D4.1 RRI Self-Evaluation Tool

Version 2.3

Submission date:
Dissemination Level: Consortium
Author(s): Santtu Lehtinen & Henri Wiman
Peer-reviewed by: Peter Biegelbauer (AIT), Petra Wagner (AIT), Caroline Lackinger (AIT), György Pataki (ESSRG), Antonia Bierwirth (Tecnalia), Lucia Polo Alvarez (Tecnalia), Mika Nieminen (VTT), Nina Rilla (VTT)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 873112

Version log

Version	Issue Date	Authors	Notes
1.0	04/20/2022	Santtu Lehtinen (VTT)	Document creation and writing the structure
1.1	06/29/2022	Henri Wiman (VTT)	Writing Chapter 3
1.2	07/11/2022	Santtu Lehtinen (VTT)	Writing Chapter 2
1.3	07/19/2022	Santtu Lehtinen (VTT)	Writing and editing Chapters 1-5
1.4	07/27/2022	Henri Wiman (VTT)	Reviewing and editing the whole document
1.5	07/28/2022	Santtu Lehtinen (VTT)	Reviewing and editing the whole document
1.6	07/29/2022	Henri Wiman (VTT)	Reviewing and editing the whole document
1.7	08/22/2022	György Pataki (ESSRG)	Project Team review of v.1.6
1.8	08/25/2022	Caroline Lackinger (AIT)	Project Team review of v.1.6
1.8	08/25/2022	Petra Wagner (AIT)	Project Team review of v.1.6
1.9	09/02/2022	Antonia Bierwirth (Tecnalia)	Project Team review of v.1.6
1.9	09/02/2022	Lucia Polo Alvarez (Tecnalia)	Project Team review of v.1.6
2.0	09/05/2022	Nina Rilla (VTT)	Project Team review of v.1.6
2.0	09/05/2022	Mika Nieminen (VTT)	Project Team review of v.1.6
2.1	09/22/2022	Henri Wiman (VTT)	Adding Chapter 2.4
2.2	09/23/2022	Santtu Lehtinen (VTT)	Review based on the feedback
2.3	09/27/2022	Santtu Lehtinen (VTT)	Final review based on the feedback

EXECUTIVE SUMMARY

This deliverable 4.1 utilized a tailored ***systems thinking framework*** as a tool for ***Responsible Research and Innovation (RRI)*** self-evaluation and impact assessment. More specifically, our tool consisted of the utilization of ***participatory group model building*** and ***causal loop diagrams (CLD)*** to better understand the dynamics related to RRI implementation, integration, and institutionalization in three ***Co-Change Lab use cases***.

As a specific outcome of task 4.1, we developed a variety of descriptive and qualitative system models, with each model describing a specific dynamic present in a particular Lab use case. These models as well as the process preceding their development helped to create shared understanding on the system-level implementation of RRI principles as well as the drivers and barriers related to the uptake of RRI within RPOs. We also drew two generic RRI implementation models by synthetizing the relevant and analytically generalizable characters of the different individual models Lab use cases into one comprehensive model.

Our results uncover many familiar drivers and barriers of RRI institutionalization, but crucially connect them to one another. By doing so, we present ***change processes***, ***dynamics and conditions*** that can result in successful RRI institutionalization. An overarching finding is that even smaller scale or informal initial attempts at RRI practices can generate the conditions for more large scale and formalized change.

The utilization of systems thinking approach as a tool for understanding and evaluating RRI implementation, was seen to help to increase the capacity for organizational self-understanding, reflexivity, and social learning by contributing to the development and improvement of the contextual pre-conditions for organizational RRI implementation. Finally, our systems thinking -based approach also aligns well with RRI practices by creating understanding on complex systems, catalyzing collaboration and by stimulating the practice of continuous learning.

Co-funded by the Horizon 2020 programme of the European Union

Table of contents

Co-funded by the Horizon 2020 programme of the European Union

1-INTRODUCTION

Co-Create Change in Research Funding and Performing (CO-CHANGE) project is aimed at building transformative capacity and leadership for ***Responsible Research and Innovation (RRI)*** through systemic change coalitions. These coalitions are built around eight ***Change Labs*** that will test the uptake of RRI practices in selected organizations and their broader ecosystems.

This deliverable 4.1 is built on the utilization of ***systems thinking***, ***group model building*** and ***causal loop diagrams (CLD)*** as methods for understanding the drivers and barriers of RRI implementation, institutionalization, and integration within the project and among organizations more broadly. More specifically, our analysis is based on three selected Co-Change Lab use cases. The organizations behind the selected Labs were: ***Austrian Institute of Technology (AIT)***, ***Tecnalia*** and ***VTT Technical Research Centre of Finland Ltd (VTT)***. We ended up selecting three organizations instead of the initial estimation of four to five organizations because, in contrast to the other Co-Change partner organizations, the three selected Labs were relatively similar, technically oriented, leading European research performing organizations (RPOs), which created important commensurability between them[1].

The objective of the task 4.1 and the resulting deliverable 4.1 at hand was to develop a tool for systemic RRI self-evaluation and impact assessment. As per the task description, we utilized the systems thinking framework along with participatory group model building and causal loop diagrams to create reflexive capacity for understanding how the complex systemic relationships, interactions and feedback loops either strengthen or weaken the potential for responsible actions and impacts in organizations. These objectives were the essence of our group modeling exercises with the three Lab use cases.

We developed a variety of descriptive and qualitative system models, with each model describing a specific dynamic anticipated in a particular Lab use case. These models, along with the process preceding them, helped to create shared understanding on the system-level implementation, institutionalization, and integration of RRI principles as well as the drivers and barriers related to the uptake of RRI within RPOs. The models were created together with experts and stakeholders from the organizations participating in the Co-Change Labs.

As a result of the modeling process, we also created two generic RRI implementation models by synthetizing the relevant and analytically generalizable characters of the different individual models Lab use cases into one comprehensive model. Finally, we validated the generalized models of RRI implementation in Co-Change with the relevant project and lab partners in a separate session.

[1] Similarly, in terms of maturity, the three Labs were advanced enough to fruitfully conduct systemic analysis on the progress of RRI implementation in their organization. Finally, in terms of resources the selected three use cases had the available time and resources to fully commit to the modelling exercises.

The task description promised the development of plausible and alternative impact paths, which would help organizations to anticipate the potential impacts of their actions. These impact paths were elaborated as a part of the modeling exercises, which aimed to gauge the potential dynamics and impacts of RRI implementation within the organizations. In essence, each of the models we created presents either a generalized or a Lab-specific impact pathway. Nonetheless, we discovered that systems modeling itself was not the most suitable tool for an explicit impact assessment as such, since it would have required additional and complementary approaches, explicitly oriented towards evaluation and foresight practices. The inclusion of these complementary practices was out of the scope of our investigation.

In the next chapter, we present the systems thinking framework that guided our work during the task 4.1. The chapter also explains how the collaborative system modeling can be utilized by anyone for the purpose of organizational RRI implementation. The third chapter presents our own modeling process with the three Co-Change Labs together with relevant findings. Finally, we engage in a discussion between the results of our modeling process and the broader framework of the Co-Change project, including the concepts of organizational and transformative change.

2-SYSTEMS THINKING FRAMEWORK

Responsible research and innovation (*RRI*) is a part of a broader tradition consisting of various approaches drawing attention towards the ethical and social aspects, implications and impacts of science, technology and innovation in society (e.g. Rip, 2014). As detailed by the Stocktaking deliverable 1.1 (Tabares Gutierrez et al., 2020, 17), while the literature related to RRI has grown substantially (e.g. Florin, 2019; Stilgoe et al., 2013; Owen et al. 2012), the literature on the implementation, integration, and institutionalization of RRI into practice remains more elusive.

Despite, or precisely because of this deficiency, the idea of ***organizational change*** forms the core of the Co-Change project. Accordingly, the objective of the task 4.1 and resulting deliverable 4.1 was to help create insight into the question of how to change organizational practices, institutional frameworks, and people's mindsets in order to integrate, implement and institutionalize RRI practices into an organization.

Importantly for our work, the stocktaking deliverable 1.1 (Tabares Gutierrez et al., 2020) conducted a review of previous RRI-related EU-projects and academic articles relevant for the Co-Change project. As a result of the review, the stocktaking report identified ***societal challenges*** and ***the distribution of responsibilities between stakeholders in research and innovation*** as the two main drivers facilitating the adoption of RRI in general, while the issue of ***funding incentives*** was also seen as an important common driver for RRI. Moreover, the stocktaking deliverable formulated five pillars for promoting the uptake of RRI in organizational and institutional contexts, which were: ***Contextualization***; ***Ecosystems***; ***Organizational theory***; ***Metrics and indicators***; ***Communication, culture, and trust***. Finally, the stocktaking report also identified **eight specific barriers** and **nine separate drivers** for RRI implementation and institutionalization in organizations.

The **barriers** were: ***Diverging views of science and society relations*****;** ***Fear of the loss of scientific autonomy*****;** ***Difficulty to apply RRI in practice*****;** ***Tendency to outsource RRI*****;** ***Lack of incentives*****;** ***The unpredictability of scientific enterprise*****;** ***Insufficient resources and capacity for RRI*****;** ***Unclear added value of RRI***.

The **drivers** were: ***Pursuit of good society****;* ***Responsible Scientists****;* ***Alignment of Science and Society****;* ***Response to Societal Challenges; Participatory Science****;* ***Risk Governance****;* ***Social Innovation****;* ***Social License to operate****;* ***The gap between the implementation of RRI and the theory of RRI*****.**

As highlighted by the deliverable 1.1 (Tabares Gutierrez et al., 2020, 11), the issue of organizational change can be approached from many different perspectives. For example, the systemic view of organizational change, which draws from Complex Adaptive Systems Approach (CAS), sees organizations through characteristics such as, non-linearity, self-organization, interaction, emergency and path dependence (e.g. Mitleton-Kelly 2007). Similarly, the ecosystem concept, highlighted in deliverable 1.2 (Rilla et al., 2020) is linked to the idea of complex adaptive systems (Philips and Ritala 2019).

Our approach builds on these perspectives of change, as well as on to the idea of interactive learning, which emphasizes continuous learning, engagement, and communication (Tabares Gutierrez et al., 2020, 12-13). Nonetheless, despite these linkages, our approach towards analyzing organizational change in the project Co-Change is more akin to theory agnosticism. We did not adopt any specific theory of change as such, but rather chose to approach change as a context-sensitive phenomena.

Indeed, according to deliverables 1.1 (Tabares Gutierrez et al., 2020, 28) and 1.2 (Rilla et al., 2020, 2), RRI-driven transformations are always highly context-dependent, which makes it difficult to create any "off-the-shelf" or "one-size-fits-all" approaches to RRI implementation. Since every organization perceives the concept and practice of RRI differently, the implementation of RRI should be based on an organizational self-understanding and adapted to fit the relevant organizational context. Consequently, we chose to build on a context-specific empirical analysis by focusing on three specific Co-Change Lab case studies which provided contextual knowledge and insight into RRI implementation.

In short, while our own work is situated in the framework of ***systems thinking*** and focuses on the empirical evidence of RRI implementation, we're also mindful of the broader organizational change theories relevant to the Co-Change project, particularly ***the concept of transformative change*** (Wolfram 2016). According to Wolfram (2016, 126) ***transformative capacity*** "represents the power to change", in essence a "collective ability" to "conceive of, prepare for, initiate and perform path-deviant change towards sustainability", thus reflecting an "emergent property" of the relevant stakeholder context.

In short, specific transformative capacities can be utilized to achieve profound and sustainable change. These transformative capacities include ***Inclusive and multiform governance (TC1); Transformative leadership (TC2); Empowered and autonomous communities of practice (TC3); System(s) awareness and memory (TC4); Foresight (TC5); Diverse experimentation with disruptive solutions (TC6); Innovation embedding and coupling (TC7); Reflexivity and social learning (TC8); Working across agency levels (TC9); Working across political-administrative levels and geographical scales (TC10).***

2.1 Framework of systems thinking: theory and methodology

Our world is increasingly made up of, and dependent upon, various complex and interlinked sociotechnical systems (e.g. Geels 2004). These systems and the interdependent linkages between them have brought substantial material gains and wealth, but from the point of view of individual citizens, they have also made the world increasingly opaque and complex, making it harder to assess the impacts of one's own actions and decisions (Mulgan 2013). Moreover, in addition to the inherent opacity of the modern society, the increase in material wealth has also created societal challenges related to ecological degradation and social injustices. This has resulted in a new kind of demand for theories, methods and tools which utilize a more systemic and holistic approach towards understanding of the world.

