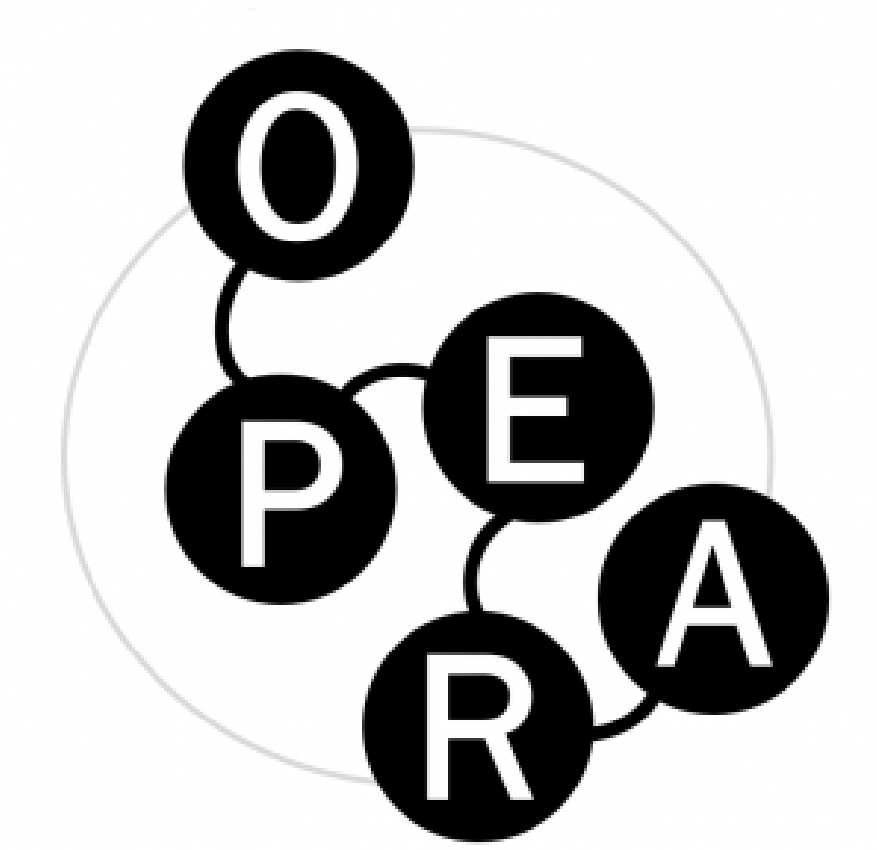
Quantitative Indicators for Assessing the Societal Impact of Research in SSH



OPERA Project Report no. 1

18 November 2020

Contributor credits

Specified using the CRediT Taxonomy contributor roles:

Conceptualization
Ideas; formulation or evolution of overarching research goals and aims

Methodology
Development or design of methodology; creation of models

Software
Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components

Validation
Verification, whether as a part of the activity or separate, of the overall replication/ reproducibility of results/experiments and other research outputs

Formal Analysis
Application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study

Investigation
Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection

Resources
Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools

Data Curation
Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later reuse

Writing – Original Draft
Preparation, creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation)

Writing – Review & Editing
Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision – including pre- or postpublication stages

Visualization
Preparation, creation and/or presentation of the published work, specifically visualization/ data presentation

Supervision
Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team

Project Administration
Management and coordination responsibility for the research activity planning and execution

Funding Acquisition
Acquisition of the financial support for the project leading to this publication.

AUTHORS

David Budtz Pedersen – ORCID ID 0000-0001-7861-7068 – email: davidp@hum.aau.dk
ROLES: Conceptualization, Methodology, Investigation, Writing Original Draft, Writing, Review & Editing, Project Administratio
AFFILIATION: Aalborg University

Rolf Hvidtfeldt – ORCID ID 0000-0002-0905-0093 – email: rolfh@hum.aau.dk
ROLES: Writing Original Draft, Resources. Writing, Review & Editing
AFFILIATION: Aalborg University

TABLE OF CONTENTS

1. INTRODUCTION

The DEFF OPERA "Also humanities" Work Package addresses different dimensions of impact assessment and research analytics in the humanities. With impact becoming the driving force for assessing the relevance and uptake of research in society, the Social Sciences and Humanities are in a unique position to contribute to and shape the concept of impact. The deliverables of this work package explore and review a selection of methodologies and indicators for impact assessment in the humanities and beyond. While there are many challenges related to applying indicators and analytics from the STEM disciplines to the SSH field, there is considerable space for re-thinking and experimenting with new metrics for publications, citations, co-authorship, co-production as and wider societal impact.

In this review, we examine some of the most dominant indicators for societal impact of research, focusing on the social sciences and humanities (SSH). Based on a larger scoping review of the literature, we focus this working paper on the examination of quantitative research indicators (Budtz Pedersen et al. 2020). We first present a number of taxonomies for impact assessment drawing on international frameworks and methodologies. Following this examination, we examine advantages and disadvantages of the selected methods inspired by a SWOT analysis (i.e. review of strengths, weaknesses, opportunities, and threats).

1.1 SELECTION METHOD

The indicators presented in this review are selected on the basis of their influence and application in real-world evaluative contexts in the European Union and member states. We have excluded indicators and assessment metrics with a purely qualitative aim, and have focused on quantitative indicators that may be of relevance to libraries and research support units that are seeking inspiration and templates for further working on data collection and data analysis of research impact activities at the institutional level. The following criteria have served as the basis of selection:

- the indicators have been tested and refined in real-world institutional settings
- the indicators are relevant for international comparison of national research systems
- the indicators are useful for impact assessment within funding agencies and universities.

The result of this selection is the following six indicators:

- UK Research Excellence Framework (REF)

- The Danish National Bibliometric Research Indicator
- SIAMPI framework to track "productive interactions"
- Researchfish to track research and evidence impact.
- Altmetrics to track scholarly content happen online
- Responsible Impact assessment (ReAct) to track knowledge exchange

2. INDICATORS

The research impact frameworks presented in this review utilise different methodological strategies for assessing the impact of research. The methods range from bibliometric metrics and analyses, through statistical databases and commercialization statistics, open data repositories and impact templates to different quantitative and qualitative methods such as surveys, interviews, workshops, and focus groups.

