



SATORI Evaluation and Reflection Principles and Criteria

*Authors: Brent Mittelstadt (De Montfort University)
Mark Coeckelbergh (De Montfort University)
Kutoma Wakunuma (De Montfort University)
Tilimbe Jiya (De Montfort University)*

December 2014

SATORI Deliverable D12.2

This publication and the work described in it are part of the project *Stakeholders Acting Together on the Ethical Impact Assessment of Research and Innovation – SATORI* which received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 612231

Contact details for corresponding author:

Prof. dr. Mark Coeckelbergh, Centre for Computing and Social Responsibility
De Montfort University, The Gateway, Leicester, LE1 9BH United Kingdom
mark.coeckelbergh@dmu.ac.uk

<http://satoriproject.eu/>



Abstract	4
Executive Summary	5
1 Introduction	6
2 DMU’s Role	7
2.1 Scope of Task 12.2.....	7
3 Classification of SATORI Activities for Evaluation	8
4 Evaluation and Reflection Criteria	9
4.1 Justification for selecting Evaluation Principles and Criteria specific for SATORI	10
4.2 Selected Evaluation Principles and Criteria specific for SATORI	11
4.2.1 <i>Principles and Criteria for Evaluating Stakeholder Engagement / Involvement</i>	11
4.2.2 <i>Principles and Criteria for Evaluating Recruitment</i>	11
4.2.3 <i>Principles and Criteria for Evaluating Surveys, Interviews and Case Studies</i>	11
4.2.4 <i>Principles and Criteria for Evaluating Recommendations/Tools</i>	11
4.2.5 <i>Principles and Criteria for Evaluating Dissemination/Impact</i>	11
4.2.6 <i>Principles and Criteria for Evaluating Evaluation</i>	12
4.2.7 <i>Principles and Criteria for Evaluating Administration</i>	12
4.2.8 <i>Principles and Criteria for Evaluating ‘Internal’ Activities</i>	12
5 Conclusion	13
6 Annexes	16
6.1 Annex A: Evaluation.....	16
6.1.1 <i>Evaluation vs. Verification</i>	17
6.1.2 <i>Subjects of Evaluation</i>	18
6.1.3 <i>Mutual Learning and Mobilisation</i>	18
6.2 Annex B: Principles of Good Practice	22
6.2.1 <i>Criteria Principles</i>	22
6.2.2 <i>Mutual Learning Principles</i>	23
6.2.3 <i>Reflection Principles</i>	24
6.3 ANNEX C: EVALUATION PRINCIPLES AND CRITERIA	25
6.3.1 <i>Principles and Criteria for Evaluating Stakeholder Engagement</i>	26
6.3.2 <i>Principles and Criteria for Evaluating Recruitment</i>	30
6.3.3 <i>Principles and Criteria for Evaluating Surveys, Interviews and Case Studies</i>	30
6.3.4 <i>Principles and Criteria for Evaluating Recommendations/Tools</i>	31
6.3.5 <i>Principles and Criteria for Evaluating Dissemination/Impact</i>	31
6.3.6 <i>Principles and criteria for evaluating evaluation</i>	34
6.3.7 <i>Principles and criteria for evaluating administration</i>	35
6.3.8 <i>Evaluating ‘Internal’ Activities</i>	35
6.3.9 <i>Summary of Evaluation and Reflection Criteria</i>	36

ABSTRACT

This report outlines the chosen SATORI evaluation and reflection principles and criteria in Task 12.2. The report builds upon the findings of ‘D12.1 – Good practice in evaluation, reflection and civil society engagement’ to specify a range of evaluation and reflection principles and criteria to be used in evaluating SATORI. The criteria are developed through a classification of SATORI tasks by ‘activity type’, and application of the principles of good practice and evaluative frameworks considered in D12.1. A set of criteria applicable to different activity types were specified and a set of criteria which we think are applicable to SATORI are offered. These are based on the analysis undertaken in D12.1. These criteria will be further ‘operationalised’ for each individual task. The principles and criteria reported here will further inform the SATORI evaluation and reflection strategy to be developed in Task 12.3. This strategy will then be implemented in evaluating SATORI in Task 12.4.

EXECUTIVE SUMMARY

This report describes evaluation and reflection principles and criteria for the SATORI Mobilisation and Mutual Learning Action (MML). Mobilisation and Mutual Learning Actions are a new type of engagement project funded by the European Commission (EC) under the 7th Framework Programme Mobilisation and Mutual Learning Action Plan (MMLAP). One point of consensus concerning necessary activities undertaken in MMLs is the need for on-going evaluation and reflection (or monitoring) on project progress, methods and impact.

Despite the general consensus on the need for evaluation in MMLs, it remains unclear how, when and according to which measures MMLs should be evaluated. The EC has not openly identified specific requirements to be met by evaluation, beyond having a separate Work Package dedicated to evaluation. Following on from a literature review and empirical study into this apparent gap, which identified 21 principles of good practice (see: Deliverable 12.1), this report specifies a set of criteria for evaluation and reflection in SATORI.

In particular, in its conclusion this report proposes to focus on evaluating stakeholder engagement (main criterion given this type of project), evaluating recruitment (in particular representativeness), evaluating surveys, interviews and case studies (in particular engagement and involvement), evaluating and monitoring the way the project produces and deals with legal and other documentation, evaluating recommendations (how are they arrived at and do they acknowledge all stakeholders' values and perspectives), facilitating the self-evaluation of impact by means of indicators of success, encouraging partners to reflect on their research (again self-evaluation, e.g. via workshops), improving and stimulating collaboration. These principles and criteria will be further developed and operationalised in Task 12.3.

1 INTRODUCTION

As an MML project, SATORI conceptualises ethical assessment as a form of engagement between research and society. In this context SATORI aims to “improve respect of ethics principles and laws in research and innovation, and to make sure that they are adequately adapted to the evolution of technologies and societal concerns.” These proposed aims will be accomplished through the development of an ethics assessment framework based upon “thorough analysis, commonly accepted ethical principles, participatory processes and engagement with stakeholders, including the public, in Europe and beyond¹.” To this end the project will first create an inventory of current practice and principles in ethics assessment. Stakeholders in ethical assessment in existing research projects will be identified, with due attention paid to the impact of globalisation and movement towards conducting research in non-European countries to benefit from ‘looser’ standards of ethical review. An initial framework will then be developed, with a subsequent review of the prospects for standardisation and sustaining a network to promote adoption of the framework in future research. Policy and other developments meriting ethical review will be monitored throughout the duration of the project, providing an initial channel for promotion and adoption of the framework beyond the SATORI project. Results of the MML will be disseminated to stakeholders identified in the project, including representatives of government and research or policy requiring ethics assessment. Dissemination activities undertaken in the project seek to engage members of society with a stake in ethics assessment. This report therefore builds upon the findings of ‘D12.1 – Good practice in evaluation, reflection and civil society engagement’ to specify a range of evaluation and reflection principles and criteria to be used in evaluating SATORI.

The report is structured as follows: Section 1 gives the introduction, Section 2 highlights DMU’s role, Section 3 outlines the classification of SATORI activities for evaluation while Section 4 provides the principles and criteria for evaluation and reflection. Finally, Section 5 concludes the report and suggests a number of principles and criteria for evaluating SATORI.

¹ SATORI, *THEME [SiS.2013.1.2-1 SiS.2013.1.2-1] Mobilisation and Mutual Learning (MML) Action Plans: Mainstreaming Science in Society Actions in Research] - Annex 1 - Description of Work.*

2 DMU'S ROLE

As leader of WP12, DMU will provide an independent evaluation of SATORI's activities for the duration of the project, advising the SATORI consortium on potentially beneficial changes to project activities when necessary (See Annex A for detailed discussion on Evaluation). The evaluators will use principles of good practice in MML evaluation and reflection that have been determined in Task 12.1 which are highlighted in Annex B. In Task 12.2 these principles will be adapted to SATORI's specific aims and activities to create a set of customised evaluation and reflection principles and criteria for the project. DMU will then develop an evaluation strategy in Task 12.3 that will employ the principles and criteria in carrying out the evaluation of the project, which will take place in Task 12.4.

2.1 SCOPE OF TASK 12.2

In Task 12.2 DMU will identify which principles from D12.1 are applicable to SATORI, and how they may be specified into particular criteria for evaluating and reflecting upon the success of project activities. The adaptation of the principles will focus on how the quality of the multitude of project activities can be assessed using criteria designed for particular Work Packages, task types or individual tasks. As SATORI will begin to develop an ethics assessment framework, the method of development and application of the framework across other project activities and beyond the consortium will be assessed, for example in terms of how project activities have supported the adoption or further development of the framework by relevant non-consortium stakeholders. Evaluative criteria will be developed according to the principles of good practice and evaluative frameworks identified in Deliverable 12.1 along with the variety of aims pursued in SATORI. For example, criteria will concern questions of quality and feasibility, the prospects for incorporation of the ethical framework into the EU system of research and innovation, support for transnational co-operation in ethical assessment provided by the project, and the contribution to sustainable and inclusive solutions to key challenges facing European society made by the framework and supporting project activities. As SATORI demands an intensive and prolific deliberation among participants (in line with being an MML), criteria focusing on the 'fairness' and 'competence' of these processes, including the quality of debate facilitated and the representativeness of stakeholders engaged, will likely feature in the final set of evaluation and reflection criteria².

Each of the above examples concerns the quality of tasks (and underlying methods) completed by the consortium. However, criteria will also be developed to assess the impact of SATORI as a whole and its ethical impact assessment framework in particular on citizens and society. In doing so the principles identified in D12.1 will also be considered to ensure a realistic range of impacts, which can be practically assessed given the timeframe of evaluation.

The output of Task 12.2 will be a SATORI-specific set of principles and criteria adapted from the 21 principles identified in Deliverable 12.1, and a set of evaluative criteria to be applied in the evaluation and reflection strategy developed in Task 12.3. The overall evaluation and reflection strategy will describe the activities to be undertaken by DMU to measure the success of SATORI against the principles and criteria identified in Deliverable 12.2.

² See: SATORI, *THEME [SiS.2013.1.2-1 SiS.2013.1.2-1] Mobilisation and Mutual Learning (MML) Action Plans: Mainstreaming Science in Society Actions in Research* - Annex 1 - Description of Work, 43.

In order to come up with much more robust principles and evaluation criteria specific for SATORI, it becomes necessary to classify SATORI activities.

3 CLASSIFICATION OF SATORI ACTIVITIES FOR EVALUATION

As suggested in D12.1, MMLs are defined as interdisciplinary activities, meaning evaluation must account for different ways of doing research and measuring success across the different disciplines represented in a project's consortium. Following on from good practice discussed in the empirical study described in D12.1, evaluation of SATORI will occur in part at a task-specific level so as to allow for discipline-specific methods.

While task-specific evaluation is desirable, it presents a much more complex challenge for the evaluator who is responsible for conducting separate evaluations and perhaps, defining task-specific evaluation criteria. Classifications or 'task types' were assigned by comparing the description of each Task in the SATORI Description of Work to the classifications defined below. The classifications are derived from D12.1 and the description of the aims and characteristics of MMLs as a new type of research and coordination action. The list is not intended as a comprehensive representation of MML activities. The classifications are instead based on the authors' interpretation of the defining characteristics and 'ends' of MMLs, and the general types of activities undertaken. Here, 'stakeholders' refer to individuals, groups and organisations engaged through the MML, whereas 'partners' refers to the consortium partners and contracted personnel responsible for organising and carrying out the activities described. The types of activities are:

- **Desk research** refers to any analysis of extant data, which does not explicitly involve empirical data collection or a structured literature review. Desk research can include analysis of empirical data and literature, so long as it does not involve structured collection of either. For example, secondary analysis or consideration of findings from a prior task or consortium partner are considered desk research.
- **Recruitment** refers to any activities which recruit stakeholder participants to MML activities (e.g. stakeholder engagement, capacity building, survey, interviews, etc.).
- **Stakeholder engagement – dialogue** refers to activities in which stakeholders (civil society organisations, members of the public, ethical committee members, regulators, etc.) are engaged in a process of dialogue or deliberation. It is expected that stakeholders and coordinators will learn from and between one another. Intended outcomes include mutual learning between all participants, facilitation of network/relationship building, and exchange of ideas/perspectives (and therefore, awareness raising).
- **Stakeholder engagement - capacity building** refers to stakeholder engagement activities in which stakeholders are engaged in training programmes, educational programmes and similar activities in which 'one-way' flow of information is expected. This does not preclude the possibility of mutual learning; rather, the purpose of stakeholder engagement - capacity building is to help stakeholders develop particular skills, awareness or knowledge. Activities focus on passing the desired information or skills **from** the partners or activity leaders **to** stakeholders. The ends are therefore distinguishable from stakeholder engagement, which focuses on mutual learning, although overlap can be expected.

- **Stakeholder engagement - feedback** refers to stakeholder engagement activities in which stakeholder engagement - feedback is specifically sought on a tool, framework or other deliverable created by the consortium. Mutual learning and dialogue between stakeholders is possible, but the explicit goal of the activity is to discover how stakeholders feel about the output considered.
- **Survey** refers to any type of questionnaire or survey instrument intended to collect data from ‘stakeholders’ (e.g. non-consortium partners) identified in the MML.
- **Interviews** refers to any type of interview instrument intended to collect data from ‘stakeholders’ (e.g. non-consortium partners) identified in the MML.
- **Case studies** refers to any type of case study protocol intended to collect and analyse data from ‘stakeholders’ (e.g. non-consortium partners) identified in the MML. Case studies may include interviews, document analysis, surveys, etc. but are distinguished from other activities by the explicit usage of a case study methodology.
- **Recommendations/tools** refers to the creation of practice-oriented recommendations or tools are created, for example for refining legal, regulatory and ethical frameworks. Restrictions are not placed on the source of recommendations, and may be developed with input from both partners and stakeholders.
- **Dissemination/impact** refers to any activity related to disseminating outputs of the MML or encouraging/supporting/developing impact of the project and its outputs. Examples include support activities related to heritage and sustainability of stakeholder networks established through the project.
- **Evaluation** refers to any activities undertaking or supporting evaluation of the project including its ongoing activities, outputs and impact. Evaluation may involve data collection from both stakeholders and consortium partners.
- **Administration** refers to administrative activities undertaken in coordination of the project.

4 EVALUATION AND REFLECTION CRITERIA

The activities that are identified in Section 3 above were significant in selecting the evaluation and reflection criteria for the SATORI project. Looking at the project holistically, there are generic activities that are relevant to most of the WPs. These tasks include interviews, desk research and surveys which are used for data collection. In addition to these generic activities highlighted in the preceding line, each WP was analysed in order to identify activities specific to each WP. This was in order to understand the various activities associated with each WP so that we could then establish cross-cutting theme(s) that would be relevant in establishing the evaluation and reflection principles and criteria specific for the project. The analysis resulted in the following:

Work Package	Activity
WP1	<ul style="list-style-type: none"> • Survey

	<ul style="list-style-type: none"> • Literature review
WP2	<ul style="list-style-type: none"> • Recruitment • Stakeholder engagement – capacity building, dialogue
WP3	<ul style="list-style-type: none"> • Case studies • Stakeholder engagement – feedback • Recommendations/tools
WP4	<ul style="list-style-type: none"> • Stakeholder engagement – feedback, capacity building, dialogue • Recommendations/tools
WP5	<ul style="list-style-type: none"> • Case studies • Stakeholder engagement – feedback • Recommendations/tools
WP6	<ul style="list-style-type: none"> • Evaluation • Stakeholder engagement – feedback, capacity building
WP7	<ul style="list-style-type: none"> • Stakeholder engagement – feedback, dialogue • Dissemination/impact
WP8	<ul style="list-style-type: none"> • Stakeholder engagement – feedback • Dissemination/impact • Recruitment
WP9	<ul style="list-style-type: none"> • Stakeholder engagement – feedback, dialogue • Dissemination/impact • Recommendation/tools
WP10	<ul style="list-style-type: none"> • Dissemination/impact
WP11	<ul style="list-style-type: none"> • Administration • Evaluation
WP12	<ul style="list-style-type: none"> • Evaluation • Stakeholder engagement – feedback

Table 1: Activities specific to each WP

4.1 JUSTIFICATION FOR SELECTING EVALUATION PRINCIPLES AND CRITERIA SPECIFIC FOR SATORI

Having assessed the WP activities highlighted in table 1 above, it became evident that there was one recurring cross-cutting theme across the WPs which would be useful in informing the selection of the principles and criteria for evaluating the SATORI project. This theme was stakeholder engagement which actually chimed with SATORI's main objective, that of 'Ethical assessment through stakeholder engagement'. As such, with regards to stakeholder engagement and as suggested in Table 1, it then becomes necessary when evaluating SATORI to look at the following three keys areas:

- i) Feedback (stakeholder input/views)
- ii) Capacity building (training sessions on ethical assessment)

iii) Dialogue (stakeholder agreement/consensus)

Having established a recurring cross-cutting theme across all WPs of SATORI, (that of stakeholder engagement and related key areas), we then looked at the 21 general principles of evaluation and criteria that were identified in D12.1 and that are outlined in Annex B in order to analyse which ones cover stakeholder engagement and therefore specific to SATORI. The analysis revealed 8 specific evaluation and reflection principles and criteria which cover stakeholder engagement. These are outlined below and are described succinctly in Annex C:

4.2 SELECTED EVALUATION PRINCIPLES AND CRITERIA SPECIFIC FOR SATORI

4.2.1 Principles and Criteria for Evaluating Stakeholder Engagement / Involvement

This criterion includes a concern for representativeness among participants in stakeholder engagement events, transparency in decision-making processes within engagement activities and accessibility of relevant information material to participants of engagement activities. In addition, this criterion addresses concerns on clarity of tasks and instruction or guidelines given to participants in relation to an event. The criterion also evaluates the extent of fair deliberation which relates to the degree to which participants are allowed to put forward their views.

4.2.2 Principles and Criteria for Evaluating Recruitment

This criterion is used to ensure that there is equal representation and that stakeholders are empowered not only through capacity building and learning but by ensuring that underrepresented stakeholders are involved in the discourse. This has the ability to help tackle large societal challenges which involve an array of stakeholders.

4.2.3 Principles and Criteria for Evaluating Surveys, Interviews and Case Studies

This criterion is used to assess the methodology used in the project such as surveys, interviews and case studies in the production of quality deliverables, engagement and application of success indicators. These indicators of success are used for specific project activities while being responsive to the main aims of the project.

4.2.4 Principles and Criteria for Evaluating Recommendations/Tools

This criterion is used to assess relevance and recognition of values and views of all stakeholders. The recommendations should be relevant to the project aims and transparent in terms of the decision-making processes that precede the recommendations. In addition, they should acknowledge all stakeholders perspectives.

4.2.5 Principles and Criteria for Evaluating Dissemination/Impact

This criterion is used to assess the impact of the project and its activities. We appreciate that it is difficult to assess impact; however it is possible to do an indicator of success questionnaires which ask stakeholders to evaluate the impact of SATORI. For example, questions that can be

asked in the evaluation could cover the following: Has the impact been positive? Has the behaviour of participants been affected by the project? Is there more recruitment? These are all relevant impact assessment criteria.

4.2.6 Principles and Criteria for Evaluating Evaluation

This criterion is used to assess challenges that evaluators can come up against. To overcome possible challenges, evaluators should be increasingly critical and identify limitations not only with the process of evaluation itself but the context in which evaluation occurs. Evaluation limitations can stem from aspects of the process or context of evaluation, such as resistance from the consortium partners or limitations established in the Description of Work (DoW). On this basis the quality of evaluation can be assessed in terms of ‘Restrictiveness’, established through critical self-assessment of limitations imposed on the evaluators and evaluation by the project’s broader context, description of work, consortium, coordinator or other relevant sources. In addition, under this criterion we assess how our evaluation has been participative in looking at the quality of stakeholder participation in evaluation.

4.2.7 Principles and Criteria for Evaluating Administration

This criterion is used to assess the quality of administration and coordination in terms of ‘Quality of Collaboration’, looking at breakdowns in communication or conflicts between partners that may reduce the quality of collaboration and thus jeopardise the project. In addition, the use of this criterion is through qualitative observations of workshops concerning any practical barriers to collaboration encountered by SATORI partners.

4.2.8 Principles and Criteria for Evaluating ‘Internal’ Activities

This criterion is used to evaluate SATORI’s activities which may be considered ‘Internal’ activities such as inter-consortium communication and collaboration including consortium meetings, peer-review and informal communication. Further, the assessment centres on partners’ critical reflection on their progress and changes to attitudes and behaviours through formal or informal methods such as interviews, project management meetings, or peer review of deliverables.

In conclusion, below is a table that gives a summary of the evaluation and reflection criteria that is specific to SATORI:

Activity Type	Criteria
Stakeholder Engagement	Representativeness Transparency Accessibility Task Definition Fair Deliberation Iteration Criticalness Ownership of Outcomes Participant Satisfaction
Recruitment	Representativeness Accessibility Criticalness
Surveys, Interviews and Case Studies	Methodological Rigour Credibility Transparency
Desk Research	Collaborative Influence
Recommendations/Tools	Transparency Relevance
Dissemination/Impact	Quantity Behaviour Adjustment Perspective Adjustment Network Expansion
Evaluation	Restrictiveness Representativeness
Administration	Quality of Collaboration
'Internal'	Stakeholder engagement criteria (as above) Reflectiveness

Table 2: SATORI Evaluation and Reflection Criteria by Activity Type

5 CONCLUSION

This deliverable has described a set of evaluation and reflection criteria which will be further applied in developing a strategy for evaluation and reflection in SATORI. The findings of this deliverable follow on from the activities of Task 12.1, and will eventually inform DMU's work in Tasks 12.3 and 12.4. Task 12.2 looked at different areas of evaluation and reflection and considered feasibility and quality of the criteria in terms of evaluating SATORI as a whole. Further, a classification of activities to be evaluated was developed which informed the recommendation for an evaluation and reflection criteria for SATORI.

Consequently, with reference to fairness, competence in evaluation together with efficiency, transparency, effectiveness of the evaluation process and reflective outcome from SATORI consortium the following criteria principles are recommended for evaluation of SATORI. It should be noted that the defined criteria principles are not prescriptive and final yet but will be discussed at the evaluation and reflection workshop in 2015 in February (Brussels) and will inform the SATORI evaluation and reflection strategy in 12.3.

Thus, we propose to select the following criteria and interpret them for SATORI in the following way:

The first and main criterion that will be used in evaluating SATORI is that which focuses on **evaluating stakeholder involvement**. This criterion will assess mutual learning among stakeholders of the SATORI project. As described in D12.1, assessing mutual learning is paramount in an MML because it can be used to evaluate the quality of mutual learning and mobilisation in stakeholder engagement activities. This criterion will include a concern for representativeness among participants in stakeholder engagement events, transparency in decision-making processes within engagement activities and accessibility of relevant information material to participants of engagement activities. In addition, this criterion will address concerns on clarity of tasks and instruction or guidelines given to participants in relation to an event. The criterion will also evaluate the extent of fair deliberation which relates to the degree to which participants are allowed to put forward their views. Relatively, the criteria will evaluate the extent to which dialogue and topics are revised in response to needs, issues and concerns expressed by participants and evaluate them in terms of criticalness which refers to the degree to which participants are encouraged to challenge and negotiate views and values of other members of the engagement process. Lastly under this criterion the evaluation and reflection will consider the level of ownership of outcomes among participants and participant satisfaction.

In relation to the above, the second criterion for SATORI evaluation and reflection will include the criterion principle for **evaluating recruitment**. The focus here is on **representativeness** and is connected to the first criterion stakeholder involvement. This criterion will be used to ensure that there is equal representation and that stakeholders are empowered not only through capacity building and learning but by ensuring that underrepresented stakeholders are involved in the discourse thereby tackling large society challenges which involve an array of stakeholders.

Thirdly, SATORI will be evaluated using the principle and criteria for **evaluating surveys, interviews and case studies**. This will use indicators of success for specific project activities while being responsive to challenges of **engagement**. The results from the evaluation process will aid reflection on SATORI's activities in terms of rigour and quality.

Fourthly, since **desk research** is not a novel activity in SATORI, we will not use this criterion in our evaluation. We prefer to focus on the novel and primary aspects of (the evaluation of) SATORI such as stakeholder involvement and the interviews and surveys.

However, since SATORI is also active with regard to legal documents and documentation (e.g. on ethical training), it may be essential to evaluate and monitor **legal and other related documentation** as a way of observing ethical considerations.

Furthermore, the **recommendations** given by SATORI should be relevant to the project aims and transparent in terms of the decision-making processes that precede the recommendations – it should be clear how the recommendations are arrived at – and they should acknowledge all stakeholders' values and perspectives.

Further, although it is difficult to **assess impact**, it is possible to do an indicator of success questionnaire which asks stakeholders to evaluate the impact of SATORI. For example, questions can cover the following: Has the impact been positive? Has the behaviour of participants been affected by the project? Has there been any expansion in their network? Is there more recruitment? These are all relevant impact assessment criteria.

Likewise, we need to facilitate and encourage partners to reflect on their research. In practice the WP leaders will play an important role in this **self-evaluation**. But as evaluators we will facilitate this during the workshops.

In addition, **collaboration** will be a necessary aspect of the project and will therefore merit evaluating.

Finally, **engagement and reflectiveness** are also important criteria but we feel that they have been covered in the previous criteria, or at least in the way we interpreted them.

This also shows that the criteria cannot and should not stand on their own; they are all related.

In conclusion, the criteria described here should not yet be considered finalised for other reasons as well; practicalities of evaluating the project and methodological difficulties with evaluating certain aspects (in particular, learning over time and impact) may preclude evaluation of particular criteria. This should not be taken as a weakness of the principles and criteria described here or the strategy to be developed; rather, as with all evaluative frameworks, the framework created here is a theoretical ideal around which evaluation activities responsive to the real-world limitations of the project context will be defined. Regardless of the eventual shape of the evaluation strategy, the primary contribution of this deliverable (and thus, D12.1) has been to provide a theoretically and empirically sound basis for the evaluation of SATORI based upon best practice as currently understood.

6 ANNEXES

6.1 ANNEX A: EVALUATION

As described in D12.1, evaluation can be defined as “the process of determining the merit, worth and value of things”³, and can be used to describe many “different kinds of judgments, from informal assessment that relies on intuition or opinion, to well-defined and systematic research that makes use of social science research methods”⁴. Concerning research and engagement projects, evaluation will at a minimum focus on the “design, implementation and effectiveness” of the project⁵, as well as outputs (e.g. reports, deliverables), outcomes (e.g. products, processes) and impacts (e.g. follow-on products, processes and research)⁶. In general evaluation aims to assist in developing research activities during the life of the project (e.g. through feedback from evaluators to partners), improve the design of future related activities, assess project impact⁷, and provide participants with a better idea of the value of their participation by tracking influence on the process⁸. Evaluation will often be required or called for by sponsors or funders of research programmes, who wish to pose certain questions or have an indication of the relative merit of the funded research⁹. Criteria against which the success of the project can be assessed are typically used¹⁰, and can be pre-defined or developed in-situ. Methods of data collection and analysis are often guided by prescriptive discipline-specific guidelines¹¹.

An initial distinction can be drawn in the evaluation of participatory research between procedural evaluation which examines the quality of procedures or methods of stakeholder engagement and mutual learning while they occur, and substantive evaluation which assesses the outcomes of procedures¹². The former can contribute to the refinement of research and participatory processes by identifying weaknesses or areas for improvement prior to the project’s conclusion¹³, thereby acting as a feedback mechanism or ‘double loop’ to refine project activities¹⁴ through assessment of deliverables, communications and events. The latter, on the other hand, evaluates the quality of a project’s outputs and outcomes to evaluate

³ Scriven, *Evaluation Thesaurus*.

⁴ Joss, “Evaluating Consensus Conferences: Necessity or Luxury”, 89.

⁵ Tuominen et al., “Evaluating the Achievements and Impacts of EC Framework Programme Transport Projects”, 61.

⁶ Ibid.; Arnold, “Understanding Long-Term Impacts of R&D Funding”; Bornmann and Marx, “How Should the Societal Impact of Research Be Generated and Measured?”; Nagarajan and Vanheukelen, “Evaluating EU Expenditure Programmes: A Guide. Ex Post and Intermediate Evaluation. XIX/02–Budgetary Overview and Evaluation”.

⁷ Research Councils UK, *Evaluation: Practical Guidelines - A Guide for Evaluating Public Engagement Activities*.

⁸ Rowe and Frewer, “Evaluating Public-Participation Exercises: A Research Agenda”.

⁹ O’Sullivan, “Collaborative Evaluation within a Framework of Stakeholder-Oriented Evaluation Approaches”, 518.

¹⁰ e.g. Rowe and Frewer, “Public Participation Methods”; Webler, “‘Right’ Discourse in Citizen Participation: An Evaluative Yardstick”; EuropeAid, *Evaluation - Guidelines*.

¹¹ e.g. O’Sullivan, “Collaborative Evaluation within a Framework of Stakeholder-Oriented Evaluation Approaches”; EuropeAid, *Evaluation - Guidelines*.

¹² Abelson and Gauvin, *Assessing the Impacts of Public Participation*, 12; Merckx et al., “Evaluation of Research in Context A Quick Scan of an Emerging Field”; Mickwitz, “A Framework for Evaluating Environmental Policy Instruments Context and Key Concepts”; Rowe and Frewer, “Public Participation Methods”, 10; Tuominen et al., “Evaluating the Achievements and Impacts of EC Framework Programme Transport Projects”.

¹³ Chess, “Evaluating Environmental Public Participation”, 773.

¹⁴ Abelson and Gauvin, *Assessing the Impacts of Public Participation*, 3.

its success¹⁵, for example in meeting its stated objectives, impacts or success indicators, or against a pre-defined framework for assessing quality. As with procedural evaluation, post-project substantive evaluation can inform the design and conduct of future participatory processes and research. In defining criteria for SATORI's evaluation and reflection activities, it is important to recognise these fundamental distinctions and prepare relevant criteria for each type.

Each type of evaluation focuses in some way on the 'quality' of research activities, participatory processes, outcomes, impact, etc. Discourse in this area tends to focus on establishing and justifying evaluative criteria to describe the quality of the process (or procedure) and its outcomes¹⁶. Many potential criteria of quality can be used, and as with evaluation of any form of public participation such criteria must be carefully chosen to ensure a match with the aims and field of the project being evaluated¹⁷. Frameworks and criteria for project evaluation should be understood as helping assess the quality (or as is sometimes said, effectiveness or success) of the activity¹⁸; while the criteria vary across disciplines, frameworks and projects, the central purpose of evaluative criteria as facilitating assessment of quality does not change.

As reported in D12.1, a universally accepted framework to define effectiveness does not exist in evaluation literature¹⁹, due perhaps to participatory processes occurring across numerous disciplines with different epistemic frameworks and standards of quality²⁰. MMLs can be thought of as interdisciplinary projects lacking a single real-world problem focus²¹. While SATORI broadly focuses on ethical assessment frameworks, the myriad of contexts in which ethical assessment is required precludes a single discipline-specific focus. Following from this, the criteria chosen to evaluate SATORI will not be based in a single discipline, and will be specified in Task 12.3 for evaluating individual tasks according to discipline-specific methods of good practice and indicators of quality.

6.1.1 Evaluation vs. Verification

Before describing evaluation criteria, an initial distinction between evaluation and verification needs to be drawn to avoid confusion. All activities and deliverables of SATORI will be 'verified' in terms of compliance with the milestones and requirements set out in the DoW. No matter the evaluative framework used a basic distinction in approaches to evaluation adopted by MMLs can be seen here, between 'verification' as an administrative or management process to ensure the requirements of the DoW are being met on time, and evaluation as a normative exercise according to which the quality of deliverables, activities and impacts are evaluated, in some cases to refine the project's ongoing activities. Verification will be carried out by the project coordinator and a European Commission project officer. Verification results will be included but only briefly discussed in forthcoming

¹⁵ Ibid.

¹⁶ Rowe and Frewer, "Evaluating Public-Participation Exercises: A Research Agenda".

¹⁷ Rowe and Frewer, "Public Participation Methods", 3; Smith, Nell, and Prystupa, "FORUM".

¹⁸ Rowe and Frewer, "Evaluating Public-Participation Exercises: A Research Agenda", 517.

¹⁹ Rowe and Frewer, "Evaluating Public-Participation Exercises: A Research Agenda".

²⁰ Klein, "Evaluation of Interdisciplinary and Transdisciplinary Research"; Bergmann et al., *Quality Criteria of Transdisciplinary Research: A Guide for the Formative Evaluation of Research Projects*; Bornmann and Marx, "How Should the Societal Impact of Research Be Generated and Measured?".

²¹ Mittelstadt, Stahl, and Wakunuma, *SATORI Deliverable D12.1 - Good Practice in Evaluation, Reflection and Civil Society Engagement*.

evaluations of the project, due to the lack of normative assessments of quality. ‘Normative’ evaluation moves beyond mere administrative concerns regarding meeting contractual objectives and deadlines, and instead deals with the procedural and substantive quality of processes and outputs. It is this normative type of evaluation which will be the primary concern of DMU as evaluators.

6.1.2 Subjects of Evaluation

For MMLs, certain subjects of evaluation seem necessitated by the types of activities undertaken (see: Section **Error! Reference source not found.**). As discovered through interviews with MML consortia in Task 12.1, evaluating mutual learning is, however, not enough by itself. Two general types of evaluation are needed in MMLs: (1) generic project evaluation akin to EuropeAid framework; and (2) evaluation of quality of participatory processes, network building, capacity building and facilitating communication, collaboration and mutual learning between stakeholders. The latter form of evaluation must necessarily look at the impacts of the project, as learning and exchange of ideas between stakeholders are continual processes occurring throughout and beyond the life of the project. The respondents also made clear that MMLs are not unique in the difficulties faced with evaluating impact. However, the attempts among respondents to begin to evaluate societal and policy impact in particular, in some cases in terms of mutual learning, shows that the methodological and practical difficulties with impact evaluation are not undermining the desire among practitioners to track the influence of research. Given the emphasis on mutual learning, capacity and network building in MMLs, attempts to evaluate project influence in these terms should be encouraged and perhaps supported through the development of evaluation frameworks and methods for MMLs.

To summarise, at a minimum generic project evaluation criteria appear to be necessary, such as those specified in the OECD framework. Criteria to assess mutual learning are also required based on mutual learning being a key characteristic of MMLs. The same may be said for assessing the quality of networks of stakeholders or communication/collaboration channels built through MMLs.

6.1.3 Mutual Learning and Mobilisation

The conception of mutual learning adopted by evaluators will significantly influence the criteria chosen to evaluate the quality of learning occurring in participatory processes, as well as the types of evidence sufficient to demonstrate changes in attitudes, beliefs or behaviour that may indicate learning has occurred. For example, a one-way ‘deficit model’ of learning would logically lead to success being evaluated against the degree of learning among societal participants while lacking reflexive assessment of learning among researchers (or evaluators).

