



SATORI Deliverable D12.1

Good practice in evaluation, reflection and civil society engagement

Brent Mittelstadt (De Montfort University)
Mark Coeckelbergh (De Montfort University)
Bernd Stahl (De Montfort University)
Kutoma Wakunuma (De Montfort University)

June 2014

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:

Dr. Brent Mittelstadt, Centre for Computing and Social Responsibility
De Montfort University, The Gateway, Leicester, LE1 9BH United Kingdom
bmittelstadt@dmu.ac.uk

<http://satoriproject.eu/>



SATORI D12.1 – Good practice in evaluation, reflection and civil society engagement report

CONTENTS

1	Abstract.....	4
2	Executive summary	5
3	Introduction	9
4	Background.....	10
5	Literature Survey	13
5.1	Results.....	13
5.1.1	Defining Public Participation.....	14
5.1.2	Methods of Evaluation	15
5.1.2.1	Data Collection and Analysis Methods	17
5.1.2.2	Who Should Serve as Evaluators?.....	18
5.1.3	Theories of Learning.....	19
5.1.3.1	Organisational Learning	20
5.1.3.2	Expansive Learning	21
5.1.3.3	Transformative and Reflective Learning	22
5.1.3.3.1	Reflection and Reflexivity	23
5.1.3.4	Social Learning.....	23
5.1.3.5	Criteria for Evaluating Mutual Learning	24
5.1.4	Evaluation Criteria	25
5.1.4.1	Inter- and Transdisciplinarity in Choosing Criteria.....	26
5.1.4.2	Whose Criteria?	26
5.1.4.3	Frameworks of Evaluation Criteria	27
5.1.4.3.1	European Commission’s EuropeAid Framework	29
5.1.4.3.2	Webler’s Framework.....	30
5.1.4.3.3	Rowe and Frewer’s Framework	33
5.1.4.3.4	Related Frameworks.....	35
5.1.4.3.5	Other Frameworks.....	37
5.1.4.4	Emerging Consensus	40
5.1.5	Impact Evaluation	41
5.1.5.1	Types of Impact in MMLs.....	42
5.1.5.2	Difficulties of Evaluating Impact	43

5.2	Discussion	45
5.2.1	Stakeholder Inclusion in Evaluation	45
5.2.2	Evaluative Framework for MMLs	46
5.2.3	The Need for Reflexivity in MMLs	46
5.3	Initial Principles of Good Practice	47
6	Study of MMLs	49
6.1	MML Publication Survey	49
6.1.1	Results.....	49
6.1.1.1	Project A	49
6.1.1.2	Project B	51
6.1.1.3	Project C	51
6.1.1.4	Other Projects	52
6.1.1.5	Discussion.....	53
6.2	Interview Study.....	54
6.2.1	Methodology	54
6.2.2	Recruitment and Sample	55
6.2.3	Data Collection	56
6.2.3.1	Interview Tool	56
6.2.4	Data Analysis	56
6.2.4.1	Method.....	57
6.2.4.1.1	The Organising System	57
6.2.4.1.2	The Steps of Data Analysis	57
6.2.4.1.2.1	Organisational Coding.....	58
6.2.4.1.3	Substantive Coding	58
6.2.4.1.3.1	Grouping.....	59
6.2.4.1.4	Discussion of Analysis Method.....	60
6.2.4.1.4.1	Reviewing Hermeneutic Interpretation	60
6.2.4.1.4.2	Theoretical Outliers in Coding.....	61
6.2.4.2	Conclusions	61
6.2.5	Results.....	61
6.2.5.1	Defining MMLs	62
6.2.5.2	Evaluation Methods and Purpose	64
6.2.5.2.1	Types of Data Considered in Evaluation.....	64
6.2.5.2.2	Consortium Involvement in Evaluation	65
6.2.5.2.2.1	Methods to Facilitate Partner Reflection.....	65