Table 1. illustrates how different methodological components are distributed across the impact assessment models, described in this paper.

	REF	**SIAMPI**	**Altmetrics**	**BFI**	**Researchfish**	**ReAct**
Ex ante		■				
Ex post	■	■	■	■	■	■
Surveys		■				
Statistical databases	■	■				■
Commercialization statistics	■	■				
Bibliometrics	■	■	■	■	■	
Logic models					■	
Impact repositories	■					■
Altmetrics	■		■			
Impact-tracking (from process)	■	■			■	
Impact-tracking (from outcomes)	■				■	
Document analysis	■	■		■		

Table 1

2.1. REVIEW OF APPLIED METHODS

Due to the complex, diverse and context-dependent nature of research impact, different assessment models often incorporate diverse and flexible methods to describe research impact and its pathways. Across frameworks and platforms, it is generally recognised that quantitative methods (e.g. citation analysis and commercialisation data) and qualitative methods (such as case studies, interviews, and field visits) are best used in combination and supplement each other (JRC 2019). In many cases, methodological strategies need to be tailored to the specific "mission" or "goals" of the research programme, department or funding instrument under assessment, and that these goals should be reflected in the way research is organised, funded and evaluated. For these reasons the international research evaluation community has issued a call for "responsible metrics" i.e. metrics that are used at the right level of assessment (e.g. research groups, units, departments, programmes rather than individual performance), the right timeframe (acknowledging that research impact takes time and have dynamics effects), and that funding decisions, allocation models and academic merit and promotion should not be based on one-dimensional or generic standards (Rafoels et al., Wilsdon et al., Hicks et al). What counts as evidence for impact in one field or discipline does not necessarily count as impact in other fields. Focusing excessively on one-dimensional indicators runs the risk of perverting incentives structures and reinforcing path-dependencies and hierarchies among disciplines. Research impact needs to be evaluated in context and assessed by using indicators and analytical frameworks that acknowledge differences among academic traditions, societal contexts (readiness) and timeframes.

Ex ante and ex post methods

The methods presented in this review have been used to assess and capture impact at the planning stage before research is initiated (*ex-ante*) as well as after research have been carried out (*ex*-post). Tools for assessing the likelihood of future research impact are most often associated with impact planning, impact toolkits, and peer review of grant applications. It has become commonplace for many leading funding agencies to include obligatory descriptions of impact pathways in grant applications in order to assess the likely impact of the suggested research. Such ex ante indicators might include data about prior impact activities, mention of research products or performances that have proven to be helpful in accelerating impact, or specific methods for knowledge exchange, knowledge mobilization, knowledge brokerage or knowledge transfer. Such ex ante indicators are used by the European Commission in calls for funding under the Challenge-pillar of "Horizon 2020", and will be used as part of the mission-

driven approach to R&D that will lay the foundation for next European framework programme for research and innovation "Horizon Europe" (2021-2027). Similar methodologies are used by national research councils and private research foundations, such as the British Research Councils, the Canadian Research Councils, the Australian Research Councils, Novo Nordic Foundation and Innovation Fund Denmark among others.

Among the frameworks reviewed in the next section, the SIAMPI model utilises ex-ante indicators. Such indicators can be used to develop impact plans and designing "theories of change" as well as to clarify the mission, identify potential partners, or develop strategies for broader dissemination and user-involvements.

In addition to ex-ante screening of research impact, most quantitative indicators focus on *ex-post* assessment. Such frameworks pay attention to the different benefits obtained throughout and after the research process have ended – typical in conjunction with project reporting, assessment of research units, evaluation of university departments, research portfolios or national research performance. In the following, we are focusing on ex-post indicators, ranging from surveys to databases and commercialisation, bibliometric performance and impact-tracking.

Surveys

Surveys are useful for collecting quantitative data on several background variables, such as motivations, perceived barriers and enablers, and engagements between research and the wider society. An advantage of this method is that surveys allow convenient comparative analysis of performance over time and across different domains of application.

As a method for impact assessment surveys have significant limitations. One limitation is that surveys assume that research impact can be measured quantitatively and can, to some degree, be captured by means of a standardized questionnaire. Since surveys primarily provide self-reported evidence of impact there is an in-build bias towards reporting involvement and activities of directly engaged stakeholders. Indirect benefits that arise from dissemination efforts, relationships and other interactions e.g. with policy-makers or civil society, are not captured by survey techniques.

Rather than accounting for demonstrable, observable impact (i.e. a change or effect of research outside academia) surveys are useful for *indicating* some level of *likelihood* of impact. Consequently, surveys often require methodological supplements to validate self-reported evidence e.g. qualitative interviews, focus groups, or workshops. Examples of this approach can be found in SIAMPI. Lastly, surveys are not very responsive to unforeseen impacts and context-specific factors (Boaz, Fitzpatrick, & Shaw 2008, Boaz, Fitzpatrick, & Shaw 2009).

Statistical databases

Statistical databases are useful for quantitative descriptions of e.g. research infrastructures, facilities, income, scientific degrees and awards, and are well-designed to provide data and information across scientific fields. It is possible to monitor developments over time and compare performance between research units, for example by combining data across administrative and statistical records. Thus, statistical databases are often considered useful for national evaluation systems where it is central to account for the relationship between distinct disciplines, the academic environment, and the academic and societal outputs and contributions (REF, 2014; Reale et al., 2014; VSNU, NWO, & KNAW, 2015). Statistical databases may include national statistical databases, population statistics or specialized comprehensive scientific databases such as the Danish National Bibliometric Indicator or similar indicators in other countries.

Within its scope and context of application statistical databases have several methodological advantages. It can shed light funding streams, publication patterns, mobility of researchers, income, and other population variables that may indicate impact-oriented activities at the aggregate level. However the method also has some disadvantages. Based on statistical search results it may prove difficult to document and derive impact from specific projects, programs or departments, especially informal engagement efforts (Abreu & Grinevich, 2013; Bekkers & Bodas Freitas, 2008; Martin & Tang, 2006; Olmos-Peñuela, Molas-Gallart, & Castro-Martínez, 2014). Further, existing archives are often incomplete and require ongoing augmentation, maintenance, documentation, and validation. On top of this, administration and user rights may hinder access to data and entail regular renewal of agreements about utilization of existing datasets. A related problem is, that the statistical databases are frequently used without paying sufficient attention to the reflexive problems mentioned above. This may lead to naïve assumptions of objectivity, which may hinder adequate analyses.