D12.1 reviewed four theories of learning: organisational, expansive, transformative or reflective, and social. In SATORI mutual learning will be conceived of within the transformative and reflective traditions (for example, transformative learning theory builds upon expansive; see: D12.1), as these theories best account for the ‘mutual learning’ and ‘mobilisation’ aims of MMLs, as explained below. The theories adopted frame the choice of criteria for evaluating mutual learning occurring throughout the project, but in particular through participatory events and in terms of impact. Prior to discussing criteria for evaluating the different types of SATORI events below, it is worth briefly revisiting these approaches.

Transformative and reflecting learning are built upon expansive learning theory, which does not conceive of learning as knowledge acquisition or “becoming an active participator of cultural practices,” but rather views learning as a form of knowledge creation²². This view of learning is equivalent to ‘double-loop learning’ described in theories of organisational learning²³, by which existing conceptions and frames of reference are broken down and re-created to better account for unfamiliar phenomena or perspectives. Expansive learning encourages dialogue between a diversity of stakeholders (including researchers and evaluators) to create ‘double-loops’ through which prevailing research strategies and accepted wisdom can be criticised and refined²⁴. In practical terms expansive learning theory suggests that learning will occur when participants in research are encouraged to engage with unfamiliar perspectives and phenomena in some sort of dialogue (which, as a necessarily two-way activity, implies ‘mutual’ learning).

Transformative and reflective learning are closely related to expansive learning; in all three, learning is initiated by encountering unfamiliar perspectives or phenomena requiring interpretation. Transformative learning is a process through which an individual’s frame of reference, understood in hermeneutic terms as the set of preconceptions and understanding which frames interpretation of phenomena such as speech²⁵, is changed²⁶. The approach is closely related to reflective learning, which can be described as a process through which a learner comes to understand the pre-analytic assumptions held by himself and others which frame how phenomena are interpreted²⁷. Transformative or reflective learning occurs when these assumptions are analysed, challenged and potentially changed, through (for example) contact with unfamiliar perspectives or values in dialogue with others.

According to these approaches, learning is possible when the ability to interpret a phenomenon fails or is incomplete due to a lack of relevant experiences, meanings, evidence or beliefs in one’s frame of reference. Learning occurs when the frame of reference is changed to accommodate new phenomena. A critical distinction here is between ‘points of view’ and ‘habits of mind’; while both are products of experience and ‘cultural assimilation’, the former are more transient than the latter²⁸. In the process of problem solving new points of view are adopted and tested to interpret an unfamiliar phenomenon or idea—in this sense a person’s point of view may frequently change, whereas a habit of mind consists of the cultural and personal ‘backdrop’ which cannot be abandoned or escaped in the act of interpretation²⁹. Transformative learning occurs when both points of view and habits of mind are changed, meaning the entire frame of reference is critically assessed to identify and change the assumptions underlying interpretation.

Transformative learning becomes reflective when the “legitimacy of other sources of knowledge and the perspectives of other actors” is acknowledged³⁰, meaning the learner’s frame of reference or assumptions are opened to modification. To understand this idea a

²² Saari and Kallio, “Developmental Impact Evaluation for Facilitating Learning in Innovation Networks”, 228.

²³ Mittelstadt, Stahl, and Wakunuma, *SATORI Deliverable D12.1 - Good Practice in Evaluation, Reflection and Civil Society Engagement*.

²⁴ Saari and Kallio, “Developmental Impact Evaluation for Facilitating Learning in Innovation Networks”, 231.

²⁵ cf. Gadamer, *Truth and Method*; Heidegger, *Being and Time*.

²⁶ Mezirow, “Transformative Learning”.

²⁷ Chilvers, “Reflexive Engagement?”.

²⁸ Mezirow, “Transformative Learning”.

²⁹ Ibid.; Gadamer, *Truth and Method*; Patterson and Williams, *Collecting and Analyzing Qualitative Data: Hermeneutic Principles, Methods and Case Examples*.

³⁰ Chilvers, “Reflexive Engagement?”.

distinction is required between reflection and reflexivity: reflection requires that attention be given to the “broadly salient attributes of the objective in question” including consideration of alternatives and consequences, whereas reflexivity requires “self-awareness and self-reflection that comes from recognizing that attributes of the subject construct and condition the object”³¹. In terms of learning, reflective learning becomes reflexive when the learner is willing to question the frame of reference behind his interpretations and not only those of others. Therefore, reflective learning, when practiced reflexively, is inherently critical because it requires analysis of the basic, potentially unacknowledged beliefs, values, judgments and structures which frame interpretation of the world.

When operationalized into criteria for evaluation of participatory processes and MMLs in general, a reflexive approach to learning requires self-critical thought and questioning, ideally among all those involved in a participatory discourse, in particular the project partners responsible for interpreting and presenting findings based on the perspectives of other stakeholders. Reflexivity, then, may need to be practiced through explicit reflection events in which dialogue between partners is constructed so as to require decisions made across the consortium to be subjected to critical questioning. Reflexivity would require participants in the dialogue to openly acknowledge “their underlying assumptions, motives, and commitments relating to the forms of public dialogue they orchestrate or are exposed to.” Such admissions would require a degree of openness and humility from participants, particularly in admitting which stakeholders have been excluded as irrelevant³², and how the desired outcomes of the participatory processes serve their interests.

While simple to describe, such reflexivity is likely difficult to achieve in practice, particularly when openness reveals weaknesses or biases that could undermine one’s position and the credibility of the outcomes of the process³³. Reflexivity requires a considerable effort in critical self-questioning (akin to philosophical inquiry), which raises questions over the appropriateness of requiring reflexivity among participants in participatory research. In particular, the limited time in which participants are typically involved in a participatory process (for instance, a focus group) likely creates a practical barrier to reflexivity. As a result, for the purposes of SATORI the quality of participatory processes will not hinge upon the reflexivity implicit in the process; reflexivity will be recognised as desirable or ‘above and beyond’ what can be expected, but not required for ‘high-quality’ stakeholder engagement. Instead, transformative learning will be used as a benchmark, evidenced by changes to behaviours and perspectives of stakeholders achieved through contact and dialogue with alternative perspectives and frameworks of understanding.

This type of learning is ‘mutual learning’ in the sense that participants learn from one another. According to one MML partner encountered in D12.1, the mutual learning aspect of MMLs is “not an expert talking to an audience, but let us say it is one of the created forms of debate, wherein knowledge from various perspectives are mutually exposed to one another.” Another commented that it is important that “all partners feel they’re contributing to the project equally but in different ways,” referring to the necessity of two-way learning rather than one-

³¹ Stirling, “9. Precaution, Foresight and Sustainability: Reflection and Reflexivity in the Governance of Science and Technology”, 226.

³² Chilvers, “Reflexive Engagement?”, 300.

³³ It is worth explaining why a social learning approach was not adopted. Social learning follows on from transformative and reflective learning, but further requires participants in a dialogue to adopt collective interests and pursue a collective good. Given the definition of MMLs provided in D12.1, such a ‘collective interest’ and focus on communitarian problem solving is not clearly required (although it is certainly possible, for example in MMLs concerning ‘global’ problems such as ocean pollution; see: D12.1 Appendix 1).

way information exchange. This suggests the success of learning in the project can be assessed by the exchange of concepts and ideas in both directions in the connections and dialogues between stakeholders fostered by the MML. Following on from such ‘mutual learning’, transformative learning theory suggests that by expanding the participants awareness of other stakeholders and alternative perspectives/interests, they may be ‘mobilised’ or more likely to collaborate with (or at least consider) stakeholders with whom they are newly brought into contact via stakeholder engagement activities, and the interests these stakeholders represent beyond their pre-existing concerns. Transformative learning theory therefore provides an appropriate framework for evaluating stakeholder interactions within MMLs, as it is responsive to the two key aims of MMLs: ‘mutual learning’ and ‘mobilisation’.³⁴

Before translating these considerations into a set of evaluative criteria for SATORI, it must be noted that the same practical limitations on reflexivity noted above cannot be said to exist for the SATORI consortium; indeed, as researchers who regularly examine bias (seen for example in research methods discourses), reflexivity may be more reasonably expected. For SATORI, while the quality of the project in terms of participant experiences, impact, etc. will only be evaluated in terms of transformative ‘mutual learning’, the quality of inter-consortium dialogue will be evaluated in terms of (reflexive) reflective learning. The same arguments for reflective learning as ‘mutual learning’ and ‘mobilisation’ apply, with the added expectation that researchers have greater abilities (or responsibilities) for self-criticism (or reflexivity).

³⁴ At this stage of planning the evaluation it is worth considering the difficulty of translating such a theory into particular evaluation activities for measuring learning. As noted by participants in the D12.1 study, evaluation of mutual learning as “very difficult” because “learning is a continual process,” meaning it is difficult to attribute attitude and behaviour change to specific project influence and activities. Additionally, “measuring it is really difficult, because things are slow to change, and obviously, changing one person can catalyse change in others through a social process that isn't actually learned first-hand.” Furthermore, mutual learning can be ‘implicit learning’ in the sense that it occurs without the explicit recognition or effort of the learner. It is therefore more than the mere exchange of ideas between two parties, and very difficult to measure in evaluation. As a result it is important to view the criteria developed to assess mutual learning as an ideal which the actual evaluation may be unable to meet due to the practical difficulties of tracking learning (e.g. changes in behaviour and perspectives) over time. Self-assessment activities by stakeholders before, during and after stakeholder engagement events may, however, offer a way forward. Additionally, changes to ‘mindset’ may be feasible, where mindset is understood as understanding, beliefs, aims, questions asked and ways to address and answer those questions. Mindset change may be evident when partners reassess the feasibility of objectives to achieve during the project, as well as the overall aims of their involvement.

6.2 ANNEX B: PRINCIPLES OF GOOD PRACTICE

D12.1 described 21 principles of good practice in evaluation and reflection of MMLS, including five principles related to evaluation criteria:

6.2.1 Criteria Principles

1. Evaluative criteria should be specified according to the context of the particular MML, including potentially engaging the consortium to identify appropriate discipline-specific or task-specific criteria for particular activities and deliverables.
2. Evaluation should address the ‘generic’ qualities of participatory processes such as those areas of consensus in evaluation literature identified by Chilvers³⁵. Evaluation should also address impacts and evidence which demonstrate that key MML activities and desired outcomes have been realised—mutual learning and the facilitation of collaboration and cooperation among stakeholders—using criteria and typologies such as those specified by Haywood & Besley³⁶ and Walter et al.³⁷.
3. The success of an MML should be ‘stakeholder oriented’, meaning evaluative criteria should be linked to factors such as the reaction of stakeholders to engagement events, the new connections established between engaged stakeholders for communication and collaboration, the effectiveness of training in building capacities, and the empowerment of underrepresented groups in MML and societal discourses.
4. Project management should be evaluated, meaning that objectives, milestones and deliverables are delivered on time and of acceptable quality according to how they are defined in the Description of Work (DoW).
5. The ability of the MML to get target stakeholder groups in attendance at engagement events may be used as an evaluative measure.

Eight ‘methodology’ principles were identified which are relevant to the principles and criteria that will guide SATORI’s evaluation strategy:

6. In general evaluation should aim to assist in developing research activities during the life of the project (e.g. through feedback from evaluators to partners), improve the design of future related activities, assess project impact³⁸, and provide stakeholders

³⁵ Chilvers, “Reflexive Engagement?”.

³⁶ Haywood and Besley, “Education, Outreach, and Inclusive Engagement”.

³⁷ Walter et al., “Measuring Societal Effects of Transdisciplinary Research Projects”.

³⁸ Research Councils UK, *Evaluation: Practical Guidelines - A Guide for Evaluating Public Engagement Activities*.

with a better idea of the value of their participation by tracking influence on the process³⁹. MML evaluation should, at a minimum, seek to meet these three generic aims.

