces beyond the well-known books traditionally associated with the field. The project ultimately produced a more refined typology of outputs, which supported the creation of a pilot database destined to cope in a more appropriate way with the wealth and variety of SSH research.

The second project, QualiSHS, looked at how evaluative reports produced by AERES reflect disciplinary representations of quality. All evaluative reports produced in 2010–2011 about the activity of all the research units in history and law from two French regions (Bretagne and Rhône-Alpes) have been scrutinized using methods and tools from corpus linguistics, in search of formulations allowing to understand how peer experts conceptualize and perceive quality in the activities and outputs they evaluate. While interviews conducted in parallel confirmed that experts from the two investigated fields diverge regarding their perceptions of quality—a finding which is in line with what other studies pointed out about the diversity of SSH disciplines when it comes to the conceptualization of research quality (see for example, Hug et al., 2013; Gogolin and Stumm, 2014; Lienhard et al., 2016)—it appears that reports do not echo these specificities adequately, since the main criteria they put forward are invariably the coherence of the research conducted in the evaluated unit and its productivity. It is not surprising, therefore, that the French SSH community found that the evaluation conducted by AERES was unsatisfactory on the whole and called for a radical modification of the exercise—a vow that was only very partially answered through the evolution of AERES towards HCERES[6].

National research evaluation practices and the SSH

There are several projects on a national level that approach (national) research assessment in the SSH from a bottom-up perspective or that have designed the model to reflect SSH specificities. The inclusion of the SSH follows different degrees, from implementation of a performance-based funding model under the lead of an SSH scholar and thus accounting for SSH research practices from the beginning (some even say that the system gives the SSH an advantage, see Aagaard et al., 2015) in Norway (Sivertsen, 2016) to a purely bottom-up approach based on research on SSH research practices and their impact on evaluation methods in Switzerland (Loprieno et al., 2016).

The so-called "Norwegian model" (Schneider, 2009) has caught considerable attention during the last years, and similar models were implemented in several countries (Belgium: Flanders, Denmark, Finland and Portugal). The Norwegian model is a performance-based funding model that should "represent all areas of research equally and properly" (Sivertsen, 2016: 80). The design of the model is a "simple pragmatic compromise" (Sivertsen, 2016: 80): one bibliometric indicator to cover all areas of research comprehensively rather than several representations of publication practices for individual disciplines. It consists of three components: a national data base that fully covers peer-reviewed scholarly output from all disciplines including books, a simple publication indicator dividing publications in level 1 and level 2 publications with a system of weights that makes discipline-specific publication traditions comparable at the level of institutions, and a performance-based funding model that reallocates a small fraction of the yearly funding according to the results of the indicator (Sivertsen, 2016: 79). Of course, the Norwegian model would also work without the third component (performance-based funding).

The indicator separates non-academic from academic publications by channels (books: publishers, journal articles: journals). The non-academic publications are not eligible for the performance indicator, while the academic publications are further divided into level 1 and level 2 publications. Level 2 publications cannot represent more than 20% of the world's publications in a field. The government selects renowned scholars (deans, representatives from learned societies), from all major areas of research to be involved in the assignment process of publishers and journals to the levels, resulting in discipline-specific lists of channels.

The system gets more attention from the SSH scholars than from scholars of other areas. While initially the reaction was negative because it turns scholarly output into measures and the system is not designed to cover all scholarly activity but only academic publications, the evaluation of the system showed that there was no major discontent about the system among the scholars (Aagaard et al., 2015). This might be well because of the fact that the indicator showed a high productivity of the SSH disciplines. In addition, while the main effect of the system is an increase of publication activity, the publication patterns did not change: book publishing, international publishing, and language use remained stable. Of course, the evaluation showed also some issues of the funding system: the fractionalizing of authorships favours the SSH, the assignment of experts in the definition of the publication levels is not transparent, and there is unintended use of the system on the individual level (Aagaard et al., 2015).

In the Netherlands, the Royal Academy of the Arts and Sciences criticized the predominance of methods for (and from) national and life sciences in assessment practices in a report called "Judging Research on its Merits" and asked for specific methods for evaluating SSH disciplines in 2005 (Royal Netherlands Academy of Arts and Sciences, 2005). In 2009, the Committee on the National Plan for the Future of the Humanities stated that the existing assessment tools are inadequate to judge the quality of humanities research and advised the Academy to develop a simple, clear and effective system of indicators for the humanities (Committee on the National Plan for the Future of the Humanities, 2009). Thus, the Academy installed a Committee on Quality Indicators in the Humanities, whose report was published in 2011 (Royal Netherlands Academy of Arts and Sciences, 2011). The committee summarizes the situation of research assessment in the humanities in the following way: some policy makers have too high expectations for a simple and purely metric system to compare research performance between research groups and even disciplines. On the other hand, there is too high an aversion against "measuring" research quality and management tools in general in the humanities disciplines. The committee thus suggests a mid-way solution and promotes applying an informed peer review process for SSH research assessments. Peer reviewers assess research along two dimensions, scholarly output and societal quality. Each of the dimensions is assessed using three criteria, that is, scholarly/societal publications or output, scholarly/societal use of output, evidence of scholarly/societal recognition. Each of these criteria can be measured by some quantitative indicators to support the peers in the decision making (for a schematic overview, see Royal Netherlands Academy of Arts and Sciences, 2011: 47). This should add some inter-subjectivity to the peer review process while at the same time recognizing that also the quantitative indicators usually find their base in peer review in the first place (Royal Netherlands Academy of Arts and Sciences, 2011: 11).