Commercialisation statistics

Commercialisation statistics are most often associated with measuring different types of economic effects, primarily in relation to the business sector and technology transfer. Standard indicators are patents, licences, joint R&D, contract research, industry funding and number of spin-out companies. Commercialization statistics can be used to study various types of entrepreneurial activities (Abreu & Grinevich, 2013; D'Este, Tang, Mahdi, Neely, & Sánchez-Barrioluengo, 2013; Martinelli et al., 2008; Perkmann et al., 2013). Advantages related to commercialisation statistics are the ability to identify formal and contractual relations between universities and societal actors, where quantitative evidence of economic and commercial uptake can be accounted for and compared. However, it is often hard to compare different types of commercial effects especially across different disciplines and national contexts. Further, outputs

from SSH may have significant effects (e.g. through education or optimisation of societal structures) which are not captured by the above measures

Bibliometric indicators

Bibliometric indicators are widely used for assessing academic output as in the case of the Danish National Bibliometric Indicator (BFI). The BFI indicator provides an overview of research publications from Danish universities. The system is a central element of the performance-based model for redistribution of block grants for universities. The indicator allocates funding based on the distribution of publications that are based on original research, peer-reviewed and published in a channel included in the BFI lists. More widely, bibliometrics are used at different levels and different institutions across the world as a statistical method to analyse books, articles and other publications. Bibliometric methods are frequently used in the field of library and information science. The sub-field of bibliometrics which concerns itself with the analysis of scientific publications is called scientometrics. Citation analysis is a commonly used bibliometric method which is based on constructing the citation graph, a network or graph representation of the citations between documents. Many research fields use bibliometric methods to explore the impact of their field, the impact of a set of researchers, the impact of a particular paper, or to identify particularly impactful papers within a specific field of research.

Bibliometrics have several limitations. As a methodology it is unfit to account for the broader communication, exchange and engagement of research in society. Bibliometrics represent a widely accepted statistical method for assessing and ranking *scientific performances*, mainly on the basis of authors or publication outlets (journals, publishers) as the unit of analysis. The H-index is an example of a bibliometric indicator that combines individual productivity with citations covered in systems such as ISI WOK and Scopus. The index is calculated by ordering the number of publications by a single researcher on the basis of the total number of citations they have received. Citations analysis such as H-index allows for studies on whether research is being pursued at the highest level of academic impact i.e. is cited by other researchers in the field (LSE Public Policy Group, 2011; Wilsdon et al., 2015). Bibliometric analyses have received criticism by the humanities and social science community, among other things, because of poor historical coverage of publications and citations. Bibliometric analysis often requires well-structured publication databases, which for the most part is focused on international journal publication and not on collecting or synthetizing data from publications in national languages or books and book chapters outside the standard realm of academic publishing (e.g. museum catalogues, design principles, educational guidelines etc). While bibliometrics in general may contribute to objective and transparent evaluations the research output and and provide an overview of publications patterns and scientific networks, this approach needs to be balanced with other approaches as well as knowledge about the specific research traditions under assessment including their

publication and citation patterns etc. In some cases, bibliometric methods have been used to track the uptake and use of research-based knowledge (e.g. publications) in real-world societal settings, for instance, by tracking references to scientific publications in official government reports or policy documents.

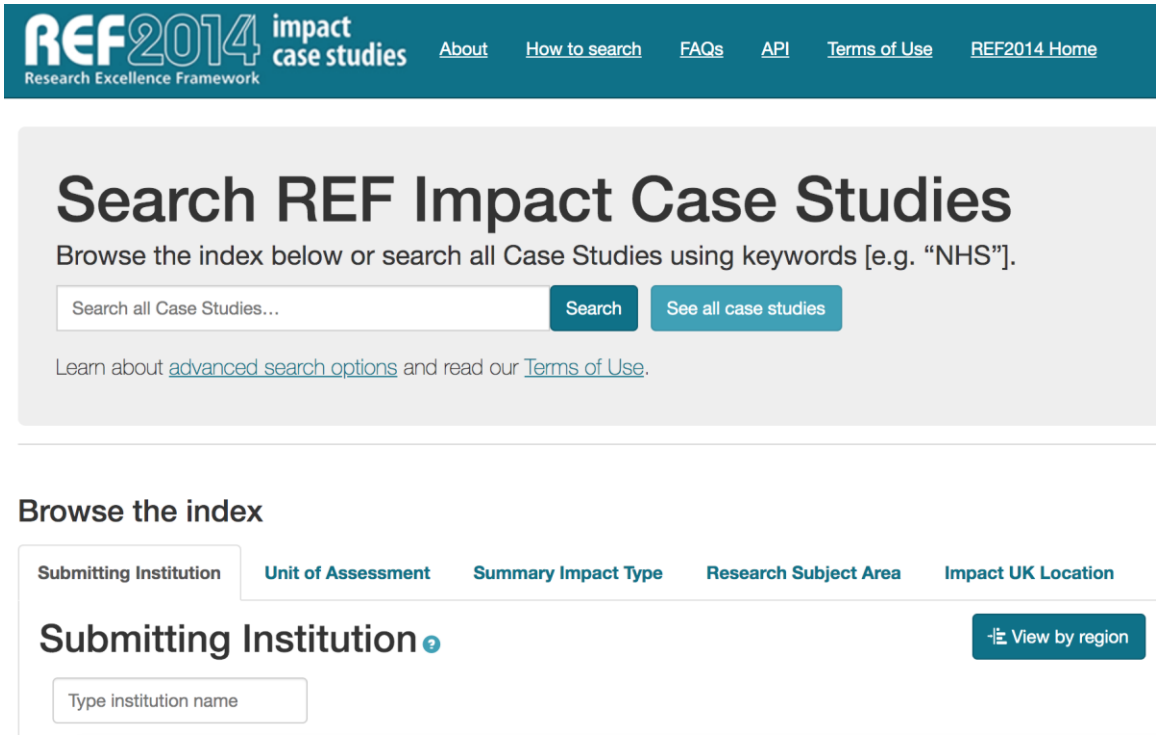
The Logic Model

The so-called logic model aims at capturing the complex processes and interactions through which knowledge is produced, exchanged, disseminated and used in society. A logic model can be used to depict input activities (ex-ante) such as impact plans, templates or case studies. A logic model may form the basis for tracking specific outcomes resulting from research (Morton, 2012; Young et al., 2014) or break down different types of engagements efforts or outcomes associated with research (Donovan & Hanney, 2011). In both cases, sensitivity towards indirect impacts and links between research and society is required.