7. Evaluation should consider data beyond the deliverables, including stakeholders in assessing the quality of dialogue facilitated by the project wherever possible. This approach is necessary because fairness, competence and learning all have an implicit component of subjectivity, requiring the perspectives of participants (or ‘learners’) to be collected and assessed.
8. Despite methodological and epistemic difficulties, an explicit method for evaluating societal impact should be adopted or designed, with particular attention paid to evidence of mutual learning (e.g. changes in stakeholder perspectives, beliefs and actions).
9. The evaluation process should be conducted transparently for the benefit of the consortium, including identifying its scope (e.g. summative/formative, technical/holistic) and the position of the evaluator in relation to the consortium (e.g. internal, external, and independent) as early as possible. This approach will help reduce resistance to recommendations made by the evaluators.
10. The entire consortium should be involved in providing data for evaluation beyond writing deliverables (e.g. interviews, surveys, reflective meetings, etc. conducted with consortium partners). Broad engagement allows for assessment of mutual learning between project partners.
11. Initial templates or indicators of success created with consortium input should be created prior to the start of each research task, and potentially added to or revised according to challenges faced. This approach can ensure that discipline-specific perspectives inform the assessment of the success or quality of project activities while being responsive to the practical challenges of engagement.
12. A clear ‘endpoint’ should be specified at which point project impacts can start to be identified and evaluated.
13. Evaluation should occur before, during and after the project to ensure all processes and impacts are evaluated to some degree.

Five further mutual learning principles were identified which are relevant to assessing the quality of mutual learning occurring within an MML, and thus the criteria necessary to qualify mutual learning:

6.2.2 Mutual Learning Principles

14. Data collection and analysis methods conducive to evaluating learning or attitudinal change over time should be employed in evaluation, meaning explicit and implicit evidence of mutual learning should be sought in evaluation by asking project partners

³⁹ Rowe and Frewer, “Evaluating Public-Participation Exercises: A Research Agenda”.

and participants to reflect on changes to their attitudes and behaviours caused by participating in the project and engaging with unfamiliar ideas and perspectives.

15. Mutual learning outcomes among project participants should be assessed, for example by monitoring changes in participant perspectives, beliefs and actions over time. Mutual learning conceived of as societal impact can also be evaluated according to the extent to which project outputs have reached and influenced them.
16. In evaluating the quality of mutual learning that has occurred, the possibility of mutual learning without absolute consensus should be recognised.
17. A participatory approach to evaluation conducive to mutual learning between stakeholders and project partners should be used. The appropriate degree of stakeholder involvement, from designing to carrying out the evaluation and reporting on its findings, must be decided on a project-specific basis according to the willingness of the stakeholders and the expertise required to perform the evaluation.
18. A reflexive account of the conception of mutual learning adapted should be provided, including its theoretical basis (where appropriate), and criteria for evaluating mutual learning should be consistent with the theoretical approach taken.

Finally, a further three principles concerning the role of reflection in a MML were described which suggest criteria for evaluating the quality of reflection in SATORI:

6.2.3 Reflection Principles

19. The evaluator transparently should report on perceived pressures and influence of project partners in the evaluation to identify, as far as possible, influence on the evaluation outcomes.
20. When conducting a formative evaluation, the evaluator should provide critical feedback and recommendations to the consortium to improve ongoing research activities.
21. The evaluator, coordinator and/or work package leaders should encourage partners to critically reflect on their progress and changes to attitudes and behaviours (e.g. implicit learning) through formal or informal methods such as interviews, project management meetings, or peer review of deliverables.

Short of prescribing specific evaluative actions or criteria to be adopted, these principles are intended to provide guidance for consortia in planning evaluation and reflection activities. The principles are therefore broad enough for specification into criteria responsive to the unique characteristics of individual MMLs. Accordingly, the next step in applying the principles is to select and specify certain ones into a project-specific set of principles and evaluative criteria, before designing or choosing evaluation tools and methods appropriate to assessing the project along these lines. Specifically, the next step of the planning undertaken by DMU in WP12 is to identify which principles are relevant to engaging stakeholders around ethical assessment frameworks, and to specify evaluative criteria on this basis in Task 12.2. An evaluation and reflection strategy can then be created in Task 12.3. Task 12.4, in which an evaluation strategy built upon these principles is applied in evaluating the success of the

SATORI project, is therefore a first step towards empirical validation of the principles recommended in this report.

6.3 ANNEX C: EVALUATION PRINCIPLES AND CRITERIA

The criteria described here are derived from the findings of D12.1, where two evaluative frameworks⁴⁰ were recommended as ideally suited to assessing the quality of MMLs. Specifically, they follow on from a key principle of good practice in MML evaluation identified in D12.1: MMLs should ensure evaluation addresses ‘generic’ evaluation criteria for participatory processes such as those identified by Chilvers (2008) (representativeness and inclusivity, fair deliberation, access to resources, transparency and accountability, learning and efficiency), while also including criteria to assess the impacts and evidence of mutual learning and the facilitation of collaboration and cooperation among stakeholders, such as those specified by Haywood & Besley (2013) or Walter et al. (2007) (network building, trust in others, understanding of others, community identification, distribution of knowledge, system knowledge, goal knowledge, transformation knowledge).

The following sub-sections detail how these frameworks can be applied to evaluate the quality of the different types of activities as described in Table 3, identifying specific criteria that will be implemented in the evaluation and reflection strategy described in D12.3. These criteria operate on the assumptions regarding the defining characteristics and purposes of MMLs established in D12.1: specifically, that mutual learning and mobilisation are primary ends and therefore key components of the ‘quality’ of MMLs. Justification for these assumptions are not repeated here, but can be found in D12.1.

An overarching theme in the criteria stems from the purpose of SATORI, understood as the development of an ethics assessment framework. The method of development and application of the framework across other project activities and beyond the consortium will therefore be assessed, for example in terms of how project activities have supported the adoption or further development of the framework by relevant non-consortium stakeholders.

For example, criteria address questions of quality and feasibility, the prospects for incorporation of the ethical framework into the EU system of research and innovation, support

⁴⁰ Haywood and Besley, “Education, Outreach, and Inclusive Engagement”; Chilvers, “Reflexive Engagement?”.

for transnational co-operation in ethical assessment provided by the project, and the contribution to sustainable and inclusive solutions to key challenges facing European society made by the framework and supporting project activities. These activities are ways of specifying the concepts ‘mutual learning’ and ‘mobilisation’ for SATORI; for example, evaluation would be concerned with the ‘mobilisation’ of stakeholders in terms of encouragement within the various social and professional networks engaged in the project to adapt the framework produced, or in terms of adaptation of the framework for feasible adoption within the various organisational and disciplinary contexts seen in recruited stakeholder ‘groups’. As SATORI demands an intensive and prolific deliberation among participants (in line with being a MML), criteria focusing on the ‘fairness’ and ‘competence’ of these processes, including the quality of debate facilitated and the representativeness of stakeholders engaged, will likely feature in the final set of evaluation and reflection criteria⁴¹.

Each of the above examples concerns the quality of tasks (and underlying methods) completed by the consortium. However, criteria also concern the impact of SATORI as a whole and the impact of the ethical assessment framework on citizens and society. For the latter, DMU will coordinate with (or receive data ‘from’) VTT (leaders of WP6), who will be assessing the impacts of the ethical assessment framework in particular, throughout the evaluation.

6.3.1 Principles and Criteria for Evaluating Stakeholder Engagement

D12.1 concluded that ‘mutual learning’ is a core concept in evaluating the quality of an MML. Starting from this assumption, a primary task of this deliverable is to develop not only traditional evaluation criteria for the project, but also criteria for assessing mutual learning (and therefore, reflection) within SATORI. Stakeholder engagement activities can be understood as the locus of MMLs which directly facilitate dialogue and network-building (or mobilisation) between stakeholders (and partners). The purpose of stakeholder engagement is to encourage mutual learning and mobilisation between all those involved (stakeholders as well as partners) (see: D12.1). As explained above (see: Section **Error! Reference source not found.**), these concepts can be operationalised in terms of the degree of exchange of perspectives between those involved, relationship or network building, etc. The criteria described here can be seen as indicators of high quality mutual learning and mobilisation in stakeholder engagement activities, achieved primarily by establishing a fair and competent dialogue⁴².

Several criteria stem from the Rowe & Frewer framework⁴³, which addresses the fairness and acceptability of participatory processes in research. Following on from Webler, these criteria focus on *fairness*, which concerns the “perceptions of those involved in the engagement exercise and/or the wider public, and whether they believe that the exercise has been honestly conducted with serious intent to collect the views of an appropriate sample of the affected population and to act on those views,” and *competence* (or for Rowe & Frewer, acceptance), which concerns “the appropriate elicitation, transfer, and combination of public and/or sponsor views.” The latter of the two is synonymous with effectiveness as used by Rowe & Frewer, understood as “maximizing the relevant information from the maximum number of all relevant sources and transferring it (with minimal information loss) to the other parties,

⁴¹ See: SATORI, *THEME [SiS.2013.1.2-1 SiS.2013.1.2-1] Mobilisation and Mutual Learning (MML) Action Plans: Mainstreaming Science in Society Actions in Research] - Annex 1 - Description of Work*, 43.

⁴² cf. Rowe and Frewer, “Public Participation Methods”; Webler, “‘Right’ Discourse in Citizen Participation: An Evaluative Yardstick”.

⁴³ Rowe and Frewer, “Public Participation Methods: A Framework for Evaluation”.

with the efficient processing of that information by the receivers (the sponsors and participants) and the combining of it into an accurate composite.” The competency or effectiveness of a public participation process can therefore be disrupted when information is suboptimal, distorted, inaccurate, incomplete, or when it is misunderstood by participants⁴⁴, perhaps due to a mismatch between the language or concepts in the information and the background expertise of the participants. Criteria derived from these authors for SATORI are intended to ensure stakeholder engagement events are fair and accessible to affected stakeholder groups, which (as discussed in D12.1) can be linked to a concern for a ‘fair discourse’ in MMLs⁴⁵. The criteria include:

- ‘Representativeness’ – The degree to which participants in stakeholder engagement events “comprise a broadly representative sample of the population of the affected public,” including both demographic and organisational groups. For SATORI, such groups will include representatives of multiple disciplines to whom the ethical assessment framework is aimed, as well as participants in research potentially subjected to review under the framework. This criterion aims to ensure that participants are exposed to a range of relevant perspectives both in making decisions within the participatory process, and to inform subsequent decision-making beyond the research project. Without a representative sample MMLs may suffer from insular and circular dialogue, in which individuals with similar backgrounds and perspectives merely reinforce one another’s views, thereby limiting the potential for transformative ‘mutual learning’. Where particular stakeholder groups are named by the organisers of an event, representativeness will be based on whether these stakeholders were actually present, as well as whether a clear bias is evident within the sample (e.g. ethics review panels only from the UK, despite the EU-wide scope of the framework). This requirement correlates with principles that suggest that evaluation should address the ‘generic’ qualities of participatory processes such as those areas of consensus in evaluation literature identified by Chilvers⁴⁶ and relatively, that a participatory approach to evaluation conducive to mutual learning between stakeholders and project partners should be used with an appropriate degree of stakeholder involvement, from designing to carrying out the evaluation and reporting on its findings.
- ‘Transparency’ – The degree to which the decision-making processes within the engagement event are transparent to stakeholders, including making the biases or underlying assumptions of the organisers or decision-makers known as far as possible to participants. This criterion is necessary to maximise the mobilising effect of the event, as stakeholders may be less likely to ‘champion’ the outputs of the event if it is unclear whether their interests have been taken seriously. This correlates with the criterion principle in D12.1 that suggests that the evaluation process should be conducted transparently for the benefit of the consortium, including identifying its scope (e.g. summative/formative, technical/holistic) and the position of the evaluator in relation to the consortium (e.g. internal, external, independent) as early as possible which helps reduce resistance to recommendations made by the evaluators.

⁴⁴ Rowe and Frewer, “A Typology of Public Engagement Mechanisms”, 262–3.

⁴⁵ cf. Habermas, *The Theory of Communicative Action: Volume 1: Reason and the Rationalization of Society*.

⁴⁶ Chilvers, “Reflexive Engagement?”.

- ‘Accessibility’ – Concerns the availability of relevant learning and information materials to the participants, so as to allow for an informed dialogue between stakeholders⁴⁷. Further, any recognised barriers to participation should be removed as far as possible⁴⁸ to ensure the resources are understandable for participants, for example language or comprehension of technical language. This criterion can be used to justify the quality of outcomes of a participatory process—the quality of judgments produced in consultation with stakeholders can be assessed for necessary understanding of a technology, technique or other relevant field of expertise; for example, do stakeholders know enough about the topic to take their views seriously? Procedurally, the arrangements made to support stakeholders in the processes can be assessed; for example, was sufficient information made available to ensure stakeholders can learn and become competent participants? It also includes a time aspect; e.g. were participants given materials near to the start of the event, or given enough time during the event to read, comprehend and ask questions? The evaluation will be dependent upon the need for such materials in the event.
- ‘Task Definition’ – Concerns the clarity of instructions or guidance given to the participants. Were participants made aware of the scope, purpose and their role in the event?

Other criteria not derived directly from Rowe & Frewer, but representing similar concerns with fairness and acceptability include:

- ‘Fair Deliberation’ – Concerns the degree to which participants are allowed to “enter the discourse and put forward their views in interactive deliberation that develops mutual understanding between participants”⁴⁹. Following on from the transformative approach to mutual learning described above with reference to ‘Representativeness’, this criterion adds a further dimension by emphasising the quality of the process in which the sample participates. Specifically, mutual learning may be hampered if dialogues are dominated by a particular perspective or stakeholder group, as it limits opportunities for learning from interaction with multiple unfamiliar perspectives and phenomena. As suggested by Chilvers⁵⁰, “While recognizing the role of consensus, the deliberative process should emphasize diversity and difference through representing alternative viewpoints, exploring uncertainties, and exposing underlying assumptions.” In evaluating according to this criterion, events may be assessed in terms of how they are structured so as to minimise opportunities for particular ‘outspoken’ stakeholders to dominate the dialogue (which can be inhibited by a moderator), or opportunities for one-to-one dialogues between stakeholders. For SATORI, it is important to note that ‘feedback’ style events may be particularly vulnerable to emphasis on the views of the organisers which will be implicit in tools or frameworks presented for review to the stakeholders.

⁴⁷ Rowe and Frewer, “Public Participation Methods”; Haywood and Besley, “Education, Outreach, and Inclusive Engagement”.

⁴⁸ Chilvers, “Deliberating Competence Theoretical and Practitioner Perspectives on Effective Participatory Appraisal Practice”.

⁴⁹ Ibid.

⁵⁰ Ibid.

- ‘Iteration’ – The degree to which dialogue and topics within the event are revised and responsive to the “needs, issues and concerns expressed by participants”⁵¹. Practically speaking this criterion can be assessed in terms of opportunities for stakeholders to propose topics for dialogue outside of a pre-defined framework of events and topics⁵². Again, this criterion emphasises opportunities for encountering multiple unfamiliar perspectives and phenomena to encourage mutual learning.
- ‘Criticalness’ – Rather than referring to questioning of participant claims by the organisers, this criterion refers to the degree to which participants are encouraged to challenge and negotiate with the perspectives and values of other members of the dialogue⁵³. This willingness to question the claims of others is critical to learning about the underlying assumptions and values held by others, both of which are necessary for transformative learning. Practically speaking, these qualities could be assessed in terms of the effectiveness of the ‘rules’ established for discourse in the process or learning materials provided to participants⁵⁴, and the degree to which this structure contributed to a high quality discourse in which social or mutual learning occurred between participants (and organisers). This is not to say participants must disagree with one another or come to a consensus, but rather that the act of questioning itself has value for the learning process. This criterion is further seen as necessary to emphasise the empowerment of participants to question the claims of others and prevent dominance of the discourse by a particular view or stakeholder⁵⁵. According to transformative learning, mutual learning can only occur when participants do more than merely agree or disagree—respect for alternative views and trust in the integrity of others is required according to which the participant feels compelled to offer reasons and counter-arguments⁵⁶. These requirements suggest specific requirements to be met in participatory discourses when mutual learning is conceived of as a type of transformative learning; specifically, participants should be ‘open-minded’ meaning they are willing to consider the views of others as legitimate, and should be seen to offer reasons of support and criticisms of particular views rather than mere opinions or ultimatums. Power relationships within a discourse need also be considered, as the perception of authority or favouring by facilitators of the views of a particular stakeholder can undermine trust among participants, respect for other views, and the overall perception of a fair discourse⁵⁷ conducive to transformative learning.
- ‘Ownership of Outcomes’ – The degree to which participants are engaged in disseminating results of the event, and thus feel that they ‘own’ or ‘control’ the decisions reached (or other outputs)⁵⁸. Ownership is seen to encourage mobilisation of stakeholders and ideas (or dissemination of ideas) by encouraging a feel of ‘belonging’ to the process and ‘responsibility’ for its outcomes.

⁵¹ Ibid.

⁵² Abelson et al., “Deliberations about Deliberative Methods”.

⁵³ Haywood and Besley, “Education, Outreach, and Inclusive Engagement”.

⁵⁴ Webler, Kastenholz, and Renn, “Public Participation in Impact Assessment”, 456.

⁵⁵ Ibid., 443.

⁵⁶ e.g. Stagl, “Multicriteria Evaluation and Public Participation”.

⁵⁷ Ibid.

⁵⁸ Haywood and Besley, “Education, Outreach, and Inclusive Engagement”.

- ‘Participant Satisfaction’ – To avoid an entirely theoretical approach to evaluation, assessing the participants’ satisfaction with the event can ensure aspects of the event not covered by the above criteria are still considered in the evaluation. Simply put, asking a participant what they found ‘good or bad’ about a particular engagement event, and to explain why, encourages identification of unforeseen aspects of the event linked to its quality. This is suggested by a criteria principle that proposes that data collection and analysis methods conducive to evaluating learning or attitudinal change over time should be employed in evaluation, meaning explicit and implicit evidence of mutual learning should be sought in evaluation by asking project partners and participants to reflect on changes to their attitudes and behaviours caused by participating in the project and engaging with unfamiliar ideas and perspectives. In terms of the aims of MML, participant satisfaction can be reasonably linked with the mobilising effects of the event, similar to ‘Ownership’.

Each of these criteria is intended to assess the quality of stakeholder engagement events, understood both procedurally and substantively. These criteria may further suggest impacts of the events, as discussed below.

6.3.2 Principles and Criteria for Evaluating Recruitment

Several of the aforementioned criteria can be used to specifically assess the quality of recruitment in a MML. The ‘Representativeness’ of the stakeholders engaged throughout the project, not only in stakeholder engagement events, is obviously relevant. Furthermore, in terms of ‘Accessibility’ and ‘Criticalness’, enhancing the ‘voice’ of underrepresented stakeholders by ensuring equal representation in the project suggests that MMLs should be empowering stakeholders involved, not only through capacity building and learning but by ensuring traditionally underrepresented stakeholders are involved in the discourse. These latter criteria can be considered generic for the MML mechanism as it is designed to tackle large societal challenges which involve a multitude of stakeholders, some of whom will inevitably, as in any discourse, have less influence or power than others⁵⁹. The criteria mentioned above relates to a criteria principle established in D12.1 which suggests that the success of an MML should be ‘stakeholder oriented’, meaning evaluative criteria should be linked to factors such as the reaction of stakeholders to engagement events, the new connections established between engaged stakeholders for communication and collaboration, the effectiveness of training in building capacities, and the empowerment of underrepresented groups in MML and societal discourses and the MML should be able to get target stakeholder groups in attendance at engagement events which can later be used as an evaluative measure.

6.3.3 Principles and Criteria for Evaluating Surveys, Interviews and Case Studies

Evaluation of these empirical activities will likely consist solely of peer-review of deliverables by partners, as suggested in the DoW. Empirical activities are referring to research activities and one of the criteria principles in D12.1 suggests that evaluation should aim to assist in developing research activities during the life of the project (e.g. through feedback from evaluators to partners), improve the design of future related activities, assess

⁵⁹ cf. Foucault, *Discipline & Punish*; Habermas, *The Theory of Communicative Action: Volume 1: Reason and the Rationalization of Society*.

project impact, and provide stakeholders with a better idea of the value of their participation by tracking influence on the process. In addition, indicators of success for evaluating the findings and methods of surveys, interviews and case studies will need to be developed by the partners conducting these activities, so as to ensure a fit between the partner's disciplinary background and the evaluation. This calls for creation of initial templates or indicators of success with consortium input prior to the start of each research task, and potentially adding to or revising according to challenges faced. This approach can ensure that discipline-specific perspectives inform the assessment of the success or quality of project activities while being responsive to the practical challenges of engagement.

Broadly speaking, quality of these activities can be described in terms of three criteria derived from the 'Quality and Rigor' criteria proposed by Haywood and Besley⁶⁰, understood as the "degree to which the project is perceived as rigorous and credible. This may be peer-review for project leaders or the plausibility of research recommendations for citizens." Specifically, these empirical activities will be evaluated according to (1) 'Methodological Rigour', meaning the empirical instrument is informed by a sound methodological basis and philosophical paradigm—these are a minimum of good practice in any type of research, (2) 'Credibility', meaning a strong connection is established between the data or interpretations of the data (as well as any theoretical or conceptual frameworks necessary to understand the data) and the findings or recommendations produced, and (3) 'Transparency', or the presence of sufficient detail in describing and analysing the data for the reader to follow the authoring partner's line of reasoning⁶¹.

6.3.4 Principles and Criteria for Evaluating Recommendations/Tools

The quality of tools, frameworks and recommendations produced by SATORI can be linked in part to the quality of stakeholder engagement events as discussed above. Specifically, recommendations/tools can be assessed in terms of the degree to which the recommendations or tools produced by the consortium reflect or address the concerns raised by stakeholders⁶², even where these concerns are rejected or otherwise dismissed. Practically, tools/recommendations can be compared with event minutes and reports. In the pursuit of mutual learning, deliverables should demonstrate at a minimum that stakeholder interests have been considered. This requirement is partially covered by the 'Transparency' criteria above in terms of decision-making processes preceding recommendations/tools; however, acknowledgement of all stakeholder values/perspectives encountered in such events which inform recommendations/tools can be described in terms of 'Relevance'. This satisfies the criteria principle that suggests that the possibility of mutual learning without consensus should be recognised when evaluating quality of mutual learning that has occurred.

6.3.5 Principles and Criteria for Evaluating Dissemination/Impact

As noted in D12.1, impact is fundamentally difficult to evaluate due to a lack of causal links between project outputs and future policy/decisions, as well as the restrictive timeline of projects versus future impacts. With this in mind, a clear 'endpoint' should be specified at which point project impacts can start to be identified and evaluated. Impact is understood here as "positive and negative, primary and secondary long-term effects produced by a

⁶⁰ Haywood and Besley, "Education, Outreach, and Inclusive Engagement".

⁶¹ See also: Kuper, Lingard, and Levinson, "Critically Appraising Qualitative Research"; Guba and Lincoln, *Naturalistic Inquiry*; Finlay, "'Rigour', 'Ethical Integrity' or 'Artistry'?"

⁶² OECD, "Evaluation of Development Programmes"; Haywood and Besley, "Education, Outreach, and Inclusive Engagement".

[project], directly or indirectly, intended or unintended”⁶³. It may initially be easier to speak of impact in terms of dissemination, which can refer to the spread of ideas, knowledge, tools, deliverables, publications and other outputs of a project. One possible criterion for measuring dissemination (and thus impact) is ‘Quantity’, which refers to the number of times project outputs are referenced in journal articles, conference proceedings, policy documents, news stories. Quantity alone is however a very crude measurement of dissemination and impact, as the mere mention of a SATORI output does not necessarily indicate uptake of the ethics assessment framework or prove that the project’s activities have influenced policy and the behaviour of stakeholders. ‘Influence’ can therefore be suggested as a complementary criteria, referring to the uptake of project ideas and tools in public and policy discourses which consortium partners expect to influence through their involvement in SATORI. This criteria would therefore require specification in the ‘Indicators of Success’ questionnaire through partners identifying specific discourses and channels where they hope to see the framework and other outputs adopted. Furthermore, it is likely only assessable through self-assessment by partners and review of news items or forthcoming policy where a direct connection can be established between (a partner in) the decision-making process behind the relevant policy or decision.

Both transformative and reflecting learning (see: Section **Error! Reference source not found.**) are useful in assessing the impact of research as a form of learning, where the research is seen as contributing unfamiliar perspectives or phenomena to be interpreted by the learner. Impact evaluation guided by this approach would require methods capable of assessing the quality of such dialogues between stakeholders whose behaviours, attitudes or mindset have been affected by participating in or coming into contact with the project. The impacts of the project can therefore be seen through changes to the participants once the project has ended⁶⁴. It follows that a link can be drawn in evaluation between *potential* impacts of the project according to the quality of outputs and events which introduce participants to unfamiliar phenomena in such a way that co-creation of knowledge is encouraged, for example through dialogue with researchers rather than a non-interactive presentation of results. Such evaluation of impact focuses on *predicting* impact rather than retrospectively assessing it, which may address practical difficulties (such as those noted in D12.1) with evaluating the impact of a project during its lifespan thus evaluation should occur before, during and after the project to ensure all processes and impacts are evaluated to some degree. This approach will be adopted in SATORI, meaning that a direct link is made between the quality of participatory processes and (publicly available) deliverables and the impact of the project.

A criterion of impact following on from this is ‘Behaviour Adjustment’, understood as the extent to which capacities, skills and knowledge have been built among stakeholders through training and dialogue and become evident in behaviour and lifestyles following participation in the project⁶⁵. An objective measure for this criterion is unlikely to exist due to the difficulty of ‘measuring’ behavioural change noted in D12.1; instead, the criterion is intended as a ‘lens’ for the subjective assessment (or perhaps self-assessment) of subsequent stakeholder behaviours. An identical criterion for change in perspectives or opinions, ‘Perspective Adjustment, will also be used. Both criteria are proposed to focus the evaluation on evidence of exchange of perspectives or ‘mutual learning’ among stakeholders, under the

⁶³ OECD, “Evaluation of Development Programmes”.

⁶⁴ For such a methodology, see: Saari and Kallio, “Developmental Impact Evaluation for Facilitating Learning in Innovation Networks”.

⁶⁵ Haywood and Besley, “Education, Outreach, and Inclusive Engagement”.

assumption that perspectives and behaviours will only change if ‘serious consideration’ is given to alternative views and interests, and that such change is a form of learning. This is in line with the principle that suggests that mutual learning outcomes among project participants should be assessed, for example by monitoring changes in participant perspectives, beliefs and actions over time. Mutual learning conceived of as societal impact can also be evaluated according to the extent to which project outputs have reached and influenced them.

Impact will also be measured in terms ‘Network Expansion’, or networking between stakeholders facilitated by the project, which may be measurable through self-assessments of changes to a participant’s social or professional networking following stakeholder engagement events. This has previously been described by Haywood & Besley as the “degree to which the project facilitates new networks and relationships among project members or reinforces existing bonds”⁶⁶. Equating quality with an increase in linkages among involved organisations also provides a useful way to assess the success of network building facilitated by an MML with quantitative measurement. On this basis MMLs can be said to be more successful as they increase the links or channels or communication between stakeholders involved. This relates to a criteria principle that propose the involvement of the entire consortium in providing data for evaluation beyond writing deliverables (e.g. interviews, surveys, reflective meetings, etc. conducted with consortium partners). Broad engagement allows for assessment of mutual learning between project partners.

As suggested in D12.1, it may be possible to evaluate the likelihood of future impacts according to the quality of the procedures through which mutual learning occurs. Impact should be distinguished from outputs and outcomes—impact refers to the mid- and long-term effects of R&D on society and ongoing research, measured for example in terms of societal products, use and benefits, or social, cultural, environmental and economic returns⁶⁷. In contrast, outputs and outcomes can be understood as short-term products, processes, recommendations and knowledge created by a project, such as meetings, reports and other deliverables^{68,69}. A basic distinction can be drawn between ‘academic’ impacts via publications and presentation of results, and ‘societal impacts’ wherein policy, social attitudes or behaviours outside the project are influenced⁷⁰. One of the criteria to consider when evaluating impact is to create initial templates or indicators of success created with consortium input prior to the start of each research task, and potentially add to or revise according to challenges encountered. This approach can ensure that discipline-specific perspectives inform the assessment of the success or quality of project activities while being responsive to the practical challenges of engagement. The evaluation of impact is therefore inherently connected to the quality of the project activities preceding the impact. For stakeholder engagement activities, impact can also be seen through change in stakeholder behaviours, attitudes or knowledge resulting from participation⁷¹. Mutual learning evident in stakeholder engagement processes and suggested by the criteria above are therefore key to evaluating impact. As suggested in D12.1, one of the good criteria principles recommends

⁶⁶ Ibid.

⁶⁷ Bornmann and Marx, “How Should the Societal Impact of Research Be Generated and Measured?”, 211.

⁶⁸ Abelson et al., “Deliberations about Deliberative Methods”; Walter et al., “Measuring Societal Effects of Transdisciplinary Research Projects”.

⁶⁹ The use of vocabulary in the literature is not consistent (see: Section **Error! Reference source not found.**). In some sources, outputs refer to products, processes, reports and deliverables, whereas in others products and processes are ‘outcomes’ and reports and deliverables are ‘outputs’. Both terms are used here to refer to all four types of results.

⁷⁰ Penfield et al., “Assessment, Evaluations, and Definitions of Research Impact”, 21.

⁷¹ Walter et al., “Measuring Societal Effects of Transdisciplinary Research Projects”.

that an explicit method for evaluating societal impact should be adopted or designed, with particular attention paid to evidence of mutual learning. If learning is understood as a process through which actors are introduced to new experiences, perspectives and knowledge which transforms or effects their ‘framework of understanding’,⁷² or future perspective, beliefs and actions, it follows that learning outcomes may influence MML participants in the long-term. While this argument does not suggest a particular criterion for impact evaluation, it does show how procedural and substantive evaluation of stakeholder engagement during the life of SATORI can inform the assessment of impact.

6.3.6 Principles and criteria for evaluating evaluation

In a recent study Chilvers noted that evaluators of public participation exercises were dubious of the possibility of operationalising the principles in evaluation (described above) due to practical institutional, cultural, political and economic constraints. It was therefore suggested that to overcome these constraints evaluators should be increasingly critical⁷³, identifying limitations not only with the process itself but the context in which it and its evaluation occur. In D12.1 one of good criteria principles was suggested in relation to limitations and it proposes that the evaluator transparently should report on perceived pressures and influence of project partners in the evaluation to identify, as far as possible, influence on the evaluation outcomes. Limitations can stem from aspects of the process or context of evaluation, such as resistance from the consortium partners or limitations established in the DoW. If evaluations of MMLs are understood as a component of mutual learning, for example by providing critical and reflective feedback to partners to revise ongoing project activities (see: D12.1), then reporting on limitations in evaluation in the same fashion as other project activities is necessary to be able to holistically evaluate the quality of mutual learning in the MML. In this light, a criteria principle described in D12.1 indicates that the evaluator, coordinator and/or work package leaders should encourage partners to critically reflect on their progress and changes to attitudes and behaviours (e.g. implicit learning) through formal or informal methods such as interviews, project management meetings, or peer review of deliverables. On this basis the quality of evaluation can be assessed in terms of ‘Restrictiveness’, established through critical self-assessment of limitations imposed on the evaluators and evaluation by the project’s broader context, description of work, consortium, coordinator or other relevant sources.

Another potentially relevant factor, depending on the strategy developed in Task 12.3, concerns the quality of stakeholder participation in evaluation. A participatory approach to evaluation can in general be advocated in MMLs due to its conduciveness to two-way learning compared to ‘objective’ approaches to evaluation. This affirms the criteria principle that propose the use of a participatory approach to evaluation conducive to mutual learning between stakeholders and project partners. It is not yet clear whether SATORI will use a participatory approach to evaluation; however, if stakeholders are included as evaluators the ‘Representativeness’ of the participants recruited for evaluation achieves new importance, as groups not participating lack a voice in the process itself and evaluation. Careful consideration is needed of who is considered a stakeholder, understood broadly as an individual or group of individuals affected by a decision, issue or other topic of research⁷⁴.

⁷² cf. Gadamer, *Truth and Method*; Heidegger, *Being and Time*.

⁷³ Chilvers, “Deliberating Competence Theoretical and Practitioner Perspectives on Effective Participatory Appraisal Practice”, 180.

⁷⁴ cf. Renn et al., “Public Participation in Decision Making”, 190–1.

6.3.7 Principles and criteria for evaluating administration

While administration is a fact of any type of collaborative research project, it should still be evaluated in MMLs. In D12.1 two criteria principles were described that points to evaluating the administration of projects. The first principle suggests that project management should be evaluated, meaning that objectives, milestones and deliverables are delivered on time and of acceptable quality according to how they are defined in the Description of Work (DoW). Poor coordination or administration can lead to a breakdown in communication and project agendas among partners, undermining the quality of all other processes. The quality of administration and coordination can be assessed in terms of ‘Quality of Collaboration’, wherein breakdowns in communication or conflicts between partners reduce the quality of collaboration and thus jeopardise the project. The second criteria principle that relates to evaluating administration, suggests that when conducting a formative evaluation, the evaluator should provide critical feedback and recommendations to the consortium to improve ongoing research activities. It may be possible to assess this criterion through qualitative observations of partner workshops and meetings (e.g. body language, atmosphere, mood, etc.) as well as self-assessment reports by partners concerning any practical barriers to collaboration they have encountered.

6.3.8 Evaluating ‘Internal’ Activities

It should also be noted that certain administrative tasks concern inter-consortium communication and collaboration, many aspects of which are considered ‘coordination’ of the project. Such activities include consortium meetings, peer-review and informal communication, all of which were not classified in Table 3 due to not being included in a particular Work Package. Additionally, consortium interaction at stakeholder engagement events may also be key to facilitating mutual learning not only among stakeholders, but the consortium itself. Recognising this, these types of activities, which may be considered ‘Internal’ activities, will also be evaluated in terms of the mutual learning that occurs between consortium partners. Criteria concerning mutual learning specified with reference to stakeholder engagement activities will therefore also be relevant for these activities. Additionally, such activities will be assessed in terms of ‘Reflectiveness’, or the degree to which partners show respect for alternative views and trust in the integrity of other partners, which is necessary if communication and collaboration are to progress beyond mere (dis)agreements on proposed actions⁷⁵. Therefore, criteria such as using a participatory approach to evaluation conducive to mutual learning between stakeholders and project partners and deciding the appropriate degree of stakeholder involvement, from designing to carrying out the evaluation and reporting on its findings, on a project-specific basis according to the willingness of the stakeholders and the expertise required to perform the evaluation could be relevant in evaluating internal activities. As suggested by Chilvers⁷⁶, participants and partners alike will need to acknowledge “their underlying assumptions, motives, and commitments relating to the forms of public dialogue they orchestrate or are exposed to,” which requires a significant degree of openness and humility. Decisions across the consortium should be systematically subjected to reflexive (critical) questioning, perhaps through scheduled events or workshops in which progress and proposed changes are discussed. This is suggested in criteria principle described in D12.1 which proposes that the evaluator, coordinator and/or work package leaders should encourage partners to critically

⁷⁵ e.g. Stagl, “Multicriteria Evaluation and Public Participation”.

⁷⁶ Chilvers, “Reflexive Engagement?”, 300.

reflect on their progress and changes to attitudes and behaviours through formal or informal methods such as interviews, project management meetings, or peer review of deliverables.

6.3.9 Summary of Evaluation and Reflection Criteria

The criteria specified above have been developed on the basis of a literature review, empirical study and application of principles developed in Task 12.1. The set of criteria are therefore both theoretically and empirically grounded, providing an evaluative framework responsive to the particular aims and activities of SATORI based upon broader principles of good practice in MMLs. While all of the criteria specified above are relevant to the evaluation of SATORI as a whole, this does not imply that all will be included in the evaluation and reflection strategy created in Task 12.3. For example, practical reasons may exist for limiting the set of criteria used for evaluating a particular stakeholder engagement event based upon the availability of stakeholders to provide feedback, the need for learning/training materials, or the necessary limitation of ‘dialogue’ to one-way communication of information or tools/recommendations for further consideration/application by stakeholders in the future.

Table 4 summarises all of the criteria specified above, showing which criteria are relevant to which types of activities. To develop practical methods of applying these criteria in the SATORI evaluation and reflection strategy developed in Task 12.3, a further ‘Indicators of Success’ questionnaire will be carried out which may provide both further criteria for evaluation, but primarily specific metrics or measures of the criteria described here.

7 REFERENCES

Abelson, Julia, Pierre-Gerlier Forest, John Eyles, Patricia Smith, Elisabeth Martin, and Francois-Pierre Gauvin, “Deliberations about Deliberative Methods: Issues in the Design and Evaluation of Public Participation Processes”, *Social Science & Medicine*, Vol. 57, No. 2, July 2003, pp. 239–251.

Abelson, Julia, and François-Pierre Gauvin, *Assessing the Impacts of Public Participation: Concepts, Evidence and Policy Implications*, Canadian Policy Research Networks Ottawa, 2006.

https://desafios2.ipea.gov.br/participacao/images/pdfs/participacao/abelson_julia_gauvin_francois_assessing_impacts_public_participation.pdf.

Arnold, Erik, “Understanding Long-Term Impacts of R&D Funding: The EU Framework Programme”, *Research Evaluation*, Vol. 21, No. 5, December 2012, pp. 332–343.

Bergmann, Matthias, Bettina Brohmann, Esther Hoffmann, M. Celine Liobl, Regine Rehaag, Englebert Schramm, and Jon-Peter Voss, *Quality Criteria of Transdisciplinary Research: A Guide for the Formative Evaluation of Research Projects*, Evalunet, Frankfurt, 2005. http://isoe.eu/ftp/evalunet_guide.pdf.

Bornmann, Lutz, and Werner Marx, “How Should the Societal Impact of Research Be Generated and Measured? A Proposal for a Simple and Practicable Approach to Allow Interdisciplinary Comparisons”, *Scientometrics*, Vol. 98, No. 1, January 2014, pp. 211–219.

Chess, Caron, “Evaluating Environmental Public Participation: Methodological Questions”, *Journal of Environmental Planning and Management*, Vol. 43, No. 6, November 2000, pp. 769–784.

Chilvers, Jason, “Deliberating Competence Theoretical and Practitioner Perspectives on Effective Participatory Appraisal Practice”, *Science, Technology & Human Values*, Vol. 33, No. 2, March 1, 2008, pp. 155–185.

———, “Reflexive Engagement? Actors, Learning, and Reflexivity in Public Dialogue on Science and Technology”, *Science Communication*, Vol. 35, No. 3, June 1, 2013, pp. 283–310.

EuropeAid, *Evaluation - Guidelines*, 2005. http://ec.europa.eu/europeaid/evaluation/methodology/guidelines/gbb_det_en.htm.

Finlay, Linda, “‘Rigour’, ‘Ethical Integrity’ or ‘Artistry’? Reflexively Reviewing Criteria for Evaluating Qualitative Research”, *The British Journal of Occupational Therapy*, Vol. 69, No. 7, 2006, pp. 319–326.

Foucault, M., *Discipline & Punish: The Birth of the Prison*, Knopf Doubleday Publishing Group, 2012.

Gadamer, H.G., *Truth and Method*, Continuum International Publishing Group, 2004.

Guba, E. G., and Yvonna S. Lincoln, *Naturalistic Inquiry*, SAGE, 1985.

Habermas, J., *The Theory of Communicative Action: Volume 1: Reason and the Rationalization of Society*, Beacon, Boston, 1984.

Haywood, Benjamin K., and John C. Besley, “Education, Outreach, and Inclusive Engagement: Towards Integrated Indicators of Successful Program Outcomes in Participatory Science”, *Public Understanding of Science*, Vol. 23, No. 1, January 1, 2014, pp. 92–106.

———, “Education, Outreach, and Inclusive Engagement: Towards Integrated Indicators of Successful Program Outcomes in Participatory Science”, *Public Understanding of Science*, Vol. 23, No. 1, January 1, 2014, pp. 92–106.

Heidegger, Martin, *Being and Time*, Blackwell, 1967.

Joss, Simon, “Evaluating Consensus Conferences: Necessity or Luxury”, *Public Participation in Science: The Role of Consensus Conferences in Europe*, Science Museum London, 1995, pp. 89–108.

Klein, Julie T., “Evaluation of Interdisciplinary and Transdisciplinary Research”, *American Journal of Preventive Medicine*, Vol. 35, No. 2, August 2008, pp. S116–S123.

Kuper, Ayelet, Lorelei Lingard, and Wendy Levinson, “Critically Appraising Qualitative Research”, *BMJ*, Vol. 337, 2008.

Merkx, Femke, Inge van der Weijden, Anne-Marie Oostveen, Peter van den Besselaar, and Jack Spaapen, “Evaluation of Research in Context A Quick Scan of an Emerging Field”, *Den Haag: Rathenau Instituut/ERiC*, 2007. <http://www.social-informatics.net/quickscan.pdf>.

Mezirow, Jack, “Transformative Learning: Theory to Practice”, *New Directions for Adult and Continuing Education*, Vol. 1997, No. 74, 1997, pp. 5–12.

Mickwitz, Per, “A Framework for Evaluating Environmental Policy Instruments Context and Key Concepts”, *Evaluation*, Vol. 9, No. 4, 2003, pp. 415–436.

Miles, M., and A. Huberman, *An Expanded Source Book: Qualitative Data Analysis*, Sage, London, 1994.

Mittelstadt, Brent, Bernd Carsten Stahl, and Kutoma Wakunuma, *SATORI Deliverable D12.1 - Good Practice in Evaluation, Reflection and Civil Society Engagement*, SATORI, De Montfort University, Leicester, 2014.

Nagarajan, Nigel, and Marc Vanheukelen, “Evaluating EU Expenditure Programmes: A Guide. Ex Post and Intermediate Evaluation. XIX/02–Budgetary Overview and Evaluation”, *Directorate-General XIX–Budgets. European Commission*, 1997.

OECD, “Evaluation of Development Programmes”, 1991. <http://www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm>.

O’Sullivan, Rita G., “Collaborative Evaluation within a Framework of Stakeholder-Oriented Evaluation Approaches”, *Evaluation and Program Planning*, Vol. 35, No. 4, *Collaborative Evaluation: Theory and Practice*, November 2012, pp. 518–522.

Patterson, Michael E., and Daniel R. Williams, *Collecting and Analyzing Qualitative Data: Hermeneutic Principles, Methods and Case Examples*, Vol. 9, Vol. 9, *Advances in Tourism Application Series*, Sagamore Publishing, Inc., Champaign, IL, 2002. <http://www.treesearch.fs.fed.us/pubs/29421>.

Penfield, Teresa, Matthew J. Baker, Rosa Scoble, and Michael C. Wykes, “Assessment, Evaluations, and Definitions of Research Impact: A Review”, *Research Evaluation*, Vol. 23, No. 1, January 1, 2014, pp. 21–32.

Renn, Ortwin, Thomas Webler, Horst Rakel, Peter Dienel, and Branden Johnson, “Public Participation in Decision Making: A Three-Step Procedure”, *Policy Sciences*, Vol. 26, No. 3, August 1993, pp. 189–214.

Research Councils UK, *Evaluation: Practical Guidelines - A Guide for Evaluating Public Engagement Activities*, 2014. <http://www.rcuk.ac.uk/RCUK-prod/assets/documents/publications/evaluationguide.pdf>.

Rowe, Gene, and Lynn Frewer, “Public Participation Methods: A Framework for Evaluation”, *Science, Technology & Human Values*, Vol. 25, No. 1, 2000, pp. 3–29.

Rowe, Gene, and Lynn J. Frewer, “A Typology of Public Engagement Mechanisms”, *Science, Technology & Human Values*, Vol. 30, No. 2, April 1, 2005, pp. 251–290.

———, “Evaluating Public-Participation Exercises: A Research Agenda”, *Science, Technology & Human Values*, Vol. 29, No. 4, 2004, pp. 512–556.

Rowe, G., and L.J. Frewer, “Public Participation Methods: A Framework for Evaluation”, *Science Technology and Human Values*, Vol. 25, No. 1, 2000, pp. 3–29.

Saari, E., and K. Kallio, “Developmental Impact Evaluation for Facilitating Learning in Innovation Networks”, *American Journal of Evaluation*, Vol. 32, No. 2, June 1, 2011, pp. 227–245.

SATORI, *THEME [SiS.2013.1.2-1 SiS.2013.1.2-1] Mobilisation and Mutual Learning (MML) Action Plans: Mainstreaming Science in Society Actions in Research] - Annex 1 - Description of Work*, 2013.

Scriven, Michael, *Evaluation Thesaurus*, Sage, 1991.

Smith, L. Graham, Carla Y. Nell, and Mark V. Prystupa, “FORUM: The Converging Dynamics of Interest Representation in Resources Management”, *Environmental Management*, Vol. 21, No. 2, March 1, 1997, pp. 139–146.

Stagl, Sigrid, “Multicriteria Evaluation and Public Participation: The Case of UK Energy Policy”, *Land Use Policy*, Vol. 23, No. 1, *Resolving Environmental Conflicts: Combining Participation and Multi-Criteria Analysis*, January 2006, pp. 53–62.

Stahl, B. C., “Responsible Research and Innovation: The Role of Privacy in an Emerging Framework”, *Science and Public Policy*, 2013, p. sct067.

Stirling, Andy, “9. Precaution, Foresight and Sustainability: Reflection and Reflexivity in the Governance of Science and Technology”, *Reflexive Governance for Sustainable Development*, 2006, p. 225.

Tuominen, Anu, Tuuli Järvi, Kirsi Hyytinen, Evangelos Mitsakis, Maria Eugenia Lopez-Lambas, Lissy La Paix, Jan van der Waard, Anne Binsted, and Anatolij Sitov, “Evaluating the Achievements and Impacts of EC Framework Programme Transport Projects”, *European Transport Research Review*, Vol. 3, No. 2, July 1, 2011, pp. 59–74.

Walter, Alexander I., Sebastian Helgenberger, Arnim Wiek, and Roland W. Scholz, “Measuring Societal Effects of Transdisciplinary Research Projects: Design and Application of an Evaluation Method”, *Evaluation and Program Planning*, Vol. 30, No. 4, November 2007, pp. 325–338.

Webler, Thomas, “‘Right’ Discourse in Citizen Participation: An Evaluative Yardstick”, *Fairness and Competence in Citizen Participation*, Springer, 1995, pp. 35–86.

Webler, Thomas, Hans Kastenholz, and Ortwin Renn, “Public Participation in Impact Assessment: A Social Learning Perspective”, *Environmental Impact Assessment Review*, Vol. 15, No. 5, September 1995, pp. 443–463.