6.2.5.2.3	Scope of Evaluation	66
6.2.5.3	Evaluator Roles.....	66
6.2.5.3.1	Internal vs. External	67
6.2.5.3.2	Feedback and Reflection	68
6.2.5.4	Evaluating Mutual Learning.....	69
6.2.5.4.1	Methods of Evaluating Mutual Learning	70
6.2.5.5	Impact	71
6.2.5.5.1	Types of Impact.....	72
6.2.5.6	Criteria and Indicators of Success	72
6.2.5.6.1	Possibility of a Generic MML Evaluation Framework.....	73
6.2.5.6.2	Consortium Criteria.....	74
6.2.5.6.3	Types of Criteria Necessary to Evaluate MMLs	75
6.2.6	Discussion.....	75
6.2.6.1	Representativeness of Impacts.....	77
6.2.6.2	Influence on Evaluation.....	78
7	Principles of Good Practice in MML Evaluation	79
8	Conclusion.....	85
8.1	Next Steps	85
9	References	87
10	Annexes.....	93
	Appendix 1 – MML Details.....	93
	Appendix 2 – Human Research Ethics Documentation.....	96
	Appendix 3 – MML Criteria	105
	Appendix 4 – Interview Schedule.....	107
	Appendix 5 – Database Search Results	109

1 ABSTRACT

This report describes the findings of research activities carried out by De Montfort University in ‘Task 12.1 – Good practice in evaluation and reflection’ of the SATORI Mobilisation and Mutual Learning Action (MML). In this task a literature survey and empirical study have been conducted to identify principles of good practice in evaluation and reflection in MMLs. The report considers a range of data including academic publications on good practice in evaluation and reflection upon public participation events and participatory research, MML publications and project documents, interviews with MML practitioners, as well as findings from a small set of interviews for SATORI Task 2.1 ‘Landscape of existing MML projects and other relevant, ethics-related projects’ prepared by Trilateral in collaboration with DMU’s completion of Task 12.1. Analysis of these data sets contributed to the identification of 21 principles of good practice in evaluation and reflection in MMLs, which are designed to be specified in designing an project-specific evaluation methods responsive to the scope and needs of specific projects. The principles therefore provide the groundwork for the creation of a common evaluative framework for monitoring and evaluating MMLs in the future. The findings reported here will feed directly into the specification of a set of SATORI evaluation principles and criteria in Task 12.2, and an evaluation and reflection strategy in Task 12.3. This strategy will then be implemented to evaluate SATORI in Task 12.4.

2 EXECUTIVE SUMMARY

This report describes the findings of research activities carried out by De Montfort University in ‘Task 12.1 – Good practice in evaluation and reflection’ of 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 under the 7th Framework Programme Mobilisation and Mutual Learning Action Plan (MMLAP). A primary aim of implementing MMLs as a new type of ‘coordination and support activity’ has been to “create mechanisms for effectively tackling scientific and technology related challenges faced by society, by proactively bringing together different actors with complementary knowledge and experiences”¹. In differentiating MMLs from other type of projects, emphasis would appear to be placed on stakeholder engagement with societal actors including engagement in mutual learning activities, and less on conducting desk research or running events such conferences as ends in themselves. However, MMLs are an evolving concept without clearly prescribed methods and activities, leading to confusion among MML consortia over what it means to ‘mobilise mutual learning’ or ‘engage citizens and civil society’². In part this may be explained by the broad ‘Societal Challenges’ addressed by different projects, as well as the wide range disciplines engaged and scientific and technological activities undertaken³.

One point of consensus concerning necessary activities undertaken in MML is the need for ongoing evaluation and reflection (or monitoring) on project progress, methods and impact. In the latest round of MML calls in the Science in Society Work Programme 2013, the EC has emphasised a need for extensive evaluation and reflection by MML partners, requiring that a separate evaluation work package must be included in all proposals in which the partners “evaluate the methodology and process put in place during the project.” Current MML practitioners have also recognised the need for a “more active role for project evaluators”⁴.

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. To investigate this apparent gap in understanding of the MML mechanism a literature survey and empirical study were conducted to identify principles of good practice in evaluation and reflection in MMLs. Analysis of academic publications on good practice in evaluation and reflection upon public participation events and participatory research, MML publications and project documents, interviews with MML practitioners, as well as findings from a small sample of interviews carried out for SATORI Task 2.1 ‘Landscape of existing MML projects and other relevant, ethics-related projects’ have led to the identification of 21 principles of good practice in evaluation and reflection in MMLs:

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 (see: Sections 3.1.4.1 and 6.2.5.6).

¹ Healy, *Mobilisation and Mutual Learning (MML) Action Plans: Future Developments*, 6.

² Healy, *Mobilisation and Mutual Learning (MML) Action Plans: Future Developments*.

³ *Ibid.*, 8.

⁴ *Ibid.*, 14.