Logic models and related methodologies for issuing 'impact plans' may be deployed retrospectively to describe the context, activities and outputs of research that have influenced society in various ways. Such plans can also be used prospectively to explain how a specific research project is expected to bring about desired results. Impact plans and templates can, for instance, be used to account for the context of individual case studies and provide additional information about the wider range of activity and its capacity for impact associated to a specific unit of assessment. This allows assessment panels to take particular circumstances into account which may have an influence on the case-studies selected for assessment.

Impact repositories

Impact repositories are used as an integral component in only few impact assessment frameworks. As we will describe in further detail in the next section, only the REF and ReACT platform utilize impact repositories. Such repositories consist of searchable, indexed impact case studies that researchers, research managers, librarians, or funding agencies can search for 'best practices' and for understanding the link between academic (bibliometric) performance and societal impact (King's College London and Digital Science 2015). The REF database contains over 6600 case studies and has an excellent research functionality, which searches within each impact case study (ICS) for results. It is possible to search for ICSs by unit of assessment, or by institution, and to analyse the rating of individual impact case studies (by looking at the grade profile of the impact for the submitting unit, not directly assessable in the database).



Impact databases are a powerful tool for indexed text-mining and in-depth analysis of impact activities. Among their strengths, impact repositories may stimulate wider interest in research and make data about research accessible to external audiences, and may as such provide comprehensive narratives about the pathway from academic performance to societal impact of research. On closer examination, impact repositories may provide insights into the strategies used by different units of assessment to establish links between academic activities and societal impact, e.g. by deploying techniques such as knowledge translation, knowledge exchange, and knowledge mobilization (Knight & Lightowler, 2010; Lightowler & Knight, 2013; Sebba, 2013). The disadvantage of repositories is mainly tied to the fact that they are time-consuming to construct, and that units of assessment (with small numbers of researchers) may be hard to anonymize, leading to breach of privacy rights or unintended managerial use of case studies (e.g. for individual reward or sanction). For these reasons, the ReACT Vivo Impact platform will release only a select number of case studies, that will be openly available for scrutiny and dissemination in an open repository, which will set an example for future attempts to document and describe impacts pathways. Because of ethical considerations in relation to sharing and utilising data, some types of research impact case studies are less appropriate to share openly.

Alternative metrics (Altmetrics)

Different terms have been used to characterise a new generation of metrics and indicators that are able to trace the spread and dissemination of research beyond academia. Prominent among these are new alternative metrics abbreviated “altmetrics”. Originally, this term was used to describe a cluster of impact metrics beyond academic publications and citations. But in recent years the term has received a more technical definition as a tool to trace the uptake and impact of published research on social media and the digital sphere including mentions, likes, downloads, page views etc. that indicate social interest in published academic writings, first and foremost academic journal articles.

In its original meaning, alternative metrics has been developed to capture the ‘broader impact’ of research by tracking communication and sharing of research content through digital platforms such as Twitter, Facebook, blogs, Mendeley, Cite U Like, Altmetric.com, and Impact Story. Various forms of data can be collected from these sources, e.g. citations, views, downloads, clicks, tweets, shares, likes, bookmarks, and comments (Bornmann 2014a, Zahedi, Costas, & Wouters, 2014; Holmberg & Thelwall, 2014; Hammarfelt, 2014). The capacity to harvest big data about the dissemination of research and the effects of scholarly content on social media and conversations in the digital sphere is the main advantage of altmetrics. Altmetrics methods work well as supplements to case studies and narrative approaches. Despite these strengths, altmetrics have several drawbacks. Data is not easily compared between disciplines, or even across research topics. In addition, the data used by altmetrics is often biased towards specific users and it is rarely possible to access precise user-statistics or samples for different types of social media. From a deeper methodological point of view, it is also unclear what a mention or download indicates in reality. In contrast to citation analysis, altmetric data is far from standardized (Bornmann and Daniel 2008; Neylon, Willmers & King, 2014). Consequently, a reflexive and responsible use of altmetrics is recommended (Wilsdon et al., 2015; Hicks, Wouters, Waltman, Rijcke, & Rafols, 2015). There is broad consensus that the sharing and uptake of research do take place in a more comprehensive environment than captured by traditional scientific metrics, and altmetrics may conveniently be used to measure broader societal outputs than traditional bibliometrics (Waltman & Costas, 2014). Indeed, a number of studies have focused on comparing altmetrics to bibliometrics or have focused on disciplinary differences in relation to using digital media and resources such as Mendeley or Twitter (Bornmann 2014a, Zahedi, Costas, & Wouters, 2014; Holmberg & Thelwall, 2014; Hammarfelt, 2014).

Backward-tracking approaches

Backward tracking approaches seek to explore impact pathways by tracing societal changes attributed to research projects or teams *backwards* from tangible outputs. The method is used to trace the return of investment of research based on a range of empirical sources, which makes it possible to explore "how" as well as "why" specific outcomes or impacts have occurred. Such methods allow research projects to be measured up against goals defined by research institutions or funding organizations. The drawbacks of the method are that backwards-tracking is heavily dependent on the quality and availability of relevant data and documentation. Furthermore, it may be difficult to *attribute* specific outcome to specific research efforts – especially when accounting for more indirect pathways and impacts that are influenced by complex networks of contributions, institutions, infrastructures, and societal readiness.

Forward-tracking approaches

Forward tracking approaches are used to identify links and productive interactions that may potentially lead to societal effects (ex ante). The techniques can help establish a connection between 1) research objectives, framing and design, 2) the research processes and outputs and 3) later outcome and impact. It is also able to identify barriers and enablers of achieving research impact. The methods can be based on both qualitative (e.g. interviews, impact logs) and quantitative approaches (e.g. social network analysis, geo-referencing, contextual response analysis etc.). It may be difficult, though, for researchers to systematically describe ways to achieve impact already during, or even at the beginning of a project. Due to the relatively short time horizon of most research projects, many societal effects will not occur during the project lifecycle.

Document analysis

Finally, document analysis covers the review and interpretation of existing documents such as books, policy reports and whitepapers. Review of documents can be done qualitatively as well as quantitatively in combination with computational text analysis (text mining, topic models, semantic text analysis etc.) or traditional coding strategies (categorised coding, thematic syntheses etc.), and can lead to comprehensive insights into both content and the context of impacts and outputs. However, document analyses require considerable effort and is highly dependent on the quality and availability of relevant texts and whether they can be systematically collected. As a consequence of these limitations, document analysis has little to say about the non-textual outputs from research.

This concludes our discussion of the methodological elements represented along the vertical axis in figure (a) above. In the next session, we proceed to describe the six selected frameworks, and their primary focal points.

3. IMPACT FRAMEWORKS

In this section, we provide a description of a number of frameworks that have implemented and tested the above methodological strategies. Common among the frameworks described in this section is their commitment to track and capture the broader impact of research in society, including research originating from the social sciences and humanities. To varies degrees, the examined frameworks have been used to analyse research impact with the purpose of gaining deeper insight into impact making, funding allocation and funding decision-making. The frameworks presented constitute a mix of policy and academic efforts. The SIAMPI and ReACT project is decisively research-driven. REF, Researchfish, Altmetrics and BFI are policy- or commercially driven. Researchfish was originally funded as a project under the UK Medical Research Councils with a deliberate bias towards impact activities originating from the health science disciplines but has later been expanded to include a taxonomy which is more inclusive towards social sciences and humanities.

3.1 SIAMPI

SIAMPI (Social Impact Assessment Methods for research and funding instruments through the study of Productive Interactions) is a research project and methodology that aims to uncover indicators for social impact through a conceptual framework based on the notion of "productive interactions" (Molas-Gallart and Tang 2011, De Jong et al. 2011, Spaapen and Drooge 2011, De Jong, Barker, Cox, Sveinsdottir, & Van den Besselaar, 2013). The model is built around a non-linear understanding of social impact as a result of dynamic interactions and joint efforts by multiple actors inside and outside the academic community. SIAMPI distinguishes between three different types of productive interactions: (1) Direct interactions (2) Indirect interactions and (3) Financial interactions (SIAMPI, 2011).

Jointly these types of interactions reflect that knowledge can be disseminated and applied in different ways (e.g. research publication, political report, guideline, website, membership of a committee or through meetings or by financial contributions). The distinction between direct interactions, indirect interactions and financial interactions serve as a useful heuristic for identifying different types of impact. In practice, most productive interactions involve combinations of various forms of connective interactions. A publication may, for instance, lead to

consultancy contracts and further long-term collaboration (de Jong, Barker, Cox, Sveinsdottir, & Van den Besselaar, 2013; Molas-Gallart & Tang, 2011; Spaapen & van Drooge, 2011).

As an evaluation tool, SIAMPI aims to synthesize multiple data sources and provide an overall assessment of the impact of research. It is a strength of the approach that data and enriched metadata on the context of research are included in analyses, which may help establish a connection between objectives, research designs, processes and outputs. Conversely, it may be difficult for researchers to describe ways to achieve impact already during or at the end of a project. In addition, very detailed impact strategies may be difficult to define and compare to corresponding documentation from other research programs. Table 2 highlights the most important elements in the model.

Direct interactions **(Personal interaction between stakeholders)**	**Indirect interactions** **(Interaction between stakeholders through media or 'carriers')**	**Financial interactions** **(Material interaction between stakeholders)**
Face-to-face interactions	Articles	Research contracts
Phone	Books	Facility, instruments, sharing
E-mail	Annual plans	Start ups
Video conferencing	Reports	Contribution 'in kind' (people)
Radio, tv or internet	Web pages	IPR arrangements
Mobility arrangements	Clinical guidelines	Project grants
Meetings	Designs	Lump sum grants
Conferences	Models	Professional training
Chance encounters,	Musical arrangements	Patents
Old friendships	Exhibitions	Licenses
…	…	…
Social impact		
Behavioural change	Uptake, use	Collaboration

Methods		
Interviews, focus groups	Quantitative data	Reports or similar documentation

Table 2

SIAMPI is funded through a research-led project consortium under the European Union Seventh Framework Programme (FP7/2007-2013), and further developed by the Rathenau Institute in the Netherlands and the pan-European League of Research Universities in 2017 (SIAMPI 2011, LERU, 2017). See more: http://www.siampi.eu/

3.2 RESEARCH EXCELLENCE FRAMEWORK (REF)

The Research Excellence Framework is a national evaluation system which aims to assess the impact and quality of research carried out at British universities. REF is based on extended peer review. Assessments are conducted by teams of academics and experts who are assigned to rank research from organizations other than their own. Assessment is carried out within 36 subject-based Units of Assessment (UOA), such as “Clinical medicine”, “Law”, “Chemistry”, and “Philosophy” to name a few. The Research is evaluated along three different dimensions: Research output, impact, and research environment, which are weighted in the following way:

> **Research output** accounted for 65% of the overall score in REF 2014 (will be adjusted to 60% in the upcoming REF 2021). Research output is defined as products of any form. This includes traditional publications such as scientific articles, monographs and book chapters, but also more unconventional outputs such as designs, performances and exhibitions. As a basis for evaluation, all research institutions have submitted up to four research outputs for each employee included in the evaluation. These outputs are evaluated based on criteria of originality, influence and stringency.
>
> **Research impact** accounted for 20% of the overall score in REF 2014 (will be adjusted to 25% in the upcoming REF 2021). For the purposes of the REF, impact is defined as “an effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia.” The impact assessment is based on case studies and so-called "impact templates". Case studies briefly describe the impact activities and impacts that have taken place in relation to specific research projects. These impacts are evaluated based on criteria that deal with scope and significance. In addition, a template explains how the research unit has strategically worked to create impact through research as well as the unit's strategy.

Research environment accounts for 15% of the overall score, and is defined as strategies, resources and infrastructure that support activities in the research unit and contribute more widely to the research discipline. The assessment is based on submitted templates and statistical data. The templates describe the research strategy, support for researchers and students, revenue for research, infrastructure and facilities, research collaboration and broader contributions to the discipline. Institutions also provide data on the amount of research income and number of scientific degrees and prizes. In the end, the research community is assessed based on its overall vitality and sustainability (REF 2012, 2014, 2015).

REF is performed every five years in the UK and covers a five-year performance account. It is not permitted to submit material or examples of impact that precedes the ongoing evaluation interval, and consequently the model has a quite limited time horizon. REF has been criticized for being time- and resource-consuming (Greenhalgh, Raftery, Hanney, & Glover, 2016; Martin, 2011). Further, the model has been criticized for cultivating fierce competition among British universities since universities may benefit from recruiting successful researchers right up to the submission of evaluations. Another weakness relates to the very open-ended impact templates and the lack of standardized ways evidence is collected and assessed (Donovan, 2011). This may create bias towards projects with certain types of more easily measured impacts with more sudden and extraordinary effects as opposed to effects that build up slowly or accumulate in less noticeable ways.

The strength of the framework is that it provides a comprehensive, nuanced, quantitative and qualitatively enriched image of the academic and societal impact of research (ex post). With the REF, British decision makers and authorities have the opportunity to orientate themselves to a wide range of indicators and parameters in scientific research. REF has led to a general change of behaviour at UK universities where merit of impact-oriented research today features more prominently than in comparable European countries. Another strength of the model is that the collected data on publications and impact can be used for a wide range of analyses that may help to shed light on roles and values in research (King's College London and Digital Science, 2015) and, thereby, contribute to creating a higher community commitment in the different phases of the research process. See more: https://re.ukri.org/research/research-excellence-framework-ref/

3.3 RESEARCHFISH

Researchfish is a tool for data-collection and monitoring of scholarly activities and progress and is mainly used to monitor and assess funding of individual research projects. Researchfish enables research funders and administrators to track the progress and output from researchers in real time, by having researchers record data about their projects. In practice Researchfish serves as a more advanced (and in principle) less burdensome reporting platform for grant holders. Instead of submitting interim and final progress reports, the tool collects data throughout the project lifecycle based on self-reporting and validation by the funding agency. Researchfish was originally designed to document the effects of health science and pharmaceutical research, but it is currently used in a wide range of other disciplines, including the natural, social, and human sciences.

Researchfish has been implemented by more than 150 organisations, which include research foundations such as the Novo Nordisk Foundation, the Bill & Melinda Gates Foundation, The Velux Foundations, and Wellcome Trust as well as the Oxford University, University of Cambridge, and Alberta University to name a few. More than 100.000 individuals are involved in the use of Researchfish for measuring and reporting impact.

From the perspective of funders, enriched metadata provides the opportunity to perform ongoing evaluation of the research conducted by grant recipients and assess whether performance goals are met. This allows foundations to refocus their strategy, portfolio and instruments. As an example, the Novo Nordisk Foundation states that their research should contribute positively to education, life quality, improved treatment and have high micro- and macroeconomic impact in society. The primary strength of Researchfish is, thus, that one can follow and assess projects as they unfold and, on this basis, engage in continuous dialogue about objectives, results and follow-up. Another strength of the instrument is the data that accumulates over time. It is rarely possible to determine the final value creation or value realization of specific research projects in the short term. But by collecting and analysing enriched metadata in Researchfish, it is possible to follow appropriations and their different outputs and imprints over a sustained period of time – even after the projects have been completed.

One possible difficulty in using Researchfish is that research councils and funds, which may initially have widely differing objectives, through the use of standardized data begin to compare their performances. Such a situation can create inappropriate competition and minimize pluralism and heterogeneity in the overall research funding ecosystem. Similarly, a weakness of the model is that options for altering categories to suit special needs and performing alternative analyses are limited. See more: https://researchfish.com/

3.4 BIBLIOMETRIC RESEARCH INDICATORS

Bibliometrics is the use of statistical methods for analysis of scientific publications and patterns of citations. Bibliometrics are widely used as a tool for assessing the academic impact of research and constitute, through the reliance on integrated, linked data, a straightforward and widely accepted way of assessing and ranking scientific performances, mainly based on authors as unit of analysis (Wouters & Costas, 2012).

Traditionally, bibliometrics has been operationalised by policy-makers and university management systems to account for performance, and several measures have been developed to capture performance in a single number. One such example is the *Hirsch*-index (*h*-index) in which *h* is defined as the number of papers with citation number ≥*h*. Originally, the *h*-index was designed to measure the relative quality of a scientist (Hirsch 2005), but comparison between disciplines shows considerable problems which extend into bibliometrics in general. Naturally, disciplines with fewer researchers will publish less and therefore cite each other less often. This is also affected by traditions regarding types of publication and citation patterns where a short paper, for example in physics, can result in a higher number of citations. Or similar, that a long and critically acclaimed book-length monograph in the humanities can result in only few citations. To counter this problem, field normalization has been explored as a pathway to better and more comparable data.

Nevertheless, bibliometrics is central to a number of national evaluation and funding systems. In Denmark, for instance, the Danish Bibliometric Research Indicator (BFI) is used to allocate funding to research institutions based on the production of peer-reviewed publications. Bibliometric indicators have several advantages. Citations analysis such as *h*-index allow for studies on whether research is being pursued at the highest level and cited by other researchers in the field (Wilsdon et al., 2015). Evaluation system such as the BFI may contribute to the objectivity and transparency of the research evaluation process and provide an overview of publications patterns and research networks that cannot be seen from the perspective of the individual researcher. This may provide useful information on how specific research fields are connected, growing or declining over time (KNAW, 2005). Bibliometrics are generally considered to be a useful method in the assessment of the quality of research, when used carefully.

However, it is important to notice that bibliometric indicators only cover part of the written communication between researchers, and that broader communication and engagement of researchers in society is not accounted for by standard bibliometrics (Bornmann, 2014). In addition, it is worth emphasizing that an exclusive use of bibliometric indicators may have negative performative effects on an evaluation system as a whole. Performing well on

bibliometric scales is not the same as producing research that is useful, powerful and relevant in a broader societal context. Moreover, several studies discuss how researchers are not passive recipients of research assessment systems. Narrow indicators and incentives may lead to strategic behaviours and gaming of research e.g. by goal displacement or the establishment of citation cartels. This is a potential consequence if the production of scientific outputs becomes more or less an end in itself (Dahler-Larsen, 2012, 2014). The influence of other methods and indicators is growing, however, based on the acknowledgement that bibliometrics do not adequately capture the full impact of research. See more: https://ufm.dk/forskning-og-innovation/statistik-og-analyser/den-bibliometriske-forskningsindikator

3.5 ALTMETRICS

The term "altmetrics" has been captured in the company name Altmetric, which is part of the Digital Science portfolio of companies. The company and their underlying metrics are committed to help researchers and funding agencies to get a better understanding of the "reach" of scholarly research outputs on digital platforms, for instance by analysing and documenting the attention that articles are receiving in real-time, and what other researchers and media participants are saying about the work. By using altmetrics, publishers are able to show authors and readers conversations surrounding their content; researchers are able to see which research their peers think is interesting, and funders are able to understand how the work they have funded is being received amongst a broader audience. Different types of data are collected and monitored by Altmetric, ranging from citations, views, downloads, clicks, tweets, shares, likes, bookmarks, and comments. The main advantage of Altmetric is the ability to harvest big data that makes it possible to quantify wider dissemination efforts and effects. In an evaluation context, such methods can work well as supplements to case-studies and narrative approaches. Further, the effects of media presence can be measured in a very short timeframe (Mohammadi & Thelwall, 2014).

However, Altmetric data is not easily compared between disciplines or even across themes or topics (the problems of field normalisation). Additionally, it is often not possible to access exact user statistics or samples for different types of social media data. Furthermore, it is unclear what an online citation, mention or download actually indicates. In contrast to citations in scientific journal articles, which are fairly standardised and transparent (though by no means perfectly so, of course), citations and references in policy reports or on social media are far less clear. Indeed, research used in such contexts is not always cited and not everything cited is actually used (Bornmann and Daniel 2008; Neylon, Willmers & King, 2014). Consequently, there is still a lack of knowledge concerning the reliability, validity, and context of specific metrics and altmetrics. It is recommended that Altmetric is used carefully and not as the central method for assessing societal

impact of research (Wilsdon et al., 2015; Hicks, Wouters, Waltman, Rijcke, & Rafols, 2015). However, increased attention towards wider communication efforts is critical when studying broader dissemination effort – including the uptake and circulation of findings and scholarly output from the humanities and social sciences. See more: https://www.altmetric.com/

2.6 REACT (RESPONSIBLE IMPACT)

The research project "Responsible Impact" (2016-2020), hosted by Aaborg University, is an interdisciplinary, explorative study aiming at developing new metrics and impact indicators for the humanities and social sciences. The project is based on a data-driven and participatory approach to research assessment – mining impact activities in partnership with 45 practising researchers from the humanities and social sciences at University of Aalborg in Denmark. For this purpose, the project has developed a new taxonomy of academic outputs and activities beyond journal articles and monographs, including non-traditional publishing formats such as reports, white papers, blogs, reviews, data, software, audio-visual output in addition to in-person activities and appointments etc. The ReAct taxonomy serves as the background for designing Responsible Impact Assessments by allowing researchers and universities to have significant influence on how their research outputs are represented and communicated.

ReAct is driven by the ambition to develop new approaches to assessing impact and evaluating knowledge exchange among a large web of actors and institutions. The ReAct team has built a sandpit model of the taxonomy into VIVO, which serves as the main platform for data collection and validation. Beyond recording real-time research activities on the VIVO platform, the project is experimenting with novel ways of representing and visualising outcomes, e.g. by deploying techniques from infographics and graph databases. The resulting aggregated data are featured in five complementary case studies, enriched by qualitative interviews with faculty members and further augmented by bibliometric records and profiles. Together, VIVO impact data, bibliometric profiles and case studies creates a portfolio for the involved research groups that can create fertile ground for career development, profiling research, and provide visibility and information vis-à-vis external stakeholders.

ReAct aims to develop a model of impact assessment, which is based on the number of relations, networks and connectivity with actors outside and beyond academia. By explicitly drawing on a *responsible* approach to research assessment, ReAct takes into account the diversity of impacts by creating a participatory taxonomy that allows researchers to create their own impact typology and redefine existing categories. By using the notion of "productive interactions" as the baseline for tracing the impact of OS activities, it is possible to explore a wide range of collaborative activities, such as policy interactions, citizen interactions, industry interactions and media

interactions (Spaapen & Drooge 2011; Budtz Pedersen et al. 2020). In effect, the ReAct does not lay out a universal framework for assessing SSH activities, but provides input on how to manage impact assessments in diverse interdisciplinary and institutional settings.

The main advantages of ReAct are 1) the very comprehensive basic taxonomy, which means that pathways to impact are captured with unparalleled adequacy, 2) the real-time recording of interactions and activities, which makes it possible to study researcher interactions as they unfold and ameliorates the bias towards what has already turned out to be successful, and 3) the possibility of distinguishing between different "impact profiles", and thereby document how researchers contribute in different ways to societal effects without assuming that some types of contributions (e.g. peer reviewed publications) are more important than others (such as informal advice for policy makers). The main drawbacks are 1) the existing VIVO is merely a sandpit model, and, thus, far from a plug-and-play software package. 2) ReAct has a strong focus on data regarding interactions and contributions, which is validated by case studies, but not by further external databases about societal end-effects, and 3) collecting ReAct data requires dedicated resources. See more: react.aau.dk

4. CONCLUSION

In this report we have shown that there exist a number of relevant quantitative indicators for studying and capturing the societal impact of research, focusing on humanities and social sciences. While it is unlikely that any of the presented frameworks will capture all impacts, or pay homage to the diversity of different pathways to impact, they each have their strength and weaknesses. A key finding of the report is that different frameworks and tools tend to focus on different aspects of impact-making, from the initial research design to co-creation and planning of research to interim and final results and observable real-world change. Depending on the approach, the cognitive interest and the funding system, each set of indicators can be used to garner new knowledge about the value of research in society and more particular about the activities leading to impact.

5. BIBLIOGRAPHY

Boaz, A., Fitzpatrick, S., and Shaw, B. (2008) *Assessing the Impact of Research on Policy: A Review of the Literature for a Project on Bridging Research and Policy through Outcome Evaluation*. London: Kings College London & Policy Studies Institute

Boaz, A., Fitzpatrick, S., and Shaw, B. (2009) 'Assessing the Impact of Research on Policy: A Literature Review', *Science and Public Policy*, 36: 255–70

Bornmann, L., & Daniel, H. D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, 64(1), 45–80.

Bornmann, L. (2014). Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics. *Journal of Informetrics*, 8(4), 1–24.

Dahler-Larsen, P. (2012). *The evaluation society*, Stanford CA: Stanford Business Books.

Dahler-Larsen, P. (2014). Constitutive Effects of Performance Indicators. Public Management Review, 16(7), 969- 986.

De Jong, S. P. L., van Arensbergen, P., Daemen, F., van der Meulen, B., & van den Besselaar, P. (2011). Evaluation of research in context: an approach and two cases. Research Evaluation, 20(1), 61–72.

De Jong, S., Barker, K., Cox, D., Sveinsdottir, T., & Van den Besselaar, P. (2013). Understanding societal impact through studying productive interactions. Den Haag: Rathenau Instituut. SIAMPI. (2011).

Donovan, C. (2011). State of the art in assessing research impact: introduction to a special issue. Research Evaluation, 20(3), 175–179.

Greenhalgh, T., Raftery, J., Hanney, S., & Glover, M. (2016). Research impact: a narrative review. BMC Medicine, 14(1), 78.

Hammarfelt, B. (2014). Using altmetrics for assessing research impact in the humanities. Scientometrics, 1419– 1430.

Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). Bibliometrics: the Leiden Manifesto for research metrics. *Nature*, 520(7548), 429-431.

Hirsch J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(46), 16569–16572. doi:10.1073/pnas.0507655102

Holmberg, K., & Thelwall, M. (2014). Disciplinary differences in Twitter scholarly communication. Scientometrics, 1–16.

King's College London and Digital Science (2015) *The Nature, Scale and Beneficiaries of Research Impact: An Initial Analysis of Research Excellence Framework (REF) 2014 Impact Case Studies*. London: Higher Education Funding Council of England (HEFCE).

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

LERU. (2013). The future of the Social Sciences and Humanities in Europe: collected LERU papers on the SSH research agenda. EU: League of european research universities, LERU.

Martin, B. R. (2011). The Research Excellence Framework and the "impact agenda": are we creating a Frankenstein monster? Research Evaluation, 20(3), 247–254.

Mohammadi, E., & Thelwall, M. (2014). Mendeley Readership Altmetrics for the Social Sciences and Humanities: Research Evaluation and Knowledge Flows. International Review of Research in Open and Distance Learning, 14(4), 90–103.

Molas-Gallart, J., & Tang, P. (2011). Tracing “productive interactions” to identify social impacts: an example from the social sciences. Research Evaluation, 20(3), 219–226.
Neylon, C., Willmers, M., & King, T. (2014). Rethinking Impact: Applying Altmetrics to Southern African Research. Ottawa, Canada: International Development Research Centre.
Pedersen, D.B., Grønvad, J.F. & Hvidfeldt, R. (2020). ”Methods for mapping the impact of social sciences and humanities—A literature review”, *Research Evaluation* 29(1): 4–21 (doi: 10.1093/reseval/rvz033)
REF (2012). Panel criteria and working methods. UK: REF2014.
REF (2014). Excellence Framework 2014: The results. UK: REF2014.
REF (2015). Research Excellence Framework 2014: Overview report by Main Panel C and Sub-panels 16 to 26. UK: REF2014, 1–123.
SIAMPI final report: Social Impact Assessment Methods for research and funding instruments through the study of Productive Interactions between science and society. SIAMPI.
Spaapen, J., & van Drooge, L. (2011). Introducing “productive interactions” in social impact assessment. Research Evaluation, 20(3), 211–218.
Waltman, L., & Costas, R. (2014). F1000 Recommendations as a potential new data source for research evaluation: A comparison with citations. Journal of the Association for Information Science and Technology, 65(3), 433–445.
Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Johnson, B. (2015). Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management. UK: Higher Education Funding Council for England, HEFCE.
Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Johnson, B. (2015). Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management. UK: Higher Education Funding Council for England, HEFCE.
Wouters, P., & Costas, R. (2012). Users, narcissism and control – tracking the impact of scholarly publications in the 21 st century. SURFfoundation.
Wouters, P., Thelwall, M., Kousha, K., Waltman, L., Rijcke, S., Rushforth, A., Franssen, T. (2015). The Metric Tide Literature Review Supplementary Report I to the Independent Review of the Role of Metrics in Research Assessment and Management. HEFCE.
Zahedi, Z., Costas, R., & Wouters, P. (2014). How well developed are altmetrics? A cross disciplinary analysis of the presence of “alternative metrics” in scientific publications. *Scientometrics*, 1491–1513.

ACKNOWLEDGEMENTS

The authors would like to thank the Danish Ministry of Higher Education and Science for funding to the OPERA project on OPEn Research Analytics, and the project leaders Karen Sofie Hytteballe Ibanez and Mogens Sandfær (Danish Technical University) for excellent management. Parts of the paper has been presented at international conferences and workshops. We thank colleagues for comments on earlier drafts of the paper. For more information about the Humanomics Research Centre see: www.mapping-humanities.dk