We chose to approach the issue of RRI implementation, integration and institutionalization in Co-Change project through a "***system lens***" (Meadows 2008). According to Donella Meadows (2008, 2) ***systems thinking*** is "a way of thinking that gives us the freedom to identify root causes of problems and see new opportunities". In short, systems thinking aims for an understanding of how the functioning of a system can be changed in ways that produce desired results.

Relevant to our project focusing on ***research performing (RPO) and funding organizations (RFO)***, it is often noted that ***organizations*** tend to "have a life of their own" (Stroh 2015, 1). According to systems thinking, the behavior of a system does not result from some explanatory factors that are 'outside' of the system, but from the internal nature of the system itself. In essence, a system consists of an interconnected set of coherently organized elements that serve a function (Meadows 2008, 2,11).

These persistent patterns of behavior within a system can often be explained by mechanisms operating through the so-called ***feedback loops***. A feedback loop is formed when changes in a variable affect that same variable, often indirectly. The concept of ***feedback*** is essential in understanding how systems can cause their own behavior. (Meadows 2008, 25, 34)

Feedback loops are structured around the purpose or a function that a system serves, either explicitly or implicitly (Meadows 2008, 11). The actual purpose of the system is often not the purpose that is desired by the people in and around the system. Thus, systems thinking is often utilized to highlight how various decisions can produce unwanted or unanticipated results through feedback loops. The holistic approach provided by the systems view can be utilized to correct these deficiencies by looking at the whole rather than mere parts of a system. In this sense, systems thinking is not only geared towards understanding the purpose that a system is serving, but also for making it possible to change the functioning of the system for the better (Stroh 2015, 16).

The theoretical background behind systems thinking approach is ***systems theory***, which has various versions and sub-fields, often divided between biological (e.g. Bertalanffy (1969) and social scientific (e.g. Luhmann 1995) approaches. Other varieties include approaches such as ***complexity theory*** and ***system dynamics***. All these strands agree on the broad system principles but differ in terms of methodologies (Stroh 2015, 16-17). ***System dynamics*** for example is a specific method developed originally in Massachusetts Institute of Technology (MIT) for the purpose of explaining unintuitive system behavior with simulation (Forrester 1961).

Conversely, at its broadest, ***systems theory*** is portrayed as a "new philosophy" which is holistic and complex, often contrasted with the reductionist and linear causal paradigm of classical science (Hammond 2005). According to Deborah Hammond (2005), the "***systems view***" of the world can be summarized as a constructivist and participatory process "that emphasizes the importance of mutual understanding, meaning and values". As a result, systems thinking can be seen as a staunchly normative paradigm, as Hammond (2005, 23) argues: "*If knowledge is indeed an interactive and collaborative process, as well as an essential part of the decision-*

making process at every level of organization, then systems thinking contains an inherent ethical bias toward democratic and inclusive forms of social organization."

This "inherent ethical bias" is evident in our ***participatory group model building***, or ***group model building***, approach, which seeks to engage and include various stakeholders. Participatory group model building is a specific sub-category of systems thinking and system dynamics, focusing on the importance of engaging with stakeholders in order to formulate a consensus about a suitable problem definition and on the required solutions. One of the main aims of group model building is to facilitate mutual learning and shared understanding through the collective modeling process. (Sterman 2000; Vennix 1996)

Through the utilization of group modeling, participants are able to explicate, expose and express their thinking and mental models to others, thus enabling the creation of shared understanding about the issues at hand (Meadows 2008, 172). Our system thinking approach aims to mobilize the collective intelligence (Mulgan 2013) of the project partners and stakeholders by bringing forth the implicit and tacit knowledge present within the project group. In terms of our specific methodological choices, we chose to apply ***systems thinking*** through ***causal loop diagrams*** in combination with a ***participatory group model building*** approach.

To our knowledge, systems thinking, systems dynamics or causal loop diagrams have not been widely applied to the field of RRI. Nonetheless, there are a few notable academic papers which discuss the utilization of system dynamics in RRI-related issue areas. The paper by Gurzawska et al. (2017) aims to demonstrate the benefit of investing in RRI from a business perspective, through causal loop diagrams, while Setiawan et al. (2018) have demonstrated that system dynamics can support RRI principles of reflexivity and anticipation in the field of energy technology. Other studies have analyzed system dynamics through the lens of ethics, particularly through the relationship between the modeler and the model (e.g. Palmer 2017). Finally, a paper by Pryut & Kwakkel (2007) engages in a substantial discussion about the linkages between ethics, responsibility, and system dynamics, thus contending that ethics is always present in the process of modeling, whether it is the choosing of the specific methodology or the assigning boundaries of the problem at hand (Pryut & Kwakkel 2007, 2-3).

In short, our approach in the Task 4.1. and the following Deliverable 4.1 at hand followed in the methodological and normative footsteps provided by systems thinking, collaborative group modeling and causal loop diagrams.

2.2 Causal loop diagrams

Causal loop diagrams (***CLD***) are a qualitative modeling method falling under the umbrella of systems analysis or systems thinking methods (Williams and Hummelbrunner 2011). They can be used as first steps in building system dynamics simulation models, or to communicate the key dynamics of system dynamics models, though they can also be used independently. CLDs consist of variables connected by positive and negative causalities. All variables are represented in a form where they can intuitively increase or decrease, though any real-world ability to measure them is

not necessary. CLDs thus contain information about the direction of change under different circumstances.

Crucially, CLDs organize causal links in a way that completes ***feedback loops***. ***Feedback loops*** fall under ***positive*** or ***reinforcing loops*** and ***negative*** or ***balancing loops***. In a ***reinforcing loop***, the direction of change of all variables in the loop is maintained or accelerated. In a ***balancing loop***, any initial direction of change is undermined and eventually stops. When CLDs contains predominantly feedback loops of one type, they can be easy to mentally simulate. When the two types of feedback combine, change in the variables of the system becomes more difficult to predict without computational simulation. In either case, what makes CLDs distinct from many other methods is their ability to explain phenomena ***endogenously***.

Endogenous explanation is a technical way of saying that systems have a 'life of their own' and cause their own change, as highlighted by Meadows above. In an ***endogenous mechanistic process***, all causes are accounted for as the effects of something else. In fact, in an endogenous causal chain all variables are both causes and effects, in contrast to ***exogenous explanations*** which rely on an outside cause. Relying an exogenous cause effectively scopes it outside of things to be explained. While the complete chain of the causes of causes can be long, if these causes do not connect back to the system being analyzed, their origin remains unexplained.

In principle, the choice of analyzing a phenomenon as ***endogenous*** or ***exogenous*** is a matter of research perspective. Pro-responsibility attitudes among research staff can be viewed as an exogenous influence towards RRI implementation. Alternatively, we view pro-responsibility attitudes and RRI implementation as causing one another.

Feedback and endogeneity can be identified in virtually all areas of social life and institutions if one views the issue with an appropriate scope. However, some problem scopes that allow an endogenous explanation can be unintuitive or weak. This is particularly the case when the causal variable is a very high-level phenomenon, e.g. EU policy, and the affected variable is a small-scale phenomenon, e.g. the grass-roots RRI initiative in a single organization. Here it may not be easy to view EU policy as strongly following from RRI initiatives, and an exogenous explanation of RRI initiatives is advantageous.

Nonetheless, successful identification of endogenous explanations and intentional problem scoping to allow for endogenous explanation also feature advantages. An endogenous explanation is more complete in the sense that all variables are consequences of other variables, and no fundamental cause is left unexplained. An endogenous explanation also contains a framework of change drivers and its limits in the form of reinforcing and balancing feedbacks. Once feedback loops are determined, it can be intuitive to identify leverage points for high-impact interventions. For example, we may consider whether balancing feedback that slows or prevents desirable change could be alleviated. Alternatively, we could propose targeting pro-responsibility interventions in those parts of the causal chain where they can trigger desirable reinforcing feedbacks.

The content of a CLD model may be uncertain. For instance, a mechanism may be somewhat speculative, the relative strengths of reinforcing and balancing feedback loops can be uncertain, or the structure of the causal chains itself can be contested or subject to change over time. Identification of these uncertainties are also useful for the research process, as it helps narrow down the precise issues that would need to be clarified to make stronger statements about change. Overall, CLDs can help articulate the nature of the problem from a systems perspective. In a participatory process this is often done over the course of multiple workshops and model iterations (Vennix 1996).

2.3 Participatory modeling

Typically, participation aims to include the knowledge held by lay persons or non-scientific experts. Participants may for instance validate model parameters or scope or they may evaluate the outputs of the model (Voinov et al. 2016). Participation can be particularly useful for problems where scientific knowledge is uncertain or there is high disagreement on knowledge or aims (Basco-Carrera et al. 2017).

Our modeling process may be described as participatory in the sense that we gathered inputs to a pre-defined research question from a panel of experts who themselves were not directly responsible for modeling that question. It is notable that our participants were themselves researchers often with an interest in RRI. The line between co-researcher and participant is therefore somewhat blurred. This issue of blurred categories has been recognized before in the philosophical literature on research in uncertain contexts (e.g. Funtowicz and Ravetz 1993). We argue that the issue here is rather that in uncertain research contexts, participants (whether “lay persons” or researchers) belong in a community of experts (or an “extended peer community”, ibid.), and that the profession of the participants is not key to deciding on methodological labelling. In any case, besides their research expertise, our participants also held the kinds of practical, tacit and contextual knowledge of their organizations that characterize “lay person” participation in participatory methods.

We highlight three benefits of a participatory modeling approach that concern this work. First, it helps dealing with uncertainty regarding the exact nature of RRI institutionalization. Several accounts of key dynamics can be reasonable, and a higher quality synthesis is more likely when several perspectives are shared and space for dialogue is given. Second, our participatory method effectively generated data of RRI institutionalization, as stemming from the perspectives and experiences of researchers. The key difference to simply conducting a survey is that participants also iteratively synthesized ideas concerning the issue into CLDs, or theories of organizational change, together with the modelers. Third, the participatory modeling process is often considered an opportunity for social learning, where also participants can better understand an issue that they have interest in. This was particularly attractive given that participants were themselves agents attempting to promote RRI institutionalization.

2.4 Collaborative system modeling as a tool to promote RRI implementation

Models are simplified representations of the real world. They can be valuable if they are ***useful*** in solving some concrete problem, for instance regarding organizational change. Here, we offer guidance for qualitative system modeling that can be carried out with little prior experience or digital tools. In essence, we present a broad heuristic for modeling RRI implementation and organizational change through collaborative and participatory system modeling approach.

When aiming at organizational change, it is crucial to build models in a collaborative way (Vennix 1996), as opposed to assigning a modeler to build a model and present it only once its finished. Regarding RRI implementation and organizational change, modeling can have for instance the following aims:

- *facilitate dialogue between different perspectives, both during the modeling process itself as well as by utilizing the finished model*
- *crystallize and articulate what is important*
- *articulate an explicit theory or a framework of how desired organizational change can be achieved*
- *explain why desired organizational change is currently failing*

Influential systems modeler John Sterman argues there are no recipes to guarantee a useful model: “Modeling is inherently creative” (Sterman 2000, 87). To promote usefulness, we want to highlight collaborative systems modeling as a ***mindset***. With the correct mindset, a useful model can emerge as a byproduct of concrete problem-solving, and the model itself takes its correct place as a tool rather than an aim.

The following practical steps are loosely based on Sterman (2000) and Vennix (1996). For each practical step, we offer additional tips and guiding questions that should help maintain the right mindset. Perhaps most importantly, the modeling process can and should **go back and forth between steps** over time.