2. Evaluation should address the ‘generic’ qualities of participatory processes such as those areas of consensus in evaluation literature identified by Chilvers (2008). 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 & Besley (2013) and Walter et al. (2007) (see: Sections 6.2.5.4).
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 (see: Section 6.1.1.5).
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 DoW (see: Section 6.2.5.2).
5. The ability of the MML to get target stakeholder groups in attendance at engagement events may be used as an evaluative measure (see: Section 6.2.5.6).

Methodology Principles

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 impact⁵, and provide stakeholders with a better idea of the value of their participation by tracking influence on the process⁶. MML evaluation should, at a minimum, seek to meet these three generic aims (see: Section 6.2.5.3.2).
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 (see: Section 5.1.4.3 and 6.2.5.2.1).
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) (see: Section 5.1.5.2 and 6.2.5.5).
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, independent) as early as possible. This approach will help reduce resistance to recommendations made by the evaluators (see: Sections 6.2.5.2.2, 6.2.5.2.3 and 6.2.5.3).
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 (see: Section 6.2.5.2.2).
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

⁵ Research Councils UK, *Evaluation: Practical Guidelines - A Guide for Evaluating Public Engagement Activities*.

⁶ Rowe and Frewer, “Evaluating Public-Participation Exercises: A Research Agenda”.

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 (see: Sections 6.2.5.2.2 and 6.2.5.6).

12. A clear ‘endpoint’ should be specified at which point project impacts can start to be identified and evaluated (see: Section 5.1.5).
13. Evaluation should occur before, during and after the project to ensure all processes and impacts are evaluated to some degree (see: Section 5.1.2).

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 and participants to reflect on changes to their attitudes and behaviours caused by participating in the project and engaging with unfamiliar ideas and perspectives (see: Sections 5.1.2.1 and 6.2.5.4).
15. Mutual learning outcomes among project participants should be assessed (see: Section 5.1.3), 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 (NB: self-reported data) (see: Section 6.2.5.5.1).
16. In evaluating the quality of mutual learning that has occurred, the possibility of mutual learning without absolute consensus should be recognised (see: Section 5.1.4.3.2).
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 (see: Section 5.1.2.2).
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 (see: Section 5.1.3).

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 (see: Sections 6.2.5.3.2 and 6.2.6.2).
20. When conducting a formative evaluation, the evaluator should provide critical feedback and recommendations to the consortium to improve ongoing research activities (see: Section 6.2.5.3.2).
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 (see: Sections 6.2.5.3.2 and 6.2.5.4.1).

The principles are designed to be specified in designing an project-specific evaluation methods responsive to the scope and needs of specific projects, meaning they provide the groundwork for the creation of a common evaluative framework for monitoring and evaluating MMLs in the future. In moving forward with the evaluation of SATORI and empirical ‘testing’, the principles identified here

will feed directly into the specification of a set of SATORI evaluation principles and criteria in Task 12.2, and an evaluation and reflection strategy in Task 12.3. This strategy will then be implemented to evaluate SATORI in Task 12.4.

3 INTRODUCTION

This report describes the findings of research activities carried out by De Montfort University in ‘Task 12.1 – Good practice in evaluation and reflection’ of the SATORI Mobilisation and Mutual Learning Action (MML). In this task a literature survey and empirical study have been conducted to identify principles of good practice in evaluation and reflection in MMLs. The report considers a range of data including academic publications on good practice in evaluation and reflection upon public participation events and participatory research, MML publications and project documents, interviews with MML practitioners, as well findings from a small set of interviews for SATORI Task 2.1 ‘Landscape of existing MML projects and other relevant, ethics-related projects’ prepared by Trilateral in collaboration with DMU’s completion of Task 12.1. Analysis of these data sets contributed to the identification of 21 principles of good practice in evaluation and reflection in MMLs, which are designed to be specified in designing an project-specific evaluation methods responsive to the scope and needs of specific projects. The principles therefore provide the groundwork for the creation of a common evaluative framework for monitoring and evaluating MMLs in the future. The findings reported here will feed directly into the specification of a set of SATORI evaluation principles and criteria in Task 12.2, and an evaluation and reflection strategy in Task 12.3. This strategy will then be implemented to evaluate SATORI in Task 12.4.

