THREE PRINCIPLES FOR EVALUATING CO-PRODUCED CLIMATE SERVICES

MATHILDA ENGLUND | DIVISION OF RISK MANAGEMENT AND SOCIETAL SAFETY | LTH | LUND UNIVERSITY, SWEDEN

Three Principles for Evaluating Co-produced Climate Services

Mathilda Englund

Lund 2021

Three Principles for Evaluating Co-produced Climate Services

Student: Mathilda Englund

Number of pages: 88
Illustrations: Mathilda Englund (unless stated otherwise)

Keywords
climate change adaptation, transdisciplinary research, participatory research, evaluation method, research impact evaluation

Abstract
As climate change impacts unfold across the globe, growing attention is paid towards producing climate information that supports evidence-based adaptation efforts. The use of such scientific knowledge remains limited in practice. Researchers and practitioners have started to co-produce climate services to ensure that information is contextual, timely and relevant to support climate change adaptation decision-making. Recent research shows a pressing need to better understand how to evaluate co-produced climate services, in particular their impact on policy and action. This thesis explores this issue through a systematic literature review, and an online questionnaire that targeted actors with previous experience in co-producing climate services. Three principles are proposed, which outline methodological choices applicable when evaluating co-produced climate services: 1) Theory of Change methodology, 2) participatory evaluation, and 3) mix-methods with a focus on visual products.

© Copyright: Division of Risk Management and Societal Safety, Faculty of Engineering
Lund University, Lund 2021
Avdelningen för Riskhantering och samhällssäkerhet, Lunds tekniska högskola, Lunds universitet, Lund 2021.

Riskhantering och samhällssäkerhet
Lunds tekniska högskola
Lunds universitet
Box 118
221 00 Lund
Sverige

http://www.risk.lth.se

Telefon: 046 - 222 73 60

Division of Risk Management and Societal Safety
Faculty of Engineering
Lund University
P.O. Box 118
SE-221 00 Lund
Sweden

http://www.risk.lth.se

Telephone: +46 46 222 73 60

Acknowledgements

First, I would like to express my gratitude to the Stockholm Environment Institute for making me feel welcomed as an intern despite the many hinders imposed by the Covid-19 restrictions. I am looking forward to meeting you all in the office once the pandemic finally is over. A special thanks to Katrin Danerlöv and Åsa Gerger Swartling for sharing your expertise.

The study provided input to a larger research project at the Stockholm Environment Institute, namely "UNCHAIN: Unpacking climate impact chains". Hopefully, I have produced some useful insights.

I was fortunate enough to have two supervisors, and each of them deserves my heartfelt gratitude for their eager support and encouragement.

I would like to extend my sincere thanks to Karin André at the Stockholm Environment Institute, for introducing me to the topic and agreeing to supervise me while writing my thesis. Without you, this research would not have been possible.

I would also like to thank Jenny Iao-Jörgensen at the Division of Risk Management and Societal Safety at Lund University, for her inspiring advice, comments, and expertise.

To my family and friends, thank you for all the love and support. To my grandmother, thank you for making the effort of trying to understand what climate services actually are at the age of 85. Last but not least, I would like to thank Pelle for buying me comfort food, doing the dishes, and saying all the right things when the workload seemed unbearable.

Summary

As climate change impacts unfold across the globe, growing attention is paid towards producing climate information that support adaptation efforts. The use of such information, however, remains limited in practice. Researchers and practitioners have therefore started to co-produce climate services to ensure that information is contextual, timely and relevant to support climate change adaptation decision-making. Recent research however shows a pressing need to better understand how to evaluate co-produced climate services, in particular their impact on policy and action. This thesis aims to address this research gap, by answering the following research question:

What methods can be used to evaluate both the process and effects of co-produced climate services?

The research question is answered through a systemic literature review, and online questionnaire that targeted actors with previous experience in co-producing climate services. The findings are presented in the form of principles to provide practitioners and policymakers with a flexible guide to draw inspiration from when designing their evaluation framework. The findings support the following three principles:

1. **Theory of Change methodology,** which is a tool for project management, stakeholder engagement, and evaluation that can help stakeholders to capture the complexity which emerges when co-producing climate services. It engages stakeholders in discussions on objectives, causal linkages, external factors, and underlying assumptions. The Theory of Change is easy to update as outcomes unfold over time.
2. **Participatory evaluation,** which actively engages stakeholders in the evaluation process. Previous research shows that participatory evaluations can yield many benefits like building a sense of ownership over the evaluation findings; generating a shared vision; increasing motivation among stakeholders to participate in the evaluation; capturing unexpected effects; and validating findings among the affected stakeholders.
3. **Mix-methods - with a focus on visual products,** which create a robust data set for analyses while capturing both the tangible and intangible effects which emerge when co-producing climate services. Visual products can help stakeholders to better understand complex change pathways.

The three principles add to the body of research that seeks to evaluate research impact, and pave the way to a more advanced approach that can assess the many outputs, outcomes, and impacts which emerge when co-producing climate services. The principles are designed to offset previously identified challenges and shortcomings encountered when evaluating co-produced climate services, and are expected to yield many benefits. Key benefits include helping stakeholders to map complex change pathways; capturing external factors; measuring intangible outcomes and impact; bridging conflicting interests; identifying potential unexpected effects; and many other.

The principles draw on evidence from a systematic literature review and an online questionnaire, where early signs for practical applicability surfaced. As a way forward, the principles should be applied and tested in practice. The principles should be applied in various constellations to further stimulate the debate on co-produced climate services.

Content

Chapter 1: Introducing the Research Topic, Aim, and Concepts

Chapter 2: Hypothesis development

Chapter 3: Hypothesis testing

Chapter 4: Conclusion

List of tables

List of figures

CHAPTER 1: INTRODUCING THE RESEARCH TOPIC, AIM, AND CONCEPTS.

1 Introduction

As climate change impacts unfold across the globe, growing attention is paid towards producing climate information that supports adaptation efforts (Wall et al. 2017). Despite recent advancements in climate projections, climate impact studies, and adaptation research (Ernst et al. 2019), the use of such climate information remains limited in practice (André et al. 2020). There is thus a usability gap between the researchers producing climate information, and the end-users who apply the knowledge in practice (Bremer and Meisch 2017). As an attempt to bridge this gap, the World Meteorological Organization in collaboration with various UN agencies initiated the World Climate Conference-3 (WCC-3) in 2009 to develop a global framework for climate services (GFCS 2009). Climate services seek to produce contextual, timely and relevant information to support climate change adaptation decision-making. Decision-relevant climate information plays a significant role in promptly adapting to the effects of climate change (Daniels et al. 2020).

Climate services, and climate change adaptation, are situated in a context characterized by complexity, cross-scale drivers, diverging stakeholder interests, and high uncertainty. Researchers and practitioners have adopted knowledge co-production as an attempt to navigate within this complex space (Norström et al. 2020). Knowledge co-production is a collaborative research process that transcends traditional disciplinary boundaries and engages non-academic stakeholders to produce context-tailored knowledge, solve pressing societal issues, and boost scientific progress (e.g., Belcher et al. 2020; Blackstock et al. 2007; Heink et al. 2015). Some scholars argue that co-produced climate services can generate new knowledge simultaneously as it strengthens relationships, trust, confidence, and capacity (Daniels et al. 2020). Allegedly, knowledge co-production can enhance ownership, legitimacy, credibility, and accountability (Lang et al. 2012), ultimately increasing the likelihood of research findings being implemented in practice (Reed et al. 2018).

There is however little existing evidence showing if co-produced climate services deliver their potential benefits. Most climate services are not systematically evaluated. This, as researchers tend to develop new co-production tools rather than evaluating and refining existing ones (Van der Molen et al. 2019). In this regard, concerned parties lack information to make sound decisions regarding where, or if, to spend their often limited budgets to improve climate services (Vaughan and Dessai 2014). For this reason, some scholars have recently started to outline evaluation practices that are fit for appraising co-produced climate services (Van der Molen et al. 2019). Evaluation can play a significant role in quality improvement, as it allows involved actors to reflect, learn, and adapt based on previous project performance (Fazey et al. 2014). Thus, evaluations can support and improve climate change adaptation decision-support, in the long run increasing the efficiency and effectiveness of evidence-based adaptation actions (Vaughan and Dessai 2014).

During the WCC-3, it was acknowledged that evaluation serve a key role in establishing best practices (GFCS 2009). Yet there is still little academic guidance on how such evaluation should look like, ultimately making it increasingly difficult to collect plausible empirical

evidence on the effectiveness, quality, and effects of co-produced climate services (Ernst 2019; Van der Molen et al. 2019). Traditionally, research evaluations employ procedures solely focusing on assessing academic outputs, thus failing to capture the broad array of effects associated with knowledge co-production (e.g. Sarkki et al. 2015; Schuck et al. 2017). Thus, evaluating the co-production of climate services require new innovative approaches (Wall et al. 2017) that accounts for the many challenges in climate change adaptation decision-making. Long-term impacts and intangible effects are especially difficult to assess (Daniels et al. 2019).

There is a pressing need to better understand how to evaluate co-produced climate services, in particular their impact on policy and action. Many methods, frameworks, and approaches for evaluating research impact exist, but little research is devoted to understand what methods best capture the many effects which emerge when co-producing climate services (Reed et al. 2021). This thesis seeks to address this research gap.

1.1 Research aim

This thesis aims to develop a set of principles that outline methods applicable when evaluating co-produced climate services. The principles are designed to fit many different contexts, and seek to capture a broad array of effects. Focus lies on understanding the co-production process itself and the effects which may emerge from such a process. The intended user is anyone involved in co-producing climate services, regardless of their background. This includes researchers, project stakeholders, funders, decision-makers, and more. The research question is the following:

What methodological principles can be used to evaluate both the process and effects of co-produced climate services?

The findings are presented in the form of principles. A standard set of indicators is consider inflexible, and is likely to fail adapting to the many different contexts that co-produce climate services (Fazey et al. 2014). Broader evaluation principles are in comparison more adaptable. Principles can be adapted to fit particular situations, while providing practitioners with a flexible guide to draw inspiration from when designing their evaluation frameworks (O'Connor et al. 2019).

To answer the research question, a hypothesis is developed. The hypothesis outlines the methodological principles considered fit for evaluating co-produced climate services. Liedtka (2015) identifies two types of hypotheses: The *scientific hypothesis* that seeks to explain casual linkages, and the *design hypothesis* that employs abductive reasoning to find a probable solution to an identified problem. Here, the latter is applied. The design hypothesis provides a powerful tool for abduction, and can facilitate the formulation of new ideas, concepts, and methods. Also, the design hypothesis is considered robust in light of complexity, uncertainty, and ambiguity (Liedtka 2015), all which are prevalent when trying to evaluate co-produced climate services (Bremer et al. 2019). For simplifying purposes, the design hypothesis is hereinafter referred to as hypothesis.

1.2 Outline

To answer the research question, the thesis is divided into four chapters. The first chapter introduces the thesis topic, research aim, and conceptual framework. The second chapter reviews the existing literature to form a hypothesis, which then is tested in the following chapter. The fourth and final chapter concludes the research findings.

2 Conceptual framework

This section outlines the main concepts used in the thesis, whilst placing the study in a broader academic context. Looking at the research question, the following three concepts stand out: 1) climate services, 2) knowledge co-production, and 3) evaluation. Figure 1 shows how the concepts relate to each other.

The concepts are placed on a three-level ladder. Each level is designed to target the former. First, the area is defined. Focus lies on climate services, which is approached through the second level using knowledge co-production. At the broadest level lies the intervention that the thesis aims to improve, in which this case refers to evaluation practices.

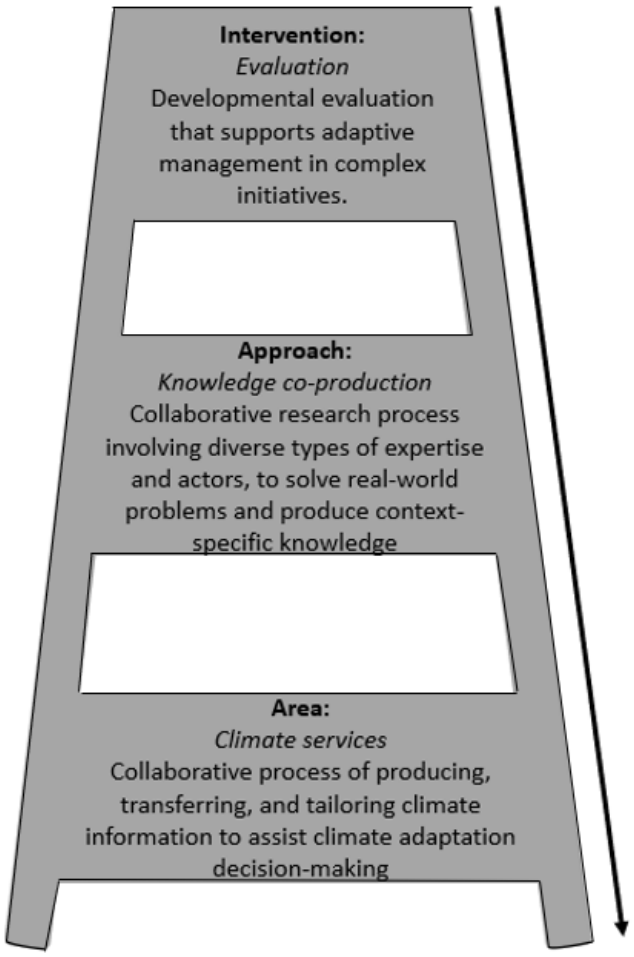

Figure 1: Conceptual framework

2.1 Area: Climate services

Climate services convey credible weather and climate information, commonly in the form of seasonal and decadal climate forecasts (Christel et al. 2018). Other forms of climate services include, but not limited to, maps, risk assessments, and narratives, ultimately equipping decision-makers with scientific information to adapt to climate change (Daniels et al. 2020). Climate services differ from climate research, as the former seeks to respond to user-needs whereas the latter intends to enhance our understanding of the climate system as a whole. Climate services emerged as a concept during the WCC-3 held in 2009 (Vaughan and Dessai 2014). As a rather new concept, climate services are yet to adopt a universal definition. Regardless of the definition, climate services have received increasing recognition over the past 15 years at a global, regional, and national scale. Many scholars consider climate services as essential for bridging the gap between research and practice, thus ensuring that science effectively supports evidence-based climate change adaptation decision-making (Bremer et al. 2019).

Until more recently, climate services have been defined as merely a mean to deliver a final product. Such supply-driven delivery of climate services from providers to users often falls short in practice, as it fails to account for local needs, capacities, and context (Daniels et al. 2020). An example of such definition is presented by the European Commission (2015), in which climate services are considered as the transformation of climate information into usable products like climate forecasts, financial assessments, advices on best practices, and others.

Some scholars have tried to respond to this need for a more process-centric approach. One notable example is the Tandem Framework, which is applied in this thesis. The Tandem Framework has been empirically tested in Southeast Asia, Latin America, Sub-Saharan Africa, and Europe. Using the analogy of a tandem bicycle, the framework sets out to involve relevant stakeholders in a collaborative process to co-produce climate services (Daniels et al. 2019). The Tandem Framework proposes a process-centric, co-producing approach which seeks to enhance collaboration in order for climate services to fulfill their potential and bring about societal change. Focus lies on the process *itself* rather than its final outputs, where the involved stakeholders work together to develop a common understanding of the problem at hand. In the Tandem Framework intended outcomes extend beyond the traditional singular products to encompass a wider range of possible impacts, such as improved institutional capacity, increased confidence, and expanded networks. Evaluation plays a key role in the Tandem Framework, as it helps to foster a culture of learning and reflection among the involved stakeholders (Daniels et al. 2020)

Arguably, such approach, and its emphasis on collaboration, nicely brings together two of the main branches of interest explained in the research aim: climate services and knowledge co-production. The Tandem Framework largely builds on knowledge co-production theory (Daniels et al. 2020), as previous research shows that co-production makes climate services more effective, usable, and contextual (Vaughan and Dessai 2014). Knowledge co-production is often cited as a mean to address complex systematic challenges, thus making it well-fit for developing climate services (Harvey et al. 2019).

2.2 Approach: Knowledge co-production

There is no universal definition of knowledge co-production (Jahn and Keil 2015). The many meanings, terms, and aims attached to knowledge co-production allow for application in many different disciplines (Bremer and Meisch 2017) such as health research, rural development, sustainability science, and others (Vincent et al. 2018). However, terminological ambiguity arises when diverse disciplinary and professional jargons merge (Jahn and Keil 2015). Many terms are used interchangeably (Lang et al. 2012; Norström et al. 2020). Knowledge co-production can be referred to as transdisciplinary (e.g., Belcher et al. 2016), participatory research (e.g., Blackstock et al. 2007), communities of practice (Fulgenzi et al. 2020), knowledge exchange (e.g., Fazey et al. 2014), joint knowledge production (e.g., Hegger et al. 2012), science-policy interfaces (e.g., Heink et al. 2015), and knowledge integration (e.g., Hitziger et al. 2019). Such research methods are hereinafter referred to as knowledge co-production. Terminological ambiguity can negatively impact evaluation practices, as it makes it difficult to share and transfer findings across cases (Bremer and Meisch 2017). Therefore, it is vital to understand how knowledge co-production is conceptualized to effectively evaluate co-produced climate services.

Looking at the co-production literature, some key defining characteristics emerge. Knowledge co-production cuts traditional disciplinary and professional boundaries by involving various stakeholders, thus multiple perspectives, in a collaborative research process to solve pressing societal issues and produce context-specific knowledge (e.g Carew and Wickson 2010; Fulgenzi et al. 2020; Roux et al. 2010). Knowledge co-production initiatives can have many different purposes, commonly at least one of the following: bridging the usability gap; strengthening adaptive capacity in government institutions; embracing complexity; facilitating social learning; empowering traditional environmental knowledge systems; and increasing public participation (Bremer and Meisch 2017).

In knowledge co-production, stakeholder refers to "any person or group who has an interest in the research topic and/or who stands to gain or lose from a possible policy change that, directly or indirectly, might be influenced by the research findings" (Slunge et al. 2017, p. 4). Stakeholders can be divided into three groups when co-producing climate services: the *users* of climate information; the *providers* who produce information to support climate change adaptation decision-making; and *intermediaries* who facilitate or support the co-production process (Salamanca and Biskupska 2021). In addition, funding agencies play a critical role when co-producing climate services, as they set the agenda by allocating funds to whatever they see as a priority (Roux et al 2010). It is however worth noting that roles are fluid, meaning that stakeholders are likely to adopt many different roles as the co-production initiative progresses (Bremer et al. 2019).

The way knowledge co-production is conceptualized and implemented, including its aims and terms, affects what effects that emerge and thus what is being evaluated (Fazey et al. 2014). The literature shows that knowledge co-production has numerous effects, many which lies beyond traditional academic outputs (Sarkki et al. 2015; Zscheischler et al. 2018). Possible effects include, not limited to, improved networks (Wall et al. 2017), enriched decision-making (Heink et al. 2015; Sarkki et al. 2015), and social learning (Blackstock et al. 2007). It is worth

noting that some of these effects may emerged after the end of a project, thus imposing additional challenges for evaluation.

To conclude, the evaluative principles should be broad enough to capture the diversity in knowledge co-production, thus fit well with many definitions, aims, and effects. Therefore, a broad definition of knowledge co-production is employed that captures key characteristics, while allowing for a great level of flexibility to support many different purposes and contexts. Drawing inspiration from Norström et al. (2020), knowledge co-production is hereby defined as a collaborative research process involving diverse types of expertise and actors, to solve real-world problems and produce context-specific knowledge.

2.3 Intervention: Evaluation

Like knowledge co-production, the concept of evaluation cuts across disciplinary and professional boundaries. Evaluations can take many different forms depending on what context needs to be assessed, and for what purpose (Madaus and Stufflebeam 2000). Cambridge University Press (2021, np.) presents a definition of evaluation in its simplest form: "the process of judging or calculating the quality, importance, amount, or value of something". Reviewing the evaluation literature, two broad typologies emerge: summative and formative.

The summative approach takes place at the end of an intervention to assess its overall merit (Fazey et al. 2014). Summative evaluations tend to employ a linear cause-effect model and rigid methodologies to measure change, making the summative approach unfit for endeavors like co-produced climate services (Patton 2010). Climate services are facing an inherent uncertainty, meaning that they need to be flexible and adapt as change unfolds (Salamanca and Biskupska 2021). The summative approach fails to provide this flexibility (Patton 2010). Moreover, Fazey et al. (2014, p.206) note that the summative approach "focus more on the outcomes rather than the processes that led to them", meaning that it fails to capture the process-centric element in co-produced climate services (Daniels et al. 2020).

Formative assessments are embedded into the project cycle to enhance learning to improve project performance (e.g. Madaus and Stufflebeam 2000; van Tulder and Keen 2018). The formative approach fails to provide the flexibility that is required when evaluating co-produced climate services (Salamanca and Biskupska 2021). Formative assessments support improvements towards a pre-defined objective, whereas it is common for co-production initiatives to experience changes in objectives as the process evolves (Blackstock et al. 2007). The evaluation framework must adjust to such changes in conditions (Patton 2010), to capture the true value of co-produced climate services.

Developmental evaluation emerged as a response to the limitations imposed by the summative and formative typologies (Patton 2010). The developmental evaluation corresponds to the assumptions that underpin knowledge co-production initiatives. Co-production builds on the assumption that change is complex, non-linear, and emergent (Norström et al. 2020). Developmental evaluation is designed to understand such complexity. Drawing inspiration from complexity theory, the developmental evaluation sets to support adaptive management in social innovation initiatives, like the co-production of climate services, characterized by complexity, emergence, stakeholder diversity, long time-horizons, and uncertainty (e.g.,

Dozois et al. 2011; van Tulder and Keen 2018). There is no universal definition of adaptive management, but for the purpose of this thesis adaptive management is defined as an iterative process of "learning by doing" (Reever Morghan et al. 2006).

As the name entails, developmental evaluation puts emphasis on development rather than accountability or improvement. Lemon and Mitchell (2020, p.95) cleverly describe developmental evaluation as "evaluation for development rather than evaluation of development". Table 1 shows the main differences between traditional and developmental evaluation. The traditional evaluation is here considered as a contrasting heuristic device to better explain the developmental evaluation niche. The traditional evaluation is therefore described in a rather simplistic terms, meaning that some of the diversity associated with such approaches is overlooked.

Table 1: Traditional and developmental evaluation (Patton 2006; Patton 2010).

	Traditional evaluation	**Developmental evaluation**
Purpose	Formative-summative divide, in which formative improves and summative tests. Measure success against fixed objectives.	Support continuous development of innovation and adaptation as new insights emerge in a context characterized by complexity.
Roles and responsibilities	Evaluator is independent to ensure an impartial assessment.	Evaluator embedded in the intervention to facilitate adaptive learning.
Timing	Often one-off events.	Continuous.
Theoretical underpinnings	Linear cause-effect model.	Complexity theory.
Approaches to uncertainty	Quixotic control, aims for predictability and certainty.	High tolerance for ambiguity and uncertainty, adapting as change unfolds.
Results	Reports that cover lessons learnt, to support similar projects in other contexts.	Real-time feedback.

The co-production of climate services is highly dependent on contextual factors, making universally-adopted guidelines and practices undesirable (Jahn and Keil 2015). The initial stages of a co-production initiative are characterized by uncertainty, as practices adapt as the project evolves (Laycock et al. 2019). There is also an inherent uncertainty associated with the problem, climate change, which the co-produced climate services seek to address (Hegger et al. 2012). Goals and strategies are likely to change as the project evolves, thus making it unfeasible to try to measure success against a set of pre-defined objectives. As an attempt to handle such uncertainty, developmental evaluation practices entails rapid, real-time feedback to nurture adaptive decision-making and social learning (Patton 2010), and can thus help co-production initiatives to develop in an environment characterized by high uncertainty and complexity.

CHAPTER 2: HYPOTHESIS DEVELOPMENT

3 Methodology

The first part of the research aimed to develop a hypothesis. It therefore examined previous attempts to evaluate knowledge co-production initiatives, paying special attention to evaluation challenges and enabling factors to overcome them. A systematic literature review was performed, using the systematic snowballing approach presented by Wohlin (2014) with some methodological additions from Dawkins et al. (2019). The systematic snowballing approach allows for a comprehensive synthesis of existing literature, while clarifying any controversies or gaps. Its systematic nature, and strict series of guidelines, reduce subjectivity while ensuring transparency and replicability (Wohlin 2014).

The systematic snowballing approach consisted of five steps: 1) determine eligibility criteria, identify a start set, 3) literature search, 4) secondary screening, and 5) analysis. Each step is described in the sections below. The process is visualized in Figure 2.

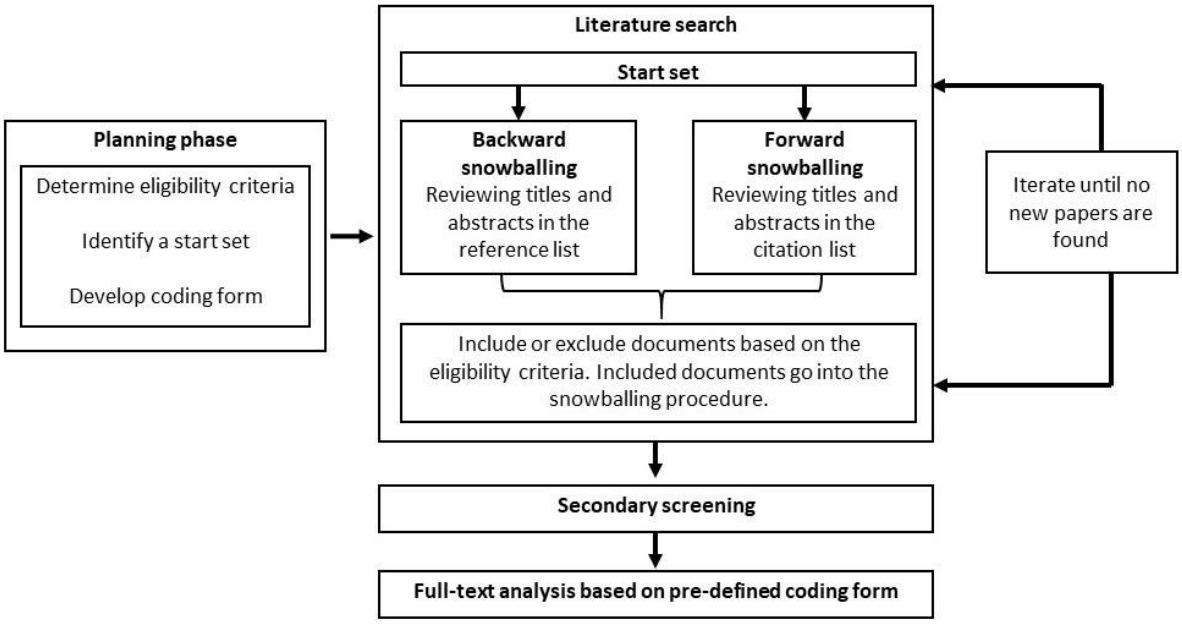

Figure 2: Snowballing approach (Wohlin 2014)

3.1 Determine eligibility criteria

The first step of the literature review was to develop a set of eligibility criteria that determined the basic conditions that a document must fulfill for inclusion in the final sample. The eligibility criteria were designed to reflect the research aim and conceptual framework, which seeks to understand what methods can be used to evaluate the process and effects of co-produced climate services. Four criteria were identified.

First, the type of *intervention* that the document must address was specified, in this case evaluations.

The second criterion determined the *approach* which the intervention addressed. The approach was specified as knowledge co-production, and included related research practices like transdisciplinary research, participatory methods, multi-actor projects, and others. Documents

that considered adjacent topics like knowledge brokers, interdisciplinary research, public participation, or research centers were not included.

The third criterion specified what *area* was considered, in this case climate services. However, solely focusing on climate services proved insufficient due to the lack of previous research. Therefore, the area was broadened to capture other adjacent topics gathered under the umbrella term "sustainability science", including, but not limited to, environmental research, climate change adaptation, and natural resource management. Although climate services and sustainability science are different research topics, they both face similar challenges invoked by their inherent wickedness. Wicked problems are characterized by "multiple and conflicting inputs and multiple possible outcomes, all of which play over time against, or occasionally with, each other" (Sun and Yang 2016 p. 2). Consequently, wicked problems have no definitive formulation or solution (Vink et al., 2013). In addition, sustainability science and climate services can be viewed as different facets of the same crisis – climate change. It is however worth noting that sustainability science is much broader discipline, thus including many other topics in adjacent to climate change.

The fourth, and final, criterion referred to the *setting*. For this study, only studies published in scientific journals were included. Thus, no grey literature was considered. It is worth noting that this may have resulted in a publication bias, as unpublished material was overlooked (Haddaway et al. 2015). Moreover, the publication needed to be in English.

3.2 Start set

The second step of the literature review entailed identifying relevant articles to form a start set (Wohlin 2014). The creation of the start set began with a preliminary literature search on Scopus. The preliminary literature search aimed to provide a very basic understanding of available research, rather than producing a comprehensive list of available literature. The search string was therefore kept simple, thus excluding the previously mentioned synonyms. The search string was structured as following: *TITLE-ABS-KEY (evaluating AND knowledge AND co-production).* The preliminary literature search yielded 32 documents, each screened on their title and abstract to assess their relevance. In total, 4 documents were tentatively included.

To increase comprehensiveness, an expert in the field was asked to share any relevant literature. The expert recommendations generated an additional 6 documents.

The final inclusion or exclusion was based on a secondary screening process, looking how many times a document was cited. Documents with many citations were prioritized, as this would provide a larger input for the snowballing. It is worth noting that this selection process unintentionally favored older documents, as they have circulated longer and thus given the opportunity to gain more citations. However, the documents excluded from the start set were included in the final sample of literature as they were identified in the literature search. The start set is presented in Appendix A.

3.3 Literature search

Stage three entailed an extensive literature search, applying backward and forward snowballing in iterations. Backward snowballing meant that the reference list of the documents in the start set were reviewed, whereas citations were considered during forward snowballing. Citations and references lists were identified using the well-regarded database Scopus in October 2020. Documents were included for a secondary screening in case their title and abstract met the eligibility criteria. Documents found during the initial iteration were added to the start set, and subject to snowballing during the next iteration. The process continued until no additional documents were found.

In total, the literature search generated 2384 documents. A total of 70 documents were included for a secondary screening. The process is visualized in Figure 3 below.

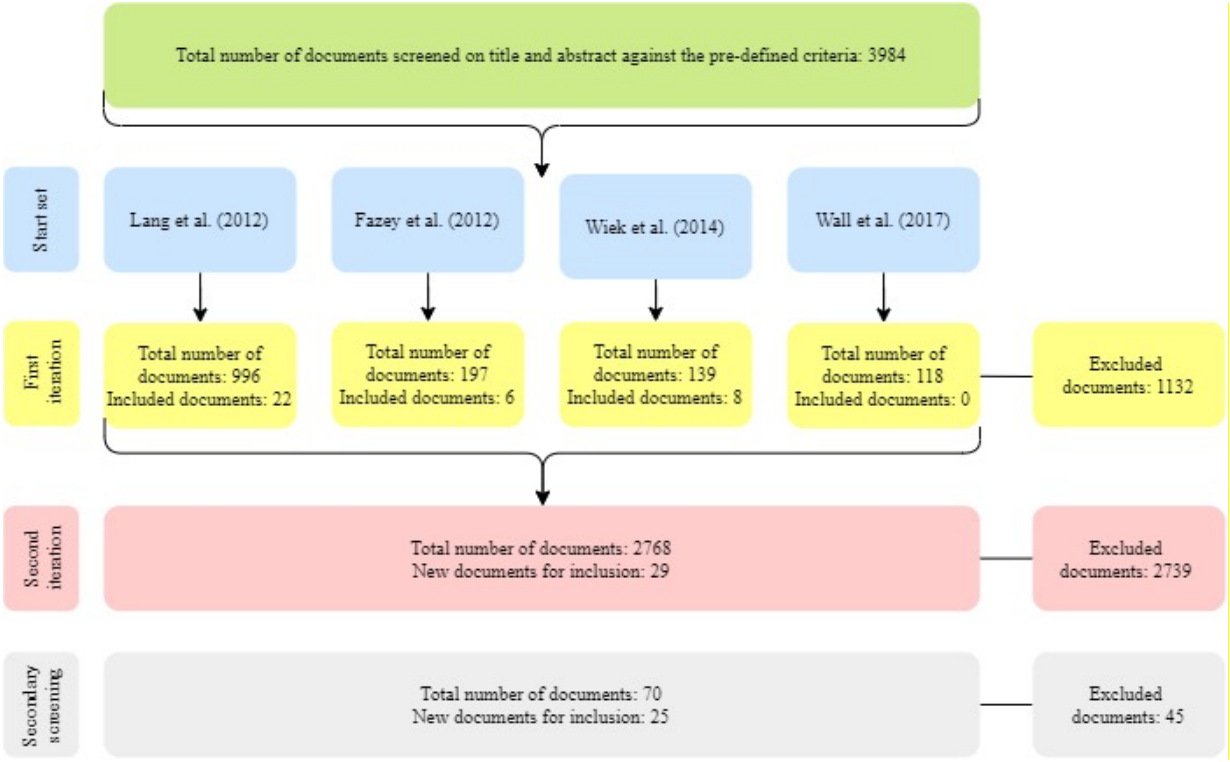

Figure 3: The literature selection process

3.4 Secondary screening

The next step was a secondary screening, in order to make the sample more manageable considering the given time and resources constraints. The secondary screening also provided an additional opportunity to verify all the documents against the criteria, thus ensure full relevance. It entailed a full-text screening. Any unfitting documents were at this stage excluded.

An additional eligibility criterion was added. The criterion specified that only conceptual papers were to be included. Conceptual papers here refer to those papers that focus on developing new evaluation approaches for assessing co-production initiatives. The conceptual papers were considered well-fit as they mimicked the research question's conceptual side. Case studies were excluded, as most focused on the co-production initiative itself rather than the development of an evaluation approach. It is however worth noting that some papers were both conceptual and

empirical, as they developed and tested a novel evaluation approach. Such papers were included, to gain a better understanding of potential practical challenges and success factors that may arise when evaluating co-production initiatives.

In total, 25 documents were included for analysis. The documents were retrieved from a range of scientific journals, and showed great diversity in terms of disciplines that use knowledge co-production approaches. This included sustainability science, natural resource management, environmental research, socio-ecological research, water management, climate change adaptation, and climate science. None of the reviewed documents explicitly considered climate services. For the full list of the included literature, see Appendix B.

3.5 Analysis

Once selected, documents were coded in Excel. A deductive approach was employed, using a pre-defined coding form (see Appendix C). The coding form was based on a preliminary literature search. The coding form ensured consistent data extraction while controlling for subjectivity and biases (Haddaway et al. 2015).

Three types of codes were considered. First, basic citation information was noted. Second, conceptualizations and approaches to knowledge co-production were registered to avoid any terminological ambiguity. Third, the proposed evaluation design was considered. Special attention was paid to challenges and enabling factors to overcome these. It is worth noting that some additional information was registered under the code "other" to make sure that no important data was lost, despite not fitting into any of the other categories.

Findings were then synthesized. A narrative synthesis was applied, which adopted a textual approach to summaries the findings (Popay et al. 2006). Tables were constructed to identify emerging themes and patterns in the data (Dawkins et al. 2019). Some simple quantitative analysis was applied to codes that were the easiest to quantify. The quantitative analysis is presented in Appendix D.

Thereafter a hypothesis was formulated, see section 5 for more information.

3.6 Limitations

Some limitations exist. There is always a risk of missing some literature. The comprehensiveness can be questioned, ultimately weakening the theoretical grounding of the principles. Drawing from inspiration from Haddaway et al. (2015), two major limitations were identified:

- The literature search was limited to one database.
- A publication bias occurred as only peer-reviewed articles were considered.

However, theoretical saturation occurred arguably showing on a sufficient sample regardless of whether it was fully comprehensive or not. The depth of the literature review was therefore considered to meet its formulated goal, as it provided a thorough overview of the topic in question.

The literature review was also limited in terms of its scope. Boundaries were established to match the given time and resource constraints, unfortunately sacrificing some of the overall

comprehensiveness. The review did not include any grey literature, thus excluding some of the available knowledge base. It is also worth noting that the excluded case studies may have provided much-needed practical insights.

4 Results: Literature review

This section presents a synthesis of previous academic literature that have developed approaches, methods, and frameworks to evaluate knowledge co-production initiatives. Potential challenges and success factors in evaluating such initiatives are also introduced.

4.1 Approaches to evaluating knowledge co-production

Evaluating knowledge co-production may handle one or more of the following: *outputs*, which are the tangible products of the process; *outcomes*, as the effects of the process on the knowledge, attitude, or behavior on the involved actors; and *impact*, as the long-term effects of the knowledge co-production process (Hassenforder et al. 2016). Traditionally, research evaluations employ reductionist procedures solely focusing on assessing the academic outputs, thus failing to capture the broad range of effects associated with knowledge co-production (Sarkki et al. 2015; Zscheischler et al. 2018). Consequently, most of the reviewed documents proposed an evaluation approach that holistically appraised outputs, outcomes, and impact.

Most of the reviewed documents suggested a formative assessment (e.g. Jones et al. 2009; Lang et al. 2012; Sarkki et al. 2015). Other documents combined the summative and formative approach (Hegger et al. 2012; Roux et al. 2010; Wickson et al. 2006). Some documents created an adaptable evaluation framework, open to both summative and formative approaches depending on the context (Carew and Wickson 2010; Holzer et al. 2018). Only two documents employed a summative approach (Hassenforder et al. 2016; O'Connor et al. 2019).

Per definition, a formative approach is integrated into the project it seeks to assess (Fazey et al. 2014). Consequently, many of the reviewed documents argued for implementing multiple evaluation interventions at different points of time. Wall et al. (2017) claimed that integrating the evaluation early in the project cycle is crucial to build trust among the involved actors. Roux et al. (2010, p.735) explained that "while post-program evaluations provide a sense of closure, they usually do not provide for reflection (…). Also, they happen too late for the associated learning to influence the direction of the program being evaluated". Embedded evaluations also offer a possibility to compare change across different points in time (van der Wal et al. 2014) Walter et al. (2007) argued for performing an evaluation after the completion of a project due to the considerable time lags between cause and effect when co-producing knowledge. Other documents valued flexibility, meaning that the timing of the evaluation is adapted to fit the given situation (e.g. Belcher et al. 2020; Blackstock et al. 2007).

Universal criteria for evaluating knowledge co-production are a contested subject, as explained by O'Connor et al. (2019, p.2): "developing evaluation criteria for knowledge co-production remains a challenge because of its variety of forms, contexts, and participants who may have differing views of what is valuable". Moreover, a universal set of criteria is generally inappropriate when evaluating complex systems. However, some scholars argued that universal

criteria can serve as a guide to evaluators, funders, and project managers (Belcher et al. 2016). A full list of criteria derived from the reviewed literature are presented in Table 2.

Table 2: Criteria for evaluating knowledge co-production

Category	Indicator	Description	References
Enabling environment	**Access to resources**	Referring to the support to participants for them to meet their responsibilities, and includes: Competence, time, scientific disciplines in research team, budget, adequate infrastructure, practical information, and staffing	Belcher et al. 2020; Blackstock et al. 2007; Fulgenzi et al. 2020; Hitziger et al. 2019; Knickel et al. 2019; Lang et al. 2012a; Roux et al. 2010; Sarkki et al. 2015; Wall et al. 2017
	Drivers	Referring to the incentives behind the project and ultimately the ownership of project outcomes – can be demand-driven or supply-driven	Blackstock et al. 2007; Hassenforder et al. 2016; Knickel et al. 2019; Sarkki et al. 2015
	External context	Referring to the political, economic, environmental, and social aspects of the system in which the project operates in, with a focus on understanding complexity, system boundaries, finding synergies and catalyzing events	Belcher et al. 2020; Blackstock et al. 2007; Hassenforder et al. 2016; Hitziger et al. 2019; Jahn and Keil 2015; Roux et al. 2010; Wall et al. 2017
	Expectations	Referring to how confident participants are that the process will yield positive effects	Hassenforder et al. 2016
	Institutional memory	Referring to safeguarding mechanisms to protect the acquired collective knowledge	Hitziger et al. 2019; Jahn and Keil 2015; Knickel et al. 2019
	Preexisting relationships	Referring to preexisting professional relationships between the involved actors	Wall et al. 2017
	Willingness to learn	Referring to the capacity and personal motivation to participate in project activities	Blackstock et al. 2007; Hitziger et al. 2019; Jahn and Keil 2015; Knickel et al. 2019; Lang et al. 2012; Roux et al. 2010; Sarkki et al. 2015; Wall et al. 2017; Walter et al. 2007
Process	**Awareness**	Referring to how participants can identify available resources and possible gaps	Fulgenzi et al. 2020
	Capacity development	Referring to the process of developing, strengthening, maintaining, and adapting skills, knowledge, and awareness to meet the defined objectives	Blackstock et al. 2007; Fulgenzi et al. 2020; Roux et al. 2010; Sarkki et al. 2015; Wall et al. 2017
	Co-location	Referring to if involved actors are willing to host PhD-students	Roux et al. 2010
	Conflict resolution	Referring to the degree of conflict between participants, and the ability to manage such conflict	Blackstock et al. 2007; Hitziger et al. 2019; Knickel et al. 2019; Lang et al. 2012; Sarkki et al. 2015; Wall et al. 2017
	Continuity	Referring to the consistency in participation, in terms of recurring participants	Wall et al. 2017
	Cost effectiveness	Referring to the relationship between costs and benefits of the project	Blackstock et al. 2007; Jahn and Keil 2015

	Effective collaboration	Referring to mechanisms promoting effectiveness, such as clearly defined project plan, continuous documentations, meeting agendas, defined roles and responsibilities, and experience of the moderator	Belcher et al. 2020; Fulgenzi et al. 2020; Knickel et al. 2019; Wall et al. 2017
	Effective communication	Referring to the appropriateness, relevance, clarity, and accessibility of the communicated information	Belcher et al. 2020; Hitziger et al. 2019; Jahn and Keil 2015; 2015; Knickel et al. 2019; Roux et al. 2010; Sarkki et al. 2015; Wall et al. 2017; Walter et al. 2007
	Ethical aspects	Referring to how the research is adhering to ethical standards	Belcher et al. 2020
	Inclusion of all relevant perspectives	Referring to the creation of a safe space for participants to voice their opinions and influence the decision-making process	Blackstock et al. 2007; Fulgenzi et al. 2020; Hassenforder et al. 2016; Hitziger et al. 2019; Knickel et al. 2019; Walter et al. 2007
	Involvement	Referring to at what time of the process different participants are engaged	Knickel et al. 2019
	Leadership	Referring to the presence of leadership figure facilitating the process	Blackstock et al. 2007; Hitziger et al. 2019; Jahn and Keil 2015; Knickel et al. 2019; Lang et al. 2012; Roux et al. 2010; Sarkki et al. 2015; Wall et al. 2017; Walter et al. 2007
	Methods	Referring to the appropriateness of the selected disciplines, epistemology, methods, approaches, and theories	Belcher et al. 2020; Hassenforder et al. 2016; Hitziger et al. 2019; Jahn and Keil 2015; Knickel et al. 2019; Lang et al. 2012; Wall et al. 2017
	Objectives	Referring to the collaborative process setting project goals	Belcher et al. 2020; Blackstock et al. 2007; Hassenforder et al. 2016; Hitziger et al. 2019; Jahn and Keil 2015; Knickel et al. 2019; Lang et al. 2012; Sarkki et al. 2015; Wall et al. 2017; Walter et al. 2007
	Practicalities	Referring to the frequency of events, number of participants, and setting of exchange	Hassenforder et al. 2016; Hitziger et al. 2019
	Reflection	Referring to the opportunities to reflect upon the collective experience and adjusting the project plan accordingly	Belcher et al. 2020; Fulgenzi et al. 2020; Hitziger et al. 2019; Jahn and Keil 2015; Knickel et al. 2019; Lang et al. 2012; Roux et al. 2010; Sarkki et al. 2015; Walter et al. 2007
	Representation	Referring to the genuine inclusion of a diverse set of actors	Belcher et al. 2020; Blackstock et al. 2007; Fulgenzi et al. 2020; Hassenforder et al. 2016; Hitziger et al. 2019; Jahn and Keil 2015; Knickel et al. 2019; Lang et al. 2012; Sarkki et al. 2015; Wall et al. 2017
	Theory of change	Referring to the development of a theory of change that matches the objectives and clarifies assumptions	Belcher et al. 2020; Hitziger et al. 2019; Jahn and Keil 2015

	Transparency	Referring to how well participants and observers can understand the process	Blackstock et al. 2007; Knickel et al. 2019
Effects	**Accountability**	Referring to how well participants satisfied their personal core responsibilities	Blackstock et al. 2007
	Inspiration	Referring to the motivation to pursue follow-up projects	Fulgenzi et al. 2020; Wall et al. 2017
	Outcomes	Referring to if the defined objectives are achieved	Lang et al. 2012; Wall et al. 2017
	Outputs	Referring to the timely delivery of the tangible products, including peer-reviewed articles, gray literature, workshops, meetings, and final reports	Hassenforder et al. 2016; Knickel et al. 2019; Roux et al. 2010; Wall et al. 2017
	Quality of decision-making	Referring to the implementation, integration, and maintenance of project findings	Blackstock et al. 2007; Jahn and Keil 2015; Knickel et al. 2019; Roux et al. 2010
	Quality of research product	Referring to the validity of the final research product, considering legitimacy, transferability, credibility, comprehensiveness, and robustness	Blackstock et al. 2007; Sarkki et al. 2015; Wall et al. 2017
	Recognized impacts	Referring the perceived change associated with the project, including unintended effects, changes in perspectives, and improved organizational performance.	Blackstock et al. 2007; Fulgenzi et al. 2020; Hassenforder et al. 2016; Jahn and Keil 2015
	Relationships	Referring to the improvements in social capital, and the development of new social networks	Blackstock et al. 2007; Fulgenzi et al. 2020; Knickel et al. 2019; Lang et al. 2012; Sarkki et al. 2015; Walter et al. 2007
	Relevance to society	Referring to the significant outcome where research findings are used in practice to solve the targeted problem	Belcher et al. 2020; Jahn and Keil 2015; Knickel et al. 2019; Lang et al. 2012; Roux et al. 2010; Sarkki et al. 2015; Wall et al. 2017
	Social learning	Referring to changes in individual's values and behavior, ultimately reflected as changes in the collective culture and values	Blackstock et al. 2007; Hitziger et al. 2019; Jahn and Keil 2015; Knickel et al. 2019; Wall et al. 2017

4.2 Evaluation design and data collection methods for evaluating knowledge co-production

A systematic review performed by Ernst (2019) showed the many methods used in evaluating knowledge co-production initiatives, such as questionnaires, interviews, document analysis, observation, and others. Around half of the reviewed documents suggested an evaluation design and data collection methods, as shown in Table 3.

Table 3: Evaluation design and data collection methods

Evaluation design	Data collection methods	Reference
Mix-method	Data collection methods were sequenced to serve a specific purpose at different points of time during the evaluation.	Holzer et al. 2018; Jones et al. 2009; Wiek et al. 2014
Quantitative	Different types of data collection that allowed for statistical analysis: • Likert scale questionnaire • Questionnaire based on descriptive statements of cultural theory archetypes	Fulgenzi et al. 2020; Hitziger et al. 2019; van der Wal et al. 2014; Walter et al. 2007; Zscheischler et al. 2018
Participatory evaluation	Methods that aim to enhance stakeholder participation.	Fazey et al. 2014; Norström et al. 2020.

Some of the reviewed documents suggested a mix-method approach that combined qualitative and quantitative data collection methods (Holzer et al. 2018; Jones et al. 2009; Walter et al. 2007; Wiek et al. 2014). Data collection methods were sequenced to serve a specific purpose at different points in time during the evaluation. Holzer et al. (2018) developed a six step approach, where each step built on the former: 1) qualitative interviews, 2) qualitative data analysis, 3) questionnaire, 4) statistical analysis, 5) focus group discussion, and 6) incorporating insights from the focus group discussion. Wiek et al. (2014) used similar methods but in another sequence: 1) documentation review, 2) questionnaire, and 3) qualitative interviews. Another approach was presented by Jones et al. (2009), who started with reviewing project documentation to form a foundational understanding of the context, followed by a mix of interviews, questionnaire, and informal discussions with the involved stakeholders.

Many of the reviewed documents argued for a quantitative approach using questionnaires. Most of the reviewed documents used a self-administered questionnaire with Likert-scale questions (e.g., Fulgenzi et al. 2020; Walter et al. 2007). Hitziger et al. (2019) made the questionnaire participatory, by engaging the involved stakeholders in a discussion that aimed towards consensus on Likert-scale scores. van der Wal et al. (2014) designed a questionnaire based on descriptive statements of cultural theory archetypes.

Other documents suggested a participatory approach (Fazey et al. 2014; Norström et al. 2020). Fazey et al. (2014, p.218) claimed that participation "enhances ownership and responsibility for the knowledge exchange process, clarifies roles, and can facilitate discussions about different perspectives of knowledge and its exchange".

4.3 Frameworks for evaluating knowledge co-production

Five types of frameworks emerged in the reviewed literature. The frameworks differed in terms of design, purpose, and scope. There are, however, some significant overlaps between the frameworks. The frameworks are described below. An overview is presented in Table 4.

Table 4: Frameworks for evaluating knowledge co-production

Foci	Framework	Reference
Academic knowledge quality in co-production processes	Applies the CRELE attributes: credibility, relevance (or saliency), and legitimacy. Some scholars add other attributes like "iteratively" and "effectiveness".	Belcher et al. 2016; Hegger et al. 2012; Heink et al. 2015; Sarkki et al. 2015
The interplay between context, process, and effects	Uses different arrangements depending on the author. Most structure the framework around a few criteria that look at the context, process, and effects.	Carew and Wickson 2010; Hassenforder et al. 2015; Jahn and Keil 2015; Knickel et al. 2019
The participatory process itself in relation to the produced outcome	Formulates different success factors needed in the co-production process in order to produce the desired effects.	Fulgenzi et al. 2020; Norström et al. 2020; O'Connor et al. 2019; Wall et al. 2017
Societal effects	Looks at different types of societal impacts.	Walter et al. 2007; Wiek et al. 2014
Cultural Theory	Assess changes in perception through descriptive statements of four archetypical perspectives.	van der Wal et al. 2014

The first group concerned evaluation frameworks that aim to measure research quality by applying the CRELE attributes: credibility, relevance (or saliency), and legitimacy (Belcher et al. 2016; Hegger et al. 2012; Heink et al. 2015; Sarkki et al. 2015). The CRELE attributes were initially developed to assess the effectiveness of sustainability science, paying little attention to knowledge co-production. However, the CRELE attributes are increasingly applied in knowledge co-production evaluations (Heink et al. 2015). Sarkki et al. (2015) introduced "iteratively" as a fourth attribute to capture the complex interactions between science and society, and Belcher et al. (2016) added "effectiveness" to assess research impact on real-word problems.

The second group referred to evaluation frameworks that appraised the interplay between context, process, and effects (Carew and Wickson 2010; Hassenforder et al. 2016; Jahn and Keil 2015; Knickel et al. 2019). These attempts used various arrangements: Carew and Wickson (2010) created an adaptable heuristic called the Transdisciplinary Wheel, where the process and product are connected through loops that are situated within a larger wheel of context. Jahn and Keil (2015) provided nine quality criteria encompassing the quality of the research problems (context), research process (process), and research results (product), corresponding to Hassenforder et al. (2016) who considered context-related characteristics, process-related characteristics, and output, outcome, and impact-related characteristics. Knickel et al. (2019) provided criteria structured along four key dimensions: context, approach, process, and outcomes.

The third group involved of evaluation frameworks that assessed the participatory process itself in relation to the produced outcome (Fulgenzi et al. 2020; Norström et al. 2020a; O'Connor et al. 2019; Wall et al. 2017). O'Connor et al. (2019) identified eight elements in the knowledge co-production process which led to enriched decision-making, whereas (Fulgenzi et al. 2020) formulated six interdependent success factors that account for knowledge co-production principles and social learning outcomes. Wall et al. (2017) on the other hand looked at context factors that related to pre-existing conditions, process indicators, and three types of project results: outputs, outcomes, and impacts. Norström et al. (2020) considered the process itself as the most important aspect to evaluate, and looked at the following four principles: 1) context-based, 2) pluralistic, 3) goal-oriented, and 4) interactive.

The fourth group of evaluation frameworks focused on measuring societal effects (Walter et al. 2007; Wiek et al. 2014). Walter et al. (2007) formulated six types of societal impacts: network building, enhanced trust, understanding of others, distribution of knowledge and transformation knowledge. Similar but different, Wiek et al. (2014) looked at societal effects in terms of usable products, enhanced capacity, network effects, and structural changes.

The fifth and last group of evaluation frameworks consists of only one scholar – van der Wal et al. (2014). Drawing inspiration from Cultural Theory, van der Wal et al. (2014 p. 3) used descriptive statements of four archetypical perspectives: 1) hierarchy, 2) individualist, 3) egalitarian, and 4) fatalist. Each archetypical perspective consisted of a "world view (how the world is seen) and management style (how the world should be managed)". The archetypical perspectives were operationalized into a scoring table, that participants were asked to complete at multiple occasions throughout the project. This, in order to identify potential changes in attitudes, beliefs, and behavior amongst the participants.

4.4 Challenges in evaluating knowledge co-production

Some challenges arise when evaluating knowledge co-production initiatives. For a full list of challenges, see Appendix H.

According to the reviewed documents, a major challenge in evaluating knowledge co-production is the complexity of the process itself and of the system in which it operates in (e.g. Fazey et al. 2014; Lang et al. 2012; Roux et al. 2010). As previously mentioned, complex systems are characterized by non-linearity, multi-pathways, emergent properties, dynamic change, and interdependencies (Zscheischler et al. 2018), which combined with the inherent uncertainty and long timeframes in sustainability science makes it difficult to establish causality (e.g. Hitzigeret al. 2019; Jahn and Keil 2015; Norström et al. 2020). Knowledge co-production take place in a larger context, making direct mono-causal connections difficult to prove due to the on-going influence of unforeseeable external factors (Zscheischler et al. 2018). Moreover, some scholars have noted that a reductionist approach that tries to fit complex systems into a few variables easily falls into oversimplification (Hassenforder et al. 2016; Jones et al. 2009). The intangible nature of many key elements in knowledge co-production, such as learning, empowerment and trust, means that many variables cannot be measured objectively,

but must be based on stakeholder's subjective perceptions (Blackstock et al. 2007; Hassenforder et al. 2016).

Some of the reviewed documents considered the timing of the evaluation as a challenge in itself, as there is a lack of an academic consensus defining the preferred timing of the evaluation. Broadly speaking, different results are captured depending on the timing of the evaluation (Fulgenzi et al. 2020; Wall et al. 2017). As Ernst (2019, p.9) said: "if a survey is conducted immediately after a participation event, certain social learning elements such as acquisition of knowledge can be measured, but other elements such as cognitive changes might not have taken place". Accordingly, Roux et al. (2010) argued that the exploration of causal relationships must extend beyond a single spatial scale or timeframe. There are also significant time-lags between causes and effects where societal impacts evolve over a long period of time (Blackstock et al. 2007; Jahn and Keil 2015; Wall et al. 2017), making them difficult to capture within the given project boundaries (Norström et al. 2020).

Many of the reviewed documents identified the diversity of the involved stakeholders as a challenge in evaluating co-production initiatives (e.g. Hitziger et al. 2019; Jones et al. 2009; Roux et al. 2010). Stakeholders can hold contrasting values, epistemological beliefs, educational background, professional jargons, or objectives. Motivation can also vary. For some stakeholders, evaluations are perceived as purely a burden that distracts from main project activities, especially considering that projects often struggle with limited resources whilst evaluations are resource-intensive endeavors (Knickel et al. 2019). Another challenge is staff-turnover, confining the evaluation in the initial stages of building trust, gaining buy-in, and establishing mutual ownership (Sarkki et al. 2015). Without a sense of trust, some actors can feel reluctant to share the full problem-picture with outsiders (Knickel et al. 2019; Wiek et al. 2014).

Many scholars noted that there is no academic consensus defining good practices in relation to knowledge co-production (e.g. Ernst 2019; Fulgenzi et al. 2020; Heink et al. 2015). However, a universal definition may not be desirable, as the quality of knowledge co-production is multifaceted and context-dependent (Jahn and Keil 2015). In the absence of alternatives, evaluators turn to traditional, discipline-specific measures to form the basis of their assessments, that are insufficient when assessing the wide range of effects associated with knowledge co-production (Belcher et al. 2016). Also, tensions arise when trying to apply linear frameworks to capture change in a messy and complex reality (Jones et al. 2009; Walter et al. 2007). Knowledge co-production is in itself a complex adaptive system (Roux et al. 2010), whereas evaluation frameworks often employ a rigid approach that assesses change through the comparisons of pre- and post-data (Ernst 2019; Walter et al. 2007).

4.5 Success factors in evaluating knowledge co-production

Reviewing the literature, some success factors emerged for evaluating knowledge co-production initiatives. For a full list of success factors, see Appendix I.

The reviewed documents showed that flexibility is a major success factor when evaluating knowledge co-production (e.g. Knickel et al. 2019; Lang et al. 2012; van der Wal et al.

2014). Evaluation frameworks must be adaptable to the needs of the intended users, considering timing, purpose, scale, and context (Belcher et al. 2016; Knickel et al. 2019). In the reviewed documents, context referred to both internal aspects such as project objectives, financial resources, and purpose, as well as the broader system characteristics like legalization, social setting, and national policy structures (Sarkki et al. 2015). Flexibility also referred to the need for evaluation frameworks to adapt as new insights emerge (e.g. Belcher et al. 2016; Blackstock et al. 2007; Carew and Wickson 2010). Walter et al. (2007) called for new innovative evaluation approaches allowing for flexibility, where participatory evaluations may serve as a promising alternative in the context of knowledge co-production. Participatory evaluations support learning (Fazey et al. 2014), and allow actors to share their perspectives in consensus-driven discussions arguably making the evaluation a learning activity in itself (Lang et al. 2012). Furthermore, participatory evaluations can "capture the true value of a co-production process for those working within the particular context or issue, reveal unexpected impacts of the work and prompt the articulation of new context-specific projects and knowledge needs" (Norström et al. 2020, p.187).

Some scholars argued that evaluations should be integrated into the main project as early as possible to ensure enough time for social learning and trust to emerge (Roux et al. 2010; Wall et al. 2017). Many of the reviewed documents proposed stakeholder engagement when determining evaluation objectives to encourage ownership and buy-in from the involved stakeholders, whilst ensuring contextual relevance (Wiek et al. 2014). This corresponds to the participatory evaluation approach. Objectives should furthermore be revisited and adapted as new information emerges (Norström et al. 2020). Similarly, Fazey et al. (2014) noted that evaluation indicators should be based on an agreement between the involved stakeholders in order to ensure the inclusion of a diverse set of effects. Arguably, the evaluation should aim to capture the co-production process itself, and potential expected and unexpected effects (Belcher et al. 2016; Fazey et al. 2014; Wall et al. 2017). Intangible aspects should also be considered, despite being difficult to measure (Norström et al. 2020).

The reviewed documents recommended the Theory of Change (ToC) methodology, which is a project management tool that outlines how change is expected to occur. The ToC methodology is designed to capture change that take place in complex adaptive systems (Fazey et al. 2014). The ToC methodology can facilitate the evaluation of co-production initiatives as it allows the involved stakeholders to identify causal linkages between the intervention and its effects. If used in a participatory setting, the ToC methodology can help the involved stakeholders to develop a shared understanding of the evaluation process in relation to its objectives, indicators, and assumptions (Knickel et al. 2019; Norström et al. 2020; Wall et al. 2017).

Lastly, the reviewed documents introduced some practical aspects to consider when evaluating knowledge co-production. First, the evaluation should be designed to allow for maximum participation, where the timing, location, and communication should fit the needs of the involved stakeholders. Second, visual products can encourage meaningful discussions among the involved stakeholders by overcoming potential barriers like differences in educational backgrounds and preferred languages (Lang et al. 2012). Third, activities purposefully targeting

memory distortion may be necessary for ex-post evaluations (Wiek et al. 2014). Fourth, questionnaires should be anonymous to make sure that the stakeholders feel safe to share their opinion without being judged (Walter et al. 2007).

5 Hypothesis development

This section seeks to develop a hypothesis based on the findings in the literature review. First, the hypothesis development process is presented, followed by the actual hypothesis. The hypothesis is presented in the form of principles to facilitate the upcoming analysis and testing.

5.1 Hypothesis development process

Although none of the reviewed documents directly addressed climate services, they still presented many lessons relevant to the study. Challenges like complexity, long time horizons, uncertainty, and lack of academic guidance cut across disciplinary boundaries. Thus, many insights and approaches can be applied in the case of climate services. The principles drew inspiration from existing evaluation approaches presented in the literature, which have been adapted and combined to account for the specific challenges related to evaluating co-produced climate services. The hypothesis development process consisted of four steps; each is presented below.

Firstly, the challenges identified in the literature review were paired with a potential solution. The potential solutions were based on findings from the literature review, with special focus on the identified success factors in evaluating knowledge co-production. See Table 5 for an overview. It is worth noting, in line with Hassel (2010), that there are an infinite number of solutions to a single problem. It was impossible to identify *all* such solutions, and consequently impossible to state with certainty what the optimal solution entails. The goal was therefore not to find the optimal solution, but to find one solution which could provide relevant actors with some guidance on how to deal with the challenges that emerge when evaluating co-production initiatives.

Secondly, some underlying assumptions for the principles were created. This, in order for the principles to correspond to the conceptual framework presented in section 2. The conceptual framework built on the developmental evaluation line of thought. Developmental evaluation has many unique characteristics, which were transformed into the following four assumptions:

- The evaluation seeks to support adaptive management as the initiative evolves.
- The evaluation employs flexible practices.
- The evaluation is integrated into the project cycle to provide real-time feedback.
- The evaluator is embedded into the project team.

Thirdly, the potential solutions were reviewed to assess if they mirrored the assumptions presented above (as shown in Table 5). This, in order to eliminate potential solutions that merely imitated basic developmental evaluation characteristics. Such potential solutions were considered as a way of thinking rather than a concrete method, thus failing to support the research aim. The potential solutions that corresponded to the assumptions were therefore not included in the final principles, but instead considered in the conceptual framework in section

2. The assumptions, and thus developmental evaluation, accounted for challenges associated with long time horizons, timing, increased burden, lack of trust towards the evaluator, and uncertainty associated to the process itself and the problem which it tries to address.

Fourthly, the potential solutions were organized into groups based on emerging themes. The process is visualized in Figure 4. Themes were generated by reviewing the methods outlined in each potential solution. Three themes emerged, which were labelled and transformed into principles. Some overlaps occurred among the principles, showing some very early signs on an interconnectedness. Once the three principles were identified, some additional literature was reviewed. This, to collect further information on the methods outlined in the principles. This included some peer-reviewed articles, guidance documents, and reports on the methodological approaches outlined in the hypothesis.

Table 5: Hypothesis development process

Challenge	Potential solution	Assumption (yes/no)
Assess both tangible and intangible effects	Employ a mix-method approach to capture different types of effects.	No
Terminological ambiguity	Take a participatory approach for stakeholders to develop a shared vocabulary.	No
Long time horizons and time lags between cause and effects	Support an on-going approach that can be updated as outcomes and impacts as they unfold.	Yes
External factors	Use a ToC methodology that takes a system perspective to monitor relevant external factors.	No
Contribution	Use a ToC methodology where stakeholders can reflect upon causal linkages.	No
Timing	On-going process that captures different effects as they emerge in order to provide real-time feedback.	Yes
Complexity	Use a ToC methodology that employs participation and visualization to make it easier for stakeholders to grasp complexity.	No
Diversity of involved actors	Take a participatory approach using visual products to develop a shared vision and understanding among the involved stakeholders.	No
Increased burden	Stakeholder-driven evaluation process.	Yes
Lack of trust towards evaluator	Embed the evaluator into the project to build trust.	Yes
Uncertainty	Adapt as change unfolds.	Yes
Staff turnover	Produce clear visual project documentation to makes it easy for new staff to catch up on previous activities.	No

Note: The column called "Assumptions (yes/no)" shows whether a potential solution mimics the assumptions presented in page 23. If yes, the potential solution is excluded from the principles.

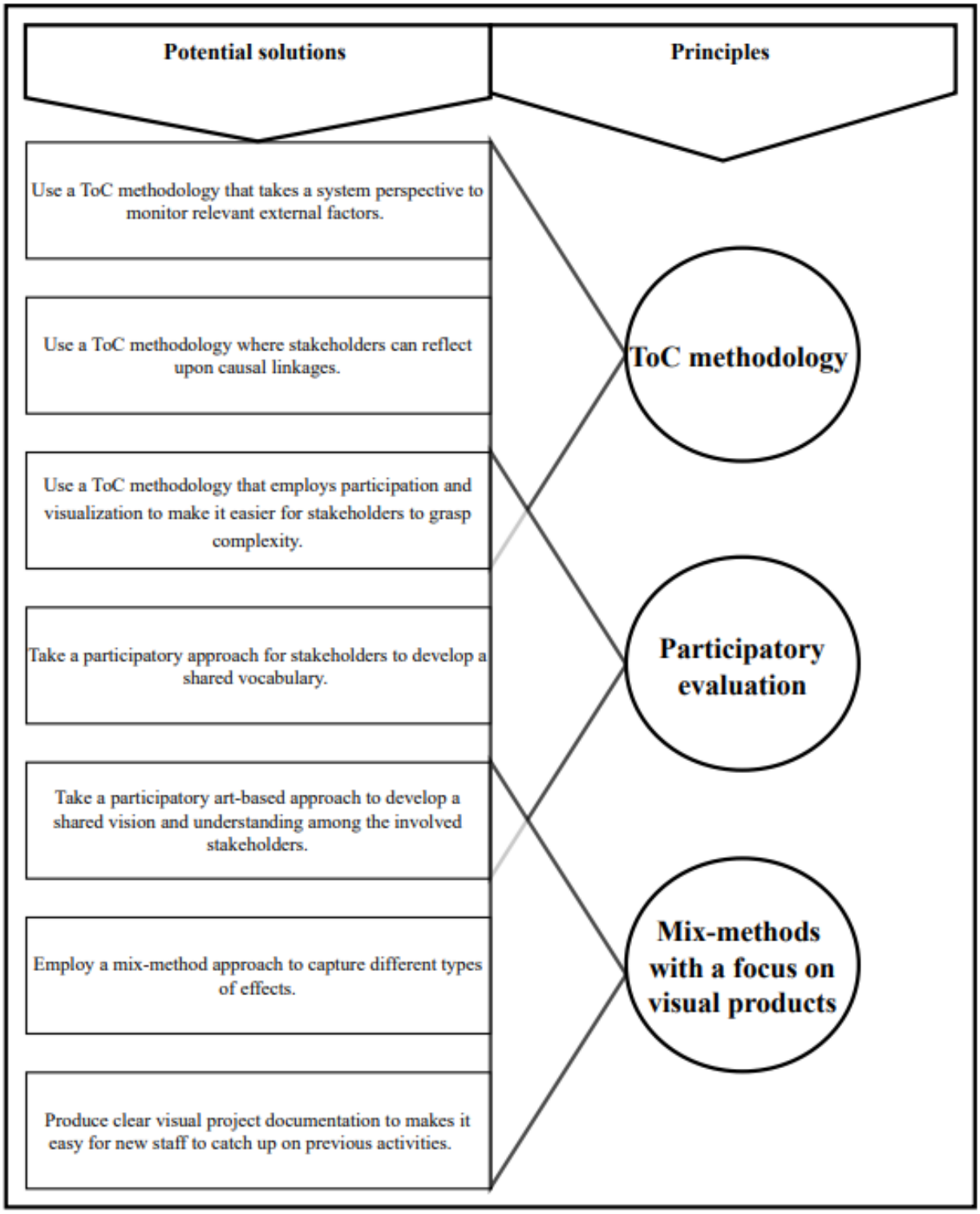

Figure 4: Hypothesis development process

5.2 Hypothesis

Based on the reviewed documents, a hypothesis is formulated that introduces three evaluation principles essential for successful evaluation of co-produced climate services. The principles are tested in the next section, to look for some early signs of practical applicability. The principles encompass: 1) ToC methodology, 2) participatory evaluation, and 3) mix-methods – with a focus on visual product. Each principle is presented below, and a brief overview is provided in Table 6.

Table 6: Hypothesis

ToC methodology	- Captures complexity. - Easy to update as outcomes unfold over time. - Takes a system perspective to capture external factors. - Allows stakeholders to reflect upon causal linkages
Participatory evaluation	- Develops a shared problem understanding. - Ensures that all perspectives are equally considered and represented. - Allows stakeholders to draw attention to unexpected outcomes. - Develops a shared vocabulary to overcome any professional jargon.
Mix-methods – with a focus on visual products	- Creates a robust data set for analysis. - Provides a more comprehensive understanding of the problem at hand. - Captures both tangible and intangible aspects. - Makes it easier to grasp complexity. - Provides good project documentation.

ToC is a logic model for project management, stakeholder engagement, and evaluation that seeks to identify change pathways in complex endeavors (van Tulder and Keen 2018). A brief comparison of the ToC methodology and other logic models is provided in Appendix E. The ToC methodology helps stakeholders to clarify objectives, causal linkages, and underlying assumptions (Fazey et al. 2014), thus encouraging stakeholders to form a shared vision of the project itself and problem at hand (Rajala et al. 2021). The ToC methodology is here used for evaluation purposes, with the aim to identify outputs, outcomes, and impact. As explained by Norström et al. (2020, np.): "tracking progress against indicators that align with the project theory of change enables monitoring against goals at different temporal scales within the project timeframe". The ToC is considered as a process that is comprised of a "living product", which engages stakeholders in continuous reflection, analysis, and discussions. Appendix F shows an example of how a ToC may look like. The ToC product enjoys regular updates to capture development that unfolds over long periods of time. The ToC can help stakeholders to continue their evaluation efforts after the project ends (van Es et al. 2015), and capture those societal impacts that evolve over a long period of time. The ToC methodology embraces complexity, non-linearity, and flexibility which makes it appropriate when, as in the case of climate services, "there are high degrees of subjectivity [*and uncertainty*] in relation to the problem focus and potential solutions" (Fazey et al. 2014, p.217). To improve its comprehensiveness and soundness, the ToC methodology encourages stakeholders consider contextual factors that

might influence the course of the project. This strengthens any causal claims (van Es et al. 2015).

Participatory evaluation approaches actively engage stakeholders in the evaluation process (Guijt 2014). Such approaches come in many different forms, including, but not limited to, role plays, matrix rankings, games, and mapping exercises (Reed et al. 2021). Participatory evaluations can yield many benefits, including:

- Helping the involved stakeholders to form a shared vision (Fazey et al. 2014; Plottu and Plottu 2011), and thus bridge any conflicting interests, contrasting values and diverging epistemological beliefs (Knickel et al. 2019).
- Enhancing motivation, buy-in, and willingness among the involved stakeholders to participate in the evaluation process (Fazey et al. 2014).
- Ensuring that the involved stakeholders feel ownership over the evaluation process (Fazey et al. 2014), as it increases the likelihood of the evaluation findings being implemented in practice (Patton and Horton 2009).
- Drawing attention to unexpected outcomes and impact (Norström et al. 2020).
- Creating a shared vocabulary among the involved stakeholders (van Es et al. 2015).
- Validating the evaluation findings among the involved stakeholders (Guijt 2014).

Arguably, many of the benefits correspond to the positive outcomes which emerge when co-producing knowledge. This, as participatory evaluation shares the same theoretical underpinnings as knowledge co-production. It is therefore considered suitable to extend stakeholder participation from the overall project activities to the evaluation process.

Mix-method approaches combine qualitative and quantitative methods, in order to take advantage of their respective strengths while counterbalancing any potential weaknesses (Ernst 2019). On one hand, qualitative methods are well fit to explore the many intangible factors associated with co-production, like learning, empowerment, and trust (Fazey et al. 2014). Qualitative methods expect the unexpected, and allow the involved stakeholders to draw attention to any unintended effects (Bryman 2012). On the other hand, quantitative methods can assess how change unfolds over time by employing longitudinal data (Fazey et al. 2014). Quantitative methods can also improve the generalizability of the evaluation findings, and thus identify transferable lessons (Reed et al. 2021). Each method fulfills its own purpose, and together they form a holistic understanding of the process itself and its outputs, outcomes, and impact (Ernst 2019). Other advantages include: increased robustness, enhanced comprehensiveness, improved credibility and validity of findings, and generation of unexpected insights (Reed et al. 2021). It is furthermore suggested to incorporate visual products as much as practically possible. This, as visualization creates an informal and inclusive setting where stakeholders can share their insights, expectations, and assumptions (van Es et al. 2015). Visualization often make the evaluation process more accessible and relevant for non-academic

CHAPTER 3: HYPOTHESIS TESTING

6 Methodology

The second part of the research consisted of an online questionnaire that targeted actors with previous experience in co-producing climate services. The survey sought to test the formulated hypothesis by gathering insights from different types of actors. By including different stakeholders, it was assumed that the principles were more likely to be usable for people with different backgrounds.

To some, the research topic can at first glance appear overwhelming, technical, and theoretical. Some respondents may therefore struggle to provide in-depth responses on the spot as required in an interview. An online questionnaire was therefore deemed appropriate, as it allowed the respondents more time to think about their responses.

The online questionnaire was selected given its many advantages like: increased response rates, broad geographical reach, and faster responses (Bryman 2012). Online questionnaires also facilitate data processing and analysis, as the data is automatically transferred to the data analysis software (Evans and Mathur 2005).

Respondents were contacted at multiple occasions via e-mail to increase the response rate. The questionnaire was open for around three weeks.

6.1 Sampling

Several sampling methods were combined. Purposive-, convenience-, and snowball sampling were employed (Bryman 2012). The aim was for the sample to reflect the diversity of stakeholders involved in co-producing climate services. Stakeholders can have different expectations on the knowledge co-production process and its outcomes, thus approach the evaluation differently (Fazey et al. 2014).

The thesis was written in partnership with the Stockholm Environment Institute (SEI), which served as an entry-point for finding respondents. People with some experience in co-producing climate services were approached. Five groups were targeted: 1) The project HazardSupport, 2) the project Unchain, 3) staff at SEI, 4) personal network, and 5) referrals.

Stakeholders from the project HazardSupport were approached. This, as HazardSupport aligned with the type of projects which the principles seek to address. HazardSupport was a five-year project, completed in December 2020. The project engaged SEI and the Swedish Meteorological and Hydrological Institute (SMHI) in a collaborative process with municipal stakeholders from three case studies: flooding in Karlstad, heatwaves in Stockholm, and coastal flooding on Swedish west coast. The aim was to "develop a new, collaborative method for tailoring information about the impacts of climate change on natural hazards to inform adaptation decisions" (André et al. 2020, p.2). In total, 17 researchers and practitioners from HazardSupport were asked to answer the questionnaire, of which six replied.

Stakeholders from the project Unchain were also approached. Much like HazardSupport, Unchain corresponded to the type of projects that the principles are designed to address. Initiated in 2019, Unchain entailed a three-year project that aimed to further develop and improve the impact chain framework, first introduced by Schneiderbauer et al. (2013), through 11 case studies in Europe (Vestlandsforskning n.d.). The international reference group,

researchers, and intermediaries involved in Unchain were asked to complete the questionnaire. The questionnaire was shared to 46 people, of which 11 responded.

Researchers at SEI were approached, as many have expertise and experience in co-producing climate services. As a research institute SEI seeks to bridge science, policy, and practice through stakeholder engagement. Many researchers at SEI co-produce climate services. SEI have offices in Asia, Europe, Latin America, and USA, and thus offered an extensive geographical coverage (SEI 2021). In total 327 people were asked to answer the survey at SEI, of which 29 replied. Although many at SEI work with co-producing climate services, most are involved in other research fields. This partly explains the rather low response rate.

Some convenience sampling was also employed to extend the sample frame beyond SEI, and thus limit potential biases by bringing in other perspectives, experiences, and opinions. The questionnaire was shared by Jenny Iao-Jörgensen, who supervised the thesis, in her personal network on LinkedIn. Anyone with experience in knowledge co-production was eligible to respond. In total, 11 responded.

Some snowball sampling was employed to further broaden the scope. The questionnaire asked respondents if they knew any other people relevant to the research and to share their contact information. The questionnaire was shared to an additional 10 people, of which 3 responded.

In total, 61 replied to the online questionnaire. For an overview of the sampling, see Appendix G.

6.2 Design

The online questionnaire was created in Microsoft Forms, and reflected key aspects found in the hypothesis. Questions were presented in many forms, including multiple choice, Likert scale, and open-ended text responses (Evans and Mathur 2005). Questions were presented in a pre-defined sequence, where most were marked mandatory to avoid any missing answers. 10 open-ended questions were included to give the respondents the opportunity to freely elaborate their answers based on what they sought appropriate. Respondents were given the option to answer the questionnaire in Swedish or English, depending on what they felt the most comfortable with.

Some aspects were considered when designing the questionnaire. The questionnaire was designed to be short and simple, as an attempt to increase the response rate. The average time to answer the survey was around 10 minutes. The questionnaire began with asking the most basic questions, progressing up to the more technical ones. Questions were designed to minimize ambiguity, as respondents had little opportunity to ask clarifying questions. Questions were therefore clearly stated, simple to understand, and presented in an unbiased way (Bryman 2012). The term climate services was avoided, as it may appear overly technical to some respondents given its scientific connotations.

The questionnaire was designed as follows. First, a brief introduction was provided. Respondents were then asked to give their informed consent. This was followed by some introductory questions to better understand the respondent's background, role, and experience in co-producing climate services. The next sections asked questions that reflected the main

points in the hypothesis, and thus covered the following three topics: 1) stakeholder participation, 2) methods, and 3) the ToC methodology. Lastly, some closing questions were asked to conclude the questionnaire. The complete questionnaire is presented in Appendix J and K, in English and Swedish.

6.3 Analysis

Once collected, the data was synthesized using Excel. First, answers in Swedish were translated to English. Thereafter the quantitative and qualitative data was analyzed separately, as explained below.

The quantitative data was summarized and visualized in different types of graphs. All numbers were rounded to the nearest integer. Statistical analysis was avoided. This, as most of the data was structured on an ordinal scale which eliminated most statistical methods (Bryman 2012). Moreover, the sample was deemed inadequate for statistical analysis as it was not representative of the whole population involved in co-producing climate services. Statistical claims would therefore have been weak, and suffered from a large margin of error (Overton and van Diermen 2014).

The open-ended question provided qualitative data that was analyzed using an inductive approach, in which themes were registered as they emerged. The qualitative data was treated as one cohesive dataset, meaning that significant patterns were identified across the entire dataset rather than in the single questions alone (Braun et al. 2020). None of the quotes used when presenting the results were modified, expect if translated.

6.4 Limitations

The online questionnaire presented some limitations and challenges, such as little oversight, missing answers, respondents unable to ask clarifying questions, and no way for the researcher to probe (Braun et al. 2020). To address such limitations, the questionnaire was piloted and tested before distributed to the respondents. Changes were made accordingly. It was also decided to distribute the questionnaire online to improve control over format and design.

The sample presented some additional limitations. The principles largely built on the views, experiences, and knowledge of the targeted respondents. This implies that the research findings reflect potential biases and other shortcomings among respondents. The sample most likely suffered from the response bias, considering the rather low response rate. As explained by Bryman (2012, p.169): “the problem with non-response is that those who agree to participate may differ in various ways from those who do not agree to participate”. The limitation was addressed by triangulating the survey data with the available academic literature. Convergence triangulation was applied (Nightingale 2009), in order to find consistency across the two separate datasets.

Generalizability across cases and contexts remained questionable due to the rather small sample size. The goal was however not to generate a sample suited for generalizability, but to gain a better understanding of how potential users perceived the principles and their practical use. The sample size was therefore considered sufficient.

7 Results: Online questionnaire

This section presents the results derived from the online questionnaire, and is structured as follows: First, the sample is presented followed by an overview of the findings that relate to the principles. Lastly, some practical barriers to applying the principles in practices are introduced.

7.1 Sample characteristics

When asked what countries they primarily worked with, 41% of the respondents said they were working with European countries. 25% worked with a mix of regions, half of which worked with European countries in combination with another region. 8% of the respondent worked on a global level. 11% worked with African countries, and 9% worked with Asian countries. The remaining 5% worked with the Americas.

As shown in Figure 5, 64% of the respondents represented a research institute or university. 20% represented a non-governmental organization or the Red Cross Red Crescent Movement. 10% represented a governmental agency, of which most worked for SMHI. It is worth noting that SMHI mainly works with research, despite being a governmental agency. Only 2% represented a municipality. 3% worked for a private enterprise. The category 'other' included 2% of the respondents, who worked as a private consultant.

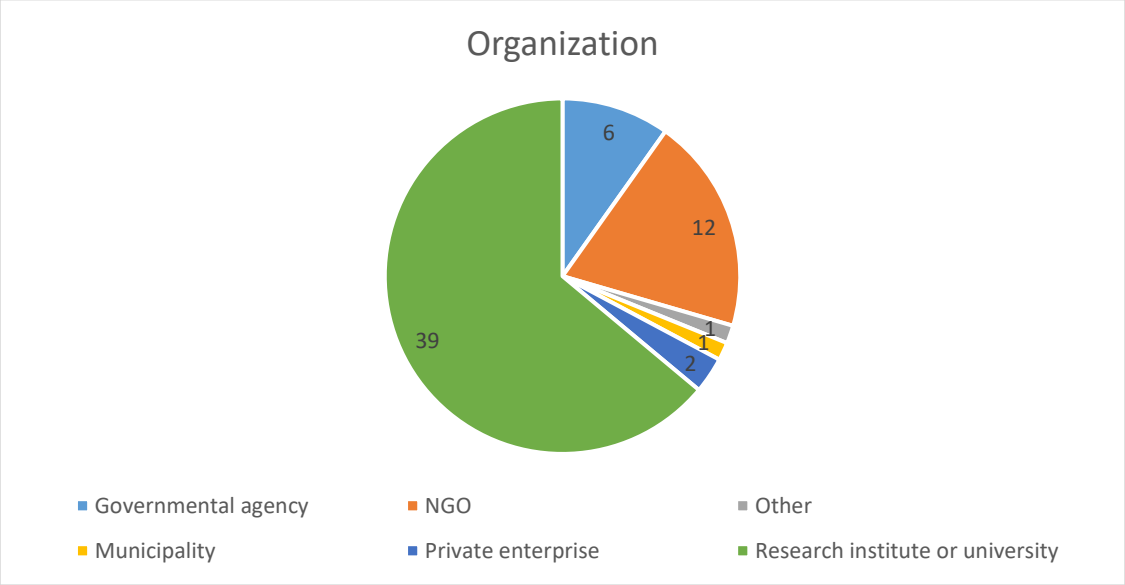

Figure 5: Organization (numbers representing respondents)

The results showed that the respondents adopted multiple roles when co-producing knowledge, as demonstrated in Figure 6. 51 respondents out of 61 thought of themselves as producers of information, and 52 considered themselves as users of climate information. 48 respondents saw themselves as intermediaries. Five respondents considered themselves as financiers, whereas 26 respondents considered themselves as evaluators.

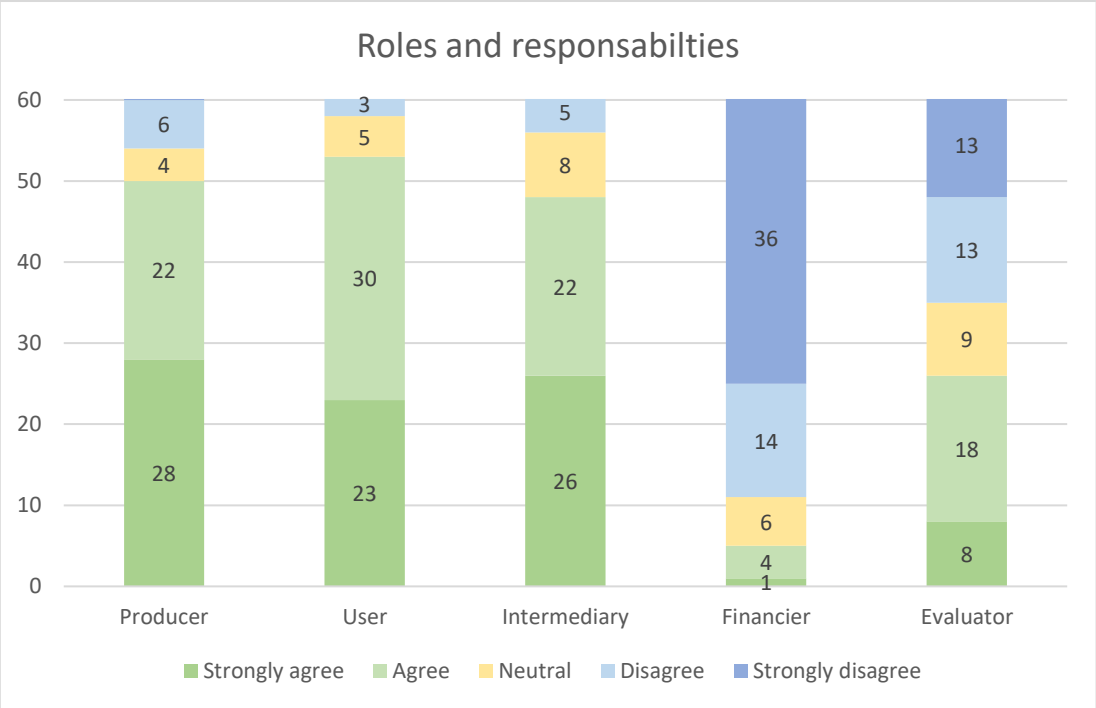

Figure 6: Roles and responsibilities (numbers representing respondents)

As demonstrated in Figure 7, respondents considered a wide range of outputs, outcomes, and impact as important when evaluating co-production initiatives. None of the presented effects was viewed as 'not important'. However, peer-reviewed articles were considered the least important. The following effects were considered as 'very important' by more than half of the respondents:

- Usable information
- Research findings implemented in practice
- Mutual learning
- Enriched decision-making

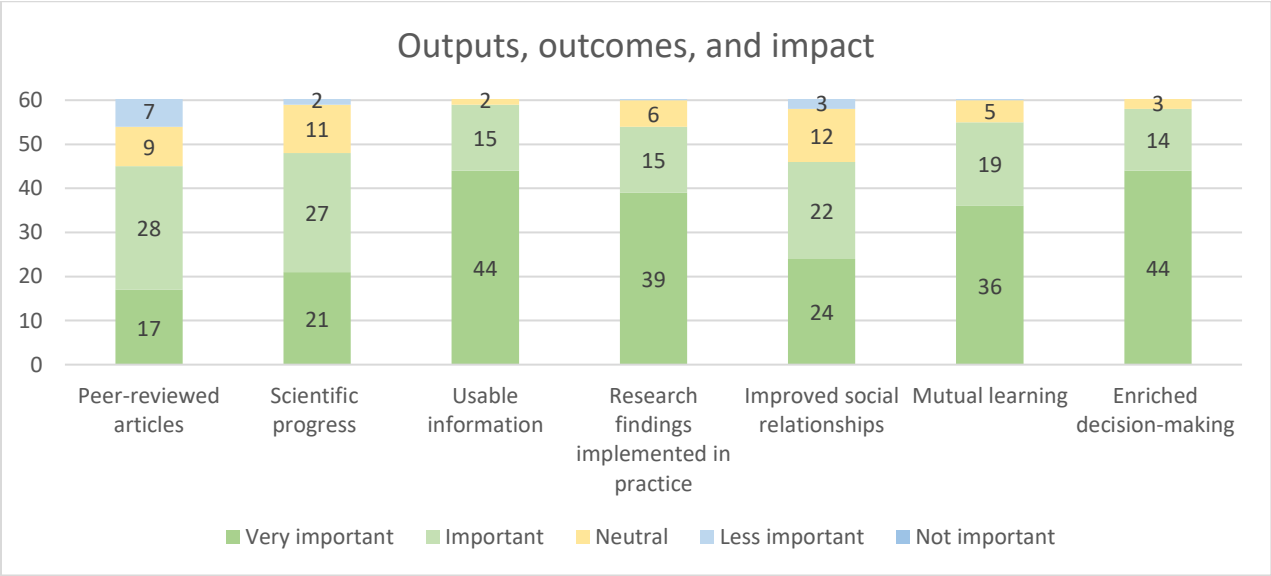

Figure 7: Outputs, outcomes, and impact (numbers representing respondents)

7.2 Principle 1: ToC Methodology

The findings showed that 40 respondents (66%) were familiar with the ToC methodology, of which 31 respondents (51%) recommended it for evaluating co-production initiatives.

Benefits using a ToC methodology when evaluating co-production initiatives are presented in Figure 8. Of those that recommended using a ToC methodology, 90% said that a ToC methodology could help to identify underlying assumptions and define the purpose of the evaluation. 74% thought that a ToC methodology could help to clarify cause-effect relationship, whereas 68% stated that it could help to grasp complexity and context. 65% said that a ToC methodology could help to define what to monitor.

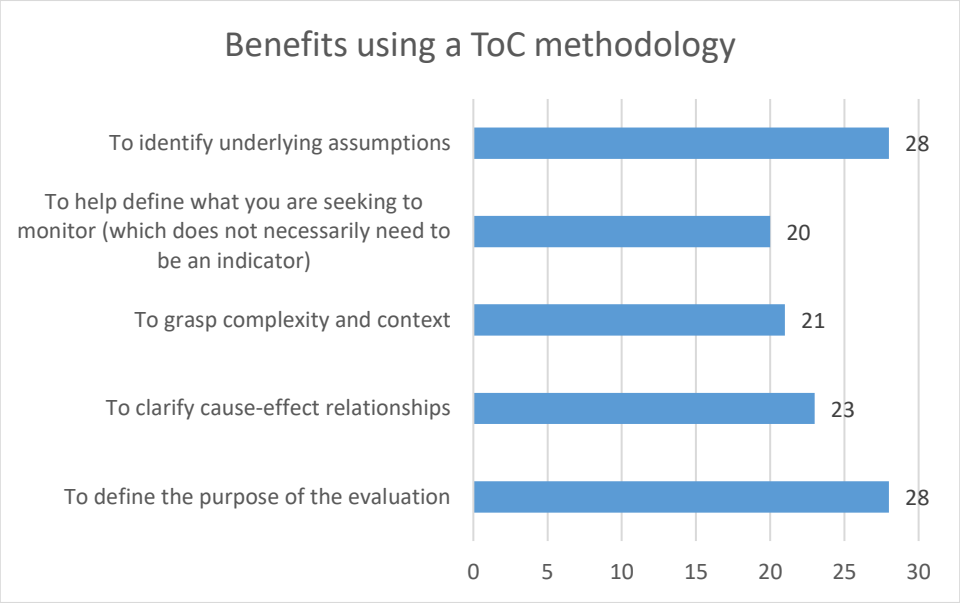

Figure 8: Benefits using a ToC methodology (numbers representing respondents)

Two respondents said they were familiar with the ToC methodology, but never used in practice. Another respondent highlighted that the ToC methodology can be difficult to apply in practice, and highlighted that:

"…it can take quite a lot of time to explain what it is and why it is important to do [but I] have had some great experiences of using it once this has happened (…) useful way to have conversations about what you really think is going to change (…) and digging out the normally tacit assumptions (…) making those visible can be really powerful and a useful conversation if this can be done with all those affected by the outcomes or who can influence the process".

Of those not recommending the ToC methodology when evaluating co-production initiatives, one respondent questioned the timing of the ToC methodology and argued that "…*using the theory of change method should already come at the beginning of the development process, not at the evaluation stage*". Four respondents argued that the ToC methodology was "*too abstract*", "*too academic*", "*bulky*", or even "*pointless*". One respondent worried that the ToC methodology focused too much on donor-reporting, and remarked that "*Theory of Change focusses too much on indicator development and reporting to the funder. It does not focus on the needs of the people*". Another respondent claimed that the ToC "… *is a really simplistic*

model (…). It is really no different than logic models which (…) are not that helpful in directing real co-creation or meaningful interaction".

7.3 Principle 2: Participatory evaluation

97% of the respondents recommended stakeholder participation when evaluating co-production initiatives. Respondents who recommended stakeholder participation when evaluating co-production initiatives recognized many benefits, as shown in Figure 9. 77% believed that stakeholder participation could help the involved stakeholders to develop a shared vision, and 75% thought it could build a sense of ownership among the involved stakeholders. 74% of the respondents considered stakeholder participation as a mean to validate findings. 61% thought that stakeholder participation could increase buy-in from the involved stakeholders.

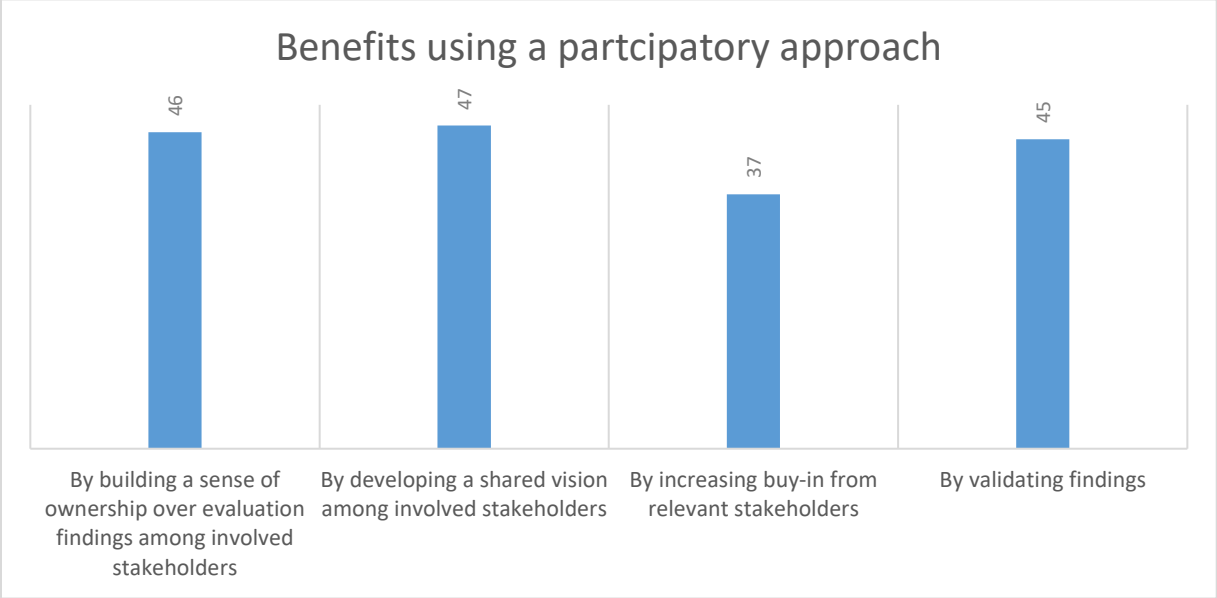

Figure 9: Benefits using a participatory approach (numbers representing respondents)

10 respondents shared some additional insights. Some respondents referred to stakeholder participation as a mean to develop a common understanding among involved actors, by understanding "…*what is valued, how and why*" and "*building trust*…". Other respondents noted that stakeholder participation could validate findings among users of climate information by "*making sure user needs are addressed*" and "*to check if the co-production initiative has really benefitted stakeholders*". Another respondent claimed that stakeholder participation "*rather than validating findings, it's more a process of reflecting initial hypotheses*". Some respondents considered stakeholder participation as way to improve the quality of the evaluation itself by "*diversifying the knowledge base/providing a practical perspective*" and "*gaining insights from users that may otherwise not be known*". Two respondents noted that stakeholder participation could improve project performance, "…*through sharing experiences from relevant stakeholders*" and "*creating a better understanding of how knowledge can be used across organizational (…) boundaries*". One respondent however acknowledged potential negative trade-offs as "*there is a tension between using stakeholder participation to validate findings or to build ownership and increase buy-in*".

The results showed that stakeholders can be involved at various stages of the evaluation process, as demonstrated in Figure 10. 74% of the respondents considered it appropriate to involve stakeholders when developing indicators and reporting the findings, and 70% suggested including the stakeholders when defining the purpose of the evaluation. 62% of the respondents suggested to include stakeholders when collecting data, 60% when designing the evaluation approach, and 44% when analyzing the data.

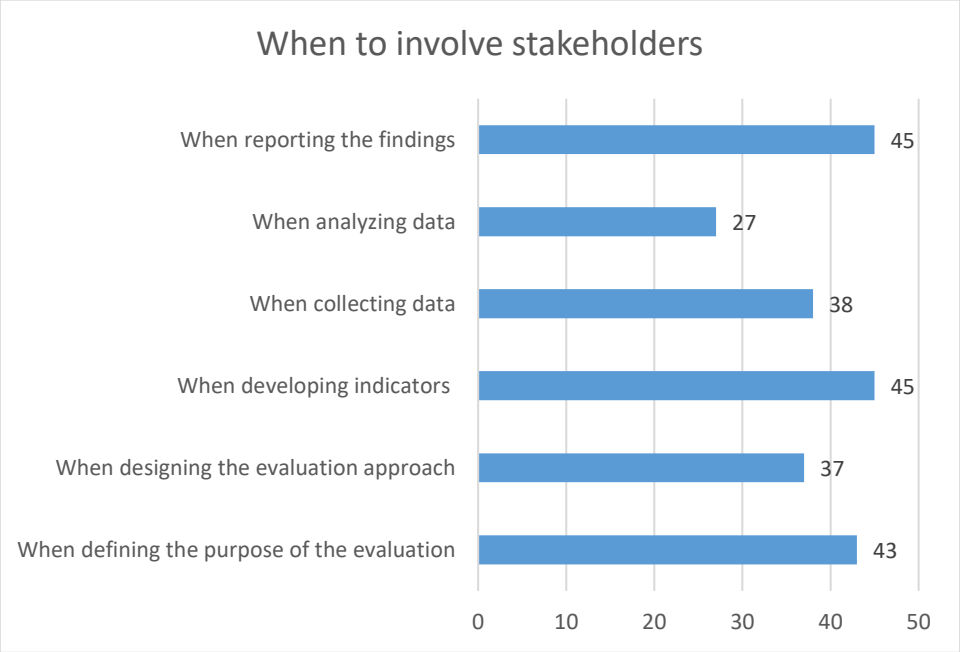

Figure 10: When to involve stakeholders (numbers representing respondents)

Two respondents acknowledged the importance of involving stakeholders in the evaluation process, as they are closely linked to the project outputs and outcomes. One respondent stated that stakeholders "*…tend to affect the output and outcome of the findings*", whereas another respondent noted that stakeholder "*contribution and criticisms are very important in taking a decision to correct a gap*". Some respondents however noted that for stakeholder participation to be fruitful, it must be designed to fit the project at hand ensuring that it is "*practically possible for the stakeholders to participate, considering their situation and capacities*". As one respondent argued: "*while I think it appropriate to involve stakeholders at any stage, including them at all stages would be too demanding on the stakeholders and likely lower participation. Where in the evaluation process it is most crucial to include stakeholders is going to be project specific*". Another respondent remarked that "*the stakeholder input has to be proportionate to the size of the initiative- it is not appropriate to waste people's time*".

Some respondents noted the importance of understanding the stakeholder group to find entry-points for engagement. One respondent claimed that "*understanding the divergence and similarities of stakeholders and how best to engage them in evaluation is critical and challenging*". Another respondent added that "*one needs to understand alignment, interest and influence in undertaking stakeholder participation. Also an appreciation of power is important*". Given this complexity, one respondent recommended to "*…involve social*

scientists who are experts in participatory methods and learning theories". Another respondent suggested that "*determining who the stakeholders are and categorizing them using certain criteria will help determine where and how to involve them in the evaluation*".

Some respondents provided some practical advice for how to optimize stakeholder participation. One respondent argued for the need "…*to have a clear purpose with the evaluation in order to achieve progress*…". Two respondents noted the importance of incorporating incentives that encourage stakeholder participation, and ensure that "*stakeholder themselves see the benefit from being included in the evaluation*". Lastly, one respondent acknowledged the role of communication, and remarked that "*language and mode of communication to stakeholders are also important*".

Other respondents argued for a utilization-focused approach, in which the evaluation is designed to enhance the usability of the findings in practice. Such approach often involves intended users in a participatory evaluation process. This, as one respondent noted, "…*as they are typically the ones who would 'practice'*". Another respondent emphasized:
"For me that is the whole point of an evaluation as I do utilization-focused evaluation generally (not evaluation just for accountability). The design of the evaluation is thus focused on learning and usability of the findings - so designed with discussion with the purpose in mind and with people who are intended to use the findings to check they are accessible, fit for purpose, timely...".

Two respondents did not recommend stakeholder participation when evaluating co-production initiatives. One respondent thought that stakeholder participation created "*biased evaluation, due to personal involvement*". The other respondent noted that stakeholder participation "… *can be useful but many of our funding agencies overdo it so we lose a lot of time being forced into stakeholder discussions that are not always useful…".*

7.4 Principle 3: Mix-methods – with a focus on visual products

Respondents considered many different methods useful when evaluating co-production initiatives, as shown in Figure 11. Interviews with the involved stakeholders were viewed as the most useful method, closely followed by mix-methods and group discussions. 52 respondents thought that questionnaires were useful, whereas 48 respondents viewed written reflections useful. Indicators and review of project documents were considered the least useful when evaluating co-production initiatives.

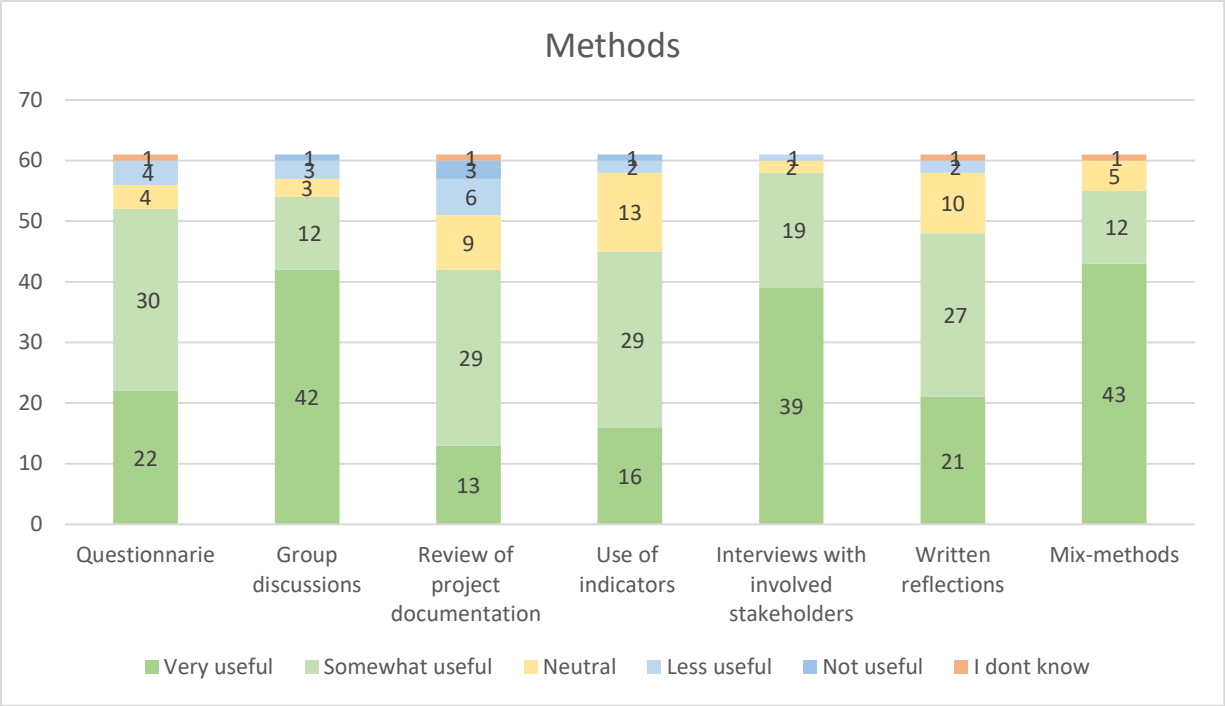

Figure 11: Methods (numbers representing respondents)

Respondents were given the opportunity to list other methods which they considered useful when evaluating co-production initiatives, as presented in Figure 12. The most suggested methods were participant observation, workshops, visual products, and living labs.

Figure 12: Other methods

98% of the respondents agreed that visual products could help to clarify complex issues. One respondent added that "*a really brief qualitative summary supplemented by visual representations of quantitative assessments*" could improve the likelihood of evaluation findings being implemented in practice.

7.5 Practical barriers

The findings also revealed structural problems associated with the short-term funding cycle that characterizes many co-production initiatives. Evaluations are typically performed at the end of a project, meaning that there is neither time nor funds left for implementing the findings. As one respondent remarked:

"Usually, programs end when evaluations are done. Unless additional funding is secured, findings from participatory evaluations cannot be integrated due to a lack of resources. Longevity of climate service projects and research is a major issue."

Another respondent problematized the "*lack of time and financing for the follow-up…*", but also noted that the *"…co-production throughout the process might enable practitioners still to have ownership and thus enable concrete usability of project findings and results*".

8 Discussion

Returning to the research question, this section seeks to discuss what methodological principles that can be used to evaluate the process and effects of co-produced climate services. The discussion builds on key findings from the study presented earlier, and is structured as follows: First, the formulated hypothesis is scrutinized, followed by a review of potential challenges and suggestions on application. Thereafter implications of the key findings are discussed. Lastly, some recommendations on future research are provided.

8.1 The principles

The findings from the questionnaire endorsed the three principles outlined in the hypothesis:

1. **ToC methodology**
2. **Participatory evaluation**
3. **Mix-methods – with a focus on visual products**

The findings thus showed some early signs of practical applicability. The principles are designed to offset previously identified challenges and shortcomings encountered when evaluating co-produced climate services, and are expected to yield many benefits. Key benefits include helping stakeholders to map complex change pathways; capturing external factors; measuring the intangible; bridging conflicting interests; identifying unexpected effects; and many other.

Building on previous research described in section 4, the three principles are a continued effort to refine evaluation practices in the context of co-produced climate services. In line with Sarkki et al. (2015) and others, the principles shed light on the need to re-think research impact evaluation. The principles support a shift from traditional evaluation practices that put heavy emphasis on academic outputs, to an evaluation approach that can capture the many, although at times intangible, effects that emerge when co-producing knowledge. This shift is supported by the findings from the online questionnaire, which showed great diversity in regards to what type of effects to consider when evaluating co-production initiatives. This, despite that only effects linked to the co-production process itself were introduced. In practice, contextual and project-specific effects are likely to further add to this diversity.

As an attempt to develop a holistic approach, the principles seeks to understand the co-production process itself and the effects which may emerge from such a process. This, to build

a good understanding of the intervention, whilst providing information about emerging effects. The iterative nature of climate services creates a significant interdependence between the process and its effects. Inspired by complexity theory as presented by Patton (2010), it can be argued that their relationship is situated in a complex feedback system in which the process is responsive to the effects that emerge and vice versa. This complexity justifies the use of developmental evaluation when evaluating co-produced climate services. Consequently, the principles draw from basic developmental evaluation design and seek to support adaptive management as the fog lifts from the complexity inherent in climate services.

As outlined in the research aim, the three principles are designed to be flexible. There is however a risk that the flexibility makes the principles too vague to actually contribute anything of value. There is thus a trade-off between being general and flexible on one hand, and on the other hand being overly specific and rigid. Flexibility was in this case prioritized, considering the many different contexts, actors, and resources involved in co-producing climate services. The flexibility allows the principles to be adapted to better fit the given context, arguably increasing their usability and effectiveness in practice.

As just highlighted, the principles are grounded in the developmental evaluation line of thought. Remarkably, only Van der Molen et al. (2018) seem to have adopted developmental evaluation for evaluating co-produced climate services. This, in spite of developmental evaluation being specifically developed for complex and emergent situations (Patton 2010) like climate services. Nevertheless, Dozois et al. (2011) state that developmental evaluation is a fairly new approach with few guidelines. Presumably, this indicates that developmental evaluation remains open to further empirical research and scrutinization. The principles recognize this untapped potential, and present a great opportunity to apply developmental evaluation in the context of co-produced climate services. In theory, as outlined in section 5, the developmental evaluation can offset many challenges associated with evaluating co-production initiatives. The principles acknowledge this opportunity, and therefore present methods that align with developmental evaluation practices. Corresponding to basic developmental evaluation design, the principles seek to support adaptive management, flexible practices, and real-time feedback. Arguably, this shows an early indication of applicability although future empirical research is recommended to investigate it further.

Given their close ties with developmental evaluation, the three principles are a powerful vehicle for those involved in co-producing climate services to engage in adaptive management and participatory engagement. Adaptive management was at large ignored in the reviewed literature, with Salamanca and Biskupska (2021) as a notable exception who established adaptive management as a necessity when co-producing climate services. Adaptive management can help stakeholders to handle the inherent uncertainty associated with climate change adaptation, as it allows stakeholders to adjust their actions as new information emerges (Salamanca and Biskupska 2021). In its essence, adaptive management is iterative and participatory. There is a continuous real-time feedback loop connecting evaluation findings and decisions, which makes adaptive management highly dependent on robust and credible data input (Patton 2010). The principles can improve the quality of evaluation findings, thus help stakeholders to collect better evidence as input to adaptive management practices. This, as the principles better represent complex change pathways, capture the intangibles, monitors external factors, and draw attention to the unexpected.

8.2 Challenges in applying the principles

The findings derived from the online questionnaires shed light upon some challenges in evaluating co-produced climate services that were overlooked in the reviewed literature. The challenges are further discussed below.

As outlined in the research aim, anyone involved in co-producing climate services should be able to apply the methodological principles in different contexts in practice. This, regardless of their academic background. Some of the findings derived from the online questionnaire indicated that the ToC methodology was not fulfilling this criterion. Many were not familiar with the ToC methodology, whilst others criticized it for being hard to use. Returning to the literature, this is a common challenge. For example, Rajala et al. (2021) point out that the ToC methodology can initially be hard for the involved stakeholders to grasp. There is thus a need for "de-mystifying the ToC methodology". In line with van Es et al. (2015), it is recommended to use visualizing tools and methods. This, as it easier to understand complicated matters when they are supported by visual products. Arguably, this indicates that the principles are closely interlinked and best used when combined.

The findings derived from the online questionnaire suggested that there is a significant risk of uncritical application of the ToC methodology, making it no different from other logic models. However, as noted by van Tudler and Keen (2016) among others, the ToC methodology is in theory supposed to better represent complexity. It seems like the ToC methodology fails to live up to this promise, indicating a gap between theory and practice. A probable reason for this, supported by findings from the online questionnaire, is the limited time and budget for reflection. According to van Es et al (2015), critical reflection is one of the key features that distinguishes the ToC methodology from other logic models. Nevertheless, in practice stakeholders face budget and time constraints which inhibits such critical reflection. Arguably, in line with developmental evaluation practices outlined by Patton (2010), there is a need to embed the reflection process into the project cycle to stimulate reflection and learning, ultimately maximizing the many benefits associated with the ToC methodology.

The likening of the ToC methodology with other logic models furthermore indicates that they are in practice used as if they were the same. Perceived theoretical benefits with using a ToC methodology simply does not matter if it continues being used as other logic models. In practice, this means that the ToC methodology adopts the same shortcomings as other logic models, including being donor-driven, reductionist, and simplistic. This, despite that the reviewed literature, for example van Es et al. (2015), make a clear distinction between the ToC methodology and other logic models. The ToC methodology must therefore be applied with caution to ensure that it used critically to fit its purpose. If misused, the ToC methodology is likely to cause a loss of richness in the evaluation findings. This, by trying to fit a messy reality into a linear-cause effect model. Such an approach clashes with the complexity theory embodied in the developmental evaluation practices (Patton 2010), which underpins the principles.

Looking at the findings from the online questionnaire, another challenge emerges concerning the time allocated to the evaluation. Participatory evaluation is time-consuming, in which extensive participation is likely to cause fatigue and lower engagement. It is vital to adjust the level of stakeholder participation to the project at hand. Yet the findings from the online questionnaire indicated that funders at times require time-consuming stakeholder participation at every stage of the evaluation, whilst many researchers and practitioners claim that they lack

time and budget for such endeavors. There is a disconnect between funders' expectations and local realities, confirming the assumption outlined in the hypothesis regarding flexible and reflexive practices. The evaluation framework should avoid being rigid, and instead adapt as new insights emerges. In line with Roux et al. (2010), such process should also include funders. Arguably, this allows funders and practitioners to learn from each other's respective contexts, accountabilities, and motivations. Funders and practitioners can form a mutual understanding of the co-production initiative and its context, thus avoid any potential disconnect.

To conclude, the reviewed literature seems to build on the assumption that there is sufficient time and budget allocated for the evaluation. Yet, this is not always the case as demonstrated by the findings derived from the online questionnaire. There is thus a significant gap between theory and practice, which future research should attend to. This, in order to generate evaluation practices that account for local and multiple realities as well as organizational practices of the involved stakeholders.

8.3 Application of the principles

The flexibility that is prevalent throughout the three principles enables smooth application in a variety of climate service initiatives, regardless of their scope, scale, or context. Some suggestions on application are presented below.

The three principles can be applied in isolation, but it is advised that they are combined as they are designed to complement each other. Each principle addresses some, but not all, identified challenges in evaluating the co-production of climate services. Combined, the principles take a holistic approach when addressing the identified challenges, and thus improve evaluation practices and adaptive management in the context of co-produced climate services.

Stakeholder participation unlocks the full potential of the ToC methodology, as it incorporates multiple sources of knowledge and realities, allow the involved stakeholders to form a shared vision, and enhance ownership over evaluation processes and use of evaluation findings (van Es et al. 2015). Ownership is crucial as it increases the likelihood of evaluation findings being implemented in practice (Patton 2010). Stakeholder participation is applicable at all stages of developing a ToC. A ToC is best presented as visual product accompanied by a narrative, which helps to visualize complex change pathways (van Es et al. 2015). Both the visual product and narrative should be updated as the co-production initiative progresses. Visualization methods are often participatory by nature (Reed et al. 2021), as they seek to create an inclusive setting that promotes knowledge exchange (van Es et al. 2015). Visualization however works the best when combined with other methods, hence the mix-method approach. Visualization helps to disentangle complex relationships, which then can be analyzed using statistical methods (Reed et al. 2021). Mix-methods can also help to capture different types of effects expressed in the ToC, including the intangible (Fazey et al. 2014).

At the moment there is no prescribed approach for how the principles are best combined. Stakeholders are encouraged to adapt the principles for their context. Here is however an example of application, in which the ToC methodology serves as the basis. It engages project-specific stakeholders in a participatory process guided by an embedded evaluator. The methodological principles are sequenced to overlap. It consists of an iterative five-step cycle:

1. Develop a **ToC product** using a **participatory approach**, in which the involved stakeholders clarify objectives, change pathways, critical external factors, and underlying assumptions. The involved stakeholders use **visualization** to support their discussions. On a blank piece of paper, the project objective is noted at the top. Thereafter, backward mapping is applied. The mapping starts at the project objective, and then works its way backwards asking what change is needed for the project to fulfill its objectives. This is visualized as pathways of change, outlining the flow between the envisioned impacts, outcomes, outputs, and activities. All these are connected by links. The involved stakeholders are encouraged to note down their assumptions and external factors which may influence the course of the project. The ToC is presented as a visual product accompanied by a supporting written narrative. See page 26 for a visual representation of a ToC product.
2. Evaluation priorities are defined, based on the change pathways hypothesized in the **ToC product**. Evaluation priorities are areas of enquiry that are important to monitor, including critical assumptions, intended outcomes, and external influences. Evaluation priorities are established through a **participatory process,** in order for them to reflect the wide range of perspectives involved when co-producing climate services.
3. Plan the learning process. Using a **participatory approach**, an embedded evaluator together with the involved stakeholders decide when and how the **ToC product** is revisited, adapted, and updated. Stakeholders are encouraged to engage with the ToC product throughout implementation. The ToC product can continue to be updated after the completion of the project. This, to capture those impacts that evolve over long periods of time.
4. Collect the data. An embedded evaluator collects data for the outlined evaluation priorities. A **mix-method approach** is applied. Findings are continuously shared with the involved stakeholders for validation.
5. Use and adaptation of the **ToC product**. It is recommended that the ToC product is revisited on a regular basis. Through a **participatory process**, the ToC is be updated as the data collection and the co-production initiative itself progresses. The ToC provides a sound basis for identifying lessons transferable between contexts.

8.4 Research implications

Despite recent advancements in climate services, little attention is paid towards how to evaluate such practices. As highlighted in the reviewed literature, many evaluators have no alternative but to continue employing output-focused evaluation measures when trying to evaluate co-produced climate services (Belcher et al. 2016; Van der Molen et al. 2019). Such measures simply fail to capture the broad array of effects that emerge when co-producing climate services. Although many methods exist for evaluating research impact (Reed et al. 2021), few are designed to assess co-produced climate services. This study filled this research gap by generating three methodological principles essential for successfully evaluating co-produced climate services.

It is, however, crucial to understand the limitations imposed by the chosen research design. The principles were never applied in practice due to time and resource constraints. As a result, the practical feasibility remained unverified. Despite this limitation, the study takes a significant step towards better understanding what methods can be used when evaluating co-produced climate services. The study shows the theoretical potential of a selection of methods when

evaluating co-produced climate services, and thus forms a strong basis for future research and testing.

The three principles are first and foremost designed to support adaptive management when co-producing climate services, but can also help to identify lessons transferable between contexts. This, as the principles improve data collection practices which supposedly increases the robustness, saliency, and credibility of key findings. A better evidence-base can help the research community to generate improved best practices for the co-production of climate services. As a relative new research field, climate services can benefit from any recommendations on best practices. Nevertheless, some caution is needed. As noted by Norström et al. (2020), knowledge co-production is a highly context-dependent process. It is therefore recommended that best practices serve as a source of inspiration which can be adjusted and adapted to the project at hand to make sure they correspond to the local reality on the ground.

It is fair to assume that the three principles are transferable to other types of co-production projects that also engage stakeholders from multiple disciplines and professional sectors. This, due to the principles being grounded in evidence collected from the broader sustainability literature. Sustainability science also tackles the complex interaction between natural and social systems, and struggle with the similar challenges as climate services. Such challenges include complexity, long time horizons, uncertainty, stakeholder and organizational diversity to only name a few. The principles are designed to offset such challenges, and can therefore prove useful for other disciplines which engages in knowledge co-production.

8.5 Recommendations on future studies

In order to better understand the principles and their applicability in practice, the following is recommended for future research:

- Future studies should apply the principles in practice to assess their practical feasibility, and thus contribute to further refinement and validation.
- An additional literature review is recommended, to cover those case studies which were excluded due to time and resource constraints. This could reveal some additional insights on more practical challenges that can arise when evaluating co-production initiatives.
- The principles should be applied in an array of contexts to test their application in various constellations. Special attention should be paid towards low-income countries, considering that most respondents worked in a European setting.
- Special attention should be paid towards the roles and responsibilities stakeholders adopt when applying the principles in practice, in order to mitigate emerging power dynamics. The principles should be adapted accordingly.
- Future research should investigate how funders perceive the principles, as few respondents self-identified as funders. Roux et al. (2010) noted that funders set the research agenda, making it vital to understand their views on how to evaluate co-produced climate services.

- In-depth exploration of how practitioners perceive the principles in their socio-ecological environment, with emphasis on the ToC methodology as it has been criticized for being difficult to use. Evaluators and ToC facilitators should be regarded as an especially important perspective to capture, given that they were underrepresented in the study.
- Future research should investigate how knowledge co-production intersects with action research, potentially identifying synergies between the two. This, in order to explore if the principles are transferable to contexts using action research.
- Lastly, future studies are recommended to contribute to emerging developmental evaluation literature by exploring practical challenges and opportunities of adopting such an approach when evaluating co-produced climate services.

CHAPTER 4: CONCLUSION

9 Conclusion

As climate change continues to alter weather patterns, there is a growing need for science to support climate change adaptation decision-making. For this reason, scientists and practitioners have started to co-produce climate services. Evaluation can further improve knowledge co-production practices, ultimately enhancing the effectiveness and efficiency of climate change adaptation actions. There is however little previous research on how such evaluation efforts should look like. The present study therefore set out to investigate what methods can be used to evaluate the process and effects of co-produced climate services.

The following three principles were proposed for a successful evaluation of co-produced climate services:

1. **ToC methodology**
2. **Participatory evaluation**
3. **Mix-methods - with a focus on visual products**

The principles draw on scientific evidence from a literature review and online questionnaires, where early signs for practical applicability surfaced. The study takes a significant step towards better understanding what methods can be used when evaluating co-produced climate services, showing on the theoretical potential of the methods outlined in the principles. The principles serve as a strong basis for future research and testing. As a way forward, it therefore recommended that the principles should be applied and tested in practice.

To conclude, this thesis can serve as a stimulus for further discussions on how to evaluate co-produced climate services. Improved evaluation practices can increase the effectiveness and efficiency of climate services, thus enhancing climate change adaptation decision-making as a whole. Most importantly, this can improve climate change adaptation efforts in a time that demands urgent climate action.

REFERENCES AND APPENDICES

References

André, K., Järnberg, L., Gerger Swartling, Å. (2020). Co-designing climate services to support adaptation to natural hazards: two case studies from Sweden. SEI Discussion Brief. Stockholm Environment Institute.

Belcher, B. M., Rasmussen, K. E., Kemshaw, M. R. and Zornes, D. A. (2016). Defining and assessing research quality in a transdisciplinary context. *Research Evaluation*, 25(1). 1–17. DOI: 10.1093/reseval/rvv025

Blackstock, K. L., Kelly, G. J. and Horsey, B. L. (2007). Developing and applying a framework to evaluate participatory research for sustainability. *Ecological Economics*, 60(4). 726–42. DOI: 10.1016/j.ecolecon.2006.05.014

Braun, V., Clarke, V., Boulton, E., Davey, L. and McEvoy, C. (2020). The online survey as a qualitative research tool. *International Journal of Social Research Methodology*, 1–14. DOI: 10.1080/13645579.2020.1805550

Bremer, S. and Meisch, S. (2017). Co-production in climate change research: reviewing different perspectives. *Wiley Interdisciplinary Reviews: Climate Change* 8(6). DOI: 10.1002/wcc.482

Bremer, S., Wardekker, A., Dessai, S., Sobolowski, S., Slaattelid, R. and van der Sluijs, J. (2019). Toward a multi-faceted conception of co-production of climate services. *Climate Services*, 13. 42–50. DOI: 10.1016/j.cliser.2019.01.003

Bryman, A. (2008). *Social research methods*. Oxford University Press.

Cambridge University Press (2021). *Evaluation*. Cambridge Dictionary. https://dictionary.cambridge.org/dictionary/english/evaluation

Carbon, M. (2017). Chapter 6: An Analytical Framework for Evaluating a Diverse Climate Change Portfolio. In Uitto, J., Puri, J. and van den Berg, R. (eds). *Evaluating Climate Change Action for Sustainable Development.* Springer Verlag. 95-110. DOI 10.1007/978-3-319-43702-6_6

Carew, A. L. and Wickson, F. (2010). The TD Wheel: A heuristic to shape, support and evaluate transdisciplinary research. *Futures*, 42. 1146–1155. DOI: 10.1016/j.futures.2010.04.025

Christel, I., Hemment, D., Bojovic, D., Cucchietti, F., Calvo, L., Stefaner, M. and Buontempo, C. (2018). Introducing design in the development of effective climate services. *Climate Services*, 9. 111–121. DOI: 10.1016/j.cliser.2017.06.002

Daniels, E., Bharwani, S., Butterfield, R. (2019). The Tandem framework: a holistic approach to co-designing climate services. SEI Discussion Brief. Stockholm Environment Institute.

Daniels, E., Bharwani, S., Gerger Swartling, Å., Vulturius, G. and Brandon, K. (2020). Refocusing the climate services lens: Introducing a framework for co-designing "transdisciplinary knowledge integration processes" to build climate resilience. *Climate Services*, 19. 100181. DOI: 10.1016/j.cliser.2020.100181

Dawkins, E., André, K., Axelsson, K., Benoist, L., Gerger Swartling Å. and Persson, Å. (2019). Advancing sustainable consumption at the local government level: A literature review. *Journal of Cleaner Production*, 231. 1450–62. DOI: 10.1016/j.jclepro.2019.05.176

Dozois, E., Langlois, M., and Blanchet-Cohen, N. (2011). *DE 201: A Practitioner's Guide to Developmental Evaluation*. J.W. McConnell Family Foundation and International Institute for Child Rights and Development

Ernst, A. (2019). Research techniques and methodologies to assess social learning in participatory environmental governance. *Learning, Culture and Social Interaction*, 23. 100331. DOI: 10.1016/j.lcsi.2019.100331

Ernst, K. M., Gerger Swartling, Å., André, K., Preston, B. L. and Klein, R. J. T. (2019). Identifying climate service production constraints to adaptation decision-making in Sweden. *Environmental Science & Policy*, 93. 83–91. 10.1016/j.envsci.2018.11.023

European Commision (2015). *A European research and innovation roadmap for climate services.* European Union. http://op.europa.eu/en/publication-detail/-/publication/73d73b26-4a3c-4c55-bd50-54fd22752a39

Evans, J. and Mathur, A. (2005). The Value of Online Surveys. *Internet Research*, 15. 195–219. DOI: 10.1108/10662240510590360

Fazey, I., Bunse, L., Msika, J., Pinke, M., Preedy, K., et al. (2014). Evaluating knowledge exchange in interdisciplinary and multi-stakeholder research. *Global Environmental Change*, 25. 204–20. DOI: 10.1016/j.gloenvcha.2013.12.012

Fulgenzi, A., Brouwer, S., Baker, K. and Frijns, J. (2020). Communities of practice at the center of circular water solutions. *WIREs Water*, 7(4). DOI: 10.1002/wat2.1450

GFCS (September 2009). World Climate Conference-3 (WCC-3). http://www.gfcs-climate.org/wwc_3/

Guijt, I. (2014). Participatory Approaches. Methodological Briefs. UNICEF Office of Research.

Haddaway, N. R., Bethel, A., Dicks, L. V., Koricheva, J., Macura, B., Petrokofsky, G., Pullin, A. S., Savilaakso, S. and Stewart, G. B. (2020). Eight problems with literature reviews and how to fix them. *Nature Ecology and Evolution*, 4(12). 1582–1589. DOI: 10.1038/s41559-020-01295-x

Haddaway, N.R., Woodcock, P., Macura, B. and Collins, A. (2015). Making literature reviews more reliable through application of lessons from systematic reviews. *Conservation Biology*, 29(6). 1596–1605.

Harvey, B., Cochrane, L. and Epp, M. V. (2019). Charting knowledge co-production pathways in climate and development. *Environmental Policy and Governance*, 29(2). 107–17. DOI: https://doi.org/10.1002/eet.1834

Hassel, H. (2010). *Risk and vulnerability analysis in society's proactive emergency management: Developing methods and improving practices*. Lund University

Hassenforder, E., Smajgl, A. and Ward, J. (2015). Towards understanding participatory processes: Framework, application and results. *Journal of Environmental Management*, 157. 84–95. DOI: 10.1016/j.jenvman.2015.04.012

Hegger, D., Lamers, M., Van Zeijl-Rozema, A. and Dieperink, C. (2012). Conceptualising joint knowledge production in regional climate change adaptation projects: success conditions and levers for action. *Environmental Science & Policy*, 18. 52–65. DOI: 10.1016/j.envsci.2012.01.002

Heink, U., Marquard, E., Heubach, K., Jax, K., Kugel, C., et al. (2015). Conceptualizing credibility, relevance and legitimacy for evaluating the effectiveness of science–policy interfaces: Challenges and opportunities. *Science and Public Policy*, 42(5). 676–89. DOI: 10.1093/scipol/scu082

Hitziger, M., Aragrande, M., Berezowski, J., Canali, M., Del Rio Vilas, V., et al. (2019). EVOLvINC: EValuating knOwLedge INtegration Capacity in multistakeholder governance. *Ecology and Society*, 24(2). DOI: 10.5751/ES-10935-240236

Holzer, J. M., Carmon, N. and Orenstein, D. E. (2018). A methodology for evaluating transdisciplinary research on coupled socio-ecological systems. *Ecological Indicators*, 85. 808–819. DOI: 10.1016/j.ecolind.2017.10.074

Jahn, T. and Keil, F. (2015). An actor-specific guideline for quality assurance in transdisciplinary research. *Futures*, 65. 195–208. DOI: 10.1016/j.futures.2014.10.015

Jones, N. A., Perez, P., Measham, T. G., Kelly, G. J., d'Aquino, P., Daniell, K. A., Dray, A. and Ferrand, N. (2009). Evaluating Participatory Modeling: Developing a Framework for Cross-Case Analysis. *Environmental Management*, 44(6). 1180. DOI: 10.1007/s00267-009-9391-8

Knickel, M., Knickel, K., Galli, F., Maye, D. and Wiskerke, J. S. C. (2019). Towards a Reflexive Framework for Fostering Co—Learning and Improvement of Transdisciplinary Collaboration. *Sustainability*, 11(23). 6602. DOI: 10.3390/su11236602

Lang, D. J., Wiek, A., Bergmann, M., Stauffacher, M., Martens, P., Moll, P., Swilling, M. and Thomas, C. J. (2012). Transdisciplinary research in sustainability science: practice, principles, and challenges. *Sustainability Science*, 7(1). 25–43. DOI: 10.1007/s11625-011-0149-x

Laycock, A., Bailie, J., Matthews, V. and Bailie, R. (2019). Using developmental evaluation to support knowledge translation: reflections from a large-scale quality improvement project in Indigenous primary healthcare. *Health Research Policy and Systems*, 17(1). DOI: 10.1186/s12961-019-0474-6

Lemon, M. and Mitchell, A. (2020). Learning How to Learn in Sustainability Transitions Projects: The Potential Contribution of Developmental Evaluation. *Journal of Multidisciplinary Evaluation*. 16(34). 91-103. ISSN 1556-8180

Liedtka, J. (2015). Perspective: Linking Design Thinking with Innovation Outcomes through Cognitive Bias Reduction. *Journal of Product Innovation Management*. 32(6). 925-938. DOI: 10.1111/JPIM.12163

Madaus, G. F. and Stufflebeam, D. L. (2000). Program Evaluation: A Historical Overview. In Stufflebeam, D. L., Madaus, G. F., and Kellaghan, T. (eds). *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*. Evaluation in Education and Human Services. Springer Netherlands. 3–18. DOI: 10.1007/0-306-47559-6_1

Nightingale, A. (2009). Triangulation. In Kitchin, R. and Thrift, N. (eds). International Encyclopedia of Human Geography. Elsevier Science. 489-492. DOI: 10.1016/B978-008044910-4.00552-6

Norström, A. V., Cvitanovic, C., Löf, M. F., West, S., Wyborn, C., et al. (2020). Principles for knowledge co-production in sustainability research. *Nature Sustainability*, 3(3). 182–190. DOI: 10.1038/s41893-019-0448-2

O'Connor, R. A., Nel, J. L., Roux, D. J., Lim-Camacho, L., van Kerkhoff, L. and Leach, J. (2019). Principles for evaluating knowledge co-production in natural resource management: Incorporating decision-maker values. *Journal of Environmental Management*, 249. 109392. DOI: 10.1016/j.jenvman.2019.109392

Overtorn, J. and van Diermen, P. (2014). Quantitative research. In Scheyvens, R. (eds). *Development fieldwork: A practical guide*. Sage. 39-58. DOI: 10.4135/9781473921801

Patton, M. Q. (2010). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. Guilford.

Patton, M. Q. (2006). Evaluation for the way we work. *The Nonprofit Quarterly*, 13(1). 28-33.

Patton, M. Q. and Horton, D. (2009). Utilization-focused evaluation for agricultural innovation. ILAC Brief. Institutional Learning and Change Initiative.

Plottu, B. and Plottu, E. (2011). Participatory Evaluation: The Virtues for Public Governance, the Constraints on Implementation. *Group Decision and Negotiation*, 20. 805–824. DOI: 10.1007/s10726-010-9212-8

Prokopy, L., Carlton, J., Haigh, T., Lemos, M., Mase, A., & Widhalm, M. (2017). Useful to Usable: Developing usable climate science for agriculture. *Climate Risk Management*, 15, 1-7. DOI: 10.1016/j.crm.2016.10.004

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M. and Britten, N. (2006). *Guidance on the Conduct of Narrative Synthesis in Systematic Reviews.* ESRC Methods Programme. https://www.lancaster.ac.uk/media/lancaster-university/content-assets/documents/fhm/dhr/chir/NSsynthesisguidanceVersion1-April2006.pdf

Rajala, E., Vogel, I., Sundin, A., Kongmanila, D., Nassuna-Musoke, M. G., et al. (2021). How can agricultural research translation projects targeting smallholder production systems be strengthened by using Theory of Change? *Global Food Security*, 28. 100475. DOI: 10.1016/j.gfs.2020.100475

Reed, M. S., Ferré, M., Martin-Ortega, J., Blanche, R., Lawford-Rolfe, R., Dallimer, M. and Holden, J. (2021). Evaluating impact from research: A methodological framework. *Research Policy*, 50(4). 104147. DOI: 10.1016/j.respol.2020.104147

Reed, M., Vella, S., Challies, E., de Vente, J., Frewer, L., et al. (2017). A theory of participation: What makes stakeholder and public engagement in environmental management work? *Restoration Ecology*, 26. DOI: 10.1111/rec.12541

Reever Morghan, K., Sheley, R., & Svejcar, T. (2006). Successful Adaptive Management—The Integration of Research and Management. *Rangeland Ecology & Management,* 59(2), 216-219. doi: 10.2111/05-079r1.1

Roux, D. J., Stirzaker, R. J., Breen, C. M., Lefroy, E. C. and Cresswell, H. P. (2010). Framework for participative reflection on the accomplishment of transdisciplinary research programs. *Environmental Science & Policy*, 13(8). 733–741. DOI: 10.1016/j.envsci.2010.08.002

Salamanca A., and Biskupska, N. (2021). Monitoring, evaluation and learning to build better climate services: A framework for inclusion, accountability and iterative improvement in tandem. SEI Discussion Brief. Stockholm Environment Institute.

Sarkki, S., Tinch, R., Niemelä, J., Heink, U., Waylen, K., et al. (2015). Adding 'iterativity' to the credibility, relevance, legitimacy: A novel scheme to highlight dynamic aspects of science–policy interfaces. *Environmental Science & Policy*, 54. 505–512. DOI: 10.1016/j.envsci.2015.02.016

Schuck-Zöller, S., Cortekar, J. and Jacob, D. (2017). Evaluating co-creation of knowledge: from quality criteria and indicators to methods. *Advances in Science and Research*, 14. 305–312. DOI: 10.5194/asr-14-305-2017

SEI (2021). About. Stockholm Environment Institute. https://www.sei.org/about-sei/

Slunge D., Drakenberg O., Ekbom A., Göthberg M., Knaggård Å. and Sahlin U. (2017). *Stakeholder Interaction in Research Processes – a Guide for Researchers and Research Groups*. University of Gothenburg. https://gmv.gu.se/digitalAssets/1619/1619929_stakeholder-interaction-in-research-processes---guide---final-march-15-2017.pdf

Sun, J., & Yang, K. (2016). The Wicked Problem of Climate Change: A New Approach Based on Social Mess and Fragmentation. *Sustainability*, *8*(12), 1312. DOI: 10.3390/su8121312

van der Wal, M., De Kraker, J., Offermans, A., Kroeze, C., Kirschner, P. A. and van Ittersum, M. (2014). Measuring Social Learning in Participatory Approaches to Natural Resource Management: Measuring Social Learning in Participatory Approaches. *Environmental Policy and Governance*, 24(1). 1–15. DOI: 10.1002/eet.1627

van Es, M., Guijt, I. and Vogel, I. (2015). *Hivos ToC Guidelines: Theory of Change Thinking in Practice. A Stepwise Approach.* Hivos. https://hivos.org/document/hivos-theory-of-change/

van Tulder, R. and Keen, N. (2018). Capturing Collaborative Challenges: Designing Complexity-Sensitive Theories of Change for Cross-Sector Partnerships. *Journal of Business Ethics*, 150(2). 315–332. DOI: 10.1007/s10551-018-3857-7

Van der Molen, K., Wall, T. U. and Daudert, B. (2019). A Call for the Evaluation of Web-Based Climate Data and Analysis Tools. *Bulletin of the American Meteorological Society*, 100(2). 257–268. DOI: 10.1175/BAMS-D-18-0006.1

Vaughan, C. and Dessai, S. (2014). Climate services for society: origins, institutional arrangements, and design elements for an evaluation framework. *Wiley Interdisciplinary Reviews. Climate Change*, 5(5). 587–603. DOI: 10.1002/wcc.290

Vestlandsforskning (n.d.). *Unpacking climate impact CHAINs. A new generation of action – and user-oriented climate change risk assessments (UNCHAIN).* https://www.vestforsk.no/en/project/unpacking-climate-impact-chains-new-generation-action-and-user-oriented-climate-change-risk

Vincent, K., Daly, M., Scannell, C. and Leathes, B. (2018). What can climate services learn from theory and practice of co-production? *Climate Services*, 12. 48–58. DOI: 10.1016/j.cliser.2018.11.001

Vink, M. J., Dewulf, A., & Termeer, C. (2013). The role of knowledge and power in climate change adaptation governance: a systematic literature review. *Ecology and Society* 18(4).

Wall, T. U., Meadow, A. M. and Horganic, A. (2017). Developing Evaluation Indicators to Improve the Process of Coproducing Usable Climate Science. *Weather, Climate, and Society*, 9(1). 95–107. DOI: 10.1175/WCAS-D-16-0008.1

Walter, A. I., Helgenberger, S., Wiek, A. and Scholz, R. W. (2007). Measuring societal effects of transdisciplinary research projects: Design and application of an evaluation method. *Evaluation and Program Planning*, 30(4). 325–338. DOI: 10.1016/j.evalprogplan.2007.08.002

Wickson, F., Carew, A. L. and Russell, A. W. (2006). Transdisciplinary research: characteristics, quandaries and quality. *Futures*, 38(9). 1046–1059. DOI: 10.1016/j.futures.2006.02.011

Wiek, A., Talwar, S., O'Shea, M. and Robinson, J. (2014). Toward a methodological scheme for capturing societal effects of participatory sustainability research. *Research Evaluation*, 23(2). 117–132. DOI: 10.1093/reseval/rvt031

Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *8th International Conference on Evaluation and Assessment in Software Engineering (EASE) 2014* 321–330.

Zscheischler, J., Rogga, S. and Lange, A. (2018). The success of transdisciplinary research for sustainable land use: individual perceptions and assessments. *Sustainability Science*, 13(4). 1061–1074. DOI: 10.1007/s11625-018-0556-3

Appendices

Appendix A: Start set for literature review

Table 7: Start set for literature review

Document information	Citations	Inclusion in start set	Expert recommendation or literature search
Lang, D., Wiek, A., Bergmann, M., Stauffacher, M., Martens, P., and Moll, P. 2012. Transdisciplinary research in sustainability science: practice, principles, and challenges. *Sustainability Science*, *7*(S1), 25-43.	917	Yes	Expert recommendation
Fazey, I., Bunse, L., Msika, J., Pinke, M., Preedy, K., Evely, A., Lambert, E., Hastings, E., Morris, S. and Reed, M., 2014. Evaluating knowledge exchange in interdisciplinary and multi-stakeholder research. *Global Environmental Change*, 25, 204-220.	126	Yes	Literature search
Wiek, A., Talwar, S., O'Shea, M., and Robinson, J., 2014. Toward a methodological scheme for capturing societal effects of participatory sustainability research. *Research Evaluation*, 23(2), 117-132.	77	Yes	Expert recommendation
Wall, T., Meadow, A., and Horganic, A., 2017. Developing Evaluation Indicators to Improve the Process of Coproducing Usable Climate Science. *Weather, Climate, And Society*, *9*(1), 95-107.	49	Yes	Expert recommendation
Norström, A.V., Cvitanovic, C., Löf, M.F., West, S., Wyborn, C., Balvanera, P., Bednarek, A.T., Bennett, E.M., Biggs, R.,l, de Bremond, A., Campbell, B.M., Canadell, J.G., Carpenter, S.R., Folke, C., Fulton, E.A., Gaffney, O., Gelcich, S., Jouffray, J.-B., Leach, M., Le Tissier, M., Martín-López, B. and Louder, E. 2020. Principles for knowledge co-production in sustainability research. *Nature Sustainability*, 3(3), 182-190	30	No	Expert recommendation
Hansson, S., & Polk, M. (2018). Assessing the impact of transdisciplinary research: The usefulness of relevance, credibility, and legitimacy for understanding the link between process and impact. *Research Evaluation, 27*(2), 132-144. doi: 10.1093/reseval/rvy004	27	No	Expert recommendation
Van Epp, M. and Garside, B., 2019. Towards an evidence base on the value of social learning-oriented approaches in the context of climate change and food security. *Environmental Policy and Governance*, 29(2), 118-131	8	No	Expert recommendation
Restrepo, M.J., Lelea, M.A. and Kaufmann, B.A., 2018. Evaluating knowledge integration and co-production in a 2-year collaborative learning process with smallholder dairy farmer groups. *Sustainability science*, 13(5),1265-1286Top of FormBottom of Form	3	No	Literature search
O'Connor, R. A., Nel, J. L., Roux, D. J., Lim-Camacho, L., van Kerkhoff, L. and Leach, J. (2019). Principles for evaluating knowledge co-production in natural resource management: Incorporating decision-maker values. *Journal of Environmental Management*, 249. 109392.	2	No	Literature search
Fulgenzi, A., Brouwer, S., Baker, K. and Frijns, J. (2020). Communities of practice at the center of circular water solutions. *WIREs Water*, 7(4).	0	No	Literature search

Appendix B: Documents included in the literature review

Belcher, B. M., Rasmussen, K. E., Kemshaw, M. R. and Zornes, D. A. (2016). Defining and assessing research quality in a transdisciplinary context. *Research Evaluation*, 25(1). 1–17. DOI: 10.1093/reseval/rvv025

Blackstock, K. L., Kelly, G. J. and Horsey, B. L. (2007). Developing and applying a framework to evaluate participatory research for sustainability. *Ecological Economics*, 60(4). 726–742. DOI: 10.1016/j.ecolecon.2006.05.014

Carew, A. L. and Wickson, F. (2010). The TD Wheel: A heuristic to shape, support and evaluate transdisciplinary research. *Futures*, 42(10). 1146–1155. DOI: 10.1016/j.futures.2010.04.025

Ernst, A. (2019). Research techniques and methodologies to assess social learning in participatory environmental governance. *Learning, Culture and Social Interaction*, 23. 100331. DOI: 10.1016/j.lcsi.2019.100331

Fazey, I., Bunse, L., Msika, J., Pinke, M., Preedy, K., et al. (2014). Evaluating knowledge exchange in interdisciplinary and multi-stakeholder research. *Global Environmental Change*, 25. 204–220. DOI: 10.1016/j.gloenvcha.2013.12.012

Fulgenzi, A., Brouwer, S., Baker, K. and Frijns, J. (2020). Communities of practice at the center of circular water solutions. *WIREs Water*, 7(4). e1450. DOI: 10.1002/wat2.1450

Hassenforder, E., Smajgl, A. and Ward, J. (2015). Towards understanding participatory processes: Framework, application and results. *Journal of Environmental Management*, 157. 84–95. DOI: 10.1016/j.jenvman.2015.04.012

Hegger, D., Lamers, M., Van Zeijl-Rozema, A. and Dieperink, C. (2012). Conceptualising joint knowledge production in regional climate change adaptation projects: success conditions and levers for action. *Environmental Science & Policy*, 18. 52–65. DOI: 10.1016/j.envsci.2012.01.002

Heink, U., Marquard, E., Heubach, K., Jax, K., Kugel, C., et al. (2015). Conceptualizing credibility, relevance and legitimacy for evaluating the effectiveness of science–policy interfaces: Challenges and opportunities. *Science and Public Policy*, 42(5). 676–689. DOI: 10.1093/scipol/scu082

Hitziger, M., Aragrande, M., Berezowski, J. A., Canali, M., Del Rio Vilas, V., et al. (2019). EVOlvINC: Evaluating knowledge integration capacity in multistakeholder governance. *Ecology and Society*, 24(2). DOI: 10.5751/ES-10935-240236

Holzer, J. M., Carmon, N. and Orenstein, D. E. (2018). A methodology for evaluating transdisciplinary research on coupled socio-ecological systems. *Ecological Indicators*, 85. 808–819. DOI: 10.1016/j.ecolind.2017.10.074

Jahn, T. and Keil, F. (2015). An actor-specific guideline for quality assurance in transdisciplinary research. *Futures*, 65. 195–208. DOI: 10.1016/j.futures.2014.10.015

Jones, N. A., Perez, P., Measham, T. G., Kelly, G. J., d'Aquino, P., Daniell, K. A., Dray, A. and Ferrand, N. (2009). Evaluating Participatory Modeling: Developing a Framework for

Cross-Case Analysis. *Environmental Management*, 44(6). 1180. DOI: 10.1007/s00267-009-9391-8

Knickel, M., Knickel, K., Galli, F., Maye, D. and Wiskerke, J. S. C. (2019). Towards a Reflexive Framework for Fostering Co—Learning and Improvement of Transdisciplinary Collaboration. *Sustainability*, 11(23). 6602. DOI: 10.3390/su11236602

Lang, D. J., Wiek, A., Bergmann, M., Stauffacher, M., Martens, P., Moll, P., Swilling, M. and Thomas, C. J. (2012). Transdisciplinary research in sustainability science: practice, principles, and challenges. *Sustainability Science*, 7(1). 25–43. DOI: 10.1007/s11625-011-0149-x

Norström, A. V., Cvitanovic, C., Löf, M. F., West, S., Wyborn, C., et al. (2020). Principles for knowledge co-production in sustainability research. *Nature Sustainability*, 3(3). 182–190. DOI: 10.1038/s41893-019-0448-2

O'Connor, R. A., Nel, J. L., Roux, D. J., Lim-Camacho, L., van Kerkhoff, L. and Leach, J. (2019). Principles for evaluating knowledge co-production in natural resource management: Incorporating decision-maker values. *Journal of Environmental Management*, 249. 109392. DOI: 10.1016/j.jenvman.2019.109392

Roux, D. J., Stirzaker, R. J., Breen, C. M., Lefroy, E. C. and Cresswell, H. P. (2010). Framework for participative reflection on the accomplishment of transdisciplinary research programs. *Environmental Science & Policy*, 13(8). 733–741. DOI: 10.1016/j.envsci.2010.08.002

Sarkki, S., Tinch, R., Niemelä, J., Heink, U., Waylen, K., et al. (2015). Adding 'iterativity' to the credibility, relevance, legitimacy: A novel scheme to highlight dynamic aspects of science–policy interfaces. *Environmental Science & Policy*, 54. 505–512. DOI: 10.1016/j.envsci.2015.02.016

van der Wal, M., De Kraker, J., Offermans, A., Kroeze, C., Kirschner, P. A. and van Ittersum, M. (2014). Measuring Social Learning in Participatory Approaches to Natural Resource Management: Measuring Social Learning in Participatory Approaches. *Environmental Policy and Governance*, 24(1). 1–15. DOI: 10.1002/eet.1627

Wall, T. U., Meadow, A. M. and Horganic, A. (2017). Developing Evaluation Indicators to Improve the Process of Coproducing Usable Climate Science. *Weather, Climate, and Society*, 9(1). 95–107. DOI: 10.1175/WCAS-D-16-0008.1

Walter, A. I., Helgenberger, S., Wiek, A. and Scholz, R. W. (2007). Measuring societal effects of transdisciplinary research projects: Design and application of an evaluation method. *Evaluation and Program Planning*, 30(4). 325–338. DOI: 10.1016/j.evalprogplan.2007.08.002

Wickson, F., Carew, A. L. and Russell, A. W. (2006). Transdisciplinary research: characteristics, quandaries and quality. *Futures*, 38(9). 1046–1059. DOI: 10.1016/j.futures.2006.02.011

Wiek, A., Talwar, S., O'Shea, M. and Robinson, J. (2014). Toward a methodological scheme for capturing societal effects of participatory sustainability research. *Research Evaluation*, 23(2). 117–132. DOI: 10.1093/reseval/rvt031

Zscheischler, J., Rogga, S. and Lange, A. (2018). The success of transdisciplinary research for sustainable land use: individual perceptions and assessments. *Sustainability Science*, 13(4). 1061–1074. DOI: 10.1007/s11625-018-0556-3

Appendix C: Coding form for literature review

Table 8: Coding form

Codes		Type of code
Basic information	Author(s)	Descriptive text
	Title	Descriptive text
	Year	Descriptive text
	Abstract	Descriptive text
	Research design	Descriptive text (case study, literature review, etc)
	Discipline	Descriptive text (sustainable development, natural resource management, climate change adaptation, etc)
	Country/Region/Sub-national	Country/Region/Sub-national area
	Empirically tested	Yes, no
Knowledge co-production	Term used	Descriptive text (transdicplinary, partcipatory, co-producing, etc).
	Definition	Descriptive text
	Theoretical approach	Descriptive text
Evaluation	Effects	Output, outcome, impact
	Typology	Summative, formative
	Timing	Pre-assessment, monitoring, retrospective, all
	Design	Descriptive text (e g. qualitative, quantitative, mix-method, participatory, etc)
	Data collection	Descriptive text (workshops, surveys, interviews, knowledge tests, etc)
	Evaluation framework	Descriptive text explaining how the evaluation is approached
	Evaluation criteria	Descriptive text presenting and explaining evaluation criteria
	Success factors for evaluation	Descriptive text
	Challenges for evaluation	Descriptive text
	Other useful information	Descriptive text

Appendix D: Quantitative analysis in the literature review

For the systematic literature review, some simple statistical analysis was applied to codes that were most easily quantified. The results are presented below.

Figure 13: Most of the reviewed documents did not specify what is to be evaluated. Some of the reviewed documents suggested an evaluation approach appraising outputs, outcomes, and impact, whereas some others focused on either outcomes and impact, or outputs and outcomes. None of the reviewed literature exclusively considered outputs or impacts

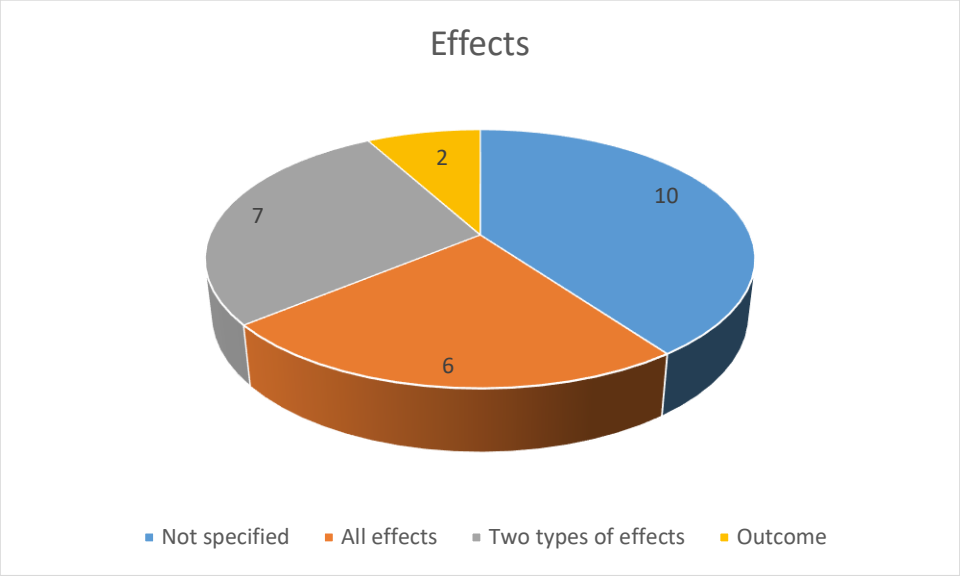

Figure 13: Effects.

Figure 14: Nine documents did not specify the purpose of their assessment. Other documents used a formative assessment. Some documents incorporated both summative and formative aspects. Only two of the documents employed a summative approach.

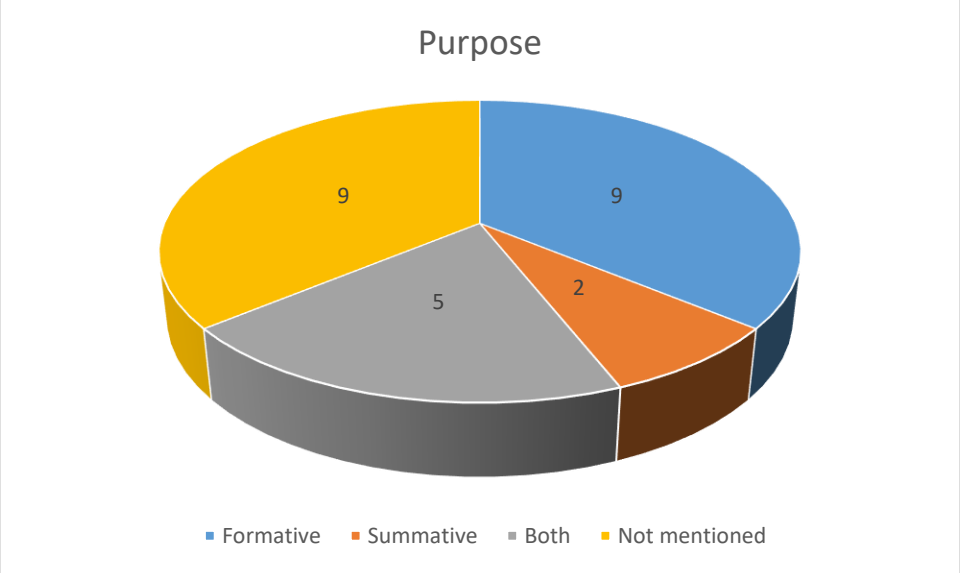

Figure 14: Purpose

Figure 15: Most documents argued for implementing multiple evaluation interventions at different points of time during a research project. Eight documents did not suggest a specific timing. Four documents presented a retrospective assessment.

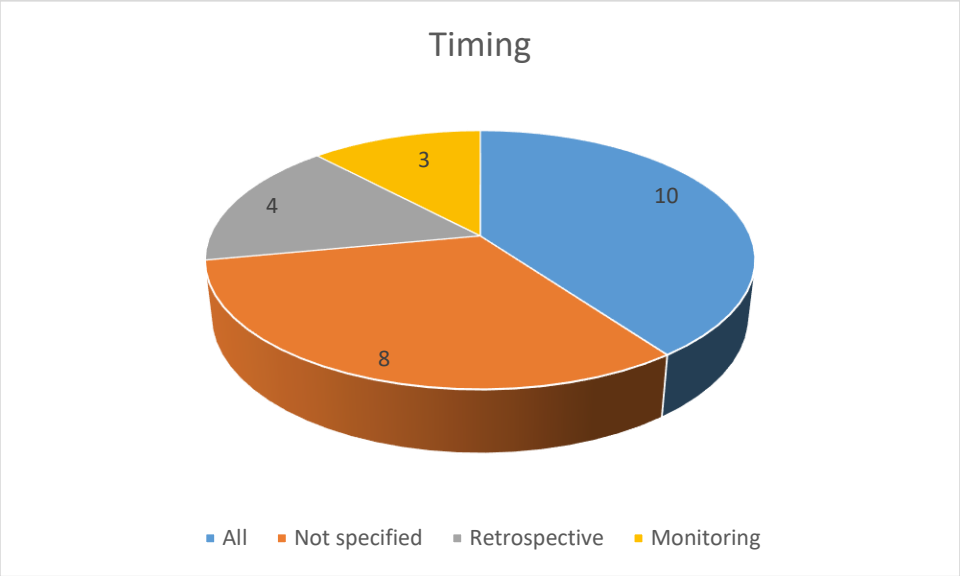

Figure 15: Timing

Appendix E: ToC methodology versus Logical Framework Approach

In short, logic models are roadmaps that hypothesizes the relationship between inputs, activities, outputs, outcomes, and impact in a project. Many types of logic models exist (Fazey 2014), of which the Logical Framework Approach is the most common (van Tulder and Keen 2018). As noted by van Tulder and Keen (2018, p.316), the ToC methodology "came from a dissatisfaction with the evaluation practices of the time, the limited understanding of complexity". There are some key differences that distinguishes the ToC methodology from the Logical Framework Approach, as presented in Table 9.

Table 9: The ToC methodology versus the logical framework approach (van Es et al. 2015).

ToC methodology	**Logical framework approach**
Complexity theory	Linear cause-effect model
Flexible, adaptive, and iterative	Rigid
Pathways of change	Three result levels

Why is the ToC methodology better at capturing complexity? In short, the ToC methodology largely builds on complexity theory (van Tulder and Keen 2018). Similar to developmental evaluation (Patton 2010), the ToC methodology acknowledges that flexibility is key when operating in contexts characterized by complexity. The ToC methodology is iterative, implying that it "needs regular revisiting and updating to the real situation as knowledge emerges from implementation experience". The ToC methodology thus encourages adaptation rather than prediction, meaning that it adapts as complexity unfolds.

Appendix F: ToC visualization

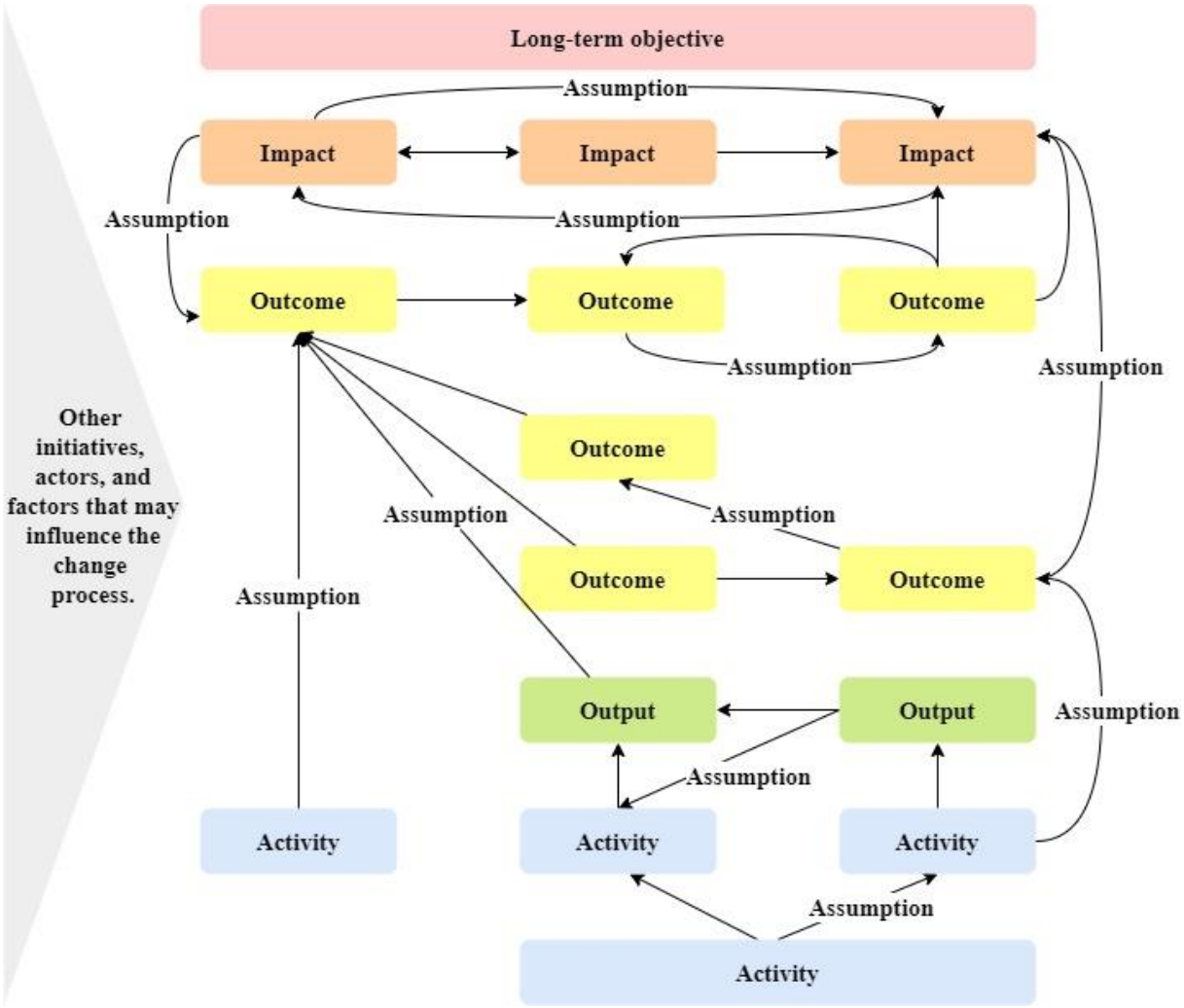

Figure 16: ToC visualization

Appendix G: Sampling

Table 10: Sampling

Group	Type of sampling	Description	Responses	Number of people targeted	Response rate
Stakeholders from HazardSupport	Purposive	The project engaged practitioners and researchers in a collaborative process to produce risk-based decision support for adaptation to future natural hazards.	6	17	35%
Stakeholders from Unchain	Purposive	The project aimed to improve the impact chain framework by integrating a knowledge co-production methodology along with some other methodological innovations.	11	46	24%
Staff at SEI	Purposive	SEI is a research institute that seeks to bridge science and policy, and have therefore incorporated knowledge co-production into many of its research projects.	29	327	9%
Personal network	Convenience	The questionnaire was publicly shared by Jenny Iao-Jörgensen who supervised the thesis.	11	Unknown	Unknown
Referrals	Snowball	Respondents were asked if they knew any other people relevant to the research and to share their contact information.	3	10	30%
Total			61	400	15%

Appendix H: Challenges in evaluating knowledge co-production

Table 11: Challenges in evaluating knowledge co-production

Challenge	**Reference**
Assess both tangible and intangible effects	Blackstock et al. 2007; Lang et al. 2012; Wiek et al. 2014
Terminological ambiguity	Belcher et al. 2016; Ernst 2019; Fulgenzi et al. 2020; Heink et al. 2015; Jahn and Keil 2015; van der Wal et al. 2014; Wiek et al. 2014; Zscheischler et al. 2018
Long time horizons and time lags between cause and effects	Blackstock et al. 2007; Ernst 2019; Jahn and Keil 2015; Jones et al. 2009; Lang et al. 2012; Roux et al. 2010; Wall et al. 2017; Wiek et al. 2014; Zscheischler et al. 2018
External factors	Ernst 2019; Jones et al. 2009; Zscheischler et al. 2018
Contribution	Blackstock et al. 2007; Fulgenzi et al. 2020;
Timing	Fulgenzi et al. 2020; Wall et al. 2017
Complexity	Fazey et al. 2014; Hassenforder et al. 2015; Lang et al. 2012; Roux et al. 2010; Wall et al. 2017; Wiek et al. 2014; Zscheischler et al. 2018
Diversity of involved actors	Hitziger et al. 2019; Jones et al. 2009; Knickel et al. 2019; Roux et al. 2010; Zscheischler et al. 2018
Increased burden	Knickel et al. 2019
Lack of trust towards evaluator	Knickel et al. 2019; Wiek et al. 2014
Uncertainty	Roux et al. 2010; Zscheischler et al. 2018
Staff turnover	Sarkki et al. 2015

Appendix I: Success factors in evaluating knowledge co-production

Table 12: Success factors in evaluating knowledge co-production

Success factor	**References**
Flexibility	Belcher et al. 2016; Blackstock et al. 2007; Holzer et al. 2018; Knickel et al. 2019; Lang et al. 2012; van der Wal et al. 2014; Walter et al. 2007
Assess various outputs, outcomes, and impact	Belcher et al. 2016; Fazey et al. 2014; Wall et al. 2017; Wiek et al. 2014
Reflexivity	Belcher et al. 2016; Blackstock et al. 2007; Carew and Wickson 2010; Hegger et al. 2012
Contextual	Belcher et al. 2016; Blackstock et al. 2007; Carew and Wickson 2010; Fazey et al. 2014; Heink et al. 2015; Holzer et al. 2018; Jones et al. 2009; Knickel et al. 2019; Lang et al. 2012; Sarkki et al. 2015; van der Wal et al. 2014; Walter et al. 2007
Clear objectives	Blackstock et al. 2007; Carew and Wickson 2010; Heink et al. 2015; Knickel et al. 2019; Roux et al. 2010; Wiek et al. 2014
Theory of Change	Fazey et al. 2014; Heink et al. 2015; Knickel et al. 2019; Norström et al. 2020; Wall et al. 2017
Embed evaluation into research process	Fazey et al. 2014; Wall et al. 2017
Mix-methods	Fazey et al. 2014; Wiek et al. 2014
Trust	Roux et al. 2010; Sarkki et al. 2015; Wall et al. 2017
Stakeholder participation	Fazey et al. 2014; Lang et al. 2012; Norström et al. 2020; Walter et al. 2007; Wiek et al. 2014
Visual products	Lang et al. 2012

Appendix J: Online questionnaire – English

How to evaluate co-produced climate information

Thank you for taking your time to answer this survey that is part of a master thesis project at Lund University and the Stockholm Environment Institute (SEI). Your answers are important regardless of your previous experience in co-producing climate information.

The aim is to develop a set of principles outlining methodological choices for assessing co-produced climate information. Co-production involves researchers and practitioners in a collaborative research-process to produce climate information tailored to local contexts.

The goal of this survey is to gather your experience and insights on this subject.

***** Mandatory

Informed consent

The survey is expected to take around 10 minutes. Your participation is completely voluntary. The answers will be analyzed, synthesized, and reported in the final thesis and possibly in other related research at SEI. Your name and contact details will be visible for the researcher only. All information you supply during the research will be held in confidence, and your name will not appear in any reports. The collected data will be stored for one year.

In the case of any questions, feel free to contact Mathilda Englund, Msc candidate at Lund University and intern at SEI, at +467260432 or mathilda.englund@sei.org.

1. I have read the information above, and I hereby agree to participate in the study. *

- ○ Yes
- ○ No

2. Please enter your e-mail address. *

Introductory questions

3. What type of organization do you represent? *

- Municipality
- County administrative board
- Governmental agency
- Research institute or university
- Private enterprise
- Funding agency
- Non-governmental organization
- Other

4. What countries are you primarily working with? *

Roles and responsibilities

5. What statement best describe your role when co-producing climate information? *

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
Producer of information on climate or context	○	○	○	○	○
User of climate information	○	○	○	○	○
Intermediary, translating climate information into a more usable format	○	○	○	○	○
Finanser	○	○	○	○	○
Evaluator	○	○	○	○	○

Effects: outputs, outcomes, and impacts

6. Which outputs, outcomes, and impacts do you consider important when evaluating co-production initiatives? *

	Very important	Important	Neutral	Less important	Not important
Peer-reviewed articles and final reports	○	○	○	○	○
Scientific progress	○	○	○	○	○
Usable information	○	○	○	○	○
Improved social relationships	○	○	○	○	○
Mutual learning	○	○	○	○	○
Enriched decision-making	○	○	○	○	○

Participation

Participation means that stakeholders are actively involved in the evaluation process.

7. Would you recommend stakeholder participation when evaluating co-production initiatives? *

- Yes
- No

8. If no, why do you advise against stakeholder participation when evaluating co-production initiatives? *

9. If yes: In your view, how can stakeholder participation improve evaluations? Please select all that apply. *

- By building a sense of ownership over evaluation findings among involved stakeholders

- By developing a shared vision among involved stakeholders
- By increasing buy-in from relevant stakeholders
- By validating findings
- None of the above
- I don¨t know
- Other

10. If yes: In your view, when is it appropriate to involve stakeholders in the evaluation process? Please select all that apply. *

- When defining the purpose of the evaluation
- When designing the evaluation approach
- When developing indicators
- When collecting data
- When analyzing data
- When reporting the findings
- None of the above
- I don’t know
- Other

11. Do you have any additional comments on stakeholder participation?

Methods

12. In your view, how useful are the following methods when evaluating co-production initiatives? *

	Very useful	Useful	Neutral	Less useful	Not useful	I don’t know

Questionnaires	○	○	○	○	○	○
Group discussions	○	○	○	○	○	○
Review of project documentation	○	○	○	○	○	○
Use of indicators	○	○	○	○	○	○
Interviews with involved stakeholders	○	○	○	○	○	○
Written reflections	○	○	○	○	○	○
Mix of different methods	○	○	○	○	○	○

13. Are there any other methods you would like to recommend?

14. Do you think visual products can help clarify complex issues? Visual products include, but not limited to, network maps, mind maps, and drawings. *

- Yes
- No

15. Are you familiar with the Theory of Change methodology. If no, please proceed to question 20. *

- Yes
- No

16. Would you recommend using a Theory of Change methodology when evaluating co-production initiatives? *

- Yes
- No

17. If no, why would you advise against using a Theory of Change methodology when evaluating co-production initiatives? *

18. If yes: In your view, what are the benefits of using a Theory of Change methodology? Please select all that apply. *

- To help define what you are seeking to achieve
- To clarify cause-effect relationships
- To grasp complexity and context
- To help define what you are seeking to monitor (which does not necessarily need to be an indicator)
- To identify underlying assumptions
- None of the above
- I don't know
- Other

19. Are there any additional comments you would like to provide considering the Theory of Change methodology?

Applicability of findings

20. Do you feel like you can apply evaluation findings in practice?? *

- Yes
- No

21. If yes, what enable you to apply evaluation findings in practice? *

22. If no, what prevent you from applying evaluation findings in practice? *

Closing questions

23. Is there anything else you would like to say, point out or comment on?

24. Would you like to receive a digital copy/link to the published thesis?

- Yes
- No

25. Do you know someone else that should answer the survey? If yes, please share their e-mail address.

26. Would you be interested taking part in a follow-up interview? *

- Yes
- No

End of survey

Thank you for partcipating.

If you have any questions, please contact Mathilda Englund at +467260432 or mathilda.englund@sei.org

Appendix K: Online questionnaire – Swedish

Hur utvärderar man gemensam kunskapsproduktion för klimatanpassning?

Välkommen att delta i en enkät kring utvärderingen av gemensam kunskapsproduktion för klimatanpassning. Dina svar är lika viktiga oavsett tidigare erfarenheter.

Studien är del av en master-uppsats vid Lunds Tekniska Högskola och Stockholm Environment Institute (SEI). Målet med studien är att ta fram en vägledning för val av metoder i utvärdering av gemensam kunskapsproduktion för klimatanpassning. Gemensam kunskapsproduktion innebär att forskare och praktiker samarbetar för att utveckla klimatinformation som är skräddarsydd efter lokala behov.

Målet med enkäten är att samla in dina erfarenheter och tankar kring detta.

* Obligatoriskt

Samtycke

Enkäten kommer ta ungefär 10 minuter. Ditt deltagande är frivilligt. Svaren kommer analyseras, sammanställas, och presenteras i en master-uppsats samt liknande forskningsprojekt hos SEI. Ditt namn och kontaktuppgifter kommer bara synas för forskaren. All information du tillhandhåller är konfidentiell, och ditt namn kommer inte synas i den slutliga uppsatsen. Datan som samlas in sparas i ett år.

Om du har några frågor, kontakta Mathilda Englund, student vid Lunds Tekniska Högskola och praktikant hos SEI, via +467 260 432 eller mathilda.englund@sei.org.

1. Jag har läst och förstått informationen ovan, och samtycker till att delta i studien *

- ◯ Ja
- ◯ Nej

2. Vänligen dela din e-postadress. *

Introduktion

3. Vilken typ av organisation jobbar du i? *

- Kommun
- Länsstyrelse
- Statlig myndighet
- Forskningsinstitut eller universitet
- Privat företag
- Forsknings- och utvecklingsfinansiär
- Ideell organisation
- Annat

4. Vilka länder jobbar du huvudsakligen med? *

Roller

5. Vilket påstående stämmer bäst överens med den roll du brukar ha i gemensamma kunskapsproduktionsprocesser för klimatanpassning? *

	Instämmer helt	Instämmer	Varken eller	Instämmer inte	Instämmer inte alls
Jag producerar och delar information, både gällande klimat och kontext	○	○	○	○	○
Jag använder klimatinformation	○	○	○	○	○
Intermediär, jag ombearbetar klimatinformation till ett mer användbart format	○	○	○	○	○

Jag finansierar projekt	◯	◯	◯	◯	◯
Jag utvärderar projekt	◯	◯	◯	◯	◯

Resultat och effekter

6. Vilka av följande resultat och effekter anser du viktiga när man utvärderar gemensamma kunskapsproduktionsprocesser? *

	Mycket viktig	Viktig	Varken eller	Mindre viktig	Oviktig
Forskningsrapporter och artiklar	◯	◯	◯	◯	◯
Vetenskapliga framsteg	◯	◯	◯	◯	◯
Forskningsresultat blir implementerade i praktiken	◯	◯	◯	◯	◯
Stärkta relationer	◯	◯	◯	◯	◯
Ömsesidigt lärande	◯	◯	◯	◯	◯
Förbättrat beslutsfattande	◯	◯	◯	◯	◯

Deltagande

Deltagande innebär här att projektdeltagare är aktivt involverade i utvärderingsprocessen.

7. Skulle du rekommendera att aktivt involvera projektdeltagare i utvärderingsprocessen? *

- Ja
- Nej

8. Om nej, varför skulle du avråda från att aktivt involvera projektdeltagare i utvärderingsprocessen? *

9. Om ja, hur kan involvering av projektdeltagare förbättra utvärderingsprocessen? Du kan ange flera svar. *

- Genom att skapa en känsla av gemensamt ägandeskap över utvärderingsresultaten
- Genom att utveckla en gemensam vision och målbild
- Genom att öka motivationen för deltagare att delta i utvärderingen
- Genom att bekräfta utvärderingsresultatet bland projektdeltagarna
- Inget av ovanstående
- Jag vet inte
- Annat

10. Om ja, när tycker du att projektdeltagare ska vara involverade i utvärderingsprocessen? *

- Under beslutet av syfte
- Under utformningen av utvärderingsdesign
- Under utvecklingen av indikatorer
- Under datainsamlingen
- Under dataanalysen
- Under rapporteringen av resultatet
- Inget av ovanstående
- Jag vet inte
- Annat

11. Har du något du skulle vilja lägga till kring att aktivt involvera projektdeltagare i utvärderingsprocessen?

Metoder

12. Hur användbara tycker du de följande metoderna är vid utvärderingen av gemensamma kunskapsproduktionsprocesser? *

	Mycket användbar	Användbar	Varken eller	Mindre användbar	Inte användbar	Vet inte
Enkäter	○	○	○	○	○	○
Gruppdiskussioner	○	○	○	○	○	○
Granskning av projektdokumentation	○	○	○	○	○	○
Användning av indikatorer	○	○	○	○	○	○
Intervjuer med projektdeltagare	○	○	○	○	○	○
Skriftliga reflektioner	○	○	○	○	○	○
En blandning av olika metoder	○	○	○	○	○	○

13. Skulle du vilja rekommendera någon annan metod?

14. Tycker du att visuella hjälpmedel kan göra det enklare att förstå komplexa frågor? Visuella hjälpmedel inkluderar bland annat nätverkskartor, mindmaps, och teckningar. *

- Ja
- Nej

15. Känner du till förändringsteori? På engelska känd som Theory of Change. Om nej, kan du direkt gå till fråga 20. *

- Ja
- Nej

16. Skulle du rekommendera att använda en förändringsteori vid utvärderingen av gemensamma kunskapsproduktionsprocesser? *

- Ja
- Nej

17. Om nej, varför skulle du avråda från att använda en förändringsteori? *

18. Om ja, vilka fördelar ser du med användningen av en förändringsteori? Du kan ange flera svar. *

- Att definiera vad projektet och utvärderingen försöker uppnå
- Att klargöra sambandet mellan orsak och verkan
- Att förstå komplexitet och kontext
- Att klargöra vad som ska utvärderas (vilket inte nödvändigtvis behöver vara indikatorer)
- Att identifiera underliggande antaganden
- Inget av ovanstående
- Jag vet inte
- Annat

19. Har du något du skulle vilja lägga till gällande förändringsteori?

Tillämpning av resultat

20. Upplever du att du kan använda utvärderingsresultat i ditt arbete? *

- Ja
- Nej

21. Om ja, vad gör det möjligt för dig att använda utvärderingsresultat i ditt arbete? *

22. Om nej, vad stoppar dig från att använda utvärderingsresultat i ditt arbete? *

Avslutande frågor

23. Har du något du skulle vilja lägga till?

24. Vill du få en digital kopia av uppsatsen när den är färdig?

- Ja
- Nej

25. Känner du någon annan som borde besvara enkäten? Om ja, dela gärna deras e-post.

26. Skulle du vara intresserad av en uppföljningsintervju? *

- Ja
- Nej

Slut på enkät

Tack för du deltagit.

Om du har några frågor, kontakta gärna Mathilda Englund via +46739260432 eller mathilda.englund@sei.org.