The German Council of Science and Humanities (Wissenschaftsrat) reacted in 2004 to the growing importance of university rankings criticizing their methodology and validity

with recommendations on research rankings (Wissenschaftsrat, 2004). It established a comprehensive pilot study for developing and testing a national research rating in the disciplines chemistry and sociology. While such exercises rarely provoke strong reactions in the natural and life sciences, it is more controversial in SSH disciplines. Nevertheless, the research rating in sociology worked out well but met also criticism, especially the non-transparency of the plenary discussions in the panel annihilating the independency of the judgements of the two peers per research unit was pointed out as a danger to the objectivity and validity of the rating (Riordan *et al.*, 2011). In 2008, the *Wissenschaftsrat* decided that pilot studies in other disciplines are to be conducted to improve the procedure (Mair, 2016). History was selected for the pilot study in the humanities. However, the rating for history spurred strong resistance and ended with a boycott by the Association of German Historians (Plumpe, 2009). Mair (2016) suggests that the resistance of the historians was mainly due to miscommunication of the *Wissenschaftsrat* leading to a perception of a top-down-imposed assessment. To make the bottom-up intentions more explicit, a working group was created that worked out modifications to adapt the procedure to the characteristics of humanities research (Wissenschaftsrat, 2010: 203–205). In 2012, a pilot study in the humanities was eventually conducted. While still against the notion of quantifying research performance, the associations of English and American Studies decided to take part in the exercise (Stierstorfer and Schneck, 2016). The *Wissenschaftsrat* qualified the exercise as a success that showed that such a rating is possible in the humanities; the humanities scholars involved in the exercise acknowledged the effort by the *Wissenschaftsrat* to adapt the procedure to the humanities but also identified some negative aspects and consequences of the exercise, such as a division into different sub-disciplines instead of a focus on commonalities (Hornung *et al.*, 2016).

In Switzerland, the Rectors' Conference of the Swiss Universities (CRUS, since 1 January 2016 called swissuniversities) published in 2008 a position paper on research assessment entitled "The Swiss Way to University Quality", which includes ten recommendations for quality monitoring (CRUS, 2008). According to the CRUS, each Swiss university has its own specialization. Therefore, quality assurance has to be accustomed to the mission of each university. A national assessment procedure would therefore not make much sense. Instead, each university should build its own quality assurance system. A potential analysis for bibliometric indicators for research monitoring showed that these procedures are not fitted for use in the SSH. Therefore, a project entitled "Mesurer les performances de la recherche" was initiated that focused on the diversity of SSH research because research "includes a wide array of aspects, from the discovery of new knowledge and promoting young researchers to potential impacts on the scientific community and society" (Loprieno *et al.*, 2016: 14). Since the relevance of these aspects differs between disciplines and university missions, the project paid particular attention to such differences and particularities of the disciplines. The project lasted from 2008 to 2012 and was followed by a second project during the time period of 2013 to 2016. In these two projects, several bottom-up initiatives were funded that researched such diverse topics as, amongst others (for a complete overview of the projects, see Loprieno *et al.*, 2016), profiling in communication sciences (Probst *et al.*, 2011), cooperation of research teams with university partners as well as external stakeholders (Perret *et al.*, 2011), notions of quality of literature studies and art history scholars (Ochsner *et al.*, 2016), evaluation procedures and quality conceptions in law studies (Lienhard *et al.*, 2016),

academic reputation and networks in economics (Hoffmann *et al.*, 2015).

At the same time, the Swiss Academy of Humanities and Social Sciences (SAGW) started a bottom-up initiative on reflections on research assessment in SSH disciplines. Following a conference on the broader topic entitled "For a New Culture in the Humanities" (SAGW, 2012b), the SAGW published a position paper on new developments in the humanities, including recommendations on assessment practices (SAGW, 2012a: 32–36) that emphasizes the importance of bottom-up definitions on quality criteria and methods. The SAGW subsequently funded projects within their member associations to develop their recommendations or standards for research assessments in their disciplines. The resulting report features statements from Asian and Oriental studies, area studies, cultural and social anthropology, peace research, political sciences, art history and environmental humanities accompanied by a synthesis report by the SAGW (Iseli, 2016).

Bottom-up initiatives at the European level
The different assessment procedures applied at the university or national level, the initial exclusion of SSH research in the ERC Grant-schemes as well as the initial concerns of severe cut-backs for the SSH in the Horizon 2020 program (König, 2016: 154–155) led to a higher interest of SSH scholars in the topic of research assessment. As the sections above show, there is a rise in SSH research on research assessment and evaluation, leading to sessions or even tracks dedicated to SSH research assessment at international scientometric conferences like the ISSI 2015 (www.issi2015.org) or the STI 2016 (sti2016.org) conferences, or to an international conference dedicated exclusively to SSH research evaluation, RESSH 2015 (www.ressh.eu). Even more important, SSH scholars team up with scientometricians concerned about the state of SSH research assessment (often SSH scholars themselves) in a European association called EvalHum initiative (www.evalhum.eu). EvalHum sets out to motivate and support bottom-up work on research evaluation in the SSH and encourages best practices in research evaluation in SSH that ensure adequate assessment procedures for the respective disciplines. EvalHum is also a forum on this topic and will strive for an accurate recognition of SSH research at the European level.

Currently, there is a COST Action entitled "European Network for Research Evaluation in the Social Sciences and Humanities (ENRESSH)" (CA-15137) that brings together SSH scholars from 30 European countries working together to improve assessment procedures in and for the SSH (http://www.cost.eu/COST_Actions/ca/CA15137). The idea behind the action is "evaluating to valorize" because applying ill-adapted methods lead to under-valuation of SSH research. Participants in the Action share data about SSH research and confront methodologies, resulting in co-authored publications but also in policy briefs, collections of best practices and, ultimately, guidelines for SSH research evaluation. ENRESSH seeks also to involve the different stakeholders having a say in assessment principles and processes, to progress towards adequate frameworks and practices of SSH research. The Action consists of 4 Work Groups. The first Work Group focuses on the conceptual frameworks for SSH research assessment and studies the SSH knowledge production processes and strategies as a basis for developing adequate assessment procedures reflecting the SSH research practices. It investigates SSH scholars' perceptions of research quality, peer review practices and national assessment practices. The second Work Group is about societal impact and relevance of SSH research. It observes the structural requirements needed for a smooth transfer of SSH research to the society, the national policies towards transfer to socio-economic or NGO

partners, proposes procedures to collect data about engagement with the society and measures to better value the SSH. The third Work Group concerns databases and the use of data for understanding SSH research. It builds standards for the interoperability of, and methods for integrating data from, current research information systems and repositories dedicated to the SSH, to allow comparability of SSH publishing practices in various countries. It analyses the characteristics of SSH dissemination channels, develops common rules for building databases, designs a roadmap for a European bibliometric database and develops alternative metrics for the SSH. The fourth Work Group is concerned with the dissemination of the results of the Action. It builds a list of relevant European stakeholders in SSH research assessment and interacts actively with them and organizes conferences.