The report is structured as follows: Section 2 discusses the background of Mobilisation and Mutual Learning Action Plans as a new type of EC-funded project is assessed to begin to identify key characteristics of the project type. Section 3 describes the literature survey, starting with the methodology used to examine academic literature discussing evaluation of public participation, participatory research and other activities with similarities to MMLs. Section 3.1 presents a narrative account of the results of the review, focusing in turn on how ‘participatory processes’ are defined, methods of evaluation, theories of (mutual) learning, frameworks of evaluation criteria, the problem of impact evaluation, and finally the relationship between reflection, reflexivity and evaluation. Section 3.2 briefly discusses the results, identifying overlap between the various methods, theories of learning and evaluation frameworks reviewed to begin to identify principles of good practice in evaluation. Section 3.3 presents an initial set of principles identified through the literature review.

In the second half of the report a study conducted in coordination with existing MMLs is described. The study involved document analysis of MML publications and interviews with MML partners including coordinators, evaluators and other various project partners. Section 4 describes the study, in particular the methodology and results of a document analysis and interviews with partners in other MMLs. A discussion considers principles that can be gleaned from the results of both modes of inquiry. Section 5 then considers the relationship between the principles identified in both halves of the report, concluding with a preliminary set of principles of good practice for evaluating and reflecting on MMLs. Section 6 concludes the report.

4 BACKGROUND

Mobilisation and Mutual Learning Actions are a new type of engagement project funded by the European Commission under the 7th Framework Programme Mobilisation and Mutual Learning Action Plan (MMLAP)⁷. Launched in the 2010 FP7 Work Programme, at present 18 MMLs have been funded⁸, with the earliest (PERARES) set to finish in October 2014. As described by the European Commission⁹ MMLs are intended to encourage a “two-way dialogue between researchers and other stakeholders,” creating “opportunities for members of the public and other groups in society (such as civil society organisations) to appropriate relevant knowledge, and for scientists to draw closer to the concerns of citizens.” MMLs are also meant to emphasise “mobilizing all relevant actors and on mutual learning in order to pool experiences and better focus their respective efforts on finding solutions that develop and use scientific and technological knowledge in the public interest”¹⁰. Even in this early stage of the MMLAP, the MML instrument has been seen by current practitioners to offer “much needed means of addressing the key societal challenges outlined in Europe 2020”¹¹. A primary aim of implementing MMLs as a new type of ‘coordination and support activity’ has been to “create mechanisms for effectively tackling scientific and technology related challenges faced by society, by proactively bringing together different actors with complementary knowledge and experiences”¹². In this sense MMLs can be seen as a continuation of the aims of the FP7 Science in Society programme, which was “designed to encourage a more meaningful engagement of citizens and civil society in research and research based policies”¹³.

While these broad aims are helpful in understanding the purpose of MMLs, they lack detail concerning how MMLs should operate, or which characteristics should be embodied in its engagement activities and outputs. MMLs are an evolving concept without clearly prescribed methods and activities, leading to confusion among MML consortia over what it means to ‘mobilise mutual learning’ or ‘engage citizens and civil society’¹⁴. In part this may be explained by the broad ‘Societal Challenges’ addressed by different projects, as well as the wide range disciplines engaged and scientific and technological activities undertaken¹⁵. In a 2012 workshop aiming to define the purpose and methods of the MML instruments with input from representatives of the EC and 9 MMLs, a ‘collaborative imperative’ was identified as key to defining the instrument. According to this imperative, MMLs must (1) be socially inclusive, (2) facilitate mutual learning and (3) be relevant to policy.

Concerning social inclusion, the MML instrument was understood as essentially a democratic instrument for developing “collaborative approaches with a diverse range of stakeholders.” In doing so MMLs value “different types of social actors” with considerable weight given to the experience brought by civil society organisations, while also “providing a platform for activism on a national level, but also on topics of international concern”¹⁶. In being socially inclusive MMLs are meant to

⁷ Healy, *Mobilisation and Mutual Learning (MML) Action Plans: Future Developments*.

⁸ European Commission, “Science in Society Home Page Research - Mobilising and Mutual Learning Action Plans”.

⁹ Ibid.

¹⁰ European Commission, *Mobilisation & Mutual Learning (MML) Action Plans on Societal Challenges: Call for Proposals - Science in Society Work Programme 2013*.

¹¹ Healy, *Mobilisation and Mutual Learning (MML) Action Plans: Future Developments*, 6.

¹² Ibid.

¹³ Ibid., 7.

¹⁴ Healy, *Mobilisation and Mutual Learning (MML) Action Plans: Future Developments*.

¹⁵ Ibid., 8.

