

Lund University
Department of Sociology

Adjusting to Impact

A study of research impact evaluation in Norway

Jo Arne Marvik

Master's thesis SOCM04 30 credits

Spring semester 2022

Supervisors: Åsa Lundqvist and Staffan Edling

Word count: 21833

Abstract

Author: Jo Arne Marvik

Title: Adjusting to Impact

Master's thesis, SOCM04 30 credits

Supervisors: Åsa Lundqvist and Staffan Edling

Department of sociology, spring 2022

Policymakers are increasingly demanding that their investments in research and knowledge production should yield returns. The interest of evaluating the societal impact of research has grown rapidly as research funders request evidence that their investments lead to public benefits. This thesis is a study of how the Research Council of Norway has operationalized such demands into new evaluation practices revolving around the concept of Impact. The material to this study consists of one evaluation report that introduced Impact into Norwegian research evaluation, the council's planning procedures in grant applications in addition to four expert interviews. Informed by practice-oriented document analysis, it highlights the transformative capacity of documents and emphasises the ways in which evaluation functions as tools to make visible phenomena, implement policies and transform researchers and their projects into alignment with Impact demands through processes of translation. Informed by actor-network analysis, I explore how evaluation stages and enacts realities into being.

Key words: research evaluation, impact evaluation, research council, practice-oriented document analysis, actor-network theory

Populærvitenskapleg presentasjon

Interessa for å evaluere samfunnseffektane til forskning har vakse med at avgjerdstakarar krev bevis for at investeringar i forskning fører til nytte for samfunnet. I denne masteroppgåva undersøker eg korleis Noregs forskingsråd har operasjonalisert slike krav om til nye evalueringspraksisar orientert mot førestillinga *impact*. Ved bruk av praksisorientert dokumentanalyse behandlar eg evaluering som eit optisk verktøy som gjer fenomen synlege, men også som eit politisk verktøy som overtøyer evalueringsobjektet om validiteten til det som blir evaluert. Videre undersøker eg korleis Forskingsrådet planlegg for *impact* gjennom å redusere prosjektforslag til logiske forhold mellom predefinerte element som interessant, verknad og effekt. Ved hjelp av aktør-nettverksteori skildrar eg desse som omsettingsprosessar. Slik syner eg at meklarrolla Noregs forskingsråd har mellom styresmakta og forskingssektoren ikkje berre gir utslag i dei tematiske programma deira, men korleis Forskingsrådet ved bruk av evaluering formidlar statlege krav til og utøver kraft også mot den basisfinansierte forskingssektoren. Til slutt diskuterer eg evaluering i lys av vitenskapssosiologiske innsikt som tilseier at røyndomar er ein konsekvens av og ikkje ei årsak til vitenskaplege sanningar.

Acknowledgements

I would like to thank the representatives in RCN and the researchers I have been in touch with regarding impact evaluations, who have been very accommodating and provided me with crucial information to this thesis.

I would like to express my sincere gratitude to my supervisors for their guidance and feedback despite the large number of emails and draft proposals I have forced them to read.

I am grateful my family who has given me shelter and support while also been taking part in formative discussions around the subject matter of this thesis during the research process.

My partner Kristina has witnessed me in moments of frustration and grief during these months but has given me love and support. I am forever grateful.

Table of contents

1. Introduction	1
<i>Outline of this study</i>	3
2. Previous research on evaluation and audit	4
<i>Research impact evaluations</i>	6
3. Theory and terminology	9
<i>Documents as tools</i>	13
4. Methods and materials	15
5. The Research Council of Norway	21
<i>The introduction of impact into Norwegian research evaluation</i>	22
<i>The evaluation of the humanities in Norway</i>	24
<i>Other developments within RCN</i>	29
6. Introducing impact	32
<i>What was made visible</i>	38
7. Intervening with evaluation	42
<i>Disseminating an evaluation culture</i>	43
8. Planning for impact	49
<i>Creating impact devices</i>	52
9. Conclusionary analysis: Enacting impact	54
References	57
Appendix A	61
Appendix B	62
Appendix C	64
Appendix D	65

1. Introduction

What mental images and impressions does the word “document” bring to mind? Most likely piles of papers grouped in folders or carried around in suitcases by faceless bureaucrats. The term paperwork signifies an activity that is routine and monotonous, often peripheral to a much more important task at hand, and perhaps even corrosive of creativity and ingenuity itself. In Norwegian, the expression “paper mover” has come to signal a type of work position which is inextricably tied to meaningless bureaucratic proceedings that does not really achieve anything else than fulfilling senseless tasks – moving papers from A to B. What these widespread notions seem to ignore is the incredible importance these piles of papers, physical or digital, have in modern states. Our institutions, departments, agencies, even parliaments, would simply not be recognizable if their documents were removed. The various documents that circulate to and from these locations are one of its most fundamental features. In many ways, these are document-created locations (Asdal and Reinertsen, 2020, p. 38). Behind this insight is an emphasis on documents not simply describing an external reality, but also taking part in modifying, transforming, and achieving reality.

The type of documents I will mainly be concerned with in this study are evaluations. Evaluation of public spending has become a cornerstone of contemporary democracies, and vast amounts of evaluations are commissioned with the objective of documenting whether desired results are being achieved as planned. The contemporary practices of evaluation influences media headlines, public debates, and all types of social and economic policies. Discussions about the practices of evaluation, however, seldom enters the public scene. And the scholarly traditions that have been preoccupied with evaluation and audit have been inattentive to their transformative capacities, and largely been categorizing them as symbols or rituals of the reform wave known as New Public Management, and as a result left the various tools, methods and practices of evaluation insufficiently analysed (Reinertsen, 2019).

The document-created location I will be directing my attention to in this study is the Research Council of Norway. This administrative body situates a mediating role between the government and the research sector in that it provides crucial advice to the Norwegian government in their design of research policies, while also transmitting governmental demands to the research sector. Policymakers are increasingly demanding that their investments in research and knowledge production should yield returns. While these expectations previously were primarily concerned with economic returns, universities are today being positioned (and positioning themselves) as delivering solutions that will produce

societal benefits, counter prominent challenges, and increase general well-being in society. The interest in evaluating impact has grown rapidly as research funders demand evidence that their investments lead to public benefits, and universities have established institutes with studying impacts of research as their main pursuit (Universitetet i Oslo, 2018). The demands for researchers to justify themselves in terms of their returns has, however, been perceived differently within different academic disciplines, and it has been argued that it poses a particular threat to the humanities and the social sciences, exacerbating a “crisis in the humanities” (Benneworth, 2015). Research evaluation is already a delicate issue within SSH research, and for the humanities four pertinent recurring objections has been voiced: (1) strong reservations against (simple) quantifications, (2) fear of negative steering effects of the indicators that are used, (3) the lack of a shared set of quality criteria between disciplines, (4) and the fact that its methods stem from the natural sciences and do not take into account humanities’ distinctiveness (Hug, Oschner and Daniel, 2014).

Whereas research evaluations have been carried out for decades, the phenomenon of research impact evaluations is relatively new. While the intensification and expansion of demands of return from research investments has been seen to aggravate a “crisis in the humanities”, the shift from a “fund and forget” style of funding to a more rigorous approach of documentation has also increased the level of ambition for such inquiries. If research evaluation is a type of inquiry with methods that primarily stem from the natural sciences directing its gaze upon SSH research, this study will reverse this relationship and direct the new practices of research impact evaluations with a lens inspired from schools of thought primary affiliated with the social sciences. By employing insights from actor-network theory and practical-oriented document analysis, the aim of this thesis is to explore how governmental demands of documenting effects have been operationalized into new evaluation practices by the Research Council of Norway. I will be attentive to how documents such as evaluations both describe reality and make something visible, but also how they simultaneously act upon the reality that is depicted. In that sense, documents and the act of documenting makes a change. They *do* things, but what? I will be guided by one overarching research question, that is further divided into three sub-questions: What does the Research Council of Norway’s impact evaluations do?

- How does evaluation make visible the societal impacts of research?
- How does evaluation processes intervene upon its objects of evaluation?
- How does evaluation work to transform research project proposals?

Outline of this study

This thesis is divided into nine chapters. The following chapter will present some general research on evaluation and audit and be followed by a section on research impact evaluations. In the succeeding chapter three, I will present the theories and literature which will enable me to approach documents as material objects and tools that may make a phenomenon visible, while also intervene and change its surroundings. Chapter four contains an overview of the materials I base my analysis on and how I have approached the topic, whereas chapter five gives a more in-depth presentation of the organisation and the evaluation report that is a centre piece of this study. In the succeeding chapters, my analysis is presented. The analysis is divided into three chapters based on the three sub-questions of my study. In the first, chapter six, I analyse the evaluation of the humanities as comprising a form of optics and present what it made visible. In chapter seven, I will treat the evaluation as a tool of politics and investigate how it simultaneously intervened upon its objects of evaluation during its formation. The last chapter of my analysis will cover how RCN has operationalized impact into its procedures for planning. Chapter nine contains a conclusionary analysis where I address implications from the scholarly tradition I have drawn from, onto the topic of this thesis.

Abbreviations in this thesis:

ANT	Actor Network Theory
Humeval	The Evaluation of the Humanities in Norway
H2020	EU's Horizon 2020 Framework Programme
LTP	The Norwegian Government's Long-term Plan for Research and Higher Education
RCN	The Research Council of Norway
REF	British Research Excellence Framework
STS	Science and Technology Studies

2. Previous research on evaluation and audit

Evaluation has been defined as “the systematic assessment of the operations and/or outcomes of a program or policy, compared to a set of explicit or implicit standards, as a means of contributing to the improvement of the program or policy” (Weiss, 1998, p. 4). Another definition states that evaluation is “a careful retrospective assessment of the merit, worth, and value of administration, output and outcome of government interventions, which is intended to play a role in future, practical action situations” (Vedung, 1997, p. 3). Definitions on evaluations vary and Dahler-Larsen (2015, p. 9) reminds us to be aware that evaluation finds different forms depending on contexts and based on different normative-ideological points of departure. He finds that the most integrative definitions include four key factors which any evaluation must deal: (1) an evaluand (the object of evaluation), (2) some assessment based on criteria, (3) a systematic approach or methodology to collect information about how the evaluand performs on these criteria, and (4) a purpose or intended use.

Once methodologically and professionally distinct fields, evaluation and audit have started to merge. This can be exemplified by the fact that Auditor Generals routinely conduct performance audits of state agencies’ achievements, that the big four auditing firms have established evaluation departments who bid for public assignments, and that the evaluation profession is struggling to balance the sometimes contradictory objectives of accountability and learning (Reinertsen, 2019).

The adverse effects of evaluation and audit have most prominently been highlighted in critical accounting studies and in sociology of evaluation (Dahler-Larsen, 2012; Lamont, 2012; Power, 1997; Vedung, 2010). An important point of reference has come from the works of Michael Power. According to his analysis, programmatic commitments to greater accountability has triggered a proliferation of documentation within public governance, what he calls the audit explosion (1994). This has resulted in a radically new type of government signalling the audit society (1997). The introduction of language and norms of audit into new domains such as health and education, has facilitated the emergence of a new ethics and politics of governance, where a new style of formalized accountability has become the ruling principle. The demands for auditing have not led to greater transparency and democracy, but rather changed the very relationship between the state and its constituents. From this emerges a new role for public servants who, instead of providing welfare services, are employed to verify that services are being delivered as expected.

Following Power's analysis, Shore and Wright (2000) argues that the introduction of audit's vocabulary into higher education is symptomatic of a new rationality of government they refer to as neo-liberal governmentality. Informed by Foucault and his emphasis on how techniques for measurement, documentation and comparison become instances of governing at a distance, they argue that the transfer of audit cultures into new public domains illustrates a wholesale shift in the role of government – using the norms of the free market as the organizing principle of both the economic life and the conduct of individuals. Audit involves disciplinary techniques and the production of self-scrutinized subjects.

A common critique of evaluation is that it often leads to unintended consequences. Dahler-Larsen objects to the concept of unintended consequences because it requires knowledge about intentions that he argues are not practically possible to determine, nor analytically useful (2015, p. 202-205). The concept also assumes a distinction between planning and outcome where ideas and actions are clearly separated. Such a distinction presumes that the phenomenon the evaluation seeks to measure (e.g., quality of schools) is constant and something we all agree on, while the indicator or measure is that which is socially constructed. This cleavage, he argues, is impossible to demonstrate because the indicator oftentimes helps define the phenomenon that it is set to measure, meaning that both should be regarded as equally constructed. Following his critique, he rather suggests the term *constitutive effects* to shed light on how the evaluative staging of reality lead to subsequent, and sometimes unexpected reactions. Informed by Power's argument that evaluation's most basic function is not to verify what goes on but rather to construct a definition of the activity so it can be evaluated (Power, 1996, p. 293), Dahler-Larsen sees constitutive effects as covering "the many subtle and not-so-subtle ways in which evaluation machines steer certain values, orientations, interpretations, and practices in the direction of a particular construction of social reality" (Dahler-Larsen, 2012, p. 199). The practices of evaluation can affect the content on some work, or practice, on the timing of practices and on the configuration of social roles and identities (Dahler-Larsen, 2015, p. 23-25). These three tend to be interconnected and enrol each other in a larger evaluative assemblage, which in turn may coalesce into a larger world view or underlying assumption about what is going on. For example, PISA-testing in education produces an assumption that education is international competition and that all countries have the same educational goals.

Research impact evaluations

Impact evaluations are part of a broader agenda of evidence-based policy making. It is marked by a shift in focus from inputs to outcomes and results. Not only is the focus on results being used to set and track targets, but the results are increasingly being used to enhance accountability, determine budget allocations, and guide policy decisions (Gertler *et al.*, 2016, p. 3). Impact evaluations are generally described as a particular type of evaluation which seeks to answer a specific cause-and-effect question: What is the impact (or causal effect) of a program on an outcome of interest? The causal dimension takes centre stage, that is the changes directly attributable to a program or an intervention. To be able to estimate the causal effect or impact of a program, impact evaluations usually include the counterfactual, what the outcome would have been for participants if they had not participated in the program. To assess the counterfactual, impact evaluations make use of comparison groups – groups with the same characteristics, but who did not participate in the program (Gertler *et al.*, 2016, p. 8).

However, the possibility of having control groups is not always admissible. When this is the case, another type of impact evaluation can be adopted to make probable the effect of the intervention by use of programme theory (Dahler-Larsen, 2016). A programme theory is an explicit theory or model of how an intervention – a project, program, strategy, initiative, policy – contributes to a chain of intermediate results, and finally to the intended or observed outcomes (Funnell and Rogers, 2012). In principle, the programme designer sets out hypotheses or expectations (theories) about how the program or intervention will achieve its impacts, and the evaluator then tests those hypotheses against subsequent events. The control groups can be located prior to the implementation of the program, and the potential impact can be tested and measured as such.

There are two main categories of impact evaluations. Prospective evaluations are put in place while the programme is designed and built into its implementation, while retrospective evaluations assess impact after the program has been implemented. The former is considered to produce stronger and more credible results, mainly because the treatment and potential comparison groups are identified before the intervention is implemented, in contrast to retrospective evaluations (Gertler *et al.*, 2016, p. 9-10).

The challenges of evaluating impacts of research have been voiced, where the question of *attribution* is particularly highlighted (Martin, 2007 in Muhonen and Tellmann, 2021, p. 4). Attributing impacts to certain causes such as research have been challenging, and it remains

difficult to differentiate between impacts stemming from research and that what is coming from other inputs.

The first country to elaborate a systematic approach to impact evaluation of research was Australia in 2006 (Donovan, 2008). The framework that was used was abandoned for political reasons, but later became an important basis for the development of the most important instance of research impact evaluation to date: the impact element in the British Research Excellence Framework (REF) (Wroblewska, 2019). The 2014 REF represented the first formal, ex-post assessment of how research had an impact beyond academia linked to the allocation of research funding (Derrick, 2020). Its use of the impact criterion is making impact into a serious notion of academic excellence on the individual level, as well as of strategic importance on the organisational level (Derrick, 2020). The assessment was based on researchers' own descriptions and documentations, dubbed the "impact cases". Several countries have now adopted the definition and design into their own research evaluations, including Australia, Hong Kong, Sweden, Italy, Poland, and Norway. Together with Excellence and Open science, Impact represents the most important research political priorities within the EU and in several other countries (Derrick, 2020).

This evaluation approach to research impact has sparked objections. Some voice that the final effects of research are too difficult to evaluate in a meaningful way (Gulbrandsen and Sivertsen, 2018). As an alternative, some have presented a model that does not seek to document the final effect or impact, but rather seek to map the productive interactions the research is involved in (Spaapen and van Drooge, 2011). Others have introduced the term "normal impact" as counter to concrete and often extraordinary examples that are most often used in examples of impact (Sivertsen and Meijer, 2018).

While the British REF has sparked a lot of critical discussions on evaluation methods, some empirical studies of this model have also been carried out. Derrick (2018) offers an analysis of how the dominant definition of impact was constructed within the evaluation/peer review panels. By partly stepping into the black box of peer review and interviewing the British REF's evaluators, she investigates the social dynamics between the panellists in their pursuit for coming to a consensus on how to understand and how to assess research impact. By primarily studying "the evaluator's eye" and viewing the evaluation of impact as a dynamic process, she suggests peer review can be a feasible tool if used in a smarter way.

In her PhD-thesis, Wróblewska (2018) investigates the discursive aspects of the introduction of the British REF's impact agenda and its consequences for academic identities. Informed by Foucault's theory of governmentality, she argues that its implementation was a

response to socio-economic tendencies such as the knowledge-based economy, the rise of the third mission of universities, as well as class issues and tensions between academic disciplines. The impact agenda facilitated the introduction of an “impact infrastructure” – an apparatus composed of positions, procedures, forms etc. which subtly guides the conduct of academics and shapes their professional vision to make impact an element of academic activity. She also highlighted how academics struggled with the task of subjectivation – incorporating impact into their presentation of selves.

Wróblewska has also conducted a comparative study based on REF and RCN’s evaluation of the humanities (2019). Through a linguistic approach, she analyses the impact case study as a genre of writing. Such new linguistic practices, she argues, do not simply reflect certain social practices, but actively create and shape them. She suggests viewing them as important interventions into academic culture. Although RCN adopted the same evaluation design and its definition of impact from the British REF, she observed through a study of the impact cases that the two evaluations were establishing different approaches. Illuminative for the Norwegian impact cases were the documents’ honesty and reflexivity, but also its lack of empirical substance and a well-established case genre. She argues that a firm foundation for an evaluation system could build on the best aspects of the two, striving to establish a coherent genre, while also securing a mutual trust between policy makers and the academic community, rather than a top-down “control”.

3. Theory and terminology

Where other studies have highlighted the gradual formation of impact among evaluators, and the discursive aspects around the impact agenda, I will primarily be attentive to the tools and practices that are operationalized around impact pursuits in a Norwegian setting. To do so, I will be informed by a practice-oriented approach to documents.

In the introduction I highlighted the transformative capacity of documents. This emphasis has been theoretically developed through what Asdal and Reinertsen has termed practice-oriented document analysis (Asdal, 2015; Asdal and Reinertsen, 2020). Practice-oriented document analysis takes part in discussions affiliated with a scholarly turn to materiality, which has been considered as a way of taking seriously the materiality of nature objects. This “new materialism” was developed in response to the “linguistic turn”, which was dominated by primarily textual accounts. These were deemed insufficient for an adequate understanding of the complex and dynamic interplay of meaning and matter (Lemke, 2015, p. 4). Asdal’s emphasis, however, is that if we are to analyse and take nature objects into account, it is also necessary to take seriously the extent to which materiality are made available to us by way of documents and how documents in turn take part in enacting realities (Asdal, 2015). “New materialism” and practice-oriented document analysis draws inspiration from the interdisciplinary field of science and technology studies (STS). As I will be primarily drawing from practice-oriented document analysis alongside works that are often rubriced under the tradition of actor network-theory (ANT), I see the need to familiarize the reader with some of the aspects and terminology of this tradition that I will employ.

In the 19070’s, a group of philosophers, historians and sociologists based in Edinburgh developed what is called the “strong programme” in the sociology of knowledge (Sismondo, 2010, p. 47-56). A central figure was David Bloor who summarized their approach into a list of methodological tenets. One main tenet of this school of researchers was the structure of symmetry. This was a refusal to distinguish between scientific statements which by the scientific community were deemed true and those which were deemed as false. The programme was opposed to the common scientific and naturalistic attitude saying that “true” scientific statements were solely a result of the observed natural phenomena. Logically, this implied that only false scientific statements could be analysed sociologically, because these were instances where the scientists were distracted by other (social) factors, which then could be studied. The strong programme’s emphasis on symmetry meant that all science work

should be analysed in the same manner and thus opened the terrain for social scientists to study scientific practices and cultures.

Informed by this tenet, a series of researchers began visiting laboratories and directly observing scientists, leading to works with titles such as *The Manufacture of Knowledge* (Knorr-Cetina, 1981), *Laboratory Life* (Latour and Woolgar, 1986) and *Science in Action* (Latour, 1987). These early studies of science and technology emphasized the *construction* of scientific facts, and asked questions related to how facts were made. The methods were inspired by anthropologists who until now had predominantly travelled into the depths of rainforests or to tropical islands far away from the Western audience. As Latour and Woolgar famously remarked in their introduction to *Laboratory Life*, relatively few attempts had been made to penetrate the intimacy of life among tribes which were much nearer at hand. This was puzzling, they argued, considering the importance of the products these tribes of scientists produced (1986, p. 17). Whereas philosophers had approached science and its methods theoretically, the anthropologists who ventured into the laboratories were more attentive to the distinct practices of laboratory scientists. Latour and Woolgar described the laboratory members as “compulsive and almost manic writers” (1987, p. 48), and they turned their attention to how the objects of which the scientists studied, moved from its area of origin and into the laboratory, how it turned into an object for investigation, and furthermore into a scientific paper. The laboratories were seen to contain tools that make objects human in scale, and therefore easier to observe, analyse and manipulate. They called these tools *inscription devices* – any item of apparatus or configuration of such items which can transform a material substance into a figure or diagram (Latour and Woolgar, 1987, p. 51). The strength of the scientists’ arguments was seen to rest upon practices revolving around the apparatuses, and the laboratories were described as factories of literary inscriptions. The implications of Latour and Woolgar’s work was that the making of particular realities, the making of particular statements about those realities, and the creation of inscription devices, could not be separated. They were all produced together (Law, 2004, p. 37). The controversial conclusion was that the realities the scientists depicted was not a cause, but a consequence of their work.

The important ontological commitment driving the analysis is a refusal of the division between word and world, object and subject or referent and reference (Latour and Woolgar, 1987, p. 236). Latour later substantiated this point in an ethnographic study of a team of scientists studying soil samples to establish the precise borders between the Amazon rainforest and the savannah (Latour, 1999). He pays great attention to the many small and pragmatic steps the researchers carried out to develop their analysis.

From forest to expedition report, we have consistently rerepresented the forest-savanna transition as if drawing to isosceles triangles covering each other in reverse. Stage by stage, we lost locality, particularity, materiality, multiplicity, and continuity, such that, in the end, there was scarcely anything left but a few leaves of paper... But at each stage we have not only reduced, we have also gained or regained, since, with the same work of rerepresentation, we have been able to obtain much greater compatibility, standardization, text, calculation, circulation, and relative universality, such that by the end, inside the field report, we hold not only all of Boa Vista (to which we can return), but also the explanation of its dynamic (Latour, 1999, p. 70).

Latour uses the twin concepts of *reduction* and *amplification* to describe the process and results from the scientists' expedition. The object of study is first moved from its environment into a confined space where the original sample can be *reduced*, meaning that various particular facets are removed in favour of the facets that are thought to be beneficial for the aim of the expedition. Next, *amplification* refers to the gains of such a process, where what is left of the original sample is extended and weaved together with a range of other translations. The objects of study lose some properties, but at the same time gain others, which render them useful for comparison with others. The results of this process can then be moved back into the world to convince others about its validity. Latour's idea counters one where science is depicted as a process where a subject attains knowledge about an object. What he suggests is a model built on many chains of translations, a model he refers to as *circulating reference*. The term encompasses science's important ability to retrace its steps – to rewind the chain of translations. If all the various steps connecting the field through the laboratory and to the scientific article is not documented thoroughly, the claim is weakened. The establishment of facts thus relies upon the circulating reference being intact.

The concept of the circulating reference should be considered as an alternative to “correspondence theory” common to epistemologists – the idea that science only mirrors reality. Circulating reference refers to the many small and pragmatic translations that are established between word and world in practice. Instead of treating such an opposition as given or viewing scientific results simply as a result of a negotiation between scientists, ANT considers the strength of a statement is based on a chain of stabilized translations between the statement, other statements, inscriptions, instruments, materials etc. It is these chains which are the actor-network of which an organisation, a statement, a report, or a scientific paper may

speak on behalf of. The establishment of facts are dependent on a rigorous and stabilized network of mobilized allies, human or non-human.

The central term used to describe the scientists' endeavours was *translation*. Originally a concept derived from the philosopher Michel Serres who used it to describe a kind of mediation that simultaneously transfers and distorts a signal, ANT's version of translation denotes a process whereby two actors become related in such a way so that one actor borrows strength and may speak on behalf of the other (Blok and Jensen, 2011). For Latour, it has primarily signalled a process through which phenomena are transformed from sample to data, from data to analyses, analyses to scientific articles, and finally, from articles into scientific facts. In a seminal article about a group of scientists attempt to cultivate scallops in Brittany, Callon (1984) further develops the term. The scientists' project is treated as a venue where both human and non-human entities are in a process of forming an actor network. Callon's attention is directed at the tactical mechanisms which the scientists employ to enrol the local community, including fishermen and scallops, into the project. Translation is depicted as involving four moments, where the scientists target certain actors needed for the project, recruit them through various mechanisms, coordinate them and make themselves as spokespersons for the entire network. In the end, however, the scientists' project failed. The representatives of the various actors the scientists enrolled, turned out to be unreliable. The scallop larvae did not attach to the anchors as first assumed, and the fishermen began catching scallops in the area counter to the project's plans, both disavowing their spokespersons promises. Consequently, the proposed network deteriorated. What should be noticed with this approach, and with ANT in general, is that Callon does not alter his registers when describing people or scallops, the social, the biological, or the technical. He does not analytically distinguish between human and non-human entities, so both the fishermen and the scallops were subject to attempts of translation, and they both showed dissidence to their spokespersons and the project plans.

There are two ontological points I want to draw out as inspirations to this study. The first is a refusal to regard the relationship between word and world as given, and rather convey these as accomplishments achieved through a process of *translation*. The second is the commitment to describe processes involving human and non-human entities in the same manner. The concept of translation encompasses both points, denoting the process of inscriptions that speak on behalf of a certain phenomenon and a process where certain actors may end up speaking on behalf of others. The result is a flat form of ontology, where the reality *out there* and the discursive description of that reality is seen as equally real.

Documents as tools

Most scholars who work within actor network-theory have moved from an emphasis on construction to a post-constructivist position, emphasising the *enactment* and *performativity* of social arrangements (Law, 2009). The combination of people and artefacts form heterogenous assemblages which may intervene and create situations of reality. It assumes nothing has reality or form outside the enactment of those relations, hence the other often used term of *material-semiotics*. ANT seeks to describe the enactment of materially and discursively heterogenous relations that produce and reshuffle all kinds of actors including subjects, objects, human beings, machines, animals, organisations, ideas and so forth (Law, 2009).

ANT's general refusal of essential foundations has led to inquiries investigating how realities are *enacted* into being through various tools and devices. Callon (1998) has famously argued that the theories of markets are crucial in determining the realities they allegedly only describe; market theories enact markets into being. Following, Doganova and Eyquem-Renault suggest viewing business models as market devices (2009). Business models contribute to this enactment by playing a part in the emergence of individual calculative agencies, mediating the relationships between these agencies, and coordinating their actions (2009, p. 1561). Business models are not only neutral descriptions of a phenomenon *out there*, but actively contribute to create and change the phenomenon that is described. Deriosières (2001, p. 352) has similarly emphasised how statistical work not only reflects reality but establishes it by providing the players with a language to put reality on stage and act upon it. This implies that the act of measurement, by putting reality on stage in a particular way, has a constitutive element alongside the potential interventions installed upon that reality.

The performative abilities of documents are encapsulated in practice-oriented document analysis, as put forward by Asdal and Reinertsen (2020). It involves an approach to documents that sheds light on what documents do, enable, or achieve. In their words, documents are tools, devices, and technologies. They develop three different groups of technologies that documents may take part in: knowledge, politics, and markets. As tools of knowledge, documents may make something visible. Through the acts of documenting, a phenomenon may be made observable and analysable. As tools of politics, documents formulate, suggests, decides, and move a subject matter into a certain direction and from one place to another. Asdal (2011) emphasises how such document technologies may reshape objects of nature to be quantifiable, comparable, and thus also rutable. These analytical

groups are closely connected, but what they share is an approach to documents that emphasise their ability to achieve and accomplish something.

In this light, evaluation reports with its various tools can be understood as material-semiotic artefacts that embody performative abilities. Evaluation may be seen as a tool to enable certain forms of vision and certain forms of action. This also brings evaluation in line with other apparatuses that have been important for the STS-literature, such as microscopes, telescopes, x-rays and so forth. Evaluation is analytically depicted as taking part of a range of technologies that makes a phenomenon and its results visible and eligible for analysis. This emphasis on how technologies enable certain forms of vision resonates with central themes in Donna Haraway's thought (1997). She has popularized the term of *situated knowledge* to account for how vision cannot help not to be partial. Her critique of scientific knowledge as claiming to see everything from nowhere, or performing the "god trick", opens a territory where one may ask specific questions to the partial gaze that is studied.

A more substantiated view of evaluation in this study stems from Reinertsen's PhD-thesis (2016), where she investigates how Norwegian aid was sought to become an evaluative object through practices within the Norwegian Agency for Development Cooperation (Norad). She developed the term *optics of evaluation* to refer to the technologies, infrastructures, and practices that evaluation make use of in order to make something visible from a far, temporarily and spatially. Central for her analysis were the planning tools introduced by Norad in a handbook on aid evaluation. The handbook introduced new monitoring practices, but also *inscription devices* in the form of a goal pyramid and an assessment matrix which the aid staff were to use in their documentation. These tools enabled a new way of articulating an aid project, its staff and site, and furthermore a new way of ordering these three in relation to each other. She suggests that the employment of these tools was critical in establishing a *circulating reference* of a project idea, through a process of *reduction* and *amplification*. By enabling a circulating reference aid would be become evaluable (Reinertsen, 2016, p. 90-92). Although these meticulous practices constituted a new form of optics, slipperiness and interpretive flexibility did not wither away and much more work was needed if one were to attain clarity, overview, and certainty. Rather than illuminating the precise contributions of Norwegian aid, she suggests that they moved away from the practical and the pragmatic, and rather succeeded to illuminate the uncertainties. Paradoxically, making aid an evaluable object did not show how Norwegian aid made a change, but rather contributed to the notion that "we still do not know enough" (Reinertsen, 2016, p. 317).

4. Methods and materials

I fumbled into the theme of evaluation primarily as an intern in the research project led by Hilde Reinertsen, with the name Evaluation. One of the reasons I applied for an internship in this project was that she was part of the not-so-large group of researchers who engage in the scholarly tradition of STS in Norway, a tradition I over time had developed an interest in. One of my tasks as an intern was to provide an overview of the content and historical development of the planning and evaluation tool known as programme theory. During my internship, Reinertsen informed me that RCN had recently implemented a version of programme theory into their organisation, and she asked if this could be an interesting theme for my upcoming master thesis. I agreed.

After the internship I established contact with the department for statistics and evaluation in RCN and had a conversation with a representative there. This person informed me about the various measures of which RCN had implemented to better trace the societal impacts of research and passed some of the evaluations which so far had inquired into the societal impacts of research. One of these was called “Long traces in welfare research” (Solberg *et al.*, 2019). In this report, the authors stated that the societal effects of particularly SSH research have historically been difficult to trace. It is therefore not coincidental, it was said, that it was the evaluation of humanities which first introduced the method of self-reported narratives as a new method to trace the societal effects of research (Solberg *et al.*, 2019, p. 42). I was somewhat intrigued by this and decided to adjust my thesis to address how RCN worked to document the societal effects of SSH research. I was also interested in what role programme theory might have had in this work.

One report I had used as source of information in my tasks as an intern had been an evaluation report about the implementation of programme theory into RCN, conducted by a science consultancy firm. When I started to engage with the topic of my thesis, I discovered that the firm of this report also had been involved in RCN’s evaluation of the humanities. I established contact with a person involved in these proceedings and was luckily admitted employing them as interviewee to my study.

During the first period of working with the thesis, I had originally intended to use two or three different reports as material for analysis. While writing and investigating, I experienced that the evaluation of the humanities, the Humeval report, became the main source of material for my analysis, and that I was treating the other succeeding reports comparatively to Humeval. Several reasons made this report of particular interest for my

analysis. One reason was that this was the evaluation that introduced the impact element in RCN's evaluations, and I found that several succeeding evaluations were mentioning and basing their designs upon this particular report. The collected impact cases from these reports were in turn reanalysed in other evaluations (Solberg *et al.*, 2017; Solberg *et al.*, 2019). It was also mentioned in Humeval that it had the potential to spearhead and introduce a new and broader field of research evaluations in Norway, which captured my interest. In addition, I discovered that the report had been published around a time when the humanities were subject to public debate. I could recall instances of a debate revolving around a "crisis in the humanities" and a certain fright about the humanities being fixed into having auxiliary purposes. I was curious to what role RCN may have had in defining, materializing, and realizing research policies that may have substantiated these alarms. For these reasons, I decided to use this particular evaluation report as a centre piece for this study.

During the course of my study, I was made aware of that RCN also had implemented changes and included impact as an assessment criterion in their applicant forms. In line with impact evaluations in general, this is a method to plan for impact or to evaluate impact prospectively. I was curious to how this process occurred and how the concept of impact was addressed in these applicant forms. By using the first evaluation that had retrospectively evaluated impact in addition to the applicant forms that seeks to evaluate impact prospectively, my idea was that these two in combination could strengthen my analysis and provide a wider understanding of how RCN operationalize their pursuits for impact.

The primary material underlying my analysis is the evaluation of the humanities also known as Humeval, an interview study of evaluated researchers in Humeval, the applicant forms for researchers that seek funding from RCN, in addition to four expert interviews. The evaluation of the humanities consisted of one principal report, eight sub-reports and one document with the collected impact cases. I have primarily used the principal report, two sub-reports, the collection of the impact cases and the associated analysis in this latter document as material for the analysis (see overview in Figure 1). To gather insights to the respondents' perspectives, I used the comparative study I mentioned in chapter two (Wróblewska, 2019).

The expert interviews I carried out were primarily with representatives in RCN who work with research evaluation, whereas one interviewee was not working in RCN, but rather worked with research evaluation in a consultancy firm. This person and firm had been assisting in the proceedings of Humeval. Experts can be understood as people who possess specific knowledge that relates to a clearly demarcated range of problems. Experts plays an authoritative role in decision-making of different kinds, and they exert influence by

determining the way people understand and interpret the world (Bogner, Littig and Menz, 2018). The first interview served a function as exploratory data collection (Bogner, Littig and Menz, 2018, p. 10-11). The interviewee served as a guide in my pursuit to establish an orientation into the field and helped me to gather material of interest and limit my subject matter. The other three were also exploratory, but for these I had achieved to further delineate my research interest, so these interviews were to a larger extent structured around formulated themes. Two of the interviewees forwarded me relevant information which they advised me to read before the interview. Some of this material has been used as data or informed this study in a complementary manner. Informed by practical document analysis, my interests have not primarily revolved around how the experts think and resonate about their work, but rather how they are working, how they organise their work, what this work is comprised of and what tools they employ (Asdal and Reinertsen, 2020, p. 172).

The three latter interviews were to a larger degree semi-structured, where I had formulated a handful of themes for the interviews with specific questions related to each theme (See rough interview guide in Appendix D). The interviews were conducted in Norwegian and carried out through video calls. For two of the interviews, I sent my interview guide to the interviewee prior to the interview taking place. The interviews lasted from 45 to 90 minutes and were recorded with permission from the interviewees. After the interviews were conducted, I watched and listened over the recordings and wrote transcripts of the parts I found relevant for the study. After transcribing, these recordings were deleted. Because the interviews were carried out in Norwegian and being translated into English by me, I wanted the interviewees to read and verify the way I had transcribed and translated the passages. The transcripts I have decided to include in the thesis were sent to the interviewees for verification. Only minor changes in the English translation of certain concepts were suggested to be changed by the interviewees. Some of the information I was provided in the interviews have been important, but not of a character I deemed relevant to provide in full quotation. I will therefore refer both to transcripts and these bits of information as following: (IP1, IP2, IP3 or IP4), to indicate the particularity of each interview. Furthermore, this is also a way to maintain the interviewees' confidentiality. The information the interviewees has provided is not of a character that I, nor them, deemed to be of private character, nor sensitive in that they may lead to unfortunate consequences for them as individuals or research evaluators. Nonetheless, I promised the interviewees partial confidentiality in that I would not reveal their identity, but that their work organisation could be revealed. None of the interviewees had any

objections to these proceedings. I judge my proceedings to be in line with research ethics as formulated by the Swedish Research Council (Vetenskapsrådet, 2021).

Figure 1: Primary material

Source	Detail
Evaluation of the Humanities. Report from the Principal Evaluation Committee (RCN, 2017b)	The main report of the evaluation included overall findings and recommendations of the report in addition to information on logistics and methods.
Evaluation of the Humanities in Norway. Impact Cases (RCN, 2017a)	The document was a collection of the collected impact cases and included a description of RCN's analysis.
Evaluation of the Humanities in Norway. Report from Panel 1 – Aesthetic Studies (RCN, 2017c)	One of the eight sub-reports. Provided information to the field panels' assessments.
Evaluation of the Humanities in Norway. Report from Panel 2 – Nordic Languages and Linguistics (RCN, 2017d)	One of the eight sub-reports. Provided information to the field panels' assessments.
Impact evaluation in Norway and in the UK. A comparative study, based on REF and Humeval (Wróblewska, 2019)	A study that provided insight to the perspectives of the respondents in Humeval.
Template for Project Description Scientific Renewal and Young Research Talent 2022 (RCN, 2022a)	Applicants to RCN's funds are expected to use this template.
Semi-structured expert interviews (four in total)	Three interviewees were representatives in RCN working with research evaluation. One interviewee was a representative from a science consultancy firm.

When I have been analysing the material, my main interests have been related to the novel concept of impact and how RCN and the evaluators operationalized, addressed, and assessed this concept. Opposed to other types of analyses that are oriented towards the textual, my

approach has been characterized by an interest towards the practical. Both the interviews and the written material were primarily treated as sources to the practical proceedings in the evaluation.

Bowen (2009, p. 32) suggests a three-step model of document analysis, involving skimming (superficial examination), reading (thorough examination), and interpretation, and although I have loosely followed a similar sort of approach when identifying pertinent passages and information, I am more inclined to say that the “interpretation” has been informative for the entire process. With this I mean that the terminology and basic insights I have received from the works I have rubriced under the theory-section in this study, has informed not only what often is referred to as the interpretation part of the analysis, but also the process of formulating research questions, identifying pertinent passages, formulating interview questions etc. When working through the written material, I already knew that I was interested in the practical proceedings of the evaluators regarding the novel impact-element, so my codes were affected by this interest. In the report from the principal committee, I coded passages into “findings” and “recommendations” which I gathered in separate folders. Some of these “findings” I further coded as “shortcomings”, and I tried to link these to the passages I had coded as “recommendations”. Furthermore, I collected information on how the evaluation had been carried out and gathered this in a separate folder of “methods”.

When collecting relevant information and discussions regarding research evaluation, I have used the software Zotero. My use of the software has primarily involved collecting information, creating folders, and keeping track of large numbers of websites and PDF files. Unfamiliar to this software prior to this study, I have found it extremely helpful to keep an overview of the many evaluation reports, articles, presentations, news articles etc. which have informed this study.

Throughout the proceedings of organising my sources and during the general research process of this study, I implicitly operated with a distinction between articles and reports that I categorized as being located within the “evaluation community” and those that I rather considered as research on the topic of evaluation. One article I judged to fit to the latter category was the comparative study between REF and Humeval (Wróblewska, 2019), which gave me valuable insights to the respondents’ perspectives of the evaluation I was studying. Throughout the research process, and after the supervisor suggested me to include this article in my list of materials, I have discovered that my preliminary distinction was not very useful. Both RCN and the “evaluation community” are often highly educated in social scientific

methods, and RCN not only read research articles on the subject, but also commissions research and adjust and replace their evaluation methods and models based on such research. Important sources of information for this study such as Derrick and Benneworth (forthcoming) and Wróblewska (2019) are informing RCN's approach to research evaluations. Point being is that my initial categorization seemed to not be very useful, and that research and evaluation are inextricably interwoven to the point that picturing research on evaluation as being positioned "outside" or "independent" of the field it is studying seemed to be most certainly a misconception.

My analysis is divided into three chapters, where this division is consistent with the division between retrospective and prospective evaluations. I could have done this differently, and I do not claim that it is necessarily advantageous or helpful to follow distinctions and understandings from RCN or the evaluation literature in general, when studying evaluation practices. Nonetheless, for me it seemed helpful to divide it in this way because the pursuit for impact from research that already had been carried out and the planning for impact in future projects, seemed to entail quite different practical proceedings. There is also the case that the evaluation in the former case targets a specific field, while in the latter it is being applied to all applicants. Another reason is the fact that the retrospective evaluations of Humeval is published and easily accessed whereas the prospective evaluation to a larger extent is happening behind peer review's closed walls.

To end this chapter, I would like to state some methodological "rules of thumb" that I have striven to uphold. One aim of ANT research is that it does not seek to formulate theories or explanations of why actors, human or non-human, act as they do. One of its main features has been its great reluctance to follow macro-sociological accounts that seek to formulate theories and explanations to actions, often in contrast to the ideas and intentions of the actors that are being studied. Instead of assuming that the world and society can be generalised or explained by macro-sociological generalisations, ANT seeks to describe rather than to predict (Latour, 2005). Although I have employed a particular terminology that may be foreign to the partakers in the evaluation practices I address, I have striven towards describing the practices as close to the way I have been informed through the written sources and the interviews. If the study includes parts where I have failed to uphold this principle, it is exclusively a result of my own misunderstandings.

5. The Research Council of Norway

Before the analysis, I will present the organisation that is the main object of research in this thesis. This will be followed by a presentation of the evaluation report that I have used as centre piece of my analysis, and a section on other developments within RCN which I have found relevant to this study. I have ordered the ones I regard as most pertinent to my subject matter in a timeline (Figure 2).

The Research Council of Norway is an administrative body that was established in 1993 through a fusion of five different research councils. It is Norway's biggest and most important organisation for financing research in Norway. Its core task is to delegate funds to research- and innovations projects based on applications, but it is also responsible for the base funding of around 50 different research institutes in Norway. RCN's budget is approximately 10 billion NOK (Li, 2022), and comes from all 15 Norwegian Ministries. The Ministry of Education and Research has the overall coordinating responsibility upon the ruling of RCN, but all Ministries provide leads on what research they want to be carried out. RCN have since its creation made extensive use of its research programmes, in part to demonstrate to the funding Ministries the direct link between its research and departmental policy goals (Derrick and Benneworth, forthcoming). Another important task for RCN is to act as an advisory body for the Norwegian Government in the design of research policies, illustrating the mediating role RCN situates between the government and the research sector in Norway.

RCN has routinely conducted research evaluations of disciplines and institutions since the 1990's. The interval of these is approximately 10 years, so each unit has been evaluated twice (Holm and Askedal, 2019). The evaluations are important sources of knowledge for political decision makers regarding the design of research policies, but they are also formative for the institutions and research environments themselves, who often initiate such evaluations.

A range of different evaluations are carried out (RCN, 2019a). Field evaluations provide a critical review of certain subjects and disciplines in an international perspective. Thematic evaluations assess various subjects and disciplines within a limited area. Institute evaluations assess the research institutes with the aim of identifying potential for improvement. Effect evaluations provide an analysis and increase the understanding of the relationship between an activity or intervention and its effects on research and society at large. Effect evaluations are often, but not always, integrated into other evaluations. RCN also commissions evaluations. Commissioned evaluations have targeted RCN's organisation and its activities, but also carried out effect evaluations of its own thematic programmes. These so

called “long trace”-approaches have been carried out by the Nordic Institute for Studies in Innovation, Research and Education (NIFU). Adopting various methods of inquiry such as analyses of research careers, bibliometric analyses, interviews etc., these evaluations seek to trace and document the influences of RCN’s thematic funded research in politics, administration, and wider society. So far, these evaluations have targeted the programmes of drug and mental research (Ramberg *et al.*, 2015), development research (Solberg *et al.*, 2017) and welfare research (Solberg *et al.*, 2019). For the two latter reports, relevant impact cases collected by RCN in their evaluations have been reanalysed and informed the results of these reports.

There are some distinguishing features in Norwegian research evaluations that should be mentioned. Norway does not have any single all-encompassing system of research evaluation that could compare to the British REF (Wróblewska, 2019). Instead, subjects have historically been assessed separately. Research institutes, that are more oriented towards the applied sciences, are assessed in separate evaluations, and are not divided into sub-units. In later years, however, RCN has replaced its subject-specific model with a model that targets entire fields, allowing for a more strategic approach where disciplines are assessed in an aggregated way. Another important feature of Norwegian research evaluations is that they are not directly tied to funding, so the main objective is above all formative (Holm and Askedal, 2019; Wróblewska, 2019; Muhonen and Tellman, 2021). Higher education institutions are expected to be evaluated, whereas for research institutes it is more of an option to take part. Some institutes consider the willingness to take part as a means to secure their academic status (Wróblewska, 2019).

The introduction of impact into Norwegian research evaluation

The pursuit for valorising the role of SSH research was highlighted after RCN published a strategy for research in 2008 (RCN, 2008). The strategy was commissioned by the Ministry of Education and Research after several deans within the humanities had been distraught about the invisibility of the humanities in the Ministry’s 2004 white paper to the Storting (Morgenbladet, 2008). RCN’s strategy emphasised the importance of applying humanities research and making visible its contributions and relevancy. They sought to identify new and existing areas where the humanities had a potential to increase their contribution to society. They proposed to increase humanities’ share of the grants which were thematic and problem-oriented (RCN, 2008).

The strategy was controversial among the research community and sparked a cultural debate about humanities' contributions to society, as well as the general role of research in Norway (Morgenbladet, 2008). The first reaction was seen to be dominated by a fear for instrumental ideas about the legitimacy of the humanities (Holm, 2018). Some raised concerns that RCN's strategy conveyed a "crisis-narrative" about the humanities and that it was reduced to having an auxiliary role towards other disciplines. When the evaluation of the humanities was in its preparational stage, the debate had shifted from a general fear of reducing humanities into auxiliary purposes, to a question of what societal benefits the humanities can bring about (Holm, 2018).

The impact element was to inform on how the evaluated research unit performed on influencing society outside of academia, and it first entered Norwegian research evaluation in two overlapping evaluations (See Figure 3). One of these evaluations were the evaluation of the Norwegian humanities, dubbed Humeval (RCN, 2017b), while the other was an evaluation of Norwegian social science research institutes (RCN, 2017e). The reasons behind RCN's inclusion of impact were primarily related to governmental expectations of accountability – that the wider society should harvest societal benefit from increased public investments in research, as well as be informed of what these contributions were (Holm and Askedal, 2019). Including the impact element was, however, also wanted by members within RCN. One notion was that the opacity of research's contribution to society would increase the trust in research among the wider public and appraise the importance of research (IP1).

While previous discussions and evaluations had sometimes addressed relevancy, usefulness, commercialization, or other similar concepts, they were primarily applied to address the economical and instrumental benefits of research. Relatively simple inquiries were carried out to investigate potential users and clients, the emergence of patents, licenses, and spin-off companies, as well as collaborations between research milieus and businesses. Results from these inquiries could provide indications to whether researchers were in contact with others who may benefit from their work, and that research had a societal benefit. The inclusion of the impact element thus not only signals a wider approach to what types of contributions research may give, but also a higher ambition related to the evaluations' ability to document these issues (Gulbrandsen and Sivertsen, 2018).

The British Research Excellence Framework (REF) published in 2014 became an important source of inspiration for RCN, and its design, methods, definition of impact, case template and guideline were adopted. REF is judged to include the most comprehensive impact evaluation of research to date (Derrick, 2020). The data material for the impact

element were primarily based on an impact case methodology, where respondents from institutions and research groups reported and documented the impacts of their research outside academia. The reasons why RCN adopted the REF design were that it was conceived to be well documented, tested and evaluated, and the definition of impact was judged to be sufficiently broad to include most of the expected societal benefits of SSH research. RCN considered REF's definition to be more open to account for disciplinary differences and compatible with the multitude of pathways to impact as documented in empirical studies (Holm and Askedal, 2019).

The evaluation of the humanities in Norway

As mentioned, the evaluation of the humanities, also known as the Humeval report, was one of the evaluations that first introduced the impact-element into Norwegian research evaluation. The evaluation covered 36 institutions, 2.300 researchers and 97 research groups. In the introduction it was said that it could “potentially spearhead a new and broader field of evaluation practice in Norway” (RCN, 2017b, p. 14). The follow-up report described it as giving for the first time in 30 years an overall picture of Norwegian humanities research (RCN, 2019b, p. 6).

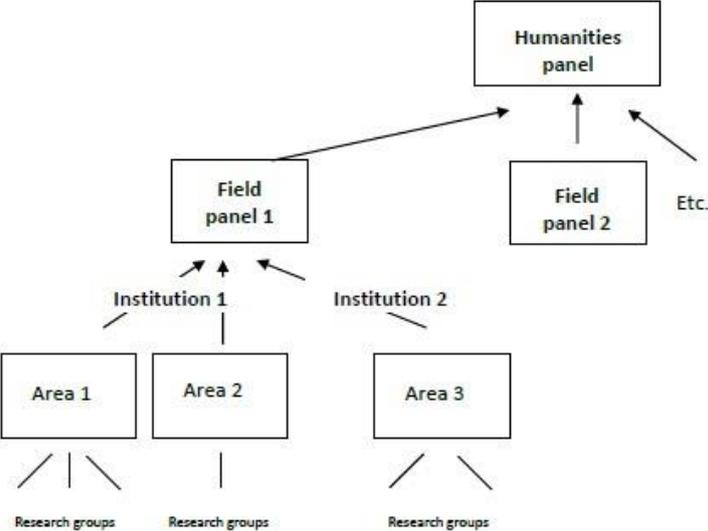
The mandate of the evaluation was given in its Terms of reference. Some of these were to review the scientific quality of Norwegian research within the humanities, provide a critical review of the strengths and weaknesses of the field, and identify research groups of high international level. Regarding the novel element of impact, its mandate was described as following: “Discuss the societal impact of humanities research in Norway in general, and in particular, its potential to address targeted societal challenges as defined in the Norwegian Government's Long-term plan for Research and Higher Education (LTP), and the EU framework programme Horizon 2020” (RCN, 2017b, p. 14).

Evaluations of the humanities had previously targeted subjects such as linguistics, or philosophy and history of ideas, one by one. Humeval was novel in that it targeted the entirety of the humanities, opposed to a few disciplines. The intention behind this move was that it enabled the evaluators to assess the disciplines in an aggregated way, which in turn allowed for a more strategic approach (Wróblewska, 2019). Another much-discussed novelty in Humeval was the addition of research groups as units of assessment (Wróblewska, 2019). Prior to this evaluation, subject-specific evaluations had primarily used disciplines within universities as assessment units. Since then, administrative units had grown and there was a sense of a growing connection across disciplines. To reflect a situation that was deemed to be

more complex and interdisciplinary, the evaluation introduced the concept of research groups as a level of analysis between the individual researcher and the institute (IP4).

To account for the larger scope of the evaluation, the main evaluation report was a synthesis of in total eight sub-reports written by independent field panels (Figure 2). All the subjects within the Norwegian humanities were grouped together in eight based on disciplinary similarities, dubbed ‘research areas.’ Each field panel were responsible for evaluating its designated research area. The field panels consisted primarily of scholars in equivalent disciplines as the disciplines they were evaluating, but a consultancy firm was also assigned to assist logistically and with management support to the evaluation. The consultancy firm provided one member to three field panels, and one to the main panel.

Figure 2: The logistics of Humeval (RCN, 2017b, p. 17)



The data available to the panels were self-assessment reports provided by the evaluated institutions. Each institution was asked to submit one self-assessment for each of the research areas that were relevant to them. The self-assessment included a description of their research activities and results within each research area, the interplay of research and teaching, and an impact statement, also dubbed the “impact cases”. RCN also provided the panels with a bibliometric analysis of all publications by listed researchers for each panel. The institutions were also asked to put individual research groups forward for evaluation within each research area. These groups were evaluated individually, and this assessment was used to support their area evaluations. Because the institutions themselves decided which parts of their organisation

to put forward, the coverage of the evaluation was not complete, but nonetheless judged to include the most significant sections of the humanities in Norway.

In total, the evaluation collected 165 impact cases. The impact cases made up one of the assessed criteria for both the institutions and the evaluated research groups. Impact was assessed by the field panels based on two criteria:

- Reach. The extent and/or diversity of the organisations, communities and/or individuals who have benefited from the impact.
- Significance. The degree to which the impact enriched, influenced, informed, or changed the policies, practices, understanding or awareness of organisations, communities, or individuals.

For each unit of assessment (the institution and the research group) the field panels wrote one full-text evaluation. This text was separated into sections based on each criterion, where impact was one. Each criterion informed the overall assessment for each unit and each unit were given a score using a series on a 5-point Likert scale.

In addition to the panels' assessments, RCN inductively developed an analytical framework to summarize the information that was collected from the 165 impact cases. While the assessment of weaker and stronger cases was a task for the evaluation panels, RCN's analysis informed and facilitated the panels' assessments. This framework consisted of several categories and were applied unison across the entire analysis. The categories used were:

- Research. Did the research stem from a sole researcher or a group of researchers? This was the only category which operated with a single, mutually exclusive dimension of either "individual" or "group".
- Channel. How was the pathway from research to impact described? Several paths were identified, and oftentimes one case included several pathways. RCN identified several sub-categories such as (1) research dissemination, such as media, exhibitions, public debates etc., (2) collaboration with external partners outside of academia, (3) policy advice, (4) professional training and (5) artistic production.
- Beneficiary. Who benefitted from the reported impact? This was the most detailed category and included many different beneficiaries, from "general public", "politics", "school", "health", "minority group" to "cultural heritage". The aim was to provide mutually exclusive sub-categories, but this was not achievable across all the reported

impact cases. Some of them were thus labelled according to the category that was identified to be most prominent.

- Reach. The impact's reach was simply divided into three sub-categories, which neither were mutually exclusive. The sub-categories were (1) local, (2) national or (3) international.
- Effect. The final category concerned the reported effects, which to some extent mirrored the reported beneficiaries of the impact. The relationship between the beneficiary and the nature of the effect is a question of further analysis, it was stated. This category included also a significantly higher number of sub-categories than the others, including labels such as “educational”, “political”, “cultural”, “empowerment”, etc.

RCN's framework enabled the impact cases to be systematized into several tables based on their five categories. In an overview of the reported impact, RCN stated that the analysis to a large extent was based on subjective interpretations, and that the impact cases were seldom written in a manner that simply could be translated into the constructed analytical categories. The resulting preliminary analysis concluded that the most common beneficiary was registered as the “general public”, the impact's reach was most commonly “national”, and that the principal effect of humanities research was registered as “cultural” (RCN, 2017a, p. 12).

In addition to this categorization, the impact cases were also analysed in terms of what type of goals the research underpinned. Two groups of areas or pillars were used for this analysis. One was the Norwegian government's Long-term plan for research and higher education (LTP), which comprised of five thematic priority areas. The other was EU's Horizon 2020 programme which had defined a set of seven societal challenges. The impact cases were sought to be allocated into one of the in total 15 sub-categories. While one case would ideally only correspond to one group of challenges, one case could be represented in both groups, symbolizing a thematic overlap between LTP and H2020. The categorization enabled the construction of several tables illustrating which categories and which group of challenges the impact cases were addressing. The preliminary analysis concluded that more than 60% of the impact cases addressed the challenges from the EU's H2020 programme, whereas less than 30% addressed the Norwegian government's LTP-areas. The main reason of this was that all 72 impact cases were judged to tailor the sub-category “Europe in a changing world” from H2020.

The last part of the analysis of impact consisted of a survey of the external references named in the impact cases. The survey was carried out by RCN, and its main aim was to add users' perspectives as supplementary information for the assessment of impact. The survey was proposed to be read both as collective auditing, as well as providing information on how users get access to humanities research as well as how they assess its relevancy to their own professions. The most significant finding was that 89% of the respondents reported that the humanities was relevant to their work. It was written that the small number of respondents made this part of analysis not eligible for generalizable conclusions, but it did serve qualitative insights into the perspectives of the users.

RCN's analyses informed the field panels, and in turn the main panel's report. In the executive summary of the main report, the panel stated that "the humanities make a strong contribution to society, culture, and the economy by engaging with a range of public and private sector organisations" (RCN, 2017b, p. 7). The panels were favourably impressed by the range and depth of humanities' collaboration with other sectors. The analysis of the impact cases showed that users strongly valued the contribution of the humanities. Impact was found to be strong in some areas but needed development in others. The panel also emphasised the need among respondents to articulate the difference between impact and engagement more clearly. Further in the report, the main panel elaborated that there was a considerable confusion among the partakers involving the meaning of impact and the criteria for assessing it. The panels found that many of the respondents did not sufficiently document their impact, but rather described their engagement and dissemination to the public. Some presented intentions or prospective plans rather than past experiences. The panel recommended that Norway could develop more sophisticated tools for gathering and articulating evidence of impact. It was proposed that RCN could play a central role in this work.

The evaluation included a section where the institutions, RCN and the Norwegian government were given its own sets of recommendations.

- RCN were advised to work with the humanities and other fields to help researchers understand the potential for greater societal impact and how to gather evidence of Impact.
- RCN and the institutions were advised to learn from successful research groups and to share the good practice of organising in groups.

- The Norwegian Government was urged to consider the value of humanities research when identifying and defining their priorities.
- The research institutions and the Norwegian Government were advised to develop stronger strategies and priorities for humanities research.

Other developments within RCN

Organisational shifts have regularly occurred in RCN. The latest was in 2018 when its programmes were abolished, and it started to plan at the higher level of portfolios. Each portfolio is responsible for a thematic priority and have formulated societal goals of which the activity of the portfolio shall work to achieve. In 2016, it implemented the planning and evaluation tool of programme theory. The intention behind this was that it would help the organisation towards a more systematic approach to planning, monitoring, and evaluating its research and innovation funding (Arnold and Åström, 2020). Programme theory was perceived to help shift the focus of its planning towards achieving societal impact.

RCN is an important source of funding for researchers who may have an idea for a project which may not be funded by their research institution. RCN have since its creation made extensive use of its research programmes, but it has also operated with an independent project scheme called Fripro. Prior to 2019, research proposals were either directed at the thematic programmes or to Fripro. This model has however been perceived to disadvantage the proposals directed at the latter, which only had a success rate for funding proposals of eight percent, compared to the thematic programmes which were at 15 percent. Consequently, SSH research was particularly disadvantaged because these were less amenable to the thematic programmes (Derrick and Benneworth, forthcoming). A new system was implemented aimed at creating a singular approach where the best proposals would be funded from either one, and the proposals are rather sorted into either a programme or Fripro at a later stage. With this change, the proposals were assessed in peer review against three novel criteria: Excellence, Impact, and Implementation. Opposed to the research evaluations that evaluate impact retrospectively, these measures are equivalent to prospective impact evaluations.

In conjunction with the Norwegian government's revised national budget presented in May 2022, the Ministry of Education of Research dismissed the entire board of RCN and replaced it with an interim board (Trædal, 2022). The reason behind the dismissal was that RCN was steering towards a serious deficit. The Ministry initiated an external investigation of RCN's economy, announced tighter control, and froze certain research allocations, which

sparked strong reactions among researchers and rectors, and a general fear of a stronger government control of the research sector (Vollan and Larsen, 2022).

Figure 3: Timeline of RCN’s development of impact evaluations¹

Date	Development	Detail
2008	RCN published its strategy for the humanities	Sparked a cultural debate about the role of humanities in Norwegian society.
2013	RCN started to design a new approach to research evaluation	Evaluate subjects simultaneously in an aggregated way. Use research groups as units of assessment.
2015	Proposal within RCN to evaluate impact systematically in succeeding evaluations	REF 2014 model chosen as basis for RCN’s approach.
2015	Impact announced as one requirement for the upcoming Humeval	
2015	Publication of a “long-trace”-evaluation of RCN’s thematic programmes (Ramberg <i>et al.</i> , 2015)	The evaluation was commissioned by RCN and carried out by NIFU. Investigated the traces of mental health and drug research in politics, administration, and society.
2016	The planning tool intevervention logics implemented	It was perceived to assist developing a systematic approach to planning, monitoring, and evaluating its research and innovation funding.

¹ This timeline is partly based on Wróblewska (2019).

2017	Publication of evaluation of the Humanities (RCN, 2017b)	First subject-specific evaluation addressing impact. Impact was used in assessment of units and for demonstrating impact from the entire field.
2017	Publication of evaluation of the social science institutes (RCN, 2017e)	Included an impact element based on the impact case methodology.
2018	Abolished programmes for portfolios	Started to plan at a higher level.
2018	Publication of evaluation of education research (RCN, 2018a)	Included an impact element based on the impact case methodology.
2018	Publication of evaluation of the social sciences (RCN, 2018b)	Included an impact element based on the impact case methodology.
2019	Updated review model	Research proposals admitted based on new criteria, one of them Impact. Proposals directed to thematic programmes or FRIPRO at a later stage in the process.
2021	Publication of evaluation of legal research (RCN, 2021)	Included an impact element based on the impact case methodology.

6. Introducing impact

“[The new ruling system for the Research Council of Norway shall] secure a system with a more strategic and overall rule of the Research Council, with more emphasis on the effects of the Research Council’s activity” (Prop. nr. 1S (2015-2016), p. 184. My translation).

This part of the thesis is where I seek to address the research question I formulated in the introduction. I will in this chapter investigate how an evaluation of the Norwegian humanities sought to make the societal impacts of the humanities research visible. I will treat the evaluation as comprising a form of optics, dismantle and examine its various components and investigate what this evaluation made visible. Another way of stating my research problem is to investigate how governmental demands of documenting effects of research, as illustrated in the quote above, was operationalized into new practices by the Research Council of Norway, now revolving around the concept of impact. My point of departure will be the Humeval report.

As I mentioned in chapter two, when preparing for the evaluation, the idea within RCN to include the element of impact came quite late (Wróblewska, 2019). The British REF had been published a year prior and its design and definition of impact was adopted. RCN considered REF’s definition to be more open to account for disciplinary differences, and thus eligible also for SSH research (Holm and Askedal, 2019). Norwegian evaluations had previously addressed relevancy, usefulness, commercialization, or other similar concepts, but these were often applied to address the economical and instrumental benefits of research (Gulbrandsen and Sivertsen, 2018), which were deemed unsatisfactory for other disciplines such as SSH research. The inauguration of the new concept of impact must therefore be seen as an innovative measure that was implemented to allow for unfamiliar disciplines to be eligible for these types of analyses, as it had been in REF where the definition was used to evaluate the entirety of British research. Impact was defined in the following way:

An effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia.

Although the definition is judged to be sufficiently broad across disciplines, RCN does not by any means consider it to be “perfect” – being able to observe the entirety of Norwegian humanities’ influence upon society. After all the definition illustrates a linear model of causality, where research is presumed to act almost as a billiard ball hitting another and

changing its trajectory. Some research evaluators have criticized the linear model, arguing that the final effects of research are too difficult to evaluate in a meaningful way (Gulbrandsen and Sivertsen, 2018). RCN are attentive to these discussions and remains open for further development of its evaluation methods. My interviewee described the definition in this manner:

It is a simplification of reality. Usually, there will be more people involved, both in the process of creating the research and for it to be applied. And there will also be a long research process that the researchers build upon. The person who publishes the article that has impact, is built upon something that previously has been researched. So, yes, it is a simplification about the entire interaction between research and society (IP4).

The primary material underlying the investigation and assessments of humanities' impact were descriptions and statements written by or on behalf of the evaluated research groups, dubbed the "impact cases". RCN invited the evaluated institutions to submit case studies documenting the impact of their research. The institutions were given a guideline (Appendix A) and a template to use when writing their testimonies (Figure 4). Each institution was invited to submit at least one impact case to each research area. The requirement was that the institution had presented at least five researchers to the field panel. However, every research group had the option to submit, but with a limitation on one impact case per ten researchers in one panel. The guideline emphasised that the research responsible for the reported impact had to be carried out within the last 10 to 15 years. Some exceptions could be given for specific research traditions if the reported impact was of "high international standing". The definition was also given a further notice, stating that it "may include, but are not limited to ...", giving the respondents more freedom to describe how they perceived the notion of impact.

Figure 4: Template for impact case (RCN 2017a, p. 2)

Name of impact case:
Summary of the impact (maximum 100 words)
Description of the research underpinning the impact: (maximum 400 words.) ⇒ include names of key researchers and, if relevant, research groups. A time frame for when the research was carried out should also be included
Details of the impact (maximum 400 words) ⇒ include a description of how the research has contributed to the impact on society
References to the research (scientific publications)
References to sources to corroborate the claims made about the impact (publications, reports, media items, policy papers, etc.)
If relevant: External references (external users or others who have witnessed the impact and could be contacted to corroborate the claims made in the reported research cases).

Prior to the evaluation, RCN invited representatives from the evaluated institutions to a workshop. The aim was to explore how the institutions could use the impact case template to describe the pathways from a research project to its impact according to the definition. A literary scholar from Cambridge University introduced the topic to the audience. She had been central in her own faculty’s submission of impact statements to the British REF evaluation. During the workshop, RCN observed that the discussion among the partakers shifted from an essentialist question of what the humanities are, or what humanities’ influences on society generally may be, to a more pragmatic attitude towards documenting impact. RCN’s attitude is that the establishment of a new practice of documenting societal impact is not the same as understanding and conceptualising a certain phenomenon such as impact (Holm and Askedal, 2019). My interviewee described the proceedings in the workshop in this manner:

I would say we went from a discussion about what humanists [humanities researcher] are, or what the humanities is, so it is the knowledge about language, history, culture, religion, everything that has anything to do with humans and therefore it is in a way relevant for everything, and there was someone who said “how could we talk to each other if we didn’t have linguistics”, which is nonsense because language is a natural thing. Anyway, we went to a question about “when, where and for whom”, right. It became concrete and situated in society. And there was a resistance to having such a small approach, an approach that would not illuminate or document the full extent of humanities’ importance for society, but it was nonetheless an approach that would make visible that the humanities sometimes actually change society in a way that, non-researchers, or non-academics, realize – or gain significance for non-academics. In the

end of the process all parties were actually satisfied with the good collection of examples. What characterized it, was a very large scope (IP4).

Treating evaluation as a form of optics implies that a range of components are assembled so that something that previously was not visible, is made visible for the human eye. It also implies a partiality of gaze. Optics need to be adjusted towards something in particular – as with its focus. Understanding evaluation in this way does not seem to counter the way RCN make out of their evaluations. There is a clear sense that the definition of impact does not perfectly reflect the world *out there* and that research may influence society in ways that this model cannot observe. In that way, the optics are not all-encompassing and able to observe the entirety of what is sought. Furthermore, there is also a sense of a pragmatic attitude towards their definition; It *works*. This pragmatic attitude can be observed in how the interviewee described the achievements of their workshop. This shift from “what is X” to “when, where and for whom” was of the essential. This pragmatic shift was judged to be crucial for the evaluation’s success (Holm and Askedal, 2019). What was once seen as somewhat of a complicated and diffuse task from the partakers had been accompanied with a more thorough and detailed manual. The workshop was a way to adjust the evaluation’s optics towards the particular. Its achievement was a consensus on how the impact cases were to be written and reported. More fundamentally, it was also a move towards the new defining features of impact. The evaluation’s perception of the term was directed at the concrete and verifiable, the contributions and influences that could be confirmed or falsified. The conception that humanities do have a fundamental influence on our language, history, culture, religion etc. which are difficult to document and to trace was not rejected, it was rather that these aspects were deemed irrelevant for this evaluation. What the workshop did was to introduce a new, unfamiliar, and perhaps pragmatic conception of how research interacts with society among the workshop’s participants.

RCN’s ambition when evaluating impact is not to document or highlight the impact of one research group, institution, or discipline, but rather illuminating the aggregated impact of Norwegian research (IP1). RCN’s analytical framework further illustrates this ambition. It also highlights the great value lying in these collected impact cases. When put together and summarized this ambition was achievable. If done correctly, the most significant, if not the entirety, of humanities’ impact could be captured. By treating the collected impact cases as *inscriptions*, the various practices and apparatuses that facilitated their emergence and their analysis I will refer to as *inscription devices* – tools and apparatuses that facilitated the

transformation of phenomena into writing (Latour and Woolgar, 1986). The central components of the optics were these devices.

Of these inscription devices were the guideline and template provided to the respondents. The respondents were all given the same two pieces of documents. The guideline included the definition of impact and gave examples to what this could include. The workshop was similarly also an inscription device that assisted to achieve a further consensus on how to write the impact cases. Together, the template, workshop, and guideline, worked to *reduce* the object of inquiry so that the testimonies could include the particular facets which were deemed useful for the aim of the evaluation. In such a way, they would enable the transformation of narrative prose or testimonies into samples or inscriptions. An important ability of inscription devices is that seek to secure that the inscriptions keep a somewhat similar nature, making them eligible for comparison. Ideally, inscriptions can possess abilities as *immutable mobiles*, they can be transported from its surroundings to an area of analysis, where they can be compared, assembled, and summarized (Latour, 1987). This seems consistent with the proceedings of Humeval.

Another group of inscription devices was set to create equal inspections and analyses of the inscriptions. These were implemented to capture the direction or trajectory of the humanities' impact. RCN created an analytical framework that facilitated the process of comparison, delineation, and synthesis. This instrument was built inductively based on the inspection of the inscriptions, and it included five domains. As a result, the impact of a research project was processed through this instrument which in turn extended and *amplified* the inscriptions to each other by translating them into numbers on tables for each domain. This process was what enabled the formulation of general findings. One important finding which it had illuminated was the wide range of beneficiaries for the Norwegian humanities. Although the bulk of it benefited the category "general public", there were 13 other categories, some of which were deemed unusual for the humanities to interact with, and which surprised the evaluation panels (Figure 5).

Figure 5: Table of beneficiaries of Norwegian humanities (RCN, 2017a, p. 10)

Beneficiary:	Panel 1	Panel 2	Panel 3	Panel 4	Panel 5	Panel 6	Panel 7	Panel 8	Total
General public	5	9	7	15	16	8	2	3	65
Politics	1	3	0	7	8	8	1	6	34
Diplomacy	0	0	0	3	2	0	1	0	6
NGO	0	0	0	1	2	0	1	1	5
School	2	12	0	6	5	2	4	1	32
Health	3	4	0	1	1	5	0	0	14
Cultural institution	4	3	1	1	3	0	0	1	13
Judicial institution	0	0	0	0	0	2	0	0	2
Private company	1	1	0	0	1	0	0	1	4
Industry	0	4	1	1	1	0	0	7	14
Religious community	0	0	0	2	0	1	4	0	7
Minority group	1	3	0	3	1	0	2	0	10
Environment	0	0	0	1	0	1	0	0	2
Cultural heritage	0	0	0	0	5	0	0	0	5
Total	17	39	9	41	45	27	15	20	213

Furthermore, the inscriptions were also processed and delineated into categories based on what type of challenge they were targeting. 44 of them were categorized to the thematic priority areas of the Norwegian government's LTP, while 102 were categorized to EU's challenges from the H2020 programme. The remaining 21 were not eligible for the categorization. The overall conclusion resulting from this process was that Norwegian humanities were surprisingly better tailored to address the societal challenges set out by the EU, than the priority areas formulated by the Norwegian government. The most relevant challenge with most of the inscriptions attached to it were EU's 'Europe in a changing world'. 70 percent of the inscriptions were delineated onto this particular challenge, a challenge which did not have any parallel in the Norwegian priority plan.

Another part of the evaluation involved an inspection and assessment of each impact case. This was the task of the field panels. Each panel ranked each collected impact case based on the criteria of reach and significance, criteria which were transferred from the British REF. Whereas the impact cases in one part of the evaluation were an inscription used to illuminate the aggregated impact of the humanities, they were in another part of the evaluation used as a piece of artefact subject for assessment and judgement. This part of the evaluation alluded to Foucault's disciplinary technique of normalizing judgement (Foucault, 1995, p. 180). The impact cases were given a score on a scale, resulting in a comparison of each impact case to the others. Furthermore, this would involve the arrangement of a standard, a mean of achieved impact. If a project did not live up to this standard, the evaluated unit could expect to have its overall score subtracted.

However, this particular part of the evaluation failed. The general perception among the evaluators were that the bulk of the impact cases were not mature enough to be rigidly assessed based on the mentioned criteria of “reach” and “significance”. The facets of which the evaluators were to assess were not included in the impact cases. An unexpected result of this was that the panels did not make public the scores that had been assigned to the individual impact cases (Wróblewska, 2019). Another weakness stated in the report was related to its use of research groups as units of assessment. The number of research groups submitted to the evaluation varied considerably. For the evaluators, it seemed like some groups were constructed artificially for the purpose of the evaluation, whereas others appeared to be groups of people who normally worked together. The approach of using groups was judged to not always reflect the ways in which humanities researchers work, a field where individuals scholarships was conceived to be more normal than in other sciences (RCN, 2017b, p. 20). Together, these two illustrated some of the weaknesses of the evaluation’s optics, as it was not shown to be perfectly adjusted towards what it was trying to capture, observe and assess.

In sum, the impact cases went through two different processors. In the latter processor, the impact cases were assessed and judged based on two sets of criteria, which in turn would enable the establishment of an impact standard or impact mean. This part of the evaluation however failed due to the “immaturity” of the assessed material, the impact cases. In the former, the impact cases were inscriptions used for purposes of synthesis and analysis of previously unknown features of the humanities. A process of *reduction* assisted to make testimonies into inscriptions whereas the inscriptions were *amplified* by extending them into the others. A range of innovative measures had made this overall achievement possible. The definition of impact, the transmitted manual to respondents, the analytical framework, the logistics of the evaluations, together comprised a new form of optics that not only considerably increased the scope of the evaluation, but also was attuned to and made visible something that previously had not been observed.

What was made visible

The inspection of the inscriptions had surprised the evaluation panels, who were favourably impressed by the humanities’ collaboration with other sectors. One notable achievement, according to one of my interviewees from RCN, was that it disproved the notion of a closed circuit regarding the humanities (IP4). Counter to widespread, derogatory ideas regarding the humanities, these disciplines do in fact influence society outside academia, and not only to small sections, but rather to a surprisingly large area of society. The evaluation also showed

that impact was found across all disciplines. It was not the case that some disciplines had impact, while others did not.

The inscription devices were essential for this achievement. What was once a descriptive testimony of a researcher's activities, had been processed by and through an analytical apparatus which turned a piece of narrative prose into a number on a graph. The various transformative steps which turned testimonies into inscriptions, inscriptions into numbers, and numbers into tables, in turn facilitated the formulation of general conclusions, such as that "the humanities make a strong contribution to society, culture and the economy by engaging with a range of public and private sector organisations" (RCN, 2017b, p. 7). This process I described with the twin concepts of *reduction* and *amplification*, made possible that these findings subsequently could be transported out from the evaluators' offices to the outside world to convince others about their validity.

A few months prior to the report was published, RCN provided the Ministry of Education and Research with some of the findings based on the preliminary analysis of the collected impact cases. The Ministry, informed by findings and recommendations in the report, emphasised in a white paper to the Storting (Norway's supreme legislature) that today's challenges are too complex to be solved by a handful of disciplines alone, and that humanities researchers possess knowledge, skills and tools that are not used enough. "To emphasise the humanities' suitability as instruments to achieve other purposes does not mean to disapprove the humanities' importance for democracy and formation, nor their value as source for understanding and delight", the letter stated, before it quoted the Bible: "These you should practice, without neglecting the others" (Meld. St. 25(2016-2017), p. 69). The evaluators' analyses illuminated in the report had been transported into the offices on a policy-making level, even before its formal publication.

A few years later, in a seminar presentation held by RCN in 2018, the representative presented the Humeval report with the title "Humanities in Norway make a strong contribution to society" (Holm, 2018). The representative from RCN cited several deans in the humanities who had issued statements in the public debate in 2011-2012. One dean emphasised that it is hard to demonstrate the direct instrumental value of the humanities. RCN's representative issued seven reasons for why it is so difficult to create lists over humanities' achievements. Some of these were: the lack of tradition to create ranking lists, the large differences between disciplines, the long period of time the humanities need to influence the wider society, and that it was difficult to measure its instrumental value. In red letters beneath these seven reasons, the presentation stated "Fortunately, we were wrong!", before

the evaluation report and its findings were presented (Holm, 2018). RCN's presentation illustrates what notions the evaluation report had undermined. The achievement of the report was that it had solved the issue regarding disciplinary differences by using a sufficiently wide definition of impact so that it worked for the entirety of the humanities. The humanities' achievements, influences, and effects upon the wider society were not only present and important, but also traceable. The common self-understanding implying that the humanities only deliver essential information which cannot be traced directly, was debunked. The transportation of the report's findings perhaps culminated in that the humanities *were* eligible for these analyses, in similar ways as other more applied sciences were.

However, the representative also presented one of the major defects the evaluation had illuminated: There was a need among respondents to articulate their differences between impact and engagement more thoroughly. This was in fact a repeated shortcoming in the report. The evaluation panels had noticed a considerable confusion among the respondents regarding the meaning of impact and the criteria for assessing it. Sometimes the impact cases only included public communication and engagement, while other times they only presented intentions or prospective plans. Although the panels were impressed by the high number of researchers who were involved in engagement among potential external users, this did not qualify as impact; dissemination was only a pathway, not a sign of actual impact. In many of the impact cases that did show signs of impact, documentation was not provided, and these failed to provide reliable evidence. An unexpected result of this was, as mentioned, that the panels did not make public the scores that had been assigned to each individual impact case, because the shared perception was that they were not mature enough to be rigidly assessed based on the used criteria (Wróblewska, 2019, p. 14). The late inclusion of impact, as well as a lack of extensive guidance and preparation among the respondents, has been pointed at as reasons to why the impact cases did not live up to the desired expectation (Wróblewska, 2019, p. 34).

I see these remarks on confusions suggesting that the process of *reduction* in parts had failed. Many of the testimonies were simply invalid data. Despite the inscription devices provided to the respondents with the aim of collecting pertinent information eligible for comparison and analysis (the template, guideline, and workshop), the evaluation had not sufficiently succeeded to extract the particular facets which the evaluation primarily had sought after. The fact that the field panels' assessments had partly failed, while RCN's inspection to a larger extent had succeeded, seemed to stem from RCN's instrument being built inductively, while the assessment criteria were predetermined and brought over from the

British REF. Did RCN's analysis achieve to illuminate humanities' impact or only its engagements? The report from the main panel did not make this entirely clear. While instances of "strong impact" had been found, the overall attention of the main report revolved around improving the reporting and documentation of impact among the evaluated units.

The invalidity of the data seemed to lead the main evaluation panel to another central conclusion: "Humanities in Norway does not have a strongly developed evaluation culture ... Parts of the community have limited experience of how to deal with an evaluation and how to communicate with the evaluators in ways that will enable positive judgements" (RCN, 2017b, p. 19). Having identified an underdeveloped evaluation culture, the evaluators recommended the institutions to develop stronger strategies for humanities research. Most institutions were found to lack an "impact strategy" (RCN, 2017b, p. 37). RCN was also envisioned as taking a central role in the improvement of impact documentation.

In sum, what had the evaluation illuminated? The report had showed that the humanities do in fact engage with and contribute to a range of beneficiaries outside of academia. While the evaluators failed to reduce all the inscriptions to the particular facets that could sufficiently inform the optics' opacity and enable the assessments of impact, instances of impact had been identified (RCN, 2017b, p. 36). The survey of external users had also confirmed that most of the respondents considered the humanities as relevant to their professions: The Norwegian humanities do show instances of impact. What the report also had shown was that the new methods, tools, and overall design of the evaluation managed to capture and make visible something that previously was not deemed to be observable: The optics worked. Alongside this, however, the evaluation also illuminated that the humanities had significant rooms for improvements when it came to reporting on this particular phenomenon. What had been observed was an underdeveloped evaluation culture. The evaluation report both described these shortcomings and identified how these could be addressed. To increase the focus of the evaluation's optics, there was a need for the inscriptions and the inscription devices to be further improved. While the evaluation report had made visible instances of impact and an underdeveloped evaluation culture, it, by having its own sets of recommendations, simultaneously attempted to intervene on its object based on its observations. The improvements were primarily identified at the hands of the respondents.

7. Intervening with evaluation

In the last chapter I was interested in how the evaluation worked as a technology of vision, as a tool to make something observable and analysable – a tool of knowledge. In this chapter I am more attentive to how it worked as a tool of politics. Evaluations encompasses both descriptive and intrusive features, which are then translated by other actors into new proposals and arrangements. The ability of evaluations to enable and implement policies opens for a more detailed understanding of how politics functions and how technical features such as evaluations are indispensable parts of political and bureaucratic action (Reinertsen, 2019, p. 21).

The Humeval report did as promised and spearheaded a new and broader field of evaluation practice in Norwegian research evaluations. Several of RCN's evaluations have subsequently adopted the impact case methodology (RCN, 2018a; RCN, 2018b; RCN, 2021). In the subsequent evaluation targeting the social sciences, the evaluators observed similar misunderstandings between engagement and impact, and many of the respondents only reported on dissemination activities, which by the evaluators only represented the first step of creating impact (RCN, 2018a, p. 33). The collected impact cases from these evaluations have nonetheless been recollected and reanalysed with other conceptual tools in commissioned evaluations (Solberg *et al.*, 2017, Solberg *et al.*, 2019). On policy-level, the expectations that the humanities should contribute to solve societal challenges increased with the mentioned white paper (Meld. St. 25 (2016-2017) and as a response, RCN has implemented a significant strengthening of humanities' share in the thematic priority areas (RCN, 2019b). Among the institutions, there are signs of that the evaluation has proliferated change in their strategies (Universitetet i Bergen, 2018), while a shift of attitudes has been observed among researchers (Wróblewska, 2019, p. 19). After Humeval was published, RCN proposed it as being in the institutions' self-interest (IP4) to improve on what the RCN conceived of an "underdeveloped impact infrastructure" (Holm and Askedal, 2019, p. 141). This concept was in turn adopted from Wróblewska who described it as a systematic approach to stimulate and document impact (2019, p. 52).² A developed impact infrastructure would involve administrative, managerial, and academic staff who worked to write the impact cases, while also train and assist researchers to improve their abilities to observe and report. The establishment of certain

² Note how research on impact evaluation feeds back into the practices of the field itself, further blurring the distinction between literature and material used in this thesis. This also illustrates that academic research about impact evaluations has performative effects, much like the impact evaluations themselves.

incentives, such as impact-focused grants, was also proposed to facilitate and increase research communities' impact.

The examples I just described makes up some of the initiatives which has been subsequently mentioned and linked to the publication of Humeval. Humeval's performative abilities might be observed in how policymakers, institutions, and individuals, and RCN themselves, are adapting to the reality which the evaluation had made visible. By discussing these issues, I am paradoxically addressing the impact of the Humeval report, and, as in impact evaluations in general, challenged by the question of *attribution*. The question whether I can attribute all these changes to the Humeval report is like the difficulties the evaluators had in their inspections of the impact cases. Where I judge it to be unlikely that these initiatives would have been possible if not a report such as Humeval had introduced the phenomenon of impact into a Norwegian setting, the pursuit for impact is, however, enforced by larger international trends, where even universities themselves have established institutes with research impact as its main pursuit (Universitet i Oslo, 2016). In that regard, attributing all this to solely one report seems absurd, considering that Humeval is but one instance of many that seeks to demonstrate research impact. I am more inclined to say Humeval was both a result of and a driver of impact pursuits.

Disseminating an evaluation culture

Instead of trying to solve the problem of attribution and determine the precise impacts of the Humeval report, I will rather analyse how the evaluation possessed features that affected its objects of evaluation. Another way to say that is what Humeval did, perhaps more illuminatingly through its transmission, it also did during its creation. If one of Humeval's achievements was the introduction of the phenomenon of impact into a new terrain of researchers, one central feature was to persuade and convince not only its readers but also its objects of evaluation about its existence and validity. The evaluation both made visible an underdeveloped evaluation culture, but as I will show in this chapter, it simultaneously sought to create and disseminate such a culture in its proceedings. In addition to trying to intervene through its recommendations, the evaluation was in this light also an intervention into the Norwegian humanities during its formation.

That evaluation also possesses elements of intervention is something that RCN are completely aware of. My interviewee described this potential in this manner:

When the Research council evaluates research, we emphasise that evaluation shall be the base for advice to the ministries – research policy advice for example. Then, we

are concerned with that side of the coin that is about documenting the societal effects of humanities research so that we can argue for more funding to strengthen humanities research, and preferably in areas that we see have societal value. But we are also concerned about how the Research council's development of policy instruments affect behaviour. We have had a huge discussion on open publication, for example. What happens with quality assurance of research if all research becomes open in journals? How the research council's requirements about open publication, which is about that research should be available for people out there, that is the goal, but on the way there our guides are affecting behaviour. And it is the same with the evaluations, but I think in the evaluations we are more aware of that we also can use them to affect behaviour and generate discussions that we think is in line with the research policy's ambitions. I will say that it has gone from a narrow focus on how to document quality in an academic context to developing a wider focus on research's importance for society. And only by us demanding research's importance for society in our evaluations, we are affecting behaviour (IP4).

If one of the Norwegian research policy's main ambitions is to shed light on how research changes Norwegian society, the challenge for RCN became to persuade the humanities researchers to adopt a particular gaze that made these causal relationships possible to observe and document. The fact that the objects of evaluation were responsible to write the inscriptions made this matter of the essential. Whereas Reinertsen remarks that the evaluation tools in Norad had the potential to not only change the way aid was being done but also how aid staff were seeing themselves in relation to their project site (2016, p. 80), the evaluated researchers in Humeval were also nudged to think and describe their work and roles as researchers in new, unfamiliar ways. Some of the Norwegian respondents described the effort of writing the impact cases as an exercise that enabled them to look at their work in a new and different manner (Wroblawska, 2019, p. 44). Some even started to appreciate previously overlooked areas of their research activity after the exercise. The respondents emphasise the experience of looking at one's own work through the prism of a new professional vision. The writing exercise was crucial in this respect.

This element that my interviewee in the last chapter elegantly described as “when, where, and for whom” is what I want to consider as the core of this particular gaze. The researchers had to reorient their own perceptions regarding the relationship between their role as researchers and the world outside academia, towards the concrete and verifiable. The

impact case writing exercise was a means to introduce the new notion of impact among humanities researchers, which had a potential to alter and transform the objects of evaluation. The fact that it was difficult for the evaluators to assess a part of the inscriptions imply that an insightful evaluation with an increased opacity, rests upon its inscription writers having adopted this gaze. The facilitation of a rigorous external view on the impact of the humanities was thus necessitated by the inscription writers having undergone this transformation. This was its major prerequisite, and although the evaluation in its preparation and its definite stages had achieved to transform some of the inscription writers, the major weakness of the report seemed to be that it had not transformed enough. I will let the report further elaborate:

Humanities in Norway does not have a strongly developed evaluation culture ... Parts of the community have limited experience of how to deal with an evaluation and how to communicate with the evaluators in ways that will enable positive judgements. This is particularly in relation to the use of impact statements, which is a novel technique everywhere. Clearly, those with a more developed evaluation culture will be better placed than others to receive a positive evaluation (RCN, 2015b, p. 19-20).

The term “evaluation culture” is not further defined in the report, but as the transcript illustrates there is a sense that within a developed evaluation culture, respondents would be more familiar and better positioned to know how and what to communicate to evaluators. It may be understood as referring to norms, values, and habits related to evaluation. A lack of willingness, or elements of resistance and refusal to carry out evaluation activities may be interpreted as evidence of a lack of an evaluation culture (Dahler-Larsen, 2015, p. 164).

For the evaluation panels, a developed evaluation culture would allow for positive judgements. For RCN, one major aim for the evaluation was to make visible the contributions of the humanities, so that policymakers and the general public to a greater degree may appreciate the contributions from humanities research (IP1). According to Callon’s four moments of translation, the first moment *problematization* is when the initiating actors make themselves indispensable in the network. The first moment has the potential to create what he calls an *obligatory passage point* (Callon, 1984). By determining a set of actors and define their identities in such a way, both RCN and the evaluation can be defined as obligatory passage points for the evaluated researchers. Together, their aim is defined as to preserve the interests of its targeted actors.

However, the targeted actors may refuse the definitions that are being offered to them. The debate prior to the evaluation, which I briefly covered in chapter two, seemed to embody

some of this refusal. A translation process may fail at any moment, and Callon employs *dissidence* to refer to conflicts or betrayals caused by the inconsistency between the spokespersons and the actors they seek to represent. The interviews of the evaluation's respondents give insight into aspects of this process (Wroblewska, 2019). While the majority of the interviewees had positive reactions to the evaluation and considered it to be boosting the profile of the humanities, one of the six interviewed academics expressed strong criticism based on fundamental objections to the nature of the evaluation. It was seen as directly contradictory to the ethos of academic work in their discipline (Wroblewska, 2019, p. 38). A somewhat similar viewpoint was also highlighted by one of the external referees that was asked to participate in the survey for the evaluation. They questioned the desire for research to be impactful as leading to "narrow and unimaginative understandings of what research should be about" (RCN, 2017b, p. 7).

On another note, after the publication of the report several institutions published their impact cases on their websites, and a number of researchers also included them in their CVs (Holm and Askedal, 2019). The impact cases were in those arrangements used for promotion purposes and for researchers' presentation of selves. What was once an inscription for an evaluation report, were suddenly transported out of the evaluation report and transformed into another piece of artefact, serving a new purpose in a completely new arrangement. According to my analysis, this may illustrate a successful translation, but it may also illustrate the potential in evaluating, where it may add to or redefine previous normative notions of what a researcher is or should strive to be.³

Whereas RCN had observed some resistance among the evaluated researchers about the impact assessment, they noticed that this resistance had been reduced after the report and its findings were published and transmitted (IP3). I want to suggest that this relates to the two different processes which the impact cases went through. As inscriptions used for synthesis, the researchers' sole testimonies were of less importance. It was the aggregated impact of the humanities that were being illuminated, which in turn had the potential to highlight humanities' role in society and increase the value of the humanities among the public and among policymakers. This was an imperative for greater respect and perhaps a larger proportion of funding for the field. A way to translate and facilitate the objects of evaluation to adopt this particular gaze was thus to emphasise the part of the evaluation that would

³ Dahler-Larsen (2015) suggests five domains where constitutive effects of evaluation may be observed. These include the content on some work or practice and the configuration of social roles and identities. These are suggested to enrol each other into a larger evaluative assemblage.

highlight the importance of the researchers' aggregated endeavours – that it was in their own self-interest to adjust their view and report on their own activity in this way. If the exercise did have incentivizing features in this part of the process, it was not to change the ways of doing research, but rather to adjust the researchers towards seeing and reporting on their research in new, unfamiliar ways.

When, however, subject to a normalizing judgement, the researchers were ranked according to a new, unfamiliar set of criteria. When put through this part of the process it not only incentivised the researchers towards viewing their research in new ways, but also possessed incentivizing measures to alter researchers to create more impact. It was perhaps not so strange that RCN had observed a reduction in respondents' scepticism after the report had been published, given that the report had removed the individual scores for each research group and that the aggregated impact of the humanities was described in such a positive light – with a “positive bias” (RCN, 2017b, p. 20). By showing that impact was not missing or completely foreign for the field but rather present and existing, it was far easier to convey the message that this was a desired feature for humanities researchers to further strive to achieve and improve on. In sum, it was in their own interest to adopt an evaluation culture.

This is not to say that the evaluations' remarks on the lack of a developed evaluation culture was only due to the dissidence I described, but rather that what I have called a translation of its objects of evaluation was not only a prerequisite of the evaluation, but also one of its features. While RCN and its evaluations are essential for the formulation of research policies, they must also be seen as essential for its realizations. In an evaluation of RCN's activity performed by a consultancy firm, this potential was expressed as RCN to be a “change agent” (Arnold and Mahieu, 2012). In this light, RCN's evaluations are tools of politics. If the humanities were to increasingly take part in solving the societal challenges of our time, as the white paper had envisioned (Meld. St. 25 (2016-2017)), the researchers had to reorient their gazes and their research to the observable and the verifiable. If not, no impact could be documented. And as a falling tree makes no sound unless there be ears to hear, undocumented impact is no impact at all.

In Latour's depictions of the famous microbiologist Louis Pasteur, Pasteur used the strengths of the laboratory to get the microbes to do what he wanted. In the laboratory, the microbes could be isolated, subjected to trials of strength and manipulated. By building an alliance with microbes, Pasteur managed to convince others outside the laboratory that his findings would be beneficial also for them – that they had to pass through his laboratory equipment to reach their own goals (Latour, 1988). Although evaluation encompasses human

researchers and not microbes, the similarities are present. Through inscription devices and acts of documentation, RCN manipulated its objects of evaluation to adopt a specific outlook on how their research interacts with the outside world. By transmitting this as being in their own self-interests, RCN succeeded to translate and mediate governmental demands among the evaluated units.

During working with this study, I have been tempted to ask: Was the evaluation's aim primarily to make visible humanities' impact, or was it primarily carried out to intervene and transform humanities researchers into improving their impact? I was, however, reminded by Dahler-Larsen's emphasis that it is practically impossible to determine a certain effect as unintended, nor is it analytically useful to do so (2012, p. 202-203). Following his terminology, I may instead argue that the transmission of an evaluation gaze, or the dissemination of an evaluation culture, was a *constitutive effect* of the evaluation. This would include nudging researchers to be more attentive to observe impact, but also to create more of it. Using ANT, I have described the evaluation's prerequisites and main features as translations.

8. Planning for impact

So far, I have described one particular evaluation report initiated by RCN as a tool that through processes of translation made particular features of the humanities field in Norway visible, while simultaneously a tool of politics that translated its objects of evaluation into adopting a particular gaze that could enable and improve both evaluations' opacity, and furthermore the humanities' impact in the future.

In this chapter I will leave Humeval and turn my attention towards RCN's applicant forms. Impact have not only been introduced in RCN through various evaluations which I have mentioned, but also in its application process. As I briefly laid out in the introduction section, RCN is an important source of funding for researchers who may have a new and innovative project which may not be covered by their home institution, and RCN have implemented a new model when reviewing these applicant forms. With the new model, research proposals were assessed in peer review against three novel criteria: Excellence, Impact, and Implementation. Applicants should prepare proposals for each criterion, and peer reviewers are asked to give scores to each. Opposed to the evaluations I have addressed in this thesis, this line of work seeks to evaluate impact prospectively, or "before the event". The template for the impact criteria in the applicant forms is provided in Appendix B, while the criterion for its assessment is provided in Appendix C.

Reducing projects to logical relations

Being at the applicant stage, the most important purpose of these impact assessments is to separate the wheat from the chaff – to seal out the best projects with the goal achievement abilities that are judged to be most plausible and desirable. Although this responsibility lies in expert panels' peer review, it is RCN who define the criteria of their assessments.

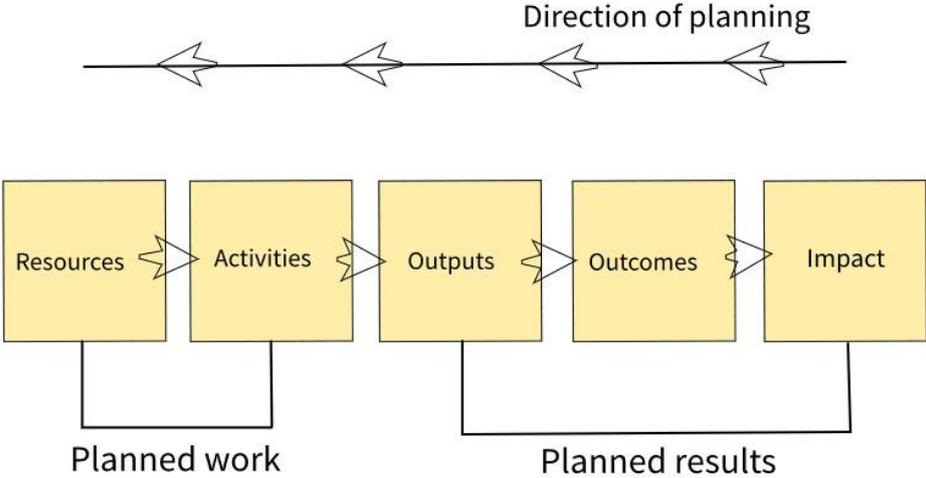
It is notable that impact is defined in another way in these assessments, compared to how it was defined in the evaluations I addressed previously. The definition is more open, in that both impact upon the scientific community and upon the wider society is deemed as relevant. The applicants are thus not obliged to describe how their projects will achieve impact outside academia, but RCN rather leaves this as optional for the applicants. This illustrates the wider conception of impact within RCN, that not all research is expected to have societal impact, but that those that do should be noticed and potentially be rewarded for carrying out research with this feature. However, RCN have observed that applicants include a description of their project's potential for societal impact possibly with an intent that this

may increase the probability of being granted, even though this is emphasised to not be a requirement for approval (IP4). What may be a constitutive effect of the novel impact criterion is that applicants regard it as beneficial for the probability of admission, and hence reorient their project to be impactful outside academia, despite RCN's emphasis that it is only optional.

If societal impact is addressed by the applicants, the extent to which they address important and/or future challenges, as well as the extent to which these are clearly formulated and plausible are to be assessed by the expert panels. The impact criterion implicitly acknowledges, perhaps as a result of reading a great number of proposals, that a proposal may not sufficiently tie their planned impact upon a defined challenge, and that the plausibility of its realization may not be sufficiently addressed by the applicant.

I regard the template as a tool to enforce the applicant into adjusting their line of resonating in a new, concretized, and verifiable way. The template operates with a specific logic of the relationship between research and society, distinguishing between “output” and “impact”, and asking the applicants to define and target specific stakeholders and beneficiaries of their planned project. This is consistent with an underlying logic and terminology from programme theory, which I briefly covered in chapter five, and which was implemented into RCN and its portfolio plans in 2015. The applicants are required to describe themselves as a project designer who plans to intervene on a specific section of society with their research project. The project's activity is understood as an intervention with a potential to change its predefined target group or beneficiaries through a change in their competence, behaviour, practice, or politics. This is understood as the project's outcome. These changes are then thought to lead to long term changes on a societal level (impact) as a result (RCN, 2022b). The applicants are thus required to describe their project according to an underlying simple form of a Logical Model (Figure 6), and to develop a theory on how their project's results will in turn lead to impact that is in line with a pertinent challenge. A proposed scenario for future application templates to further increase the plausibility of achieving impact, is that RCN provide an even more prescriptive guideline based on the Logical Model, making the applicants describe each step in the model (Derrick and Benneworth, forthcoming). The applicants would then start from what impact they envision, before they work through the model and propose what outcomes, outputs, activities, and resources which would be necessary for the intended impact to be achieved.

Figure 6: Logical Model⁴



An important feature of the impact element in the applicant forms is that it is not only the project idea that is being put forward to RCN but also the researchers’ assumptions about how the project and its activities will lead to subsequent effects. The underlying Logical Model is equivalent to an ontological framework. It presents an assumption about a series of events and implicitly addresses how this will occur and consequently can be documented, while also taking for granted that the circumstances and environment of which the research is carried out within remain static and stable throughout the duration of the project. Researchers are expected to be fortune tellers, as one postdoc described it (Bjørkdahl, 2019).

By viewing the applicant template as an inscription device, project proposals are subjected to a process of reduction where they are partially transformed into a logical relation between predefined elements, such as “beneficiary”, “outputs”, and “impact”. If granted, the researcher is furthermore required to report on the project’s achievements at successive moments throughout the duration of their project according to this new set of terminology. This enable a track record of each granted project where the proceedings of a given project are written in the same manner and with the same underlying ontological basis. By saving this track record in an archive, the traceability of a project can be saved, which allow for inspections irrespective of the project’s time span and geographical location. Such documentation practices are reminiscent of the establishment of a *circulating reference*, a chain of translations linking the project to its site at successive moments in time and allowing

⁴ I built this basic Logical Model based on Derrick and Benneworth’s (forthcoming) proposed model for the RCN.

for the project's lifetime to be retraced. This creates a whole new opportunity for evaluators to inspect and possibly identify the moment where and how a potential failure of impact occurred. Did the engagement activities fail to initiate a change in stakeholders (implementation failure), or did the change in stakeholders not lead to the intended impact on a societal level (theoretical failure)? In sum, it greatly increases the documentative ability of a project, enabling more rigorous evaluations compared to the retrospective evaluations I covered previously in this study.

What then, does RCN do with this archive of project documents? My interviewee explained that there are several purposes for this documentation.

What we use the progress and final reports from the funded projects for is to report to the Ministries and develop new tools or adjust the ones we have. What we are also working on is how we, to a greater extent, can extract the most relevant information from the progress reports. We are also running portfolio management in the Research Council now, which means that we to a larger extent have defined societal goals for each portfolio. Then, it will be interesting to investigate the projects. What are they reporting? Are they reporting effects within what is defined or are there completely different effects that are emerging? [...] The prospective in the coming evaluations will be about advice regarding the institutions' strategies: How realistic are the strategies they have, given their resources, plans and competence. The evaluations can give advice to the institution's strategies for the future (IP4).

The primary objectives are described as accountability and learning. A granted project receives public funds of which the researcher(s) must be made accountable for, while the development of tools and strategies may be informative for how the various research institutions formulate their strategies and document their own achievements. This illustrates the mutual objectives of accountability and learning in evaluation practices, but also how evaluative insights may feed into policies and research strategies.

Creating impact devices

Another supplementary view of this shift in proposals is to adopt the emphasis on the performative abilities of documents. Doganova and Eyquem-Renault (2009) suggests viewing business models as market devices, as actively contributing to create and change that which is described. By viewing the applicant forms in a similar light, the impact element may transform the applications from mere plans and ambitions, into devices or tools. If circulated

and used in a given project's impact communication strategy, a project proposal is no longer only mere descriptions of intentional plans, but an artefact with performative and intrusive features that may subdue and lay the ground clear for initiating the proposed changes before the applicants' projects are completed. While a project proposal is reduced to a logical relation between the mentioned aspects, what is gained is perhaps what I have termed an impact device, a tool that convincingly addresses how and where impact will occur, thus mediating the relationship between the proposed stakeholders and the researchers carrying out the project.

In this chapter, I have investigated RCN's planning procedures for project proposals. I have described how RCN's measures seek to reduce project proposals to a logical relation between predefined phenomena of which the researcher must adhere to. By making the researcher report on their achievements throughout the duration of their project, the project's lifetime can be recorded and retraced through the establishment of what I term a circulating reference. Furthermore, I have suggested that what is gained from this process is a much more rigorous traceability of a project, but also what I call an impact device. If circulated, this device may assist a project in its achievements.

9. Conclusionary analysis: Enacting impact

In this study I have been guided by the rather vague question of what RCN's impact evaluations *do*. I have shown how the tools and procedures used in evaluation possesses transformative capacities towards Norwegian research sector by establishing a framework revolving around the novel concept of impact. The main objectives of carrying out evaluations have been expressed as a combination of accountability and learning, but as I have shown, evaluation entails more intrusive features whereby researchers, their testimonies and project proposals are subject to processes of reduction and translation, which I claim is not fully considered when such objectives are defined. This study then highlights how RCN's role of mediating governmental demands is proliferating out of the thematic programmes into both Fripro and the basic financed institutions by use of evaluation. Furthermore, the new emerging formalizing procedures when planning for impact may paradoxically end up striving away from the pragmatic attitude that first was envisioned when searching for impact in the retrospective evaluations. If projects are reduced to theoretical assumptions and logical relations, I argue it becomes less practical – less focussed on what works.

What has motivated me throughout the workings of this thesis has been to investigate evaluation practices and how these bring about certain insights about its object of evaluation. By employing an ontological foundation and a related terminology from works that may be rubricated under the scholarly tradition of actor network-theory, my hope was that one such inquiry would provide some interesting and alternative conceptions of what evaluation is and how it works to stage reality in specific ways. A controversial and eye-opening implication of Latour and Woolgar (1988) is that reality, they use the term *out-there-ness*, is an achievement of scientific practices and its inscription devices. Rather than reality being its cause, reality is depicted as a consequence of this work. This is because, in the longer run, particular realities are brought into being with and through the arrays of inscription devices, so reality is not entirely independent of the apparatuses that produce reports of reality. Words do not simply represent phenomena, words and worlds go together (Law 2004, p. 40). This does, however, not mean that the scientists are fabricating facts as they please, that these facts are the sole end-results of some social mechanisms, nor that they are not real. "There was not the slightest doubt that the products of those artificial and costly sites were the most ascertained, objective, and certified results ever obtained by collective human ingenuity" (Latour, 2005, p. 89). The inscription devices are depicted as the keys to this achievement, but instead of facilitating correspondence between word and world, they enable translation. With ANT's shift from

construction to enactment, the prime mover of these arrangements dissolved, meaning that the social and natural worlds must be enacted into being. Science assists to enact specific realities into being.

By employing equivalent implications from Latour and Woolgar onto my own work and viewing evaluation in similar lines as they view scientific knowledge production, my analysis would go somewhere along the lines that the evaluation did not simply observe the impact of the humanities in Norway, but that the humanities' impact was a consequence of the evaluation. In some ways this is true: Before Humeval, the humanities in Norway did not have impact, while after it certainly did. Parts of the Norwegian research community have subsequently adjusted both their line of sight and line of practice along with this newly introduced feature, making impact an important denominator and a force to be reckoned with when deciding projects worthy of funding. However, if impact refers to the many influences the evaluated researchers may have had upon their objects of research, it seems absurd to suggest that this was a consequence of the evaluation itself. It seems also inconsistent with the emphasis that it is primarily science and research that affect and produce realities,⁵ not research evaluation. Another point is Law's emphasis on it being "in the long run" (2004, p. 40). The methods of evaluating impact are contested, and consequently the dissemination of this phenomenon among evaluated units is an object of discussion, possibly revision, and not at all a black boxed procedure. It is conceivable that it would not take incredible amounts of resources to introduce a quite different version and methodology when evaluating impact in any research evaluation. As I previously mentioned, research institutes have been established with conceptualizing and studying research impacts as its primary aim (Universitetet i Oslo, 2016).

What I do claim is that the various inscription devices employed by RCN may assist to finalize and enact a specific form of impact; one that is fixed towards the concrete, the verifiable and depended upon the assumptions of the actors who carry out the project. What may be the consequence of the work I have describe in this thesis is not the emergence of the causal powers of research per se, but a concretized, particularized, and *reduced* form of impact, perhaps furthest illustrated by the Logical Model. Documenting and transmitting impact relations from this point of departure have paradoxically their own impacts, and they can help bring into being that what is discovered. If humanities are found to have great impact on the educational, cultural, environmental or any other sector, these may be further

⁵ For what this entails for the social sciences, see Law and Urry (2004).

intensified by the transmission of this information. As such, realities are enacted. I do not think this should be very controversial, and it is to some extent consistent both with RCN's own depictions of evaluation, but also with the emphasis that the ways in which a phenomenon is being measured is constitutive of the phenomenon itself. In that light I may argue that the practices of impact evaluations are formative of an evaluable reality. The extent to which this reality is *realized* depends on the number and authority of future descriptions that further builds on such findings and this evaluative model – the extent to which the world is adjusting to impact. For now, it is interesting to note that the incredibly fluctuated and inconsistent terminology when addressing the concept in the Norwegian language in evaluations and research strategies is gradually being replaced by the English term of impact (IP3; IP4; Gulbrandsen and Sivertsen, 2018, p. 10).

References

- Arnesen, M. and Tønnessen, M. (2022). Statssekretæren om Forskningsrådet: – Dette handlar berre om ein einaste ting. *Khrono*, 15.05.2022. Available at: <https://khrono.no/statssekretaer-om-forskningsradet-dette-handlar-berre-om-ein-einaste-ting/688742> (Access date 19.05.2022).
- Arnold, E. and Mahieu, B. (2012). *A Good Council? Evaluation of the Research Council of Norway*. Available at: <https://www.regjeringen.no/en/dokumenter/evaluation-of-the-RCN-2012/id698794/>.
- Arnold, E. and Åström, T. (2020). *Intervention Logic. Evaluation of the introduction of theory-based planning at RCN*. Available at: <https://www.forskningsradet.no/siteassets/publikasjoner/2020/rcn-intervention-logic.pdf>.
- Asdal, K. (2011). *Politikkens Natur – Naturens Politikk*. Oslo: Universitetsforlaget.
- Asdal, K. (2015). What is the issue? The transformative capacity of documents. *Distinktion: Scandinavian Journal of Social Theory* 16(1): 74-90.
- Asdal, K. and Reinertsen, H. (2020). *Hvordan gjøre dokumentanalyse. En praksisorientert metode*. Oslo: Cappelen Damm.
- Benneworth, P. (2015). Putting impact into context: The Janus face of the public value of arts and humanities research. *Arts & Humanities in Higher Education* 14(1): 3-8.
- Bjørkdahl, K. (2019). Forskeren som spåkone. *Khrono*, 19.03. Available at <https://khrono.no/akkreditering-kristian-bjorkdahl-nfr/forskeren-som-spakone/269199> (Access date 06.05.2022).
- Blok, A. and Elgaard Jensen, T. (2011). *Bruno Latour: Hybrid thoughts in a hybrid world*. London: Routledge.
- Bogner, A., Littig, B. and Menz, W. (2018). Generating qualitative data with experts and elites. *Sage Handbook of Qualitative Data Collection*, edited by Uwe Flick. London: Sage.
- Bowen, G.A. (2009). Document analysis as a Qualitative Research Method. *Qualitative Research Journal* 9(2): 27-40.
- Callon, M. (1984). Some elements of a sociology of translation: Domestication of the scallops and the fishermen of Saint Brieuc Bay. *Sociological Review* 32 (1): 196-233.
- Callon, M. (ed.) (1998). *The Laws of the Markets*. Oxford: Blackwell.
- Dahler-Larsen, P. (2012). *The Evaluation Society*. Stanford: Stanford University Press.
- Dahler-Larsen, P. (2015). The Evaluation Society: Critique, Contestability and Skepticism. *Spazio Filosofico*, 13(1): 21-36.
- Dahler-Larsen, P. (2016). *Vejledning om at evaluere innovative tiltag*. Available at: <https://copi.dk/viden-om/materialer/vejledning-om-at-evaluere-innovative-tiltag/>.
- Derosières, A. (2001). How Real Are Statistics? Four Possible Attitudes. *Social Research* 68(2): 339-355.
- Derrick, G. (2020). Definition of societal impact and its evaluation in context. In *ENRESSH Work Group 1 Overview of Peer Review Practices in the SSH*, edited by Michael Oschner et al.
- Derrick, G. (2018). *The Evaluator's Eye*. Lancaster: Palgrave Macmillan Cham.
- Derrick, G. and Benneworth P. (forthcoming). Guidelines for the Peer Review of ex-ante Impact Evaluation. *ENNRESH*.

- Doganova, L. and Eyquem-Renault, M. (2009). What do business models do? Innovation devices in technology entrepreneurship. *Research Policy* 38(10): 1559-1570.
- Donovan, C. (2018). The Australian Research Quality Framework: A live experiment in capturing the social, economic, environmental, and cultural returns of publicly funded research. *Reforming the evaluation of research. New Directions for Evaluation*, 118: 47-60.
- Foucault, M. (1995). *Discipline and Punish. The birth of the Prison*. Translated by Alan Sheridan. 2nd edition. New York: Vintage Books.
- Funnell, S.C. and Rogers, P.J. (2011). *Purposeful Program Theory. Effective use of theories of change and logic models*. San Francisco: John Wiley & Sons.
- Gertler, P.J., Martines, S., Premand, P., Rawlings, L.B. and Vermeersch, C. (2016). *Impact Evaluation in Practice*. Second edition. Washington: World Bank.
- Gulbrandsen, M. and Sivertsen, G. (2018). *Impact i anvendt forskning: begrepsavklaring og praksis*. Arbeidsnotat, 10. Retrieved from: <https://www.nifu.no/publications/1626509/>.
- Haraway, D. (1988). Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies* 14(3): 575-599.
- Holm, J. and Askedal, A. (2019). Evaluation of societal impact in Norwegian SSH evaluations. *Fteval. Journal for research and technology policy evaluation*. 48: 139-142.
- Holm, J. and Løvhaug, J.W. (2014). Humanioras utfordringer. *Forskningsspolitikk*. Available at: <http://www.fpol.no/humanioras-utfordringer/> (Access date: 10.04.2022).
- Holm, J. (2018). *Humanities in Norway make a strong contribution to society. Om samfunnseffekten av en forskningsevaluering*. Available at: https://www.vr.se/download/18.781fb755163605b8cd21a52f/1529480564384/Evaluering+av+humanistisk+forskning+i+Norge_samfunnseffekter_180416revidert.pdf (Access date: 15.03.2022).
- Hug, S.E., Oschner, M. and Daniel, H. (2014). A framework to Explore and Develop Criteria for Assessing Research Quality in the Humanities. *International Journal for Education Law and Policy* 10(1), 55-68.
- Knorr-Cetina, K.D. (1981). *The manufacture of knowledge: an essay on the constructivist and contextual nature of science*. Oxford: Pergamon Press.
- Lamont, M. (2012). Toward a Comparative Sociology of Valuation and Evaluation. *Annual Review of Sociology* 38(21): 201-221.
- Latour, B. and Woolgar, S. (1986). *Laboratory Life. The Construction of Scientific Facts*. New Jersey: Princeton University Press.
- Latour, B. (1987). *Science in Action*. Cambridge: Harvard University Press
- Latour, B. (1988). *The Pasteurization of France*. Translated from the French by Alan Sheridan and John Law. Cambridge: Harvard University Press.
- Latour, B. (1999). *Pandora's Hope. Essays on the Reality of Science Studies*. Cambridge: Harvard University Press.
- Latour, B. (2005). *Reassembling the Social. An introduction to Actor-Network Theory*. New York: Oxford University Press.
- Law, J. and Urry, J. (2004). Enacting the social. *Economy and Society*, 33(3): 390-410.
- Law, J. (2004). *After Method. Mess in social science research*. London: Routledge.
- Law, J. (2009). Actor Network Theory and Material Semiotics. In *The New Blackwell Companion to Social Theory*, edited by Bryan S. Turner, 141-158. Malden: Wiley-Blackwell.

- Lemke, T. (2015). New Materialisms: Foucault and the 'Government of Things'. *Theory, Culture & Society*, 32(4): 3-25.
- Li, T. (2022). Norges Forskningsråd. *Store Norske Leksikon*. Available at https://snl.no/Norges_forskningsrad (Access date 05.03.2022).
- Meld. St. 25 (2016-2017). *Humaniora i Norge*.
- Morgenbladet (2008). Humaniora i krise? *Morgenbladet*. 12.09. Available at: <https://www.morgenbladet.no/boker/2008/09/12/humaniora-i-krise/> (Access date 02.03.2022).
- Muhonen, R. and Tellmann, S. (2021). Challenges of reporting societal impacts for research evaluation purposes – case of sociology. *Handbook on Research Assessment in the Social Sciences*, edited by Engels, Tim C.E. and Kulczycki, E. Cheltenham: Edward Elgar.
- Power, M. (1994). *The Audit Explosion*. London: White Dove Press.
- Power, M. (1997). *The Audit Society: Rituals of verification*. New York: Oxford University Press.
- Prop. nr. 1 S (2015-2016). *Kunnskapsdepartementet*.
- Ramberg, I., Solberg, E., Aksnes, D.W., Børing, P., Piro, F. and Wollscheid, S. (2015). *Lange spor i helseforskningen. Bidrag fra forskningsprosjekter støttet av Norges forskningsråd innenfor psykisk helse og rusmiddelfeltet på 2000-tallet*. Available at: <https://nifu.brage.unit.no/nifu-xmlui/handle/11250/2367534>.
- Reinertsen, H. (2016). *Optics of Evaluation. Making Norwegian foreign aid an evaluable object, 1980-1992*. PhD thesis. Oslo: University in Oslo.
- Reinertsen, H. (2019). *Evaluation optics of the nation state: The past, present and future of public documentation (Evaluation)*. Project description, Research Council of Norway, FRIPRO – Young Research Talents Grants (updated version 2020).
- The Research Council of Norway. (2008). *Humanistisk forskning. Nasjonal Strategi*.
- The Research Council of Norway. (2017a). *Evaluation of the Humanities in Norway. Impact cases*.
- The Research Council of Norway. (2017b). *Evaluation of the Humanities in Norway. Report from the Principal Evaluation Committee*.
- The Research Council of Norway. (2017c). *Evaluation of the Humanities in Norway. Report from Panel 1-Aesthetic Studies*.
- The Research Council of Norway. (2017d). *Evaluation of the Humanities in Norway. Report from Panel 2-Nordic Languages and Linguistics*.
- The Research Council of Norway. (2017e). *Evaluation of the Norwegian Social Science Research Institutes. Principal Report*.
- The Research Council of Norway. (2018a). *Evaluation of Norwegian education research. Report from the International Expert Committee*.
- The Research Council of Norway. (2018b). *Evaluation of the Social Sciences in Norway. Report from the Principal Evaluation Committee*.
- The Research Council of Norway. (2019a). *Våre Evalueringer*. Updated 25.03.2022. <https://www.forskningsradet.no/statistikk-evalueringer/evalueringer/vare-evalueringer/>
- The Research Council of Norway. (2019b). *Oppfølging av evaluering av humanistisk forskning i Norge. Rapport fra oppfølgingsutvalg oppnevnt av Forskningsrådet*.
- The Research Council of Norway. (2021). *Evaluation of Legal Research in Norway*.
- The Research Council of Norway. (2022a). *Template for Project Description Scientific Renewal and Young Research Talent 2022*. Available at:

- <https://www.forskningsradet.no/en/call-for-proposals/2022/researcher-project-young-talents/> (Access date 03.05.22).
- The Research Council of Norway. (2022b). *Bli kjent med søknadsskjemaet*. Updated 22.03.2022. Available at <https://www.forskningsradet.no/sok-om-finansiering/skrive-soknad/bli-kjent-med-soknadsskjemaet/> (Access date 03.03.2022).
- Shore, C. and Wright, S. (2000). Coercive accountability, in *Audit Cultures: Anthropological studies in accountability, ethics and the academy*, edited by Marilyn Strathern, 57-89. London: Routledge.
- Sismondo, S. (2010). *An Introduction to Science and Technology Studies*. 2nd edition. Malden: Blackwell.
- Sivertsen, G. and Meijer, I. (2018). Normal versus extraordinary impact. How to understand, evaluate and improve research activities in their relations to society. *Research Evaluation* 29(1): 66-70.
- Solberg, E, Tellmann, S.M., Aanstad, S., Aksnes, D.W., Ramberg, I. and Børing, P. (2017). *Pathways to global impact. Tracing the impact of development research funded by the Research Council of Norway*. Available at <https://nifu.brage.unit.no/nifu-xmlui/handle/11250/2446640>.
- Solberg, E., Schwach, V., Aksnes, D. and Børing, P. (2019). *Lange spor i velferdsforskningen. En analyse av Forskningsrådets programmer for velferd og arbeidsliv fra 2000-2014*. Available at <https://nifu.brage.unit.no/nifu-xmlui/handle/11250/2602245>.
- Spaapen, J. and van Drooge, L. (2011). Introducing ‘Productive Interactions’ in Social Impact Assessment. *Research Evaluation*, 20(3): 211–218.
- Trædal, T.J. (2022). Regjeringen kaster styret i Forskningsrådet. Stanser tildelinger til forskning og setter i gang ekstern granskning, *Aftenposten*, 12.05. Available at: <https://www.aftenposten.no/norge/i/k6grn6/regjeringen-kaster-styret-i-forskningsraadet-stanser-tildelinger-til-forskning-og-setter-i-gang-ekstern-granskning> (Access date 13.05.2022).
- Universitetet i Bergen. *Humaniorastrategi 2018-2022*. Available at: https://ekstern.filer.uib.no/ledelse/hf_strategiplan_2018_2022_norsk.pdf.
- Universitetet i Oslo. *OSIRIS – Oslo Institute for Research on the Impact of Science*. Available at: <https://www.sv.uio.no/tik/english/research/centre/osiris/> (Access date 16.05.2022).
- Vedung, E. (1997). *Public Policy and Program Evaluation*. New Brunswick: Transaction.
- Vedung, E. (2010). Four waves of Evaluation Diffusion. *Evaluation* 16(3): 263-277.
- Vetenskapsrådet. *Etik i forskningen*. Updated 16.12.2021. Available at: <https://www.vr.se/uppdrag/etik/etik-i-forskningen.html> (Access date 18.05.2022).
- Vollan, M. and Larsen, D.E. (2022) Ber Moe rydde i eget reir, *Klassekampen*, 21.05. Available at: <https://klassekampen.no/utgave/2022-05-21/ber-moe-rydde-eget-reir> (Access date 22.05).
- Weiss, C.H. (1998). *Evaluation: Methods for studying programs and policies*. Upper Saddle River: Prentice Hall.
- Wróblewska, M.N. (2018). *The making of the Impact Agenda. A study in discourse and governmentality*. PhD thesis. Warwick: University of Warwick.
- Wróblewska, M.N. (2019). Impact Evaluation in Norway and in the UK: A comparative study, based on REF 2014 and Humeval 2015-2017. *ENRESSH* working paper, 01, vol. 2019.

Appendix A: Guidelines provided to the respondents (RCN 2017a, 1)

A collection of the impact cases submitted to the evaluation of the humanities in Norway: Some introductory remarks

In December 2015, the Research Council of Norway (RCN) invited the institutions participating in the evaluation of the humanities to submit case studies documenting the societal impact of their research. The 2014 Research Excellence Framework (REF) in the UK served as a model for the inclusion of such impact case studies in a large-scale evaluation. The REF definition of impact as "any effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia" was adopted along with the model of using case studies to evaluate the impact.

The guidelines presented to the institutions for submitting the case studies included the following points (these sections are also largely borrowed from the REF specifications):

- The research underpinning the reported impact cases must wholly, or in part, have been undertaken by researchers affiliated to the submitting institution. The contribution from these researchers to the reported impact should be significant. Both the research and the impact should have been produced within the last 10 – 15 years, counting from 2015. Priority should be given to more recent examples.
- Special circumstances may allow for extending the given time interval when necessary to explain longer research traditions relevant to the reported impact. In such cases, great importance should be attached to documenting tangible impacts within the time frame provided. In all cases, the research underpinning the impact should be of a high international standing.
- Each institution is invited to submit one case per evaluation panel, provided that it has fulfilled the requirement of presenting at least five researchers for evaluation by that panel. If desired, the institution may submit further cases for evaluation, limited upwards to one case per ten researchers in one panel.
- The impacts of the research may include, but are not limited to, changes in activity, attitude, economy, capacity, attention, competence, policy, practice, or understanding. The changes may occur among individuals, groups, organisations, in public opinion, or in society at large. The changes may be local, national or international. Effects on other research, or effects within the submitting institution (for instance the effects on teaching within the institution) are not to be reported as impact cases.

Appendix B: The template and guidelines for the impact criterion in RCN's applicant forms (RCN, 2022a)

2. Impact

This chapter should describe the importance of the anticipated results in terms of the potential academic impact and, optionally, the potential societal impact of the research. The potential impact can be in the short or longer term. The chapter should also specify the planned measures for communication and exploitation of the project results.

2.1 Potential for academic impact of the research project

Building on the description of project objectives and novelty in chapter 1, describe clearly why and how the project outputs may address important present and/or future scientific challenges and have an impact on the research area/field, if successful.

Please note:

All applications must include a description of the potential for academic impact of the project.

The description of the potential impact should be project-specific and related to the planned research. General elaborations on the benefits of research in a wider context should be avoided.

2.2 Potential for societal impact of the research project (optional)

Building on the description of knowledge needs and challenges in section 1.1., describe why and how the project outputs, if successful, have the potential to meet the mentioned societal challenge(s).

Describe how new knowledge and project outputs have the potential to address one or more of the UN Sustainable Development Goals.

Please note:

The description of potential societal impact will be assessed as follows:

The panels will assess potential for societal impact if the applicant has included a description of this in the project description

For applications initiated in the context of a specific societal challenge, you should describe the potential for societal impact. If relevant for the project, this includes describing how the knowledge and outputs generated in the project can contribute to solving challenges and/or shed light on important issues related to one or more of the 17 UN Sustainable Development Goals (SDGs)

The description of the potential impact should be project-specific and related to the planned research. General elaborations on the benefits of research in a wider context should be avoided.

2.3 Measures for communication and exploitation

Describe briefly the target audiences, including stakeholders/users, of the project outputs (in or beyond the scientific community).

Outline the scope and plan for dissemination, communication and engagement activities.

Provide a brief description of planned activities that will contribute to the realisation of the potential impacts of the project outputs (in or beyond the scientific sphere).

Please note:

This part of the project description will be the basis for the assessment of communication and exploitation. Hence, you may leave the "Communication plan" section in the application form empty.

Appendix C: Impact assessment criterion (RCN 2022a)

Potential impact of the proposed research

- Potential for academic impact: The extent to which the planned outputs of the project address important present and/or future scientific challenges.
- Potential for societal impact (if addressed by the applicant): The extent to which the planned outputs of the project address UN Sustainable Development Goals or other important present and/or future societal challenges
- The extent to which the potential impacts are clearly formulated and plausible.

Communication and exploitation

- Quality and scope of communication and engagement activities with different target audiences, including relevant stakeholders/users.

Appendix D: Rough interview guide⁶

Introduction

I am interested in how the Research Council of Norway work to document the societal impact of research.

Why do you evaluate research? And why evaluate Impact?

Methods and design of Humeval

Is it correctly understood that it was Humeval that first introduced the concept of impact into Norwegian research evaluation? What other aspects were new with this evaluation?

Can you tell me about the background for Humeval? Why was it desirable to document the effects of humanities research?

How did you prepare the respondents for the evaluation? Can you tell me about what happened in the workshop you held?

Research groups as units of assessment. Why did you choose this?

Why did you adopt the impact case methodology and REF's definition of impact? What did it achieve?

The analysis

I am interested in the way you inspected and analysed the impact cases. Can you tell me about the categories which you employed? How was this analysis developed and carried out?

How did the RCN and the panels solve challenges related to their analysis?

Can you tell me about the "underdeveloped evaluation culture" which the report emphasised? What happened? Did the respondents misunderstand the task?

Were there similar problems in other subsequent evaluations? Have you sought to improve on this issue in these evaluations? If yes, how?

Do you see the evaluation as nudging researchers into thinking and reporting on their research in new ways? Is the evaluation simultaneously an intervention? How do you consider this potentiality of evaluation?

What do you mean by impact infrastructure? What do you see being implemented with this term?

Has anything happened after Humeval among institutions? What is the observed impact of the Humeval report?

How is RCN working with impact evaluations in the future? Was this a formative process also for the RCN? Are there developments occurring in the methodology?

⁶ This serves as a very rough guide for the interviews I conducted. Some of the interviewees had not been working with all the mentioned aspects in this guide, so this guide rather illustrates the entirety of the themes I had formulated for the interviewees.

Prospective evaluation

Retrospective/prospective. Is this a useful distinction for the impact evaluations in the RCN?

Can you tell me about the work and proceedings of RCN's prospective evaluations? Impact as applicant criterion. How is this assessed, and what is it used for?

What do you use the progress and final report for? Will this be used in future evaluation reports?