The future of research assessment in the humanities

While until recently research on assessment in the SSH focused on the deficiencies of the current assessment methods, such as bibliometrics and scientometrics, there is now much research going on that takes a bottom-up approach and focuses on research practices in the SSH and reflects on how to assess SSH research with its own methods instead of applying and adjusting the methods developed for and in the natural and life sciences (see also Hammarfelt, 2016: 115). This is an important development because we can learn from the examples shown in the sections above that whenever the scholars felt that the assessment procedures were imposed top-down without proper adjustments to SSH research, it resulted in boycott or resistance (see for example, Academics Australia, 2008; Andersen et al., 2009; Mair, 2016).

The projects presented in this article show furthermore that if the assessment procedures adequately reflect the SSH research practices, scholars are ready to collaborate (for example, Giménez-Toledo et al., 2013; Ochsner et al., 2014) and to accept more easily research assessment, like in the Norwegian or German case (Aagaard et al., 2015; Sivertsen, 2016; Stierstorfer and Schneck, 2016). Full-coverage databases including all relevant document types are of value for scholarly work (Gogolin, 2016; Sandor and Vorndran, 2014a, b) and increase the visibility of humanities research production (Aagaard et al., 2015). While there are some degrees of convergence in some countries regarding their databases (Giménez-Toledo et al., 2016), the conditions for full interoperability have yet to be discussed. It also has to be born in mind that universities fulfil different missions and countries face diverse challenges. Criteria and procedures for research evaluation should be adapted to the missions of the universities and to the specific aims of the evaluation (Loprieno et al., 2016).

The future of research assessment in the humanities lies therefore in bottom-up procedures that are based on the research practices in the respective disciplines. However, the projects presented in this article show that more research on the research practices in the humanities is needed. Such research has only started. If bottom-up approaches are to be followed, more knowledge is needed on how research is conducted and disseminated as well as how it is used by different stakeholders including the SSH researchers themselves.

Combining the approaches and the insights on SSH research production presented in this article, we propose the following recommendations for research assessment in the humanities (these recommendations draw on Ochsner et al., 2015):

(1) The preferred method of evaluation is *informed peer review*: peer review is accepted among scholars as an assessment procedure. However, it has several drawbacks such as, for example, poor inter-subjectivity and low reliability through dependency on the panel composition (Bornmann, 2011; Riordan et al., 2011; Royal Netherlands Academy of Arts and Sciences, 2011). Scientific and political measures can however be taken to reduce these inconveniences, such as applying a fair evaluation process that grants the evaluated scholars the possibility to comment upon the process and its results.

(2) A broad range of quality criteria has to be taken into account. The quality criteria must be developed bottom-up and reflect the notions of quality of the assessed scholars (Hug et al., 2013; Ochsner et al., 2013) as they alone can judge what quality in the discipline actually is and they do see research quality predominantly as academic quality (Kekäle, 2002). To assure that all paradigms and research traditions as well as new ways of thinking are included, quality criteria should be developed surveying all scholars to be evaluated.

(3) For the quality criteria that reach consensus among the scholars, indicators can be identified. The scholars should rate the indicators with regard to how these indicators are measuring the criterion adequately.

(4) From the quality criteria and indicators that reach consensus among the scholars, an evaluation sheet is to be created. The evaluation sheet thus includes both criteria that can be measured with indicators and criteria that cannot be measured (Ochsner et al., 2012).

(5) Other stakeholders' criteria for research performance can be included in the evaluation sheet to take into account other goals of research than academic research quality (Royal Netherlands Academy of Arts and Sciences, 2011). While not developed specifically for the humanities but in a way that allows a bottom-up approach to societal impact, the "Evaluating Research in Context"-project could serve as an example (Spaapen et al., 2007). The criteria and indicators from other stakeholders should be indicated as such to ensure the transparency to the researchers and to make visible what is important from an academic point of view and what is important from other stakeholders' view.

(6) The peers must rate every criterion on its own, which is in line with the insights of Thorngate et al. (2009) who summarize the findings of their comprehensive research on decision making the following way: judging something overall is usually inconsistent and not adequate for judging merit while judging separately according to specified criteria reveals more reliable results (Thorngate et al., 2009: 26). The peers' reading should be restricted to a reasonable amount of effort.

(7) Rankings or ratings with an overall measure should not be published. Instead, the results of every single criterion should be provided. If overall ratings are produced, the weighting procedure has to be made transparent. However, it should be kept in mind that research units have different missions to fulfil, therefore an overall rating might favour some missions over others leading to a structural discrimination of some research units.