¹⁶ Ibid., 12.

encourage “institutional, cultural and practical change and innovation, and should encourage broad stakeholder participation on a wide range of topics”¹⁷, “allowing different categories of stakeholders, and particularly marginalised social groups, to have a voice in decision-making processes”¹⁸. Concerning mutual learning, the opportunities for mutual learning provided by the MML instrument were highly valued by workshop participants. Mutual learning was seen to occur in sharing “collaborative practices and ideas,” developing “research processes that are more critically self-aware,” and in “reflection on the broader context of knowledge construction”¹⁹. More specifically, MMLs should “enable communication across the different stakeholder types...with a view to *influence* stakeholders, whether they be ‘self-interested’ or ‘defenders of the common good’”²⁰. Concerning policy relevance, MMLs were seen to provide “valuable stimulus for innovation, and the development of potentially world-changing ideas.” The instrument can also provide evidence for “knowledge based decision making processes...and [is] key to the continued development of a common European knowledge base/research area”²¹.

In differentiating MMLs from other type of projects, emphasis would appear to be placed on stakeholder engagement with societal actors including engagement in mutual learning activities, and less on conducting research or running events such conferences as ends. While helpful, these high-level descriptions of what MMLs are meant to ‘do’ and ‘be’ remain too vague to prescribe specific actions to be undertaken; indeed, what MMLs should *do* to translate these aims into practice is far from settled among current practitioners²². The responsibility to interpret and put these broad aims into practice to address specific Societal Challenges remains the responsibility of individual project teams. As a result, a broad range of aims and methods are likely to emerge as MMLs run over the coming years. As with any new type of project, opportunities and difficulties will be faced by ‘pioneering’ MMLs in defining the activities and methods necessary to meet the EC’s expectations. A key example is the necessity of a “common diagnostic for monitoring and evaluating projects,” as called for by participants of the 2012 workshop²³. The experiences of partners in early MMLs will undoubtedly be invaluable in helping future MML consortia understand their responsibilities and to define appropriate strategies. Ideally, projects should have the opportunity to learn from the experiences, recommendations and failures of other MMLs.

To contribute to this necessary process of reflection and identification of ‘good practice’ in MMLs, and to begin developing a ‘common evaluation diagnostic’ for MMLs²⁴, the Centre for Computing and Social Responsibility at De Montfort University, acting in its capacity as the internal evaluator for the SATORI MML, has carried out a study into ‘good practice’ in MML evaluation. In the latest round of MML calls in the Science in Society Work Programme 2013, the EC has emphasised a need for extensive evaluation and reflection by MML partners, requiring that a separate evaluation work package must be included in all proposals in which the partners “evaluate the methodology and process put in place during the project.” Current MML practitioners have also recognised the need for a “more active role for project evaluators, for instance in mapping the ‘stakes’ of stakeholders”²⁵. The study, consisting of a literature survey and empirical study with existing MML partners, looks at

¹⁷ Ibid., 15.

¹⁸ Ibid., 16.

¹⁹ Ibid., 12.

²⁰ Ibid., 15.

²¹ Ibid., 12.

²² Ibid., 16.

²³ Ibid., 14.

²⁴ cf. Healy, *Mobilisation and Mutual Learning (MML) Action Plans: Future Developments*.

²⁵ Ibid., 14.

how the requirement for evaluation and reflection set out by the EC should be translated into practice. Short of prescribing specific evaluative actions or criteria to be adopted, the study sought to identify principles of good practice in MMLs which can provide guidance for consortia in planning evaluation and reflection activities.

5 LITERATURE SURVEY

As the first MML has yet to end, little academic discourse exists concerning the evaluation of MMLs. In recognition of this the scope of the survey must be broader than MMLs to identify relevant literature which does not directly discuss MMLs. To define an appropriate scope it was necessary to identify characteristics shared between MMLs and other types of engagement projects with more established academic discourses. These characteristics are not intended to be a definitive list of what a MML ‘should be’, but rather an initial framework to characterize MMLs based on the vision of the EC.

Based on the description above (see: Section 4), MMLs must engage stakeholders to a significant degree, in some way promoting mutual learning or sharing of knowledge which can be ‘appropriated’ by societal actors. The literature survey therefore focused on evaluation of ‘public participation’ and ‘participatory research’, which were identified as projects with a similar emphasis on building communication channels, encouraging collaboration and facilitating mutual learning through stakeholder engagement activities’. Discussions of evaluation in other types of EU-funded projects, such as coordination and support actions, were particularly relevant.