1. **Problem articulation**: You should always model a problem. Your view and articulation of the problem can change, perhaps induced by this process itself, but one should always have a practical aim in mind. This requires asking questions such as: *What are we trying to achieve? Are different stakeholders trying to achieve different things? Does our modeling work so far suggest our initial problem articulation should be changed?*

2. **Identifying collaborators**: It is important to include relevant collaborators in the modeling process. Stakeholders inside and outside the organization often possess valuable (tacit) knowledge and perspectives of their own. More importantly, collaboration in problem-solving tends to create buy-in, which helps to increase the acceptance and desirability of the decisions and actions that can result in part from the modeling. Thus, it is important to ask questions such as: *Who has knowledge about this topic? Who is affected by our possible actions? Might someone disagree with our problem articulation, or the implications of our work so far? Can we invite those people to argue their case?*

3. **Identification of causalities and feedbacks**: This is the model building part. Think of organizational change as an *organism* or a *machine*, where one thing affects another. Begin by identifying the elements that you want to influence, following your problem articulation. Try to think of several variables that ***cause change*** in those elements. Also list some ***effects*** of the problem. Make sure to include uncertain or controversial causes and effects, which are then scrutinized before moving forward. Look for cases where ***effects influence causes***. When you form these loops, everything in the chain is a cause *and* effect! These causal loops or feedback loops can accelerate (or prevent) change, and there may lie keys to success (or failure). The following questions can help the group model building process further: *Can we question the causalities in the model? What is missing? Do multiple* causes *reinforce each other? What appears to be the actual problem? Can these things be better understood using different words or different sets of parts?*

4. **Testing the model**: The causal chains you have drawn are effectively ***theories of change***. The model should be examined step by step to make sure that the narrative makes sense. It is important to be honest in terms of whether or not the model feels useful. If you need to start over, the prior models were a necessary step to get this far. The model is not as important as the actual process of problem-solving is. ***The model is finished when it can inform the problem you want to solve, and you don't think returning to the prior steps can make it better*.** Other useful questions for testing the model include: *Do you and the collaborators find this model useful? Does the model suggest an improved problem articulation? Are there leaps of logic in the causalities? Have you or the collaborators learned anything so far?*

Regarding the actual tools for drawing the models, pen and paper or a whiteboard are the easiest to start with. Their only drawback is the lack of editing potential when the model grows in complexity through various iterations. A specialized software tool for drawing system models such as **Vensim**, which offers a free version Vensim PLE[2], are easier to edit and iterate. However, learning how to best use Vensim effectively can take a few attempts. You can also try the arrow drawing tools in **PowerPoint** or an online workshop tool like **Miro**[3], which may be easier but less visually appealing.

[2] https://vensim.com/
[3] https://miro.com/

3-GROUP MODEL BUILDING PROCESS

3.1 Selection of the Lab use cases

At the centre of the Co-Change project are eight Co-Change Labs (***Fig.1***). These Labs have different goals, although all of them are focused on institutionalising RRI related practices. For the purposes of our task 4.1, we chose three specific Lab use cases, ***VTT***; ***AIT***; and ***Tecnalia***[4]***.***

The selection of these three Labs for the modeling process was done on the basis of ***suitability***, ***commensurability*** and ***maturity*** of the Labs. ***Suitability*** was assessed in terms of resources available for participatory group modeling work, which required commitment in the course of about six months. ***Maturity*** refers to the phase of the Lab in terms of its development and the readiness to present initial results and ideas about the process of RRI implementation and institutionalization in the Lab case. Finally, ***commensurability*** refers to the similarity between the organizations, which makes it easier to compare, contrast and synthetize the results of the three different Labs. All the Labs we chose for the modeling process were Research Performing Organizations, which enable the commensurability, even though the Labs had major differences in terms of substance.

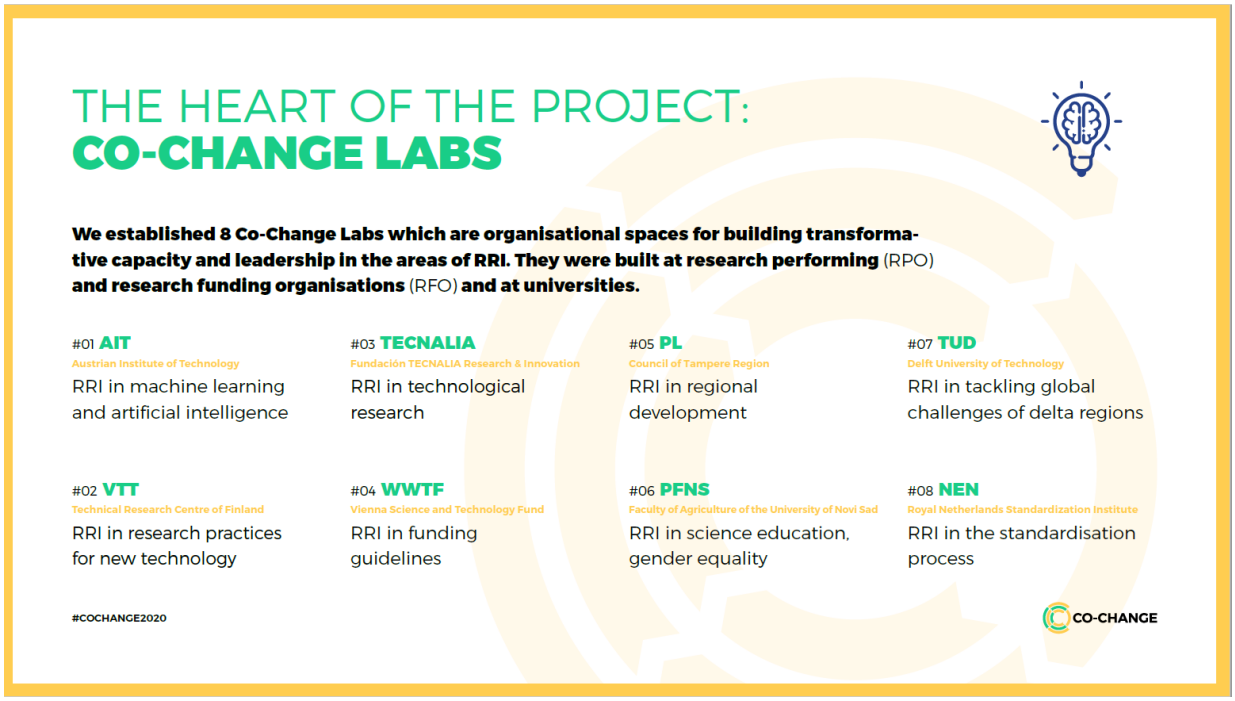

Figure 1: Co-Change Lab overview

[4] We ended up selecting three organizations instead of the initial estimation of four to five organizations because, in contrast to the other Co-Change partner organizations, the three selected Labs were relatively similar, technically oriented, leading European research performing organizations (RPOs), which created important commensurability between them.

3.2 The process of group model building

We modelled the dynamics of RRI implementation qualitatively using causal loop diagrams (CLD). The CLDs represent key dynamics in each use case and common dynamics shared by all or most cases. We drew the case-specific CLDs based on three to four 1-hour discussions with research staff from the case organizations. Each session was attended by 2-4 research staff members. The sessions were based on guiding questions concerning the drivers and barriers of RRI implementation. We interpreted the discussions from the perspective of endogeneity, i.e. by seeking to complete feedback loops that explain RRI implementation and institutionalization. We validated all CLDs with the session participants, leading to the acceptance of some and rejection of other modeler interpretations. The participants considered the end results representative of their perspectives and experiences. The CLDs representing shared aspects across cases were synthesized without a separate participatory discussion, though they were validated with the same group of participants[5].

In the modeling workshops, discussion was initiated by questions regarding the drivers and barriers of RRI institutionalization in the specific case. These separate elements were then connected with causal links based on the discussions. Much of the model building took place during discussions. In-between meetings we sought to simplify and clean up intermediate model versions without sacrificing key information. We intentionally sought to divide individual key ideas into separate smaller models since large models can work against our aim of communicating clearly and concisely. These cleaner versions were validated with workshops participants.

Figs. 2-4 show examples of unfinished draft models from each of the three cases. We present these to illustrate the iterative nature of the collaborative modeling process where new ideas can be brought forward despite uncertainty regarding their operationalization. Conversely, initially promising ideas can also be abandoned upon better understanding of the whole system. In the draft figures some of the causal arrows are colored grey to label them as tentative or uncertain, pending more careful reflection. Many variables have been left unconnected at first, as they emerged in discussion but their precise role in the system was still to be articulated.

In the ***Tecnalia*** case (***Fig. 2***), the objective is to systematically introduce and implement RRI principles within Tecnalia through the Shape Lab. The use case discussed how the system driving RRI institutionalization may be different before and after an organizational restructuring. In the draft, a wide variety of issues are visible ranging from client demand to formal project requirements and researcher habits. These were later scrutinized further, and key dynamics were separated and simplified.

The case of ***VTT*** (***Fig. 3***), looks at ***carbon handprint*** as a specific case of RRI implementation as a part of VTT's internal Sustainability Programme. In the use case discussions, relatively few endogenous dynamics were identified overall, though many individual drivers and barriers to RRI institutionalization were discussed. The draft

[5] In a workshop on June 28th 2022 we validated the models with Lab participants and selected project participants.

illustrates an attempt to try different hypothetical endogenous dynamics that would be scrutinized with the group and later, for the most part, abandoned.

The ***AIT*** case (***Fig. 4***) focuses on the risks and benefits related to machine learning and artificial intelligence in terms of socioeconomic sustainability. The case study generated surprisingly many realistic causal loops rather quickly. However, as discussions progressed, it became more clear which dynamics could be abandoned. For instance, while the dynamic related to the fact that the pool of potential lab participants reduces as people participate in the lab (they cease being a potential future source of lab growth) may be correct, it became a less important dynamic as the discussions progressed. It was emphasized that a small but diverse group can create critical mass for change, and for this reason the lab did not even attempt to reach a maximal amount of people. As a result, the dynamics concerning the number of potential participants got de-emphasized in the modeling process.

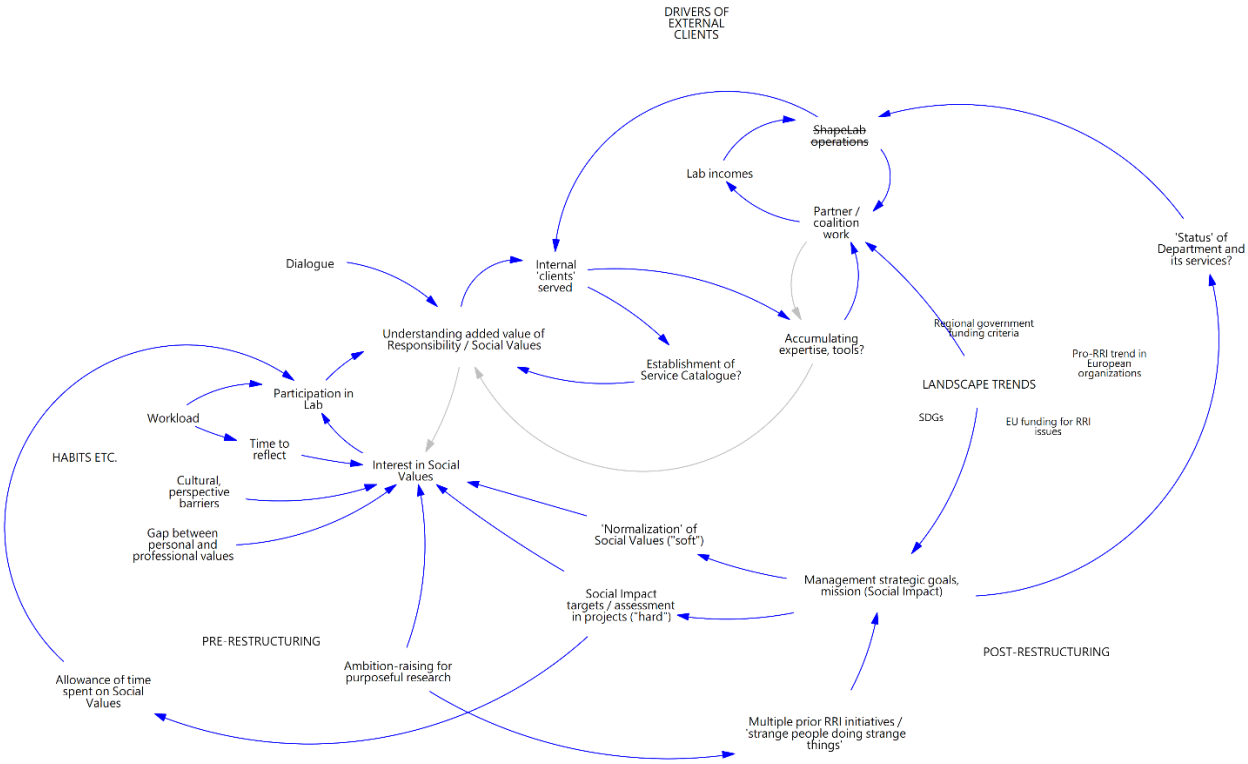

Figure 2: Tecnalia use case draft model

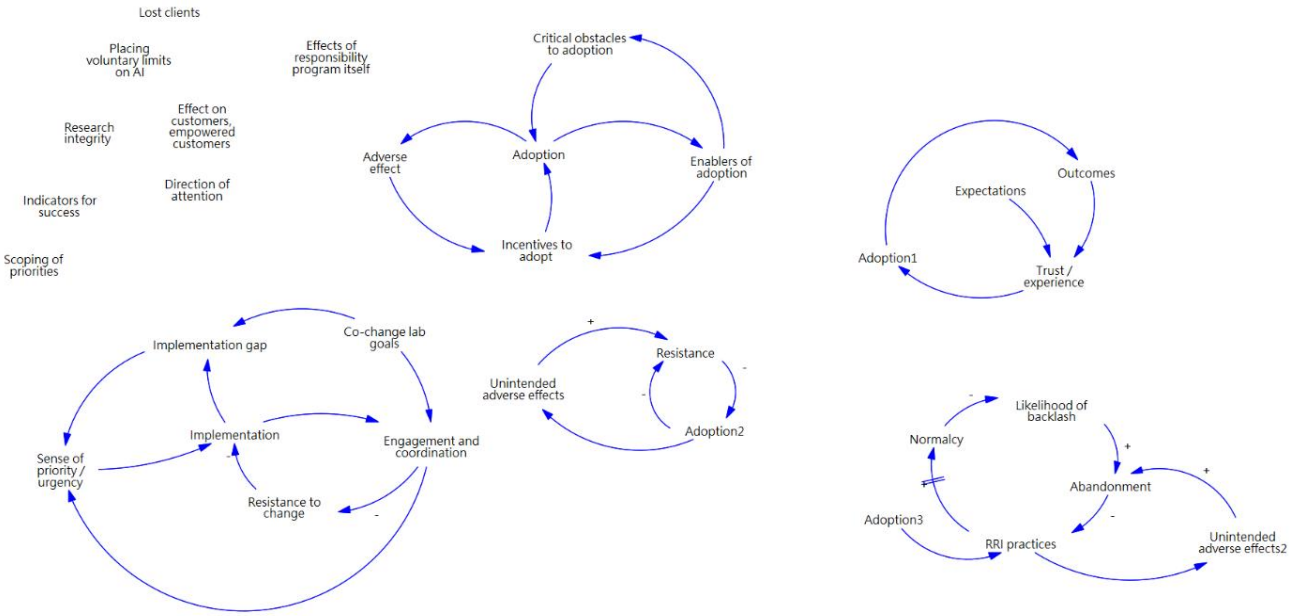

Figure 3: VTT use case draft model

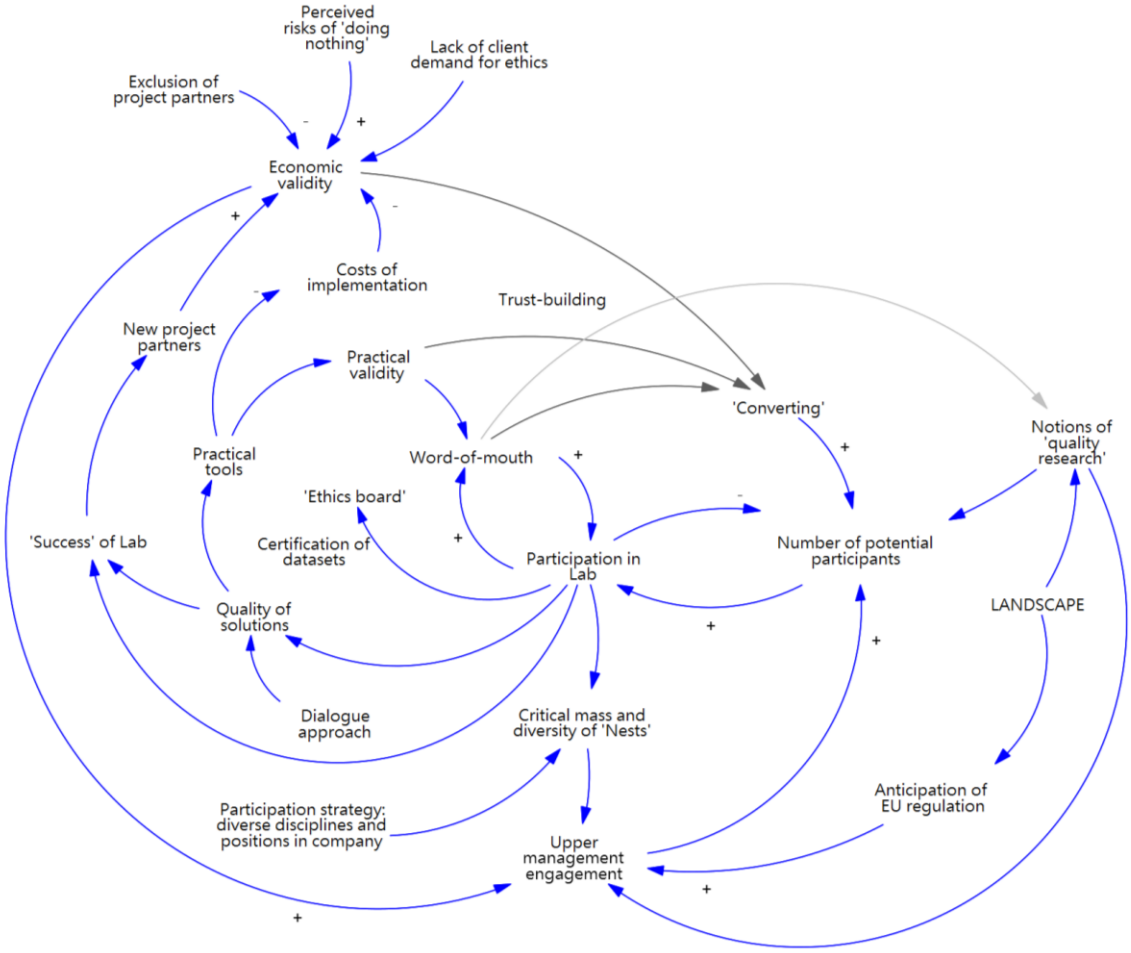

Figure 4: AIT use case draft model

3.3 Lab use cases

3.3.1 Tecnalia Lab: Shape Lab – Setting up an Internal RRI Working Group

The vision of Tecnalia's Shape Lab is to systematically introduce and implement RRI principles in the institutional discourse and the different operational units of Tecnalia, which will contribute to the development of practices, policies, and the understanding of this concept. Shape Lab also aims to promote the uptake of social and ethical elements in products and services from an early stage of the development.

The Shape Lab itself consists of a group of experts in social aspects of technology that provides integral RRI support to researchers and their working groups. Accordingly, the idea is to include RRI aspects in research and development projects (including the societal perspective) and facilitate the practical solutions. The key lies in supporting them in the use of participatory methods, educating in science, making the process and its results open, transparent, and accessible, while having an ethical attitude and promoting equal opportunities, for example through a gender plan. Tecnalia Lab team proposes an organisational solution on how to create an RRI institutional discourse and service in line with institutional/structural change.

The practical RRI concept in the case of Tecnalia was an internal service that researchers can utilize to ensure responsible practices and outcomes. The service is not forced onto any project, rather projects will voluntarily approach the service to meet their needs. Some of the identified institutionalization dynamics concerned demand creation (***Fig. 5***). Providing the service generates its own demand by expanding visibility of the service and promoting expertise within the service through experience. Since the service operates based on expertise networks (e.g. third sector organizations) outside Tecnalia, the ability to respond to given needs relies on forming those relationships. As services are provided, broader networks are established and the readily available service portfolio expands.

A recognized barrier for RRI is that it is often not very well known within the organization, which creates inertia for the implementation of RRI practices. As a response, the Department coordinating the Shape Lab activities aims to increase the visibility and awareness of the RRI-related social values work. The Shape Lab seeks to raise awareness on how researchers can connect their research towards providing sustainable and responsible solutions to societal challenges. This awareness raising is an important part of demand-driven process of RRI institutionalization.

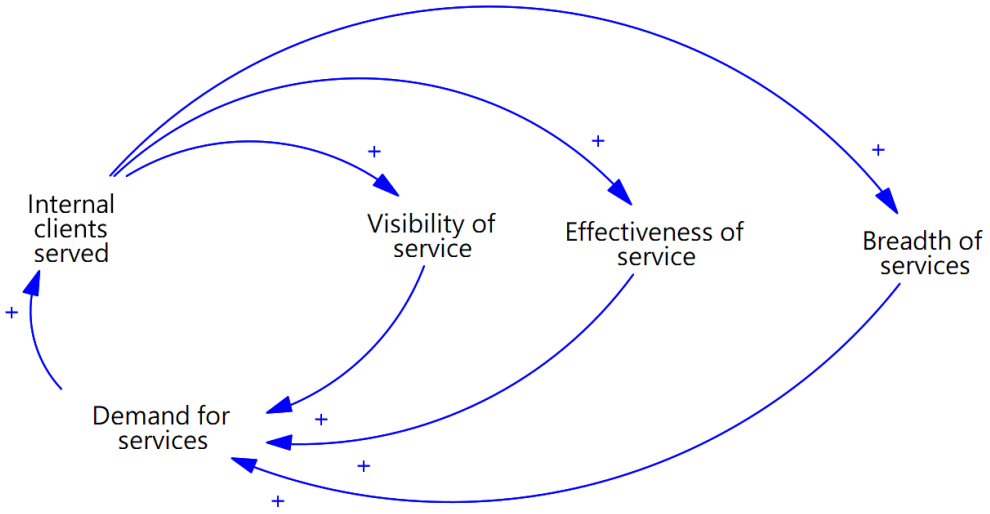

Figure 5: dynamics of institutionalization

Demand for services is also promoted by management strategy (***Fig. 6***). Explicit social targets have effect through a 'soft' channel, representing normalization and cultural change, and a 'hard' channel, representing formal impact targets and time allocation for considering social values. Management strategy can also promote the status of the POINT, which provides the social-values services (SV-services).

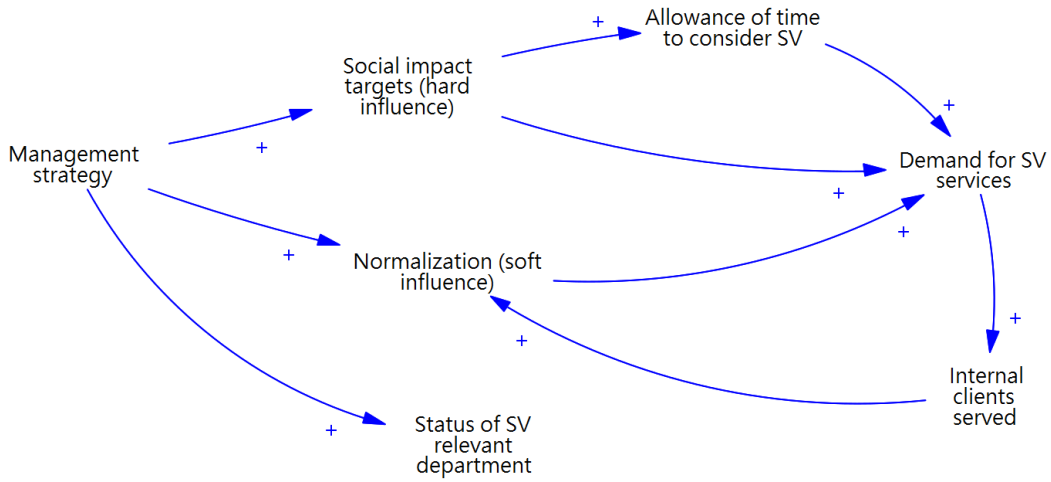

Figure 6: management strategy

Dynamics that work outside of Tecnalia can also be speculated (***Fig. 7***). These did not stem from empirical experience to the same degree as the previous models but represent an option for thinking about top-down steering (by Tecnalia management, or by public decision-makers) also as an endogenous process.

EU governance promoting responsible research can make Tecnalia more "competitive" and socially oriented as far as it has pre-emptively built capacity and experience for such practices. Management will steer company strategy in response to these competitive advantages and reinforce investments in social values work. As

Tecnalia, and the wider 'RRI community', have established capacity for responsible research practices, RTOs/RPOs overall will become a force for steering the EU agenda further in this direction. The proofs of good practice that follow responsible research projects also spread across organizations and further politically validate a responsible research agenda. In the model, some exogenous influences of EU steering are also mentioned – the inheritance from past RRI agendas, whose effects were lackluster, and demands from broader society outside only RTOs/RPOs.

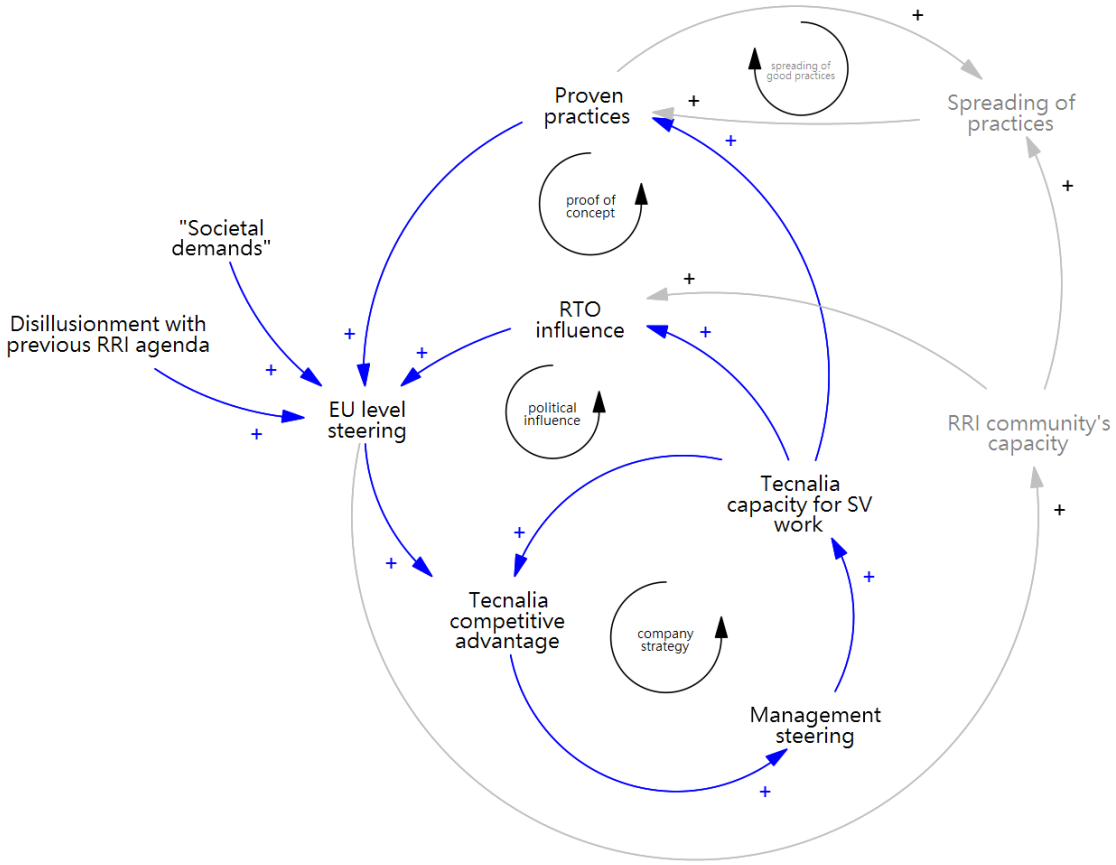

Figure 7: wider view of RRI ecosystem

3.3.2 VTT Technical Research Centre of Finland Ltd. (VTT) Lab: Creating standardised practices and defining core values for new technology (Carbon Handprint Case)

VTT Lab focuses on developing an internal Sustainability Programme. Development of the programme started in 2021 and it is designed on inclusive and participatory approach that engages different internal stakeholders from researchers to top management. The framework of the sustainability programme builds on four pillars: Sustainable foundation; Thriving professionals; Empowered customers and Resilient Society. In the systems modeling exercise, we looked at ***carbon handprint*** as a specific case of RRI implementation within VTT. ***Carbon handprint*** is a part of VTT's broader internal Sustainability Programme.

As a response to the global challenge of climate change, ***carbon footprint calculation*** has become the standard method for estimating the environmental impacts and Green House Gas (GHG) emissions of a product or a service through its life cycle. However, in contrast to the ***negative environmental footprint***, VTT together with LUT University have developed a new methodological approach for quantifying the ***positive environmental impacts***, ***the carbon handprint***, of products and services. According to VTT and LUT University, "*A **handprint** refers to the beneficial environmental impacts that organisations can achieve and communicate by offering products and services that reduce the footprints of others*." (Pajula et al., 2021)

Handprint is achieved not by reducing one's own footprint, but by improving the performance of others. The calculation and communication of the positive environmental impact of a product or a service is at the heart of the handprint approach. In essence, carbon handprint seeks shift the thinking "from negative to positive", "from producers to reducers of emissions and resource use". (Pajula et al., 2021)

Thus, the VTT use case revolved around the carbon handprint as an impact assessment method offered for customers. In practice, the carbon handprint is realized through project work. The idea of mapping and evaluating impacts in the beginning or before a project aligns well to the RRI principles of reflectiveness and anticipation.

The key drivers of the practice were largely exogenous according to our modeling workshops, stemming from public steering and customer expectations. The driving factors for Carbon handprint include combatting climate change; anticipatory compliance with future regulations; and the positive effects of voluntary responsibility practices in terms of branding.

Additionally, the barriers to implementation largely concerned the technical difficulty of conducting the impact assessment and related calculations. Moreover, it is difficult to generalize assessment methods and often incommensurable data across project cases. It is difficult to scale impact assessment without also scaling resource requirements.

Nonetheless, within the prior exogeneities, some reinforcing effects were suggested. A key effect of impact assessment is to steer client selection and project content (***Fig. 8***). As clientele and project portfolios thus improve in terms of their GHG impacts, a positive reputation effect can be expected for the company. Such a reputation effect can be thought to validate the practice, promoting its future use. In the future, Carbon Handprint might become a part of guiding criteria and indicators for selecting customer projects.

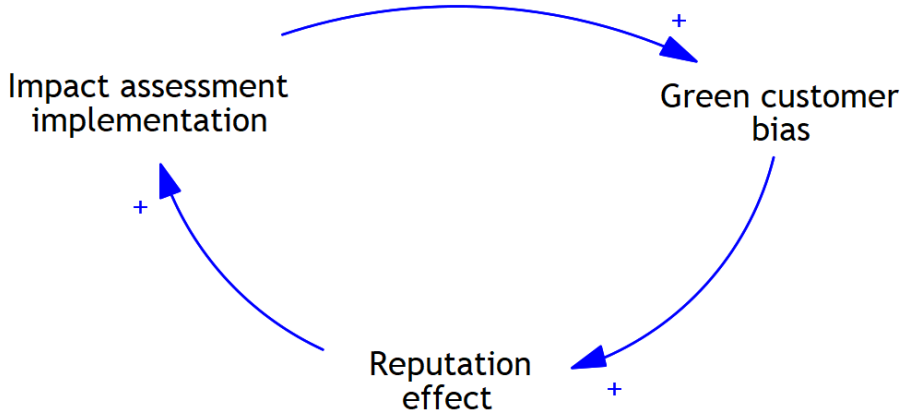

Figure 8: effects of impact assessment implementation

Some level of learning effect was also suggested (***Fig. 9***), within the confines of technical and resource limitations. As the impact assessment method is implemented, or even planned, experience tends to grow and subsequent implementations can become easier.

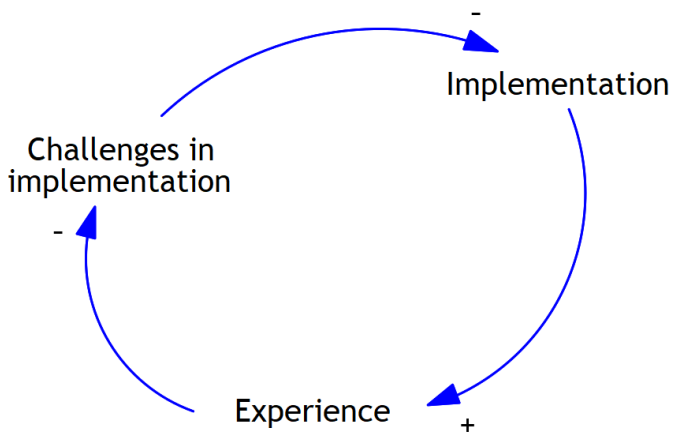

Figure 9: effects of experience to implementation

3.3.3 Austrian Institute of Technology (AIT) Lab: Establishing an Ethics Advisory Service for Artificial Intelligence (AIT AI Ethics Lab)

The Co-Change Lab of the Center for Innovation Systems and Policy of AIT focuses on the challenges that the application of artificial intelligence (AI) brings to light. The vision of the Lab is for the AI technologies to contribute to social, economic, and ecological sustainability.

Accordingly, the motivation for the establishment of the AI Ethics Lab was to understand the background of these technologies and the way that ethical problems of artificial intelligence are born. The different ways of thinking about these technologies are being investigated in the Lab by IT researchers and social scientists. The AI Ethics Lab consists of a small core team of around six committed persons, who have an interest in AI ethics broadly. Lab members are also working on raising awareness on the challenges and changing practices regarding AI at their own institute and beyond, reaching out to other organisations and ministries. The AIT Co-Change

Lab aims to find useful practices to make this technology more human-centred and less dangerous. Responsible innovation should support privacy and self-determination of humans and keep society in control over these new technologies.

Researcher engagement with AI ethics is currently constrained by a lack of time and incentive to reflect on ethical issues, though some researchers are in fact engaged due to clear ethical dimensions they recognize in their own work (***Fig. 10***). As there is simultaneously a lack of top-down steering in favor of considering AI ethics, there is risk of little change occurring. Dialogue across disciplines, departments and levels of the company can be an effective strategy in harnessing the initial pool of engaged researchers. The theory goes that this pool can form a critical mass if it is diverse in these dimensions. With a critical mass, management may begin prioritizing AI ethics issues, resulting perhaps in formal requirements for project practices. The model can also be read as an explanation of a lack of change: if the exogenous elements, such as EU pressure is lacking, then the reinforcing loop does not emerge.

The AI Ethics Lab also recognized the importance of "selling" their activities to the upper management by also discussing potential business case with AI ethics as well the importance of addressing risks related to AI development. Another important driver for AI ethics is the anticipation of potential EU regulation (AI Act for example), which might mandate certain regulatory compliance in the future.

However, a lack of demand from customers in terms of AI ethics was noted as a barrier for RRI, since the lack of demand creates challenges for resource allocation. As a response, AI ethics could be included as a standard or a modular part of customer offering. One potential avenue for customers is the creation of a responsible "fair trade" label or certification of AI ethics compliance within AIT, or even a broader European level of certification within the relevant ecosystem. This voluntary institutional compliance can help to create a brand of responsibility vis-à-vis the potential customers, thus establishing a first-mover advantage in the field of responsibility and sustainability.

The AI Ethics Lab was seen to be potentially functioning as a bottom-up driven "proto ethics board" (***Fig. 11***) for AI issues, consisting of a combination of social scientists and data scientists, who would interact and provide feedback to each other. This informal ethics board could provide a positive service by providing non-binding guidance and recommendations about AI ethics related issues, without distinct management involvement. Instead of a passive reviewer, the board would also function as a pro-active awareness-raising actor that would help create additional demand for AI ethics inside AIT by educating researchers about the potential issues related to AI. Another important facet of the AI ethics Lab is to function as a "safe-space", a format for "worry-free" discussion around AI ethics.

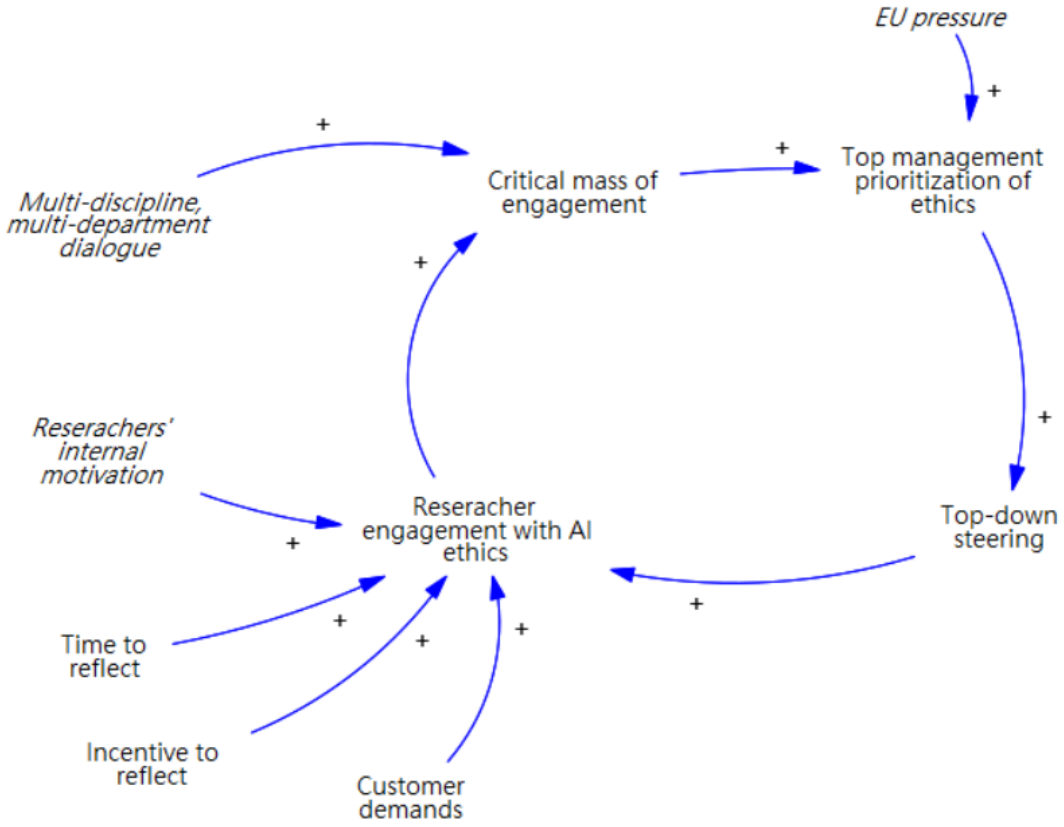

Figure 10: dynamics of researcher engagement with AI ethics

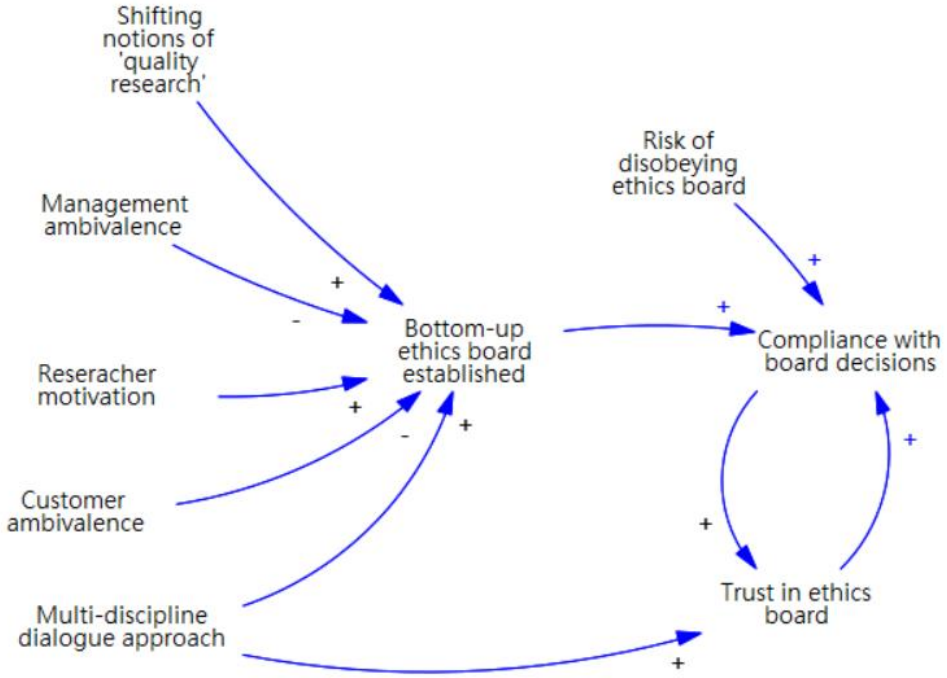

Figure 11: dynamics of the bottom-up ethics board

A lack of customer and management demands naturally hinders the creation of any type of ethics board. On the other hand, the internal motivation of some researchers and a shifting culture around what constitutes 'quality research' can promote the establishment of an ethics board, particularly if interest and engagement spans across departments and disciplines. The opinions of a bottom-up ethics board would not be binding in nature. However, researchers would have trust in the institution because of its origin in multi-disciplinary dialogue. Over time, trust can build further. Even though decisions are not binding, it would be likely that project leads follow board opinions, because a prior expression of concern from the board raises the stakes for project leads in the possible event that ethical outcomes are not achieved.

The quality of AI ethics work is dependent on active engagement by researchers (content of the work, practical and contextual knowledge), project and department leaders (implementation of good practices) and company management (formal requirements and norms) (***Fig. 12***). The quality of AI ethics work can itself drive normalization of such considerations, as good practices and tools serve as proof of concept and momentum. Better and more practical means for considering AI ethics also helps drive demand for their inclusion in projects. Demand creation can be a factor in getting project leads and management engaged in implementing and formalizing AI ethics work.

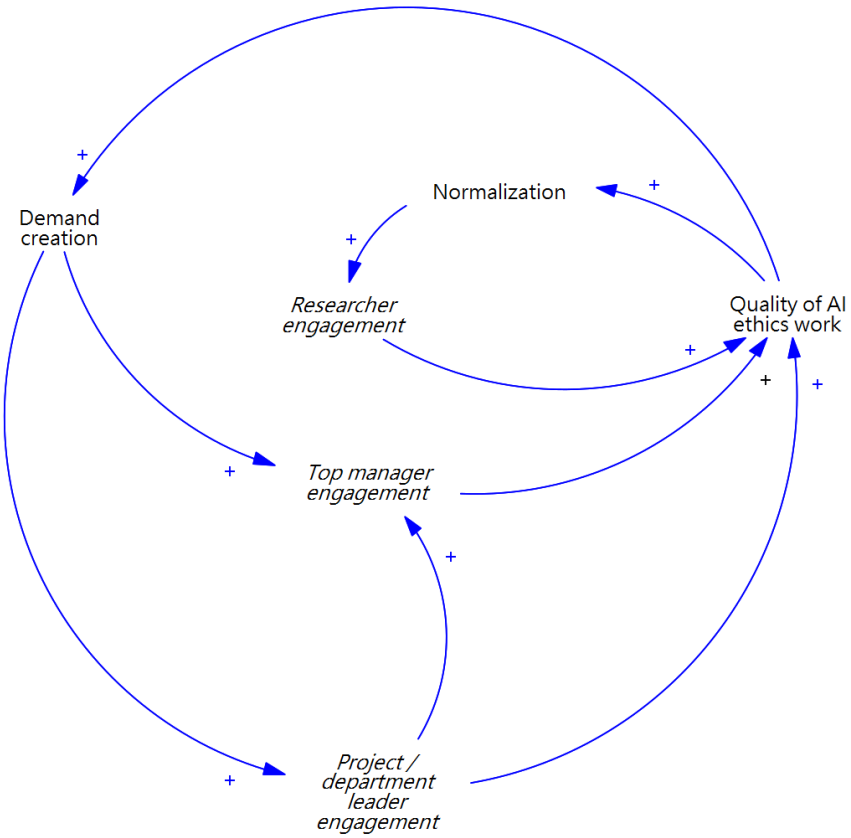

Figure 12: engagement dynamics

3.4 Synthesis of the individual Lab results

While a variety of drivers and barriers to RRI institutionalization were identified in each of the Lab use cases, all feedback loops were reinforcing in nature, and none were balancing. The implication is that no self-defeating effects of RRI institutionalization were identified. To the extent that there are barriers to RRI institutionalization, they are exogenous and could not be connected to RRI practices or institutions themselves. Examples of such barriers could have been social backlash effects or disillusionment over time. While these did not come up in discussions, we recommend being mindful of their possibility, as unrecognized balancing feedback can be a key factor preventing desired outcomes despite otherwise effective interventions.

The basic principle common to reinforcing effects across cases is that RRI activities can generate their own within-company drivers and enablers (***Fig. 13***).

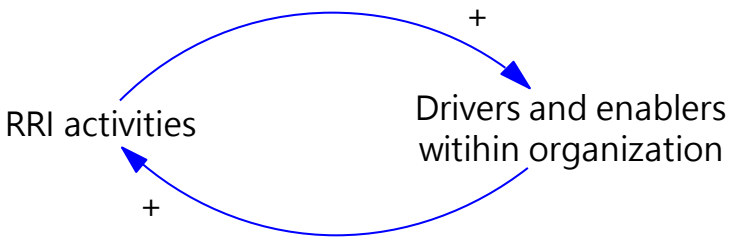

Figure 13: reinforcing effects

Disaggregating the basic reinforcing dynamic of ***Fig. 13*** into some common constituent parts yields the model in ***Fig. 14***. Most of these reinforcing effects were shared across cases, with the exception of visibility, which was mainly emphasized in the Tecnalia case. A visibility effect does not however appear to contradict any of the learnings from other cases, and it may be an intuitive dynamic to anticipate in RRI implementation generally. Other reinforcing dynamics include normalization, or RRI practices becoming the new typical activity and expectation in professional culture. RRI initiatives can also compound experience and capacities, making them more effective over time and thus more desirable to implement. Finally, initial bottom-up RRI initiatives may be required to engage upper management, rather than the other way around, while eventual management engagement can lead to formal rules, expectations, time allocations and workplace norms in favor of RRI activities.

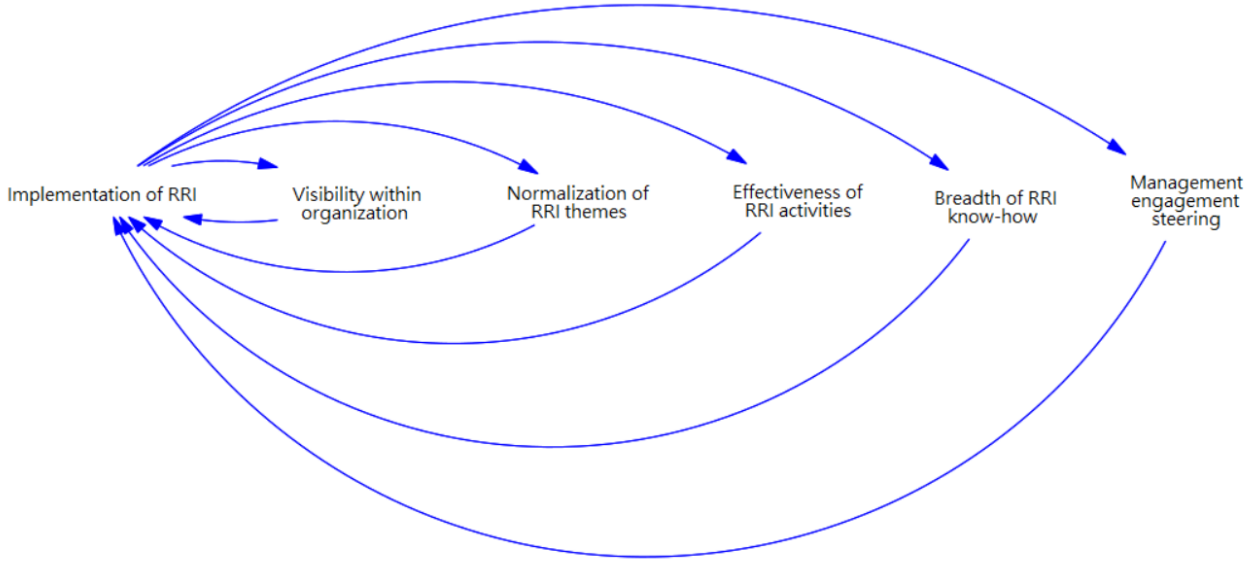

Figure 14: reinforcing effects in the implementation of RRI (all causalities are positive)

In each case we asked whether RRI dynamics can encompass wider society outside individual ***Research Performing Organizations*** (***RPOs***), or ***Research and Technology Organizations (RTOs)***. These effects were less emphasized by participants and for example EU regulation was understood as primarily exogenous. However, the possibility for endogenous societal dynamics was also not explicitly rejected in any case study. In ***Fig. 15*** we present a suggestion for how to potentially view RRI implementation in RPOs as a broader societal dynamic. RRI implementation in many RPOs can be part of a wider societal normalization trend promoting expectation for responsible research practices and recognition of responsibility issues. Together with the practical proof of concept that RPOs provide with their activities, governance bodies may be encouraged to steer their agenda in a pro-RRI direction. While this perspective relegates initiatives in *individual* RPOs to a lower level of effect, invisible in the model, it places individual RPOs into broader societal and ecosystem context. Thus, this perspective offers the possibility to understand the broader change dynamic across society which steers the operational environments and strategies of RPOs as a sector.

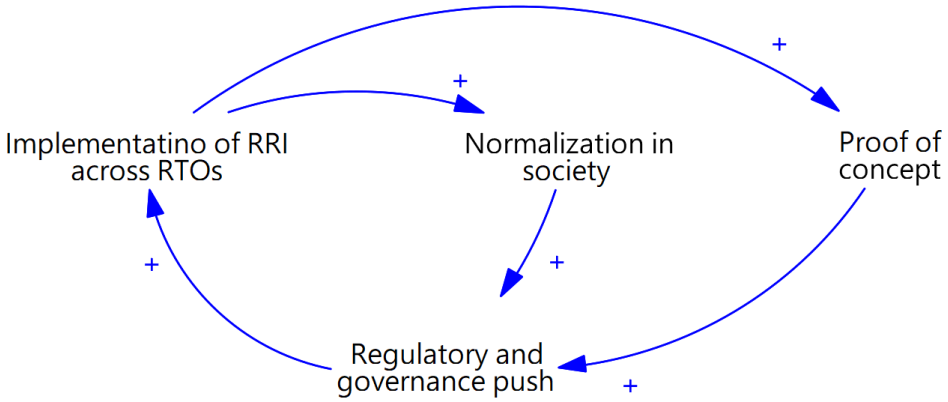

Figure 15: societal view of RRI implementation in RPOs

4- DISCUSSION

In this section we analyze the implication of our work from the point of view of organizational change theory and contrast our work with the drivers and barriers of RRI implementation, which were identified in the stocktaking report. Finally, we also discuss the viability of systems thinking methodologies as a tool for RRI institutionalization.

4.1 Theories of change

Throughout the Co-Change project (e.g. Tabares Gutierrez et al., 2020, 27), it has been argued that in order to conceptualize change, a theory of change must be established. According to the stocktaking report (ibid., 36), change agents such as institutional entrepreneurs require theoretical knowhow to successfully facilitate and implement RRI practices. Consequently, ***Organizational theory*** was identified as one of the five key pillars of the project.

Despite the emphasis on organizational theory, it was recognized that a theoretical framework of organizational change needs to be supported by context-specific empirical study. According to Tabares Gutierrez et al. (2020, 13), there is no one "universally best" mode for change, because change is always a "context-sensitive phenomenon". Following Tabares Gutierrez et al. (2020), we addressed organizational change by focusing on context-specific practices, procedures and norms of the Lab use cases. Our empirical approach relied on the ***group model building*** exercises and the utilization of ***causal loop diagrams***. As a result, we did not have an explicit or a particular theory for change, apart from the broad framework of ***systems thinking***.

Nonetheless, our mostly empirically based work did have some linkages with the earlier project deliverables working with a theory of change. With regards to organizational theory, the importance of ***the source of change*** had been noted already in the Stocktaking deliverable 1.1 (Tabares Gutierrez et al., 2020, 14). Similarly, in relation to the ***source of change***, the field of ***systems thinking*** emphasizes that systems can be driven either by external (***exogenous***) or internal (***endogenous***) forces (Meadows 2008, 2). As the result of our modeling process, we were able to produce ***endogenous explanations of change*** consisting of multiple individual drivers whose interconnections would have been ignored without a systems perspective. In short, it was possible to explain RRI institutionalization dynamics without reference to unexplained ***exogenous drivers***.

With regards to the ***transformative change theory*** (Wolfram 2016), as promoted by the deliverable 2.1 (Wagner & Wilhelmer 2020), we found linkages between our systems thinking approach and transformative change theory particularly with three ***Transformative Capacities (TC)***. These were "***Systems awareness and memory*** ***(TC4)***; "***Foresight*** ***(TC5)***; and "***Reflexivity and social learning*** ***(TC8)***. In terms of TC4, "***Systems awareness and memory***", deliverable 2.1 (Wagner & Wilhelmer 2020) argues that transformative change "presupposes awareness and understanding of the system dynamics and path dependencies that undermine sustainability" (ibid.). As a result, the implementation of RRI should be based on "institutional self-

assessments" and "collective analysis capabilities" (ibid.). In response to these needs, our tool is meant to help perform an institutional or an organizational self-assessment through the utilization of group modeling and causal loop diagrams.

Similarly, our systems thinking approach relates to "***Reflexivity and social learning***" (***TC8***) in that it enables social learning through reflection of one's own actions in an organizational context (Stroh 2015, 21). Since reflexivity is "inherently a key RRI process", our approach can assist in the creation of "diverse formal and informal reflexivity formats" (Wagner & Wilhelmer 2020). Moreover, since internal reflection often differs from an external perspective (Rilla et al., 2020, 28), the network of actors within the Co-Change project was utilized for the purpose of peer-learning, assessment and validation of the models that were created through our systems thinking and group model building approach in task 4.1.

In terms of ***Foresight*** (TC5), we argue that systems thinking can help to create dialogue and social learning on change dynamics of various policies and plans as well as their future impacts. Systems thinking -based approaches can create transformative knowledge on barriers and drivers of RRI through institutional self-assessments (TC4) along with discussions about the types of changes that are desired by the organization. Importantly, systems thinking and system dynamics can also help to create preconditions for a key RRI principle of anticipation, which can lead to the development of collective visions, simulations and alternative scenarios (Ruutu 2015, 31).

Along with the three previously mentioned transformative capacities (TC), other TC's have potential linkages to our systems thinking approach as well. For example, systems thinking methodologies such as group modeling can help to develop "***Multiform and inclusive forms of governance*** ***(TC1)***, which are essential for any participatory RRI process. As highlighted by (Wagner & Wilhelmer 2020), collective forms of group work with stakeholders can help to empower the elicitation of diverse contributions (Fraaije & Flipse, 2020). Similarly, systems thinking can also help develop "personal abilities that leverage collective energies and enable social learning" (Wagner & Wilhelmer 2020), which are required for the creation of polycentric and socially embedded "***Transformative leadership***" (***TC2***).

Moreover, our participatory group modeling work can contribute towards multi-stakeholder involvement, leading to more "***Empowered and autonomous communities of practice***" (***TC3***), by increasing the legitimacy and desirability of the work (Wagner & Wilhelmer 2020). Finally, systems thinking approaches can be combined with a more multi-level and ecosystem-based view towards "***Working across agency levels*** ***(TC9)***; and "***Working across political-administrative levels and geographical scales*** ***(TC10)***, even though these perspectives were not heavily emphasized in our modeling exercises.

4.2 Drivers and barriers of RRI

Notably, many important issues in terms of ***barriers to RRI*** that were highlighted in the Stocktaking report (Tabares Gutierrez et al., 2020, 22), did not explicitly come up in the group modeling sessions. These include issues such as "***diverging views of***

science and society relations*"; "*fear of the loss of scientific autonomy*"; "*the unpredictability of scientific enterprise". These examples of barriers are issues of great importance, but it seems that the practical orientation of modeling RRI implementation in terms of organizational dynamics meant that the emphasis on the intrinsic motivations and incentives of researchers was left more to the sidelines, even though they affect the dynamics of RRI more generally.

On the other hand, the barriers related to the "***Difficulty of applying RRI in practice***", were not prevalent in the workshops because most of the Labs had experience with RRI and many of the modelled processes included ***de facto RRI*** practices, which were already a part of the organizational culture. This is also the reason why the prevalent issue of "***Tendency to outsource RRI***", was not present in our modeling sessions.

Instead, the question of "***Lack of incentives***" and "***Insufficient resources and capacity for RRI***" were present in the modeling workshops. The implementation of RRI practices often suffers from a lack of resources, which results in a lack of awareness and understanding of RRI. Similarly, the time-consuming nature of RRI was brought up from the point of view of trying to "sell RRI" within an organization, especially when it conflicts with the incentives of the current scientific culture or with commercial interests. As a result, the "***Unclear added value of RRI***" for industries and businesses was also discussed briefly in some of the sessions.

In terms of the ***drivers for RRI*** implementation (Tabares Gutierrez et al., 2020, 22-23), the group modeling did not particularly touch upon important issues such as ***Responsible Scientists***; ***Pursuit of good society***; and the ***Alignment of Science and Society***. These broader trends towards a more reflexive science and the enhancement of democracy were implicitly present in the Lab use cases on a general level, but they were not part of the discussion about RRI implementation dynamics on an organizational level as such.

All the Lab use cases had their own linkages to specific ***RRI drivers***. **VTT's** Carbon Handprint case, which seeks to provide solutions to the challenge of sustainability, is a good example of the implementation of RRI practices related to the driver of "***Response to Societal Challenges***". **Tecnalia's** Lab case emphasized both "***Social Innovation***" and "***Participatory Science***", in which RRI acts as a participatory mechanism for inclusion and diversity in the field of research and innovation. "***Risk governance***", perceiving RRI as a way to broaden the governance, assessment and anticipation of the inherent uncertainties and risks of emerging technologies, was a particular driver for RRI in the **AIT** Lab use case focusing on AI ethics.

Finally, on an organizational level the driver of "***Social license to operate***", which emphasizes RRI as a way to enhance the competitiveness of products and services by aligning them with society and end-users, was an important facet of "selling RRI" in all the three Lab cases. Indeed, all the discussions in the group modeling sessions focused on issues related to the "selling of RRI" within an organization as well as the question of resources and recognition for RRI in terms of top management.

4.3 Systems thinking as a tool for RRI implementation and institutionalization

As a part of the task 4.1, we organized a validation session for the group model building work that we had conducted[6]. The participants in the validation session and preceding Lab-specific modeling sessions recognized and affirmed many of the factors related to RRI implementation visualized in the models. Factors such as normalization of RRI, organizational visibility and management engagement were seen as important factors in building a "critical mass" for RRI institutionalization. These notions are in line with the earlier project work, which highlights the importance of leadership commitment and support (Tabares Gutierrez et al., 2020).

Nonetheless, the feedback gathered through the validation session also emphasized that in terms of RRI institutionalization the models seem to lack an emphasis on the important factor of internal and intrinsic bottom-up motivation of individual agents for RRI practices. This was seen as a deficiency, since the personal perspectives of scientists and innovators in terms of responsibility, sustainability and ethical approaches are considered as a crucial factor in RRI implementation. It is likely that our methodological perspective, emphasizing organization-level causal chains, somewhat hindered closer attention to individual agency and motivation.

Moreover, it was also noted that our work can also be contrasted with earlier project insights. For one, the deliverable 1.2 focusing on the pillar of *Ecosystems* (Rilla et al., 2020, 2), emphasized that "*embedding and implementing RRI efficiently requires changes in the network of interlinked actors, i.e. in an ecosystem, instead of focusing only in the organization in the core of transformation*". Despite this emphasis on the importance of ecosystems perspective in terms of transformative change, our approach was more heavily focused on the organizational perspective.

The organizational focus was mainly due to practical reasons related to our framework of modeling drivers and barriers of RRI implementation in organizations. The natural starting point for any process of RRI implementation and institutionalization is organizational self-assessment. However, the organizational emphasis can be seen as a deficiency since the important role of the ecosystem view, which highlights the importance of systemic network relationships and interdependencies within which an organization is embedded into, is underemphasized (Rilla et al., 2020). Nonetheless, many of the organizational level use case models did look at ecosystem-level factors such as RPO-networks, EU-regulation, and 3rd sector partners as well as customers.

The task description described the development of plausible and alternative impact paths, which would help organizations to anticipate the potential impacts of their actions. These impacts were elaborated upon as a part of the modeling exercises, which aimed to gauge the potential dynamics and impacts of RRI implementation

[6] In the validation session on June 28th, we asked the Lab participants and selected project participants three questions: 1. *What factors do you recognize from the model? How could these factors manifest/occur in your own context? 2. What could be lacking from the model from point of view of RRI institutionalization? 3. What do you think about the utilization of systems modeling as a method of understanding and presenting the dynamics of RRI implementation & institutionalization? What might be its pros and cons?*

within the organizations. In essence, each of the models we created presents either a generalized or a Lab-specific impact pathway. Nonetheless, we discovered that systems modeling itself was not the most suitable tool for an explicit impact assessment as such, since it would have required additional and complementary approaches, explicitly oriented towards evaluation and foresight practices. The inclusion of these complementary practices was out of the scope of our investigation.

Constructive criticism aside, the utilization of systems thinking and group model building as a method of understanding the dynamics of RRI implementation and institutionalization was seen as a fruitful one. By developing visualized causal loop diagrams containing rich information, the group modeling approach can help to make implicit mental models and tacit assumptions about RRI implementation more explicit and thus more open to scrutiny and refinement. The creation of these visualized diagrams forces the modelers and the participants to think about the causalities of RRI uptake and to simplify the necessary steps required for the process of RRI implementation. In this sense, group modeling can be utilized as a tool to identify relevant leverage points and main issues related to RRI institutionalization. Moreover, the causal loop method allows understanding individual drivers and barriers as a single whole with momentum for change as opposed to isolated items that would not affect one another.

However, while the simplified visualization of change dynamics in an organizational system can create positive clarity on the complexity of RRI implementation, these simplified models can also obscure the inherent uncertainty and ambiguity of an organizational change process. Therefore, it is important to note that the visualized causal loop diagrams should not be interpreted as simple deterministic or prescriptive guides or pathways to RRI implementation. Instead, various uncertainties and different possibilities within the implementation process should be emphasized and discussed.

Indeed, many participants saw the group model building process itself as well as its discussions and self-reflections about dynamics and causalities of change as more important than the causal models we created. This importance on the modeling process itself is often emphasized in the relevant literature as well (e.g. Sterman 2000; Vennix 1996). The causal loop diagrams, or models, can be seen as ***boundary objects*** (Voinov et al. 2018), which serve as a basis for discussion, deliberation, and dialogue about RRI implementation between different actors. ***Boundary objects*** are artefacts with varying meanings in and between different social worlds (Star & Griesemer, 1989, 393), which nonetheless enable cooperation between heterogenous actors by linking them together (Haim & Zamith 2019, 80-82).

The importance of the group modeling process itself was also based on the emphasis that there is no off-the-shelf or one-size-fits-all solution to RRI implementation and institutionalization. Instead, highly contextual empirical evidence is needed. This is precisely the reason why we invested time and resources to group modeling exercises that we have performed over the course of about six months with three selected Co-change Labs.

Overall, our systems thinking-based approach, or tool, aligns well with the broader normative spirit of RRI. First, systems thinking helps to motivate change by creating

understanding on the role that people themselves play in a system, often exacerbating the very problems they intend to solve. Second, systems thinking catalyzes collaboration in emphasizing not just the individual but the collective role that people have in terms of the functioning of the systems around us, for better and for worse. Finally, systems thinking stimulates the practice of continuous learning, which is the prerequisite for achieving sustainable change in complex systems. (Stroh 2015, 21-22)

5-CONCLUSIONS

This deliverable 4.1 utilized a tailored systems thinking framework as a tool for RRI self-evaluation and impact assessment. More specifically, our tool consisted of the utilization of participatory group model building and causal loop diagrams to better understand the dynamics related to RRI implementation, integration, and institutionalization in three Co-Change Lab use cases. The organizations behind the selected Labs were three Research Performing Organizations, namely AIT, Tecnalia and VTT.

As per the task 4.1 description, we utilized the systems thinking framework along with participatory group model building and causal loop diagrams to create reflexive capacity for understanding exactly how the complex systemic relationships, interactions and feedback loops either strengthen or weaken the potential for responsible actions and impacts in organizations. These objectives were the essence of our group modeling exercises with the three Lab use cases.

As a specific outcome of task 4.1, together with experts and stakeholders from participating Co-Change organizations, we developed a variety of descriptive and qualitative system models, with each model describing a specific dynamic present in a particular Lab use case. These models as well as the preceding process that led to the development of the models helped to create shared understanding on the system-level implementation of RRI principles as well as the drivers and barriers related to the uptake of RRI within organizations. We also drew two generic RRI implementation models by synthetizing the relevant and generalizable characters of the different individual models Lab use cases into one comprehensive model. Finally, we validated the generalized models of RRI implementation in Co-Change with the relevant project and Lab partners in a separate session.

The workshop sessions with the Lab participants concluded that participatory group-based modeling in combination with causal loop diagrams can be a fruitful way of assessing and visualizing the dynamics behind RRI impacts. Moreover, the utilization of systems thinking approach as a tool for understanding and evaluating RRI implementation, can help to increase the capacity for organizational self-understanding, reflexivity, and social learning by contributing to the development and improvement of the contextual pre-conditions for organizational RRI implementation.

Our results uncovered many familiar drivers and barriers of RRI institutionalization, but crucially connect them to one another. By doing so, we present ***change processes***, ***dynamics and conditions*** that can result in successful RRI institutionalization. An overarching finding is that even smaller scale or informal initial attempts at RRI practices can generate the conditions for more large scale and formalized change.

Finally, our systems thinking -based approach also aligns well with RRI practices by creating understanding on complex systems, catalyzing collaboration and by stimulating the practice of continuous learning.

6-SOURCES

Basco-Carrera L., Warren A., van Beek E., Jonoski A. and Giardino A. (2017) Collaborative modelling or participatory modelling? A framework for water resource management. *Environmental Modelling & Software*, 91, pp. 95-110.

Bertanlaffy, L. von (1969). *General System Theory: Foundations, Development, Applications (Revised Edition)*. George Braziller: New York.

Florin, M.-V., (2019). Risk governance and "responsible research and innovation" can be mutually supportive. J. Risk Res. https://doi.org/https://doi.org/10.1080/13669877.2019.1646311

Forrester, J. W. (1961). Industrial Dynamics. Pegasus Communications.

Fraaije, A., & Flipse, S. M. (2020). *Synthesizing an implementation framework for responsible research and innovation.* Journal of Responsible Innovation, *7*(1), 113-137. doi:10.1080/23299460.2019.1676685

Funtowicz S.O. and Ravetz J.R. (1993) Science for the Post-Normal Age. *Futures*, September, pp. 739-755.

Geels, F. W. (2004). *From sectoral systems of innovation to socio-technical systems: Insights about dynamics and change from sociology and institutional theory*. Research Policy, 33, 6–7, 897–920.

Gurzawska, A., Mäkinen, M., & Brey, P. (2017). Implementation of Responsible Research and Innovation (RRI) practices in industry: Providing the right incentives. *Sustainability*, *9*(10), 1759.

Haim, Mario & Zamith, Rodrigo (2019). *Open-Source Trading Zones and Boundary Objects: Examining GitHub as a Space for Collaborating on "News".Media and Communication*, Volume 7, Issue 4, Pages 80–91 DOI: 10.17645/mac.v7i4.2249

Hammond, D. (2005). Philosophical and ethical foundations of systems thinking. *tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society*, *3*(2), 20-27.

Luhmann, N. (1995). *Social Systems*. Stanford University Press: Stanford.

Meadows D. (2008) *Thinking in Systems: A Primer*. Vermont: Chelsea Green Publishing.

Mitleton-Kelly, E. (ed.) (2007). *Complex systems and evolutionary perspectives on organizations. The application of complexity theory to organizations*. Emerald: Bingley, UK.

Mulgan, G. (2013). Thinking systems: how the systems we depend on can be helped to think and to serve us better. The STEaPP Working Paper series. Department of Science, Technology, Engineering and Public Policy, UCL. Retrieved from: https://www.ucl.ac.uk/steapp/sites/steapp/files/thinking_systems_2021_mulgan.pdf

Owen, R., Macnaghten, P., Stilgoe, J., (2012). Responsible research and innovation: From science in society to science for society, with society. Sci. Public Policy 39, 751–760. https://doi.org/10.1093/scipol/scs093

Pajula, T., Vatanen, S., Behm, K., Grönman, K., Lakanen, L., Kasurinen, H., & Soukka, R. (2021). *Carbon handprint guide: V. 2.0 Applicable for environmental handprint*. VTT Technical Research Centre of Finland. Retrieved from: https://publications.vtt.fi/julkaisut/muut/2021/Carbon_handprint_guide_2021.pdf

Palmer, E. (2017). Beyond Proximity: Consequentialist Ethics and System Dynamics. *Etikk i praksis. Nord J Appl Ethics* (2017), 11(1), 89–105. Retrieved from: https://www.ntnu.no/ojs/index.php/etikk_i_praksis/article/view/1978/2043

Phillips, M. A. - Ritala, P. (2019). A complex adaptive systems agenda for ecosystem research methodology. Technological Forecasting and Social Change, 148, 1-12.

Pruyt, E., & Kwakkel, J. (2007). Combining system dynamics and ethics: Towards more science. In *25th international conference of the system dynamics society, July*.

Rilla, N., Tomminen, J., Nieminen, M., & Lehtinen, S. (2020). *D1.2 Instititutional environment and ecosystem analysis report*. Retrieved from:

Rip, A., (2014). *The past and future of RRI*. Life Sci. Soc. Policy 10, 1–15. https://doi.org/10.1902/jop.1939.10.1.31

Ruutu S. (2015) *in Nieminen, M., & Hyytinen, K. (Eds.) (2015).* *STRADA: Päätöksenteko ja muutoksen edistäminen monimutkaisissa järjestelmissä*. VTT Technical Research Centre of Finland. VTT Technology No. 218. Retrieved from: https://publications.vtt.fi/pdf/technology/2015/T218.pdf

Senge, P. M. (1990). *The fifth discipline: The art and practice of the learning organization*. New York: Doubleday/Currency.

Setiawan, A. D., Sutrisno, A., & Singh, R. (2018). Responsible innovation in practice with system dynamics modelling: the case of energy technology adoption. *International Journal of Innovation and Sustainable Development*, *12*(4), 387-420.

Star, SL & Griesemer, JR (1989). *Institutional Ecology, `Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology*. 1907-39. Social Studies of Science. 1989;19(3):387–420. doi:10.1177/030631289019003001

Sterman, J. D. (2000). *Business Dynamics: Systems Thinking and Modeling for a Complex World*. Irwin McGraw-Hill: Boston.

Stilgoe, J., Owen, R., Macnaghten, P., (2013). Developing a framework for responsible innovation. Res. Policy 42, 1568–1580. https://doi.org/http://dx.doi.org/10.1016/j.respol.2013.05.008

Stroh, D. P. (2015). *Systems thinking for social change: a practical guide to solving complex problems, avoiding unintended consequences, and achieving lasting results.* White River Junction, Vermont: Chelsea Green Publishing.

Tabares Gutierrez, R., Arrizabalaga, E., Nieminen, M., Rilla, N., Lehtinen, S., & Tomminen, J. (2020). *D1.1 Stocktaking Report*. Retrieved from:

Vennix J. (1996) *Group Model Building: Facilitating team learning using system dynamics*. New Jersey: Wiley.

Voinov, A., Jenni, K., Gray, S., Kolgani, N., Glynn, P. D., et al. (2018) Tools and methods in participatory modeling: Selecting the right tool for the job. *Environmental Modelling & Software*, 109, pp. 232-255.

Voinov, A., Kolagani, N., McCall, M. K., Glynn, P. D., Kragt, M. E., Ostermann, F. O., Pierce, S. A., & Ramu, P. (2016). Modelling with stakeholders: next generation. *Environmental modelling & software*, *77*, 196-220. https://doi.org/10.1016/j.envsoft.2015.11.016

Wagner, P., & Wilhelmer, D. (2020). *D2.1 Guidelines for the Co-Change Platform*. Retrieved from:

Williams B. and Hummelbrunner R. (2011) *Systems concepts in action: a practitioner's toolkit*. Stanford: Stanford Business Books.

Wolfram, M. (2016). *Conceptualizing urban transformative capacity: A framework for research and policy*. *Cities, 51*, 121-130. doi:http://dx.doi.org/10.1016/j.cities.2015.11.011

Co-funded by the Horizon 2020 programme
of the European Union

CO-CHANGE