Many important issues of our times are global in nature and society has high hopes in a technical solution. The SSH, and specifically the humanities, are therefore not in the focus of the public discourse. Especially the critical questions SSH disciplines are asking are not high on the political agenda. However, complex global issues such as, for example, global warming, migration crisis, ageing or HIV cannot be sufficiently resolved without the knowledge of SSH disciplines. The critical questions challenging the blind technological faith in overcoming such problems are crucial. Not being on the top of the political agenda, however, does not mean to give in to the mainstream neo-positivist notion

of a parametrically steered research policy. Nor does it mean that SSH scholars should frown at all requests for accountability. Instead, SSH disciplines should step forward and self-confidently and openly question truisms or blind technological faith and propose alternatives to simple but misleading practices. This paper presents many bottom-up actions of SSH scholars taking research assessment in their own hands. Certainly, these bottom-up procedures will lead to a more adequate assessment of SSH research but they might also help fostering a better valorization of SSH research among policy makers and colleagues from the natural sciences. And eventually, maybe some scientists will find these approaches also fruitful for their own disciplines? At the same time, an adequate evaluation and valorization of SSH research will also help society to better understand what the SSH contribution to solving major societal challenges can be. Therefore, taking the time to encourage bottom-up evaluation initiatives should result in better solving of modern societies' issues.

Notes

1 Because it comes from SSH scholars and is clearly bottom-up in nature, we include, however, the initiative by Oancea and Furlong (2007) that was motivated by, but did not have a visible impact on, the RAE in the United Kingdom.

2 As mentioned in the introduction, we focus on European initiatives for coherence reasons and because of restrictions of space.

3 Despite the inclusion of (some) books in the commercial databases in recent years as well as the rise of networking sites also promoting bibliographic data, the under-coverage of certain disciplines and languages remains while technical challenges arise and issues of transparency persist (Gorraiz et al., 2013; Murray, 2014; Zuccala and Cornacchia, 2016)

4 See the official website https://dbh.nsd.uib.no/publiseringskanaler/erihplus/

5 Others argue that there is a difference between performance-based funding and research evaluation. The first is used to distribute scarce funds and needs not to be related to quality while the latter is formative in nature and includes an understanding of quality. However, while this might be true from the evaluator's perspective, it is misleading regarding the effect on the scholars' behaviour. If scholars are assessed by indicators, they perceive these not only as incentives but also as indications of what is expected from them (as well as what is valued as 'good' research) and they will adjust their behaviour accordingly (see for example, Hammarfelt and de Rijcke, 2015; Williams and Galleron, 2016).

6 The Evaluation Agency for Research and Higher Education (Agence d'évaluation de la recherche et de l'enseignement supérieur, AERES) was replaced by the High Council for the Evaluation of Research and Higher Education (Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur, HCERES) on 17 November 2014.

References

Aagaard K, Bloch C and Schneider JW (2015) Impacts of performance-based research funding systems: The case of the Norwegian Publication Indicator. *Research Evaluation*; **24** (2): 106–117.

Academics Australia. (2008) *Letter to Senator the Honourable Kim Carr, Minister for Innovation, Industry, Science and Research*, https://web.archive.org/web/20091221195149/http://www.academics-australia.org/AA/ERA/era.pdf, accessed 8 February 2017.

van den Akker W (2016) Yes we should; research assessment in the humanities. In: Ochsner M, Hug SE and Daniel H-D (eds). *Research Assessment in the Humanities. Towards Criteria and Procedures*. Springer International Publishing: Cham, Switzerland, pp 23–29.

Andersen H et al (2009) Editorial journals under threat: A joint response from history of science. Technology and Medicine Editors. *Social Studies of Science*; **39** (1): 6–9.

Archambault É, Vignola-Gagne E, Cote G, Lariviere V and Gingras Y (2006) Benchmarking scientific output in the social sciences and humanities: The limits of existing databases. *Scientometrics*; **68** (3): 329–342.

van Arensbergen P, van der Weijden I and van den Besselaar P (2014a) Different views on scholarly talent: What are the talents we are looking for in science? *Research Evaluation*; **23** (4): 273–284.

van Arensbergen P, van der Weijden I and van den Besselaar P (2014b) The selection of talent as a group process. A literature review on the social dynamics of decision making in grant panels. *Research Evaluation*; **23** (4): 298–311.

Arts and Humanities Research Council. (2006) *Use of research metrics in the arts and humanities*; Report of the expert group set up jointly by the Arts and Humanities Research Council and the Higher Education Funding Council for England AHRC: Bristol, UK.

Arts and Humanities Research Council. (2009) *Leading the World. The economic Impact of UK Arts and Humanities Research*. AHRC: Bristol, UK.

Barker K (2007) The UK research assessment exercise: The evolution of a national research evaluation system. *Research Evaluation*; **16** (1): 3–12.

Bornmann L (2011) Scientific peer review. *Annual Review of Information Science and Technology*; **45** (1): 197–245.

Bunia R (2016) Quotation statistics and culture in literature and in other humanist disciplines. In: Ochsner M, Hug SE and Daniel H-D (eds). *Research Assessment in the Humanities. Towards Criteria and Procedures*. Springer International Publishing: Cham, Switzerland, pp 133–148.

Burrows R (2012) Living with the h-index? Metric assemblages in the contemporary academy. *The Sociological Review*; **60** (2): 355–372.

Butler L (2003) Explaining Australia's increased share of ISI publications—The effects of a funding formula based on publication counts. *Research Policy*; **32** (1): 143–155.

Butler L (2007) Assessing university research: A plea for a balanced approach. *Science and Public Policy*; **34** (8): 565–574.

Butler L (2008) Using a balanced approach to bibliometrics: Quantitative performance measures in the Australian research quality framework. *Ethics in Science and Environmental Politics*; **8** (1): 83–92.

Butler L and McAllister I (2009) Metrics or peer review? Evaluating the 2001 UK research assessment exercise in political science. *Political Studies Review*; **7** (1): 3–17.

Butler L and Visser MS (2006) Extending citation analysis to non-source items. *Scientometrics*; **66** (2): 327–343.

Chi P-S (2012) Bibliometric characteristics of political science research in Germany. *Proceedings of the American Society for Information Science and Technology*; **49** (1): 1–6.

Chi P-S (2014) Which role do non-source items play in the social sciences? A case study in political science in Germany. *Scientometrics*; **101** (2): 1195–1213.