Academic publications discussing evaluation of stakeholder engagement, mutual learning and related topics in research were surveyed to identify principles of good practice and methods of evaluation and reflection relevant to MMLs. Sources were located through systematic searching of databases, reviewing references of returned literature, and contacting evaluators and coordinators of current MMLs directly for works informing their approach to evaluation. These multiple search techniques were used to ensure relevant papers and discussions were not missed. The title and abstract of each article identified in the database searches were reviewed to determine relevancy to the survey topic. Sources deemed initially relevant then underwent content analysis²⁶ to identify discussion of methods or principles of evaluation or reflection.

Two academic databases (Scopus, Web of Science) and one search engine (Google Scholar) were queried for publications discussing principles or examples of good practice in project evaluation. One of the many challenges of the review was in creating appropriate search queries to identify relevant discourses. Stakeholder engagement and mutual learning can theoretically be embedded in any sort of project with social aims, meaning the discourse was spread across many journals, conferences and disciplines without an obvious locus. As a result many search terms had to be tested. More often than not, reviewing dozens or hundreds of papers revealed only a relevant few. To address this difficulty numerous papers were located by reviewing the references and ‘cited by’ entries on Scopus and Google Scholar for papers identified in the database search. A complete table of database search terms and results can be found in Appendix 5 – Database Search Results.

5.1 RESULTS

A total of 73 sources were reviewed to identify discussion of good practice in evaluating and reflecting upon MMLs and related activities such as public participation in research. The following is a narrative overview of the findings of the survey, which focuses on the author’s analysis and interpretation of the literature to provide an overview of evaluation and reflection methods and criteria, and to begin to identify principles of good practice for MMLs. Types of evaluation and reflection are discussed first before moving on to theories of learning which can inform evaluation in MMLs. Frameworks of evaluative criteria are then reviewed, followed by a discussion of the

²⁶ Hsieh and Shannon, “Three Approaches to Qualitative Content Analysis”.

difficulties of assessing impact and mutual learning. Wherever possible similar recommendations are grouped into themes. Widely cited ‘seminal’ papers and frameworks are given an extended discussion given their influence on what has proven to be a fragmented discourse occurring across numerous academic disciplines.

5.1.1 Defining Public Participation

Despite the recent emergence of MMLs, the need to evaluate methods of involving the public in policy making and scientific research has been widely recognised for several decades²⁷. Historical justification for participatory processes is diverse and often attached to a desire for engaged citizens, the promotion of individualism and the realisation of democratic ideals; generally speaking, public participation is justified wherever the need for a two-way dialogue between decision-makers and stakeholders, as well as dialogue among different groups of stakeholders, is recognised²⁸. If MMLs are distinguished by the significant efforts put towards engagement of societal stakeholders, then it is worth considering briefly how such engagement can occur.

Related discourses concern processes which typically involve joint decision-making or consultation of the public concerning a problem, such as an emerging technology or policy (for example), which affects members of the public. A major discourse in this area surrounds ‘public participation’²⁹, which can be defined as “a group of procedures designed to consult, involve, and inform the public to allow those affected by a decision to have an input into that decision”³⁰, typically in the field of making or setting the agenda for policy³¹. Examples include referenda, public hearings/inquiries, public opinion surveys, negotiated rule making, consensus conferences, focus groups, citizens’ jury/panels and citizen/public advisory committees³². According to Glass³³, public participation has five objectives: (1) information exchange; (2) education; (3) support building; (4) supplemental decision making; and (5) representation input. In a review of public participation across several disciplines some of the benefits stemming from these objectives were identified as legitimisation of decision-making, enhancement of democracy and enlargement of citizenship³⁴. Public participation should therefore be understood as a set of activities in which two-way ‘mutual learning’ is implicit; whereas public communication involves ‘sponsors’ giving information to the public’, and public consultation involves the public giving information to sponsors, public participation requires information to travel in both directions³⁵ and thus can be seen as a type of ‘mutual learning’ (see: Section 5.1.3).

Another sizeable discourse in this area concerns participatory research³⁶. Participatory research can be defined as a process in which participants collaborate to “problem solve and produce new

²⁷ e.g. Chess, “Evaluating Environmental Public Participation”; Haywood and Besley, “Education, Outreach, and Inclusive Engagement”; O’Sullivan, “Collaborative Evaluation within a Framework of Stakeholder-Oriented Evaluation Approaches”; Rowe and Frewer, “Public Participation Methods”; Sewell and Phillips, “Models for Evaluation of Public Participation Programmes”.

²⁸ Abelson et al., “Deliberations about Deliberative Methods”, 239–40.

²⁹ e.g. Chess, “Evaluating Environmental Public Participation”; Rowe and Frewer, “A Typology of Public Engagement Mechanisms”; Sewell and Phillips, “Models for Evaluation of Public Participation Programmes”.

³⁰ Rowe and Frewer, “Public Participation Methods”, 6.

³¹ Rowe and Frewer, “A Typology of Public Engagement Mechanisms”.

³² e.g. Rowe and Frewer, “Public Participation Methods”, 8–9; Abelson et al., “Deliberations about Deliberative Methods”, 240.

³³ Glass, “Citizen Participation in Planning: The Relationship between Objectives and Techniques”.

³⁴ Petts and Leach, *Evaluating Methods for Public Participation: Literature Review*.

³⁵ Rowe and Frewer, “A Typology of Public Engagement Mechanisms”.

³⁶ e.g. Haywood and Besley, “Education, Outreach, and Inclusive Engagement”; Joss and Bellucci, *Participatory Technology Assessment*.

knowledge in an ongoing learning and reflective process”³⁷, which is often transdisciplinary in nature³⁸. Examples include participatory and constructive Technology Assessment³⁹, ‘Public Participation in Science and Research’⁴⁰ as well as various other discipline specific methods of research in which members of the public or affected stakeholders are involved⁴¹.

MMLs can be seen as potentially overlapping yet distinct from public participation and participatory research; whereas public participation typically aims to influence policy making and is initiated by governmental/regulatory agencies⁴², and participatory research aims to solve a particular problem⁴³ or develop a particular technology⁴⁴, MMLs seek first to foster dialogue and mutual learning between sponsors, researchers, industry, members of the public and other stakeholders, thereby treating the coordination of communication, collaboration and learning as primary ends in themselves. While the outcomes of a MML may influence policy or inform decisions on a particular problem, the emphasis on learning and communication separates MMLs from these more ‘applied’ research activities.

This is not to say public participation and participatory research do not seek to foster learning and communication or view these activities as implicitly valuable, but rather that these aims occur against a policy or research-oriented backdrop in a particular problem context. MMLs can (but need not) be broader than this as they seek to establish the communication and participatory channels through which future problem solving and knowledge exchange may occur in a particular field (e.g. climate change), even if a particular problem requiring resolution or particular solution is not yet proposed. MMLs cannot therefore be directly equated with public participation and participatory research because they do not necessarily seek public input on a particular decision, policy or solution, but rather seek to foster dialogue and mutual learning between different stakeholder groups. Despite this, discourses and approaches to the evaluation of public participation have much to contribute to MML evaluation—both activities engage a variety of ‘public’ stakeholder groups on a particular issue(s). Throughout this report the term ‘participatory processes’ is used to highlight this overlap. The term refers to all types of stakeholder engagement activities conducted in MMLs, or those public participation and participatory research processes⁴⁵ in which two-way learning and construction of communication, collaboration and knowledge exchange channels are explicit goals.

5.1.2 Methods of Evaluation

Broadly speaking, evaluation can be defined as “the process of determining the merit, worth and value of things”⁴⁶, 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

³⁷ Blackstock, Kelly, and Horsey, “Developing and Applying a Framework to Evaluate Participatory Research for Sustainability”, 728.

³⁸ Klein, “Evaluation of Interdisciplinary and Transdisciplinary Research”.

³⁹ Genus and Coles, “On Constructive Technology Assessment and Limitations on Public Participation in Technology Assessment”; Joss and Bellucci, *Participatory Technology Assessment*.

⁴⁰ Haywood and Besley, “Education, Outreach, and Inclusive Engagement”.

⁴¹ e.g. Blackstock, Kelly, and Horsey, “Developing and Applying a Framework to Evaluate Participatory Research for Sustainability”; Følstad, “Living Labs for Innovation and Development of Information and Communication Technology: A Literature Review”; Friedman, Kahn, and Borning, “Value Sensitive Design and Information Systems”.

⁴² Rowe and Frewer, “A Typology of Public Engagement Mechanisms”.

⁴³ Blackstock, Kelly, and Horsey, “Developing and Applying a Framework to Evaluