Guidelines for evidence-led evaluation of research impact in the SSH

Rationale

Research evaluation is ultimately about improving the quality of research. Policy-makers want to get the best out of their investments in science. Scientists want to do good research that makes a contribution. Research evaluation works by providing signals to scientists of what is "good research". This creates a coordinating effect amongst scientists steering them towards shared ends. Good research evaluation works at two levels: it helps those evaluated to do better science by giving feedback on what was good or bad; and provides signals to all scientists about expectations of what is "good research".

Because research evaluation has this system signalling effect, it is vital that those signals do reflect what are "good" outcomes; bad evaluation practices may encourage scientists towards bad scientific behaviour.

Science covers a very diverse set of fields and disciplines each with their own practices. These different disciplinary communities may react very differently to the same steer from policy-makers. There has been a problem in the last twenty years that the most common research evaluation practices have often been based on an implicit version of "science" reflecting science, technology engineering and maths (the "hard" STEM disciplines). SSH disciplines have particularly suffered from these assumptions, and good evaluation practice allows the reflection of diversity:

Excellent STEM may appear as short papers in English language journals produced by massive multi-author teams. But excellent SSH research also appears in single-authored monographs, years in the making, written in national languages in diverse publishers. Since the late 2000s, research funders started explicitly demanding that research creates "impact" outside the academy, with growing emphasis on evaluating that research impact.

Just as research evaluation has been dominated by STEM appropriate models, so impact evaluation has been dominated by the model that research creates impact via economic transactions.

This runs the risk of sending bad steering signals to the SSH academic community; but the problem is wider than facing just SSH alone, as there are many disciplines and researchers that create impact in a range of other ways. Good impact evaluation should therefore seek to provide a steer to researchers to do more of those activities that lead to impact, to steer the science system as a whole to creating more impact.

Good impact evaluation should match with how scientists themselves define "good impact", reinforcing it as a widely accepted norm for what good scientists do.

This process of scientists and evaluators together developing and internalising norms for more impactful research has not been widespread to date. Impact evaluation remains largely superficial, driven by the search for "extraordinary" impact, neglecting the many everyday ways in which engaged academics lay the foundations for subsequent impact.

One of the main reasons for this is that there are no clear guidelines for policy-makers that explain how evaluation works.

This short policy paper seeks to address that gap by providing a precis of work undertaken by the European Network for Research Evaluation in the Social Sciences and Humanities (ENRESSH). From 2016-20, this network has worked to bring together researchers from across Europe to create a shared knowledge base for evaluating SSH research. This note draws on a wide range of that research and for the sake of clarity does not give direct citations to project work. More information on ENRESSH sic outputs and sources is available at the website www.enreshh.eu.

How can evaluation steer research to create impact?

When policy-makers are using research evaluation, they are generally seeking to improve the overall performance of their systems. They are allocated public funds, and they have to demonstrate that those public funds are being well spent. Science policy-makers therefore use impact evaluation to demonstrate that the science funds are creating benefits for the whole state – the evaluation seeks to improve the overall performance of the system.

Research evaluation relies on the nature of science of being a highly connected and interactive field – scientists continually exchange information, make judgements upon each other's work, and science advances by achieving consensus around what is "good". "Good" things flourish, and bad things are allowed to wither. Evaluation is baked into science to improve performance, distinguishing what is good and what is not, and encouraging more scientists to do good science.

Research evaluation can only evaluate the artefacts that it has in front of it: a research proposal, a project report, a journal article. But what scientists do in evaluating artefacts is build up shared and more generalised understandings of what is "good". Other scientists then plan and shape their scientific activities on the basis of what they believe to be good, shaped by the signals they receive from evaluation. This is intuitive &

prospective – scientists write proposals to be positively evaluated, and therefore they typically try to work out what they have to write in the various sections to achieve that positive evaluation.

Research evaluation provides a set of signals about what constitutes "good" research and this allows other scientists to work out what is good and propose other research artefacts. This is equally true for research impact – impact evaluations provide a set of signals about what constitutes "good" impact.

Research evaluation can only achieve this steering effect when researchers respect and internalise those signals. The evaluation of academic journal articles identifies good papers and eliminate bad papers on the basis of what academics think is good or bad research. Papers are accepted or rejected on the grounds of their content, whether they are rigorous, whether they are logical and consistent. The way that journal articles define "good research" has a shaping effect on the way that scientists regard "good research" and the way that they plan and carry out activities to be doing good research.

This equally applies to the evaluation of research impact; the only way to achieve system level performance improvement is if it changes the way that scientists regard the value of impact. The signals that impact evaluation gives need to fit with what scientists are already doing, to be recognisable to scientists, and ultimately acceptable to scientists.

Good impact evaluation signals to all scientists what good research impact is, but is driven by what is already acceptable to good scientists as research impact creation.

A general framework for evaluating research impact

The first step of evaluation is to create an evaluation "subject" – the thing to be evaluated. Research evaluation may evaluate a publication, and the publication reports in a formalised way a messy set of research activities. The evaluation does not seek to evaluate the report, but rather to evaluate the research that goes into that report (did the questions make sense, was the sampling appropriate, was the statistical analysis correct?). The creation of "research subjects" profoundly influences the research evaluation´s effects: creating the wrong research subject undermines the whole evaluation.

There is a fundamental problem with research impact evaluation in placing the line around a "pathway" from a "excellent" research activity that creates knowledge that then leads to a visible change in society ("impact").

This framing is neat: there is a start and a finish, and a cause and effect, and everything is tied together along the pathway. The reality of how impact emerges in practice is much less clear. We do not evaluate researchers as if the world stops around them when they are working, they make their contribution and then science starts again: that is frame of impact pathway approaches, an illusion that knowledge is created, and then it is transferred to users, and then it later achieves an impact.

We know that science is a continuous process: scientists are continually exchanging findings and communicating to advance the scientific state-of-the-art. It is the same for the reality of impact – scientists interact periodically with social partners and over time

this changes what the social partners can do.

The reality of an impact pathway is that it involves the social and scientific worlds coming together for a period, and creating new knowledge that advances the state of the art. The relationship is two-way – social inputs shape scientific trajectories just as scientific inputs shape social trajectories.

We know what good ("impactful") research is like – impactful scientists enact a series of scientific practices that are carried out that make these shared pathways more accessible for users and allow them to exert greater influence.

It is extremely tempting for research evaluators to seek to create "hard" research evaluations by looking to extremely successful examples of where impact has been achieved. But that places an artificial "finish line" in impact creation that does not create in practice. Creating impact at the system level involves more scientists being for longer on these shared impact pathways.

Evaluation should account for the activities and progress along the pathway. These are bound up with the researchers´ everyday activities, and are not eye-catching.

What needs to be evaluated is the commitment that scientists show in these different processes; how far do they commit their research in their choices to take trajectories that are more useful for society. A scientist that is inspired by a societal problem is making an active choice: they are choosing to address that social problem and not a different problem: they are committing themselves to taking a "step" along a "shared impact pathway". Their subsequent research activities are oriented to understanding that social problem. This is true for the range of different kinds of activities that scientists undertake – from the inspiration, developing research questions, to planning concrete projects, carrying out those projects and disseminating findings into society.

All of these choices have produce "evidence" that can be evaluated. Being genuinely inspired by a social problem means understanding the problem in the way society sees it – this may be signalled by the use of reports from NGOs, voluntary groups or government departments, or through the use of newspaper reports, through the use of meetings and discussions with social partners, through membership of a community group. There is an "activity" and the activity brings in knowledge that acts as a constraint on the subsequent research practices.

Good research practice for impact is where researchers have longer shared impact pathways; there are a series of interactions with societal partners that shape the overall research trajectory. Partners may stay involved because they find the research useful; partners acquire new capacities, and over time they may become realised into wider societal changes.

Evidence-based evaluation of research impact allows researchers to present evidence of the ways that they have these longer shared impact pathways, where societal partners influence the direction and content of the research pursued.

The principles of research impact evaluation

A research evaluation system needs to reflect the conditions of the country in which it is being created, in terms of the academic culture, the existing research evaluation culture, its sophistication in terms of academic impact creation and indeed its openness to impact creation, particularly from the social sciences and humanities. Nevertheless, on the basis of the framework above, where research evaluation provides signals that helps coordinate scientists towards desirable common ends, there are a set of principles that research policy-makers can follow to ensure that they provide the right signals to their scientists.

- **Researcher led**: steering involves creating new norms for scientific behaviour, and encouraging more scientists to follow those norms; an impact evaluation system should be driven by researchers making claims that particular "research subjects" are good impact.
- **Making a case:** a research impact evaluation system should provide researchers with the freedom to make their own case of what is good research, marshalling their own evidence to demonstrate their research activities conforms with scientific norms.
- **Open and transparent:** to allow scientists to anticipate what constitute "good" impact behaviours, there is a need to disclose how judgements are being made about "research subjects".
- **Formative: the** impact evaluation should be a learning process by which the subject of the evaluation comes through the evaluation to understand what it means to create good impact, and how they as researchers can create good impact in their research.
- **Prepare the evaluators to learn:** the norms of impactful research are continually being made: bring your evaluators together to help them make sense of what constitutes good research, and conclude the evaluation round by attempting to define "good research impact practice" as it has emerged through the evaluation process.
- **Reflecting academic norms:** impact evaluation should seek to identify practices that are already taking place and which are accepted as good and to give additional recognition for them, so that creating good impact becomes seen as part and parcel of doing good research.
- **Supported by illustration:** impact evaluation is a sense-making process in which academic communities are trying to understand the ways in which their research can benefit society: examples of appropriate evidence can help inspire creating thinking about impact.
- **Productive interactions:** a shared impact pathway will be punctuated by moments of productive interaction between scientists and society, a genuine exchange that affects the direction of scientific research, and evaluation should follow those impacts.
- **The shared impact pathway:** good impact comes through interactions in which

scientists and societal partners influence each other, with consequences for the scientific state-of-the-art and for societal impact; together they form a "shared impact pathway", and that should form the basis of the "research subject".

- **New pathways create new opportunities:** impact is created when new societal capacities are created and they in turn are then used by social actors to achieve change; activities to create entirely new pathways take a lot of energy, may be unimpressive but ultimately create huge new opportunities for societal development.
- **Evidence driven:** a shared impact pathway is observable to a scientist from the way that they receive knowledge from societal partners and use it to influence their choices that commits them to particular directions. These activities can be corroborated by using evidence that shows (a) interaction (b) research choice (c) change of direction.
- **Impact occurs through the research cycle:** impact is not only created once new knowledge is created and transferred to users; impact evaluation should reflect interactions and exchanges throughout the research cycle, from inspiration to dissemination.
- **Give impact 'academic capital':** doing impact well should be associated with recognition, with esteem, with access and rights, and good impact evaluation should support reward and recognition processes
- **Evaluate all impact creation sites:** developing norms that impact is part of good research require that value being learned during academic training. Evaluate Ph.D. programmes and supervisors on how far they train their researchers to created shared impact pathways.
- **Context sensitive:** much of what leads to impact being realised it outside the span of control of the scientist. Where a scientist is operating in a less munificent, more hostile environment, then smaller changes might be more consequential and create new pathways.
- **Finally, research evaluation is like criminal law:** much better that your system accidentally gives credit where it is not due, than denies it where it should be due: don't worry excessively about extreme cases because effective steering is a system outcome.

10 ways to give bad steering signals

1. Use someone else's system: good research impact reflects the science systems that it emerges in, and there is no perfect system for evaluating research; even if you adopt a system from another country, it will be implemented in ways that reflect your own science system context.
2. Portray impact as "something extra": the seeds that grow into "impact" are sown throughout everyday research practices in which researchers studying the real world interact with and are impressed by real world phenomena and agents.
3. Create a dichotomy between excellence and impact: impact is created through research practices, it is not an alternative pathway for those that cannot do excellent research, and if that perception emerges, then you will never be able to influence

whole academic communities.

4. Break the light-touch performance-recognition link (I): there should be positive consequences for performing well in terms of impact evaluation in terms of academic recognition; good impact performance is part of being a good academic.
5. Break the light-touch performance-recognition link (II): giving disproportionately large rewards to those that are able to create impact runs the risk of making impact appear as super- luminary and no something for the whole research community to consider.
6. Misuse metrics: aggregating individual-level evaluations by including simple has the effect of saying that some kinds of impact are more valuable than others (e.g. financial/ economic) and to discourage and dissuade those whose good impact is not shown in the metrics.
7. Account for scale: the easiest way to create impact is to persuade a Prime Minister of the value of your work. But that is only ever applicable to a handful of researchers, and this discourages all the other scientists who do not have the luck to be eye-catching.
8. Long lead-in times: if you are evaluating things that happened twenty years ago then it operates more as a lottery rewarding people who were lucky than helping contemporary scientists to attune their research to be more impactful.
9. "Harmonise" for efficiency: comparing different disciplines and communities risks portraying certain kinds of disciplines as more intrinsically impactful than others, driving self- reinforcing hypes that end up undermining impacts scientific and societal legitimacy.
10. Create extremely rigid evaluation rules: although it can be tempting to create "clear definitions", rigid rules will discourage those that have something to say and run the risk of game playing producing whatever the rules reward.