Committee on the National Plan for the Future of the Humanities. (2009) *Sustainable Humanities: Report from the National Committee on the Future of the Humanities in the Netherlands*. Amsterdam University Press: Amsterdam, The Netherlands.

Commission of the European Communities. (2000) *Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of the Regions: Towards a European research area*. Commission of the European Communities: Brussels, UK.

Council for the Humanities, Arts and Social Sciences. (2009) *Humanities and Creative Arts: Recognising Esteem Factors and Non- traditional Publication in Excellence in Research for Australia (ERA) Initiative*; CHASS Papers. Council for the Humanities Arts and Social Sciences: Canberra, Australia.

CRUS. (2008) *The Swiss Way to University Quality*. Rectors' Conference of the Swiss Universities (CRUS): Bern, Switzerland.

Drabek A, Rozkosz EA, Hołowiecki M and Kulczycki E (2015) Polski Współczynnik Wpływu a kultury cytowań w humanistyce. *Nauka i Szkolnictwo Wyższe*; **46** (2): 121–138.

European Research Area and Innovation Committee. (2015) *European Research Area (ERA) Roadmap 2015–2020*. European Research Area and Innovation Committee: Brussels, UK.

Ferrara A and Bonaccorsi A (2016) How robust is journal rating in Humanities and Social Sciences? Evidence from a large-scale, multi-method exercise. *Research Evaluation*; **25** (3): 279–291.

Finkenstaedt T (1990) Measuring research performance in the humanities. *Scientometrics*; **19** (5): 409–417.

Fisher D, Rubenson K, Rockwell K, Grosjean G and Atkinson-Grosjean J (2000) *Performance Indicators and the Humanities and Social Sciences*. Centre for Policy Studies in Higher Education and Training: Vancouver, BC.

Genoni P and Haddow G (2009) ERA and the ranking of australian humanities journals. *Australian Humanities Review*; **46**, 5–24.

van Gestel R, Micklitz H-W and Poiares MM (2012) Methodology in the new legal world. EUI Working Papers LAW 2012/13. doi: 10.2139/ssrn.2069872.

Giménez-Toledo E (2016) Assessment of journal & book publishers in the humanities and social sciences in Spain. In: Ochsner M, Hug SE and Daniel H-D (eds). *Research Assessment in the Humanities. Towards Criteria and Procedures*. Springer International Publishing: Cham, Switzerland, pp 91–102.

Giménez-Toledo E et al (2016) Taking scholarly books into account: Current developments in five European countries. *Scientometrics*; **107** (2): 1–15.

Giménez-Toledo E, Tejada-Artigas C and Mañana-Rodríguez J (2013) Evaluation of scientific books' publishers in social sciences and humanities: Results of a survey. *Research Evaluation*; **22** (1): 64–77.

Glänzel W (1996) A bibliometric approach to social sciences. National research performances in 6 selected social science areas, 1990–1992. *Scientometrics*; **35** (3): 291–307.

Glänzel W, Thijs B and Debackere K (2016) Productivity, performance, efficiency, impact—What do we measure anyway? *Journal of Informetrics*; **10** (2): 658–660.

Gogolin I (2016) European educational research quality indicators (EERQI): An experiment. In: Ochsner M, Hug SE and Daniel H-D (eds). *Research Assessment*

in the Humanities. Towards Criteria and Procedures. Springer International Publishing: Cham, Switzerland, pp 103–111.

Gogolin I, Astrom F and Hansen A (eds) (2014) *Assessing Quality in European Educational Research*. Springer VS: Wiesbaden, Germany.

Gogolin I, Stumm V (2014) The EERQI peer review questionnaire. In: Gogolin I, Astrom F and Hansen A (eds). *Assessing Quality in European Educational Research*. Springer VS: Wiesbaden, Germany, pp 107–120.

Gorraiz J, Purnell PJ and Glänzel W (2013) Opportunities for and limitations of the Book Citation Index. *Journal of The American Society For Information Science and Technology*; **64** (7): 1388–1398.

Guetzkow J, Lamont M and Mallard G (2004) What Is originality in the humanities and the social sciences? *American Sociological Review*; **69** (2): 190–212.

Guillory J (2005) Valuing the humanities, evaluating scholarship. *Profession (MLA)*; **11**, 28–38.

Gumpenberger C, Glänzel W and Gorraiz J (2016) The ecstasy and the agony of the altmetric score. *Scientometrics*; **108** (2): 977–982.

Hamann J (2016) The visible hand of research performance assessment. *Higher Education*; **72** (6): 761–779.

Hammarfelt B (2012) *Following the footnotes: A bibliometric analysis of citation patterns in literary studies*. Doctoral dissertation. Skrifter utgivna vid institutionen för ABM vid Uppsala Universitet (Vol. 5). Uppsala Universitet: Uppsala, http://www.diva-portal.org/smash/get/diva2:511996/FULLTEXT01.pdf.

Hammarfelt B (2013) Harvesting footnotes in a rural field: Citation patterns in Swedish literary studies. *Journal of Documentation*; **68** (4): 536–558.

Hammarfelt B (2016) Beyond coverage: Toward a bibliometrics for the humanities. In: Ochsner M, Hug SE and Daniel H-D (eds). *Research Assessment in the Humanities. Towards Criteria and Procedures*. Springer International Publishing: Cham, Switzerland, pp 115–131.

Hammarfelt B and de Rijcke S (2015) Accountability in context: Effects of research evaluation systems on publication practices, disciplinary norms, and individual working routines in the faculty of Arts at Uppsala University. *Research Evaluation*; **24** (1): 63–77.

Hellqvist B (2010) Referencing in the humanities and its implications for citation analysis. *Journal of The American Society For Information Science And Technology*; **61** (2): 310–318.

Hemlin S (1993) Scientific quality in the eyes of the scientist. A questionnaire study. *Scientometrics*; **27** (1): 3–18.

Hemlin S (1996) Social studies of the humanities. A case study of research conditions and performance in ancient history and classical archaeology and English. *Research Evaluation*; **6** (1): 53–61.

Hemlin S and Gustafsson M (1996) Research production in the arts and humanities. A questionnaire study of factors influencing research performance. *Scientometrics*; **37** (3): 417–432.

Herbert U and Kaube J (2008) Die Mühen der Ebene: Über Standards, Leistung und Hochschulreform. In: Lack E and Markschies C (eds). *What the hell is quality? Qualitätsstandards in den Geisteswissenschaften*. Campus-Verlag: Frankfurt, Germany, pp 37–51.

Hicks D (2004) The four literatures of social science. In: Moed H, Glänzel W and Schmoch U (eds). *Handbook of Quantitative Science and Technology Research*. Kluwer Academic Publishers: New York, pp 473–496.

Hicks D (2012) Performance-based university research funding systems. *Research Policy*; **41** (2): 251–261.

Hoffmann CP, Lutz C and Meckel M (2015) A relational altmetric? Network centrality on ResearchGate as an indicator of scientific impact. *Journal of the Association for Information Science and Technology*; **67** (4): 765–775.

Holmberg K and Thelwall M (2014) Disciplinary differences in Twitter scholarly communication. *Scientometrics*; **101** (2): 1027–1042.

Hornung A, Khlavna V and Korte B (2016) Research Rating Anglistik/Amerikanistik of the German Council of Science and Humanities. In: Ochsner M, Hug SE and Daniel H-D (eds). *Research Assessment in the Humanities. Towards Criteria and Procedures*. Springer International Publishing: Cham, Switzerland, pp 219–233.

Hose M (2009) Qualitätsmessung: Glanz und Elend der Zahl. In: Prinz C and Hohls R (eds). *Qualitätsmessung, Evaluation, Forschungsrating. Risiken und Chancen für die Geschichtswissenschaft?*. Historisches Forum. Clio-online: Berlin, Germany, pp 91–98.

Hug SE and Ochsner M (2014) A framework to explore and develop criteria for assessing research quality in the humanities. *International Journal of Education Law and Policy*; **10** (1): 55–68.

Hug SE, Ochsner M and Daniel H-D (2013) Criteria for assessing research quality in the humanities: A Delphi study among scholars of English literature, German literature and art history. *Research Evaluation*; **22** (5): 369–383.

Iseli M (2016) *Qualitäts- und Leistungsbeurteilung in den Geistes- und Sozialwissenschaften: Prinzipien, Ansätze und Verfahren*. SAGW: Bern, Switzerland.

Johnston R (2008) On structuring subjective judgements: Originality, significance and rigour in RAE 2008. *Higher Education Quarterly*; **62** (1/2): 120–147.

Kekäle J (2002) Conceptions of quality in four different disciplines. *Tertiary Education and Management*; **8** (1): 65–80.

König T (2016) Peer review in the social sciences and humanities at the European Level: The experiences of the European research council. In: Ochsner M, Hug SE and Daniel H-D (eds). *Research Assessment in the Humanities. Towards Criteria and Procedures*. Springer International Publishing: Cham, Switzerland, pp 151–163.

Kousha K and Thelwall M (2009) Google book search: Citation analysis for social science and the humanities. *Journal of The American Society For Information Science And Technology*; **60** (8): 1537–1549.

Krull W and Tepperwien A (2016) The four 'T's: Quality indicators for the humanities. In: Ochsner M, Hug SE and Daniel H-D (eds). *Research Assessment in the Humanities. Towards Criteria and Procedures*. Springer International Publishing: Cham, Switzerland, pp 165–179.

Kwok JT (2013) *Impact of ERA Research Assessment on University Behaviour and their Staff*. National Tertiary Education Union: South Melbourne, Australia.

Lack E (2008) Einleitung—Das Zauberwort "Standards". In: Lack E and Markschies C (eds). *What the hell is quality? Qualitätsstandards in den Geisteswissenschaften*. Campus-Verlag: Frankfurt, Germany, pp 9–34.

Lamont M (2009) *How Professors Think: Inside the Curious World of Academic Judgment*. Harvard University Press: Harvard, UK.

Lariviere V, Gingras Y and Archambault É (2006) Canadian collaboration networks: A comparative analysis of the natural sciences, social sciences and the humanities. *Scientometrics*; **68** (3): 519–533.

Lauer G (2016) The ESF scoping project "towards a bibliometric database for the social sciences and humanities". In: Ochsner M, Hug SE and Daniel H-D (eds). *Research Assessment in the Humanities. Towards Criteria and Procedures*. Springer International Publishing: Cham, Switzerland, pp 73–77.

Lawrence PA (2002) Rank injustice. *Nature*; **415** (6874): 835–836.

van Leeuwen TN (2013) Bibliometric research evaluations, web of science and the social sciences and Humanities: A problematic relationship? *Bibliometrie—Praxis und Forschung*; **2**, 8.

Lienhard A, Tanquerel T, Flückiger A, Amschwand F, Byland K and Herrmann E (2016) *Forschungsevaluation in der Rechtswissenschaft: Grundlagen und empirische Analyse in der Schweiz*. Stämpfli Verlag: Bern, Switzerland.

Loprieno A, Werlen R, Hasgall A and Bregy J (2016) The "Mesurer les Performances de la Recherche" Project of the Rectors' Conference of the Swiss Universities (CRUS) and Its Further Development. In: Ochsner M, Hug SE and Daniel H-D (eds). *Research Assessment in the Humanities. Towards Criteria and Procedures*. Springer International Publishing: Cham, Switzerland, pp 13–21.

Luckman S (2004) More than the sum of its parts: The humanities and communicating the 'hidden work' of research. In: Kenway J, Bullen E and Robb S (eds). *Innovation and Tradition: The Arts, Humanities, and the Knowledge Economy*. Peter Lang: New York, pp 82–90.

MacDonald SP (1994) *Professional Academic Writing in the Humanities and Social Sciences*. Southern Illinois University Press: Carbondale, Edwardsville, IL.

Mair C (2016) Rating research performance in the humanities: An interim report on an initiative of the German Wissenschaftsrat. In: Ochsner M, Hug SE and Daniel H-D (eds). *Research Assessment in the Humanities. Towards Criteria and Procedures*. Springer International Publishing: Cham, Switzerland, pp 201–209.

Martin BR *et al* (2010) *Towards a bibliometric database for the social sciences and humanities. A European scoping project* (A report produced for DFG, ESRC, AHRC, NWO, ANR and ESF). Science and Technology Policy Research Unit: Sussex.

McCarthy KF, Ondaatje EH, Zakaras L and Brooks A (2004) *Gifts of the Muse. Reframing the Debate About the Benefits of the Arts*. RAND Corporation: Santa Monica, CA.

Mohammadi E and Thelwall M (2014) Mendeley readership altmetrics for the social sciences and humanities: Research evaluation and knowledge flows. *Journal of the Association for Information Science and Technology*; **65** (8): 1627–1638.

Mojon-Azzi SM, Jiang X, Wagner U and Mojon DS (2003) Journals: Redundant publications are bad news. *Nature*; **421** (6920): 209.

Molinié A and Bodenhausen G (2010) Bibliometrics as weapons of mass citation. *CHIMIA International Journal for Chemistry*; **64** (1): 78–89.

Moonesinghe R, Khoury MJ and Janssens A C J W (2007) Most published research findings are false—But a little replication goes a long way. *PLoS Medicine*; **4** (2): e28.

Murray M (2014) Analysis of a scholarly social networking site: The case of the dormant user. *SAIS 2014 Proceedings*. Paper 24, http://aisel.aisnet.org/sais2014, accessed 8 February 2017.

Nederhof AJ (2006) Bibliometric monitoring of research performance in the social sciences and the humanities: A review. *Scientometrics*; **66** (1): 81–100.

Nederhof AJ, Zwaan R, de Bruin R and Dekker P (1989) Assessing the usefulness of bibliometric indicators for the humanities and the social sciences—a comparative study. *Scientometrics*; **15** (5): 423–435.

Norris M and Oppenheim C (2003) Citation counts and the research assessment exercise V—Archaeology and the 2001 RAE. *Journal of Documentation*; **59** (6): 709–730.

Nussbaum MC (2010) *Not for Profit: Why Democracy Needs the Humanities*. Princeton University Press: Princeton, NJ.

Oancea A and Furlong J (2007) Expressions of excellence and the assessment of applied and practice-based research. *Research Papers in Education*; **22** (2): 119–137.

Ochsner M, Hug SE and Daniel H-D (2012) Indicators for research quality in the humanities: Opportunities and limitations. *Bibliometrie—Praxis und Forschung*; **1**, 4.

Ochsner M, Hug SE and Daniel H-D (2013) Four types of research in the humanities: Setting the stage for research quality criteria in the humanities. *Research Evaluation*; **22** (2): 79–92.

Ochsner M, Hug SE and Daniel H-D (2014) Setting the stage for the assessment of research quality in the humanities. Consolidating the results of four empirical studies. *Zeitschrift für Erziehungswissenschaft*; **17** (6): 111–132.

Ochsner M, Hug SE, Daniel H-D (eds) (2016) Humanities scholars' conceptions of research quality. In: *Research Assessment in the Humanities. Towards Criteria and Procedures*. Springer International Publishing: Cham, Switzerland, pp 43–69.

Ochsner M, Wolbring T and Hug SE (2015) Quality criteria for sociology? What sociologists can learn from the project "developing and testing research quality criteria in the humanities". *Sociologia E Politiche Sociali*; **18** (2): 90–110.

Oppenheim C and Summers M (2008) Citation counts and the research assessment exercise, part VI: Unit of assessment 67 (music). *Information Research*; **13** (2).

Ossenblok T and Engels T (2015) Edited books in the social sciences and humanities: Characteristics and collaboration analysis. *Scientometrics*; **104** (1): 219–237.

Ossenblok TLB, Engels TCE and Sivertsen G (2012) The representation of the social sciences and humanities in the web of science: A comparison of publication patterns and incentive structures in Flanders and Norway (2005–9). *Research Evaluation*; **21** (4): 280–290.

Palumbo M and Pennisi C (2015) Criteri corretti e condivisi per una valutazione buona e utile della ricerca [Correct and shared criteria for a good and useful evaluation of research]. *Sociologia E Politiche Sociali*; **18** (2): 73–89.

Perret JF, Sormani P, Bovet A and Kohler A (2011) Décrire et mesurer la "fécondité" des recherches en sciences humanies et sociales. *Bulletin SAGW*; **2011** (2): 40–42.

Plag I (2016) Research assessment in a philological discipline: Criteria and rater reliability. In: Ochsner M, Hug SE and Daniel H-D (eds). *Research Assessment in the Humanities. Towards Criteria and Procedures*. Springer International Publishing: Cham, Switzerland, pp 235–247.

Plumpe W (2009) Qualitätsmessung: Stellungnahme zum Rating des Wissenschaftsrates aus Sicht des Historikerverbandes. In: Prinz C and Hohls R (eds). *Qualitätsmessung, Evaluation, Forschungsrating. Risiken und Chancen für die Geschichtswissenschaft?*. Historisches Forum. Clio-online: Berlin, Germany, pp 121–126.

Plumpe W (2010) Der Teufel der Unvergleichbarkeit. Über das quantitative Messen und Bewerten von Forschung. *Forschung und Lehre*; **17** (8): 572–574.

Probst C, Lepori B, de Filippo D and Ingenhoff D (2011) Profiles and beyond: Constructing consensus on measuring research output in communication sciences. *Research Evaluation*; **20** (1): 73–88.

Redden G (2008) From RAE to ERA: Research evaluation at work in the corporate university. *Australian Humanities Review*; **45**, 7–26.

Riordan P, Ganser C and Wolbring T (2011) Measuring the quality of research. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*; **63** (1): 147–172.

Royal Netherlands Academy of Arts and Sciences. (2005) *Judging research on its Merits*. An advisory report by the Council for the Humanities and the Social Sciences Council. Royal Netherlands Academy of Arts and Sciences: Amsterdam.

Royal Netherlands Academy of Arts and Sciences. (2011) *Quality Indicators for Research in the Humanities*. Royal Netherlands Academy of Arts and Sciences: Amsterdam, The Netherlands.

SAGW. (2012a) *Für eine Erneuerung der Geisteswissenschaften. Empfehlungen der SAGW zuhanden der Leitungsorgane der Hochschulen, der Lehrenden, der Förderorganisationen und des Staatssekretariats für Bildung und Forschung*. SAGW: Bern, Switzerland.

SAGW. (2012b) *Für eine neue Kultur der Geisteswissenschaften?* Akten des Kongresses vom 30. November bis 2. Dezember 2011, Bern. SAGW: Bern, Switzerland.

Sandor A and Vorndran A (2014a) Enhancing relevance ranking of the EERQI search engine. In: Gogolin I, Astrom F and Hansen A (eds). *Assessing Quality in European Educational Research*. Springer VS: Wiesbaden, Germany, pp 56–59.

Sandor A and Vorndran A (2014b) Highlighting salient sentences for reading assistance. In: Gogolin I, Astrom F and Hansen A (eds). *Assessing Quality in European Educational Research*. Springer VS: Wiesbaden, Germany, pp 43–55.

Schneider JW (2009) An outline of the bibliometric indicator used for performance-based funding of research institutions in Norway. *European Political Science*; **8** (3): 364–378.

Sivertsen G (2016) Publication-based funding: The Norwegian model. In: Ochsner M, Hug SE and Daniel H-D (eds). *Research Assessment in the Humanities. Towards Criteria and Procedures*. Springer International Publishing: Cham, Switzerland, pp 79–90.

Spaapen J, Dijstelbloem H and Wamelink F (2007) *Evaluating Research in Context: A Method for Comprehensive Assessment*, Second edition, Consultative Committee of Sector Councils for Research and Development: The Hague, The Netherlands.

Stierstorfer K and Schneck P (2016) "21 Grams": Interdisciplinarity and the assessment of quality in the humanities. In: Ochsner M, Hug SE and Daniel H-D (eds). *Research Assessment in the Humanities. Towards Criteria and Procedures*. Springer International Publishing: Cham, Switzerland, pp 211–218.

Thorngate W, Dawes RM and Foddy M (2009) *Judging merit*. Pychology Press Taylor & Francis Group: New York, Hove, UK.

Unreliable research. Trouble at the lab. (2013) *The Economist*. 19 October, http://www.economist.com/news/briefing/21588057-scientists-think-science-self-cor recting-alarming-degree-it-not-trouble.

Vec M (2009) Qualitätsmessung: Die vergessene Freiheit. Steuerung und Kontrolle der Geisteswissenschaften unter der Prämisse der Prävention. In: Prinz C and Hohls R (eds). *Qualitätsmessung, Evaluation, Forschungsrating. Risiken und Chancen für die Geschichtswissenschaft?*. Historisches Forum. Clio-online: Berlin, Germany, pp 79–90.

Verleysen FT and Weeren A (2016) Clustering by publication patterns of senior authors in the social sciences and humanities. *Journal of Informetrics*; **10** (1): 254–272.

VolkswagenStiftung. (2014) *What is Intellectual Quality in the Humanities? Some Guidelines*. VolkswagenStiftung: Hannover, Germany.

Weingart P, Prinz W, Kastner M, Maasen S and Walter W (1991) *Die sogenannten Geisteswissenschaften: Aussenansichten: Die Entwicklung der Geisteswissenschaften in der BRD, 1954–1987*. Suhrkamp: Frankfurt am Main, Germany.

White HD, Boell SK, Yu H, Davis M, Wilson CS and Cole FTH (2009) Libcitations: A measure for comparative assessment of book publications in the humanities and social sciences. *Journal of The American Society For Information Science and Technology*; **60** (6): 1083–1096.

Williams G and Galleron I (2016) Bottom Up from the bottom: A new outlook on research evaluation for the SSH in France. In: Ochsner M, Hug SE and Daniel H-D (eds). *Research Assessment in the Humanities. Towards Criteria and Procedures*. Springer International Publishing: Cham, Switzerland, pp 181–198.

Wissenschaftsrat. (2004) *Recommendations for rankings in the system of higher education and research. Part 1: Research*. Wissenschaftsrat: Hamburg.

Wissenschaftsrat. (2010) *Empfehlungen zur vergleichenden Forschungsbewertung in den Geisteswissenschaften*. Wissenschaftsrat: Köln, Germany.

Zuccala AA (2016) Inciting the metric oriented humanist: Teaching bibliometrics in a faculty of humanities. *Education for Information*; **32** (2): 149–164.

Zuccala AA, Verleysen FT, Cornacchia R and Engels TCE (2015) Altmetrics for the humanities. *Aslib Journal of Information Management*; **67** (3): 320–336.

Zuccala AA and Cornacchia R (2016) Data matching, integration, and interoperability for a metric assessment of monographs. *Scientometrics*; **108** (1): 465–484.

Zuccala AA and van Leeuwen T (2011) Book reviews in humanities research evaluations. *Journal of The American Society For Information Science and Technology*; **62** (10): 1979–1991.

Data availability

Data sharing is not applicable to this article as no datasets were analysed or generated.

Acknowledgements

Additional information

How to cite this article: Ochsner M *et al.* (2017) The future of research assessment in the humanities: bottom-up assessment procedures. *Palgrave Communications*. 3:17020 doi: 10.1057/palcomms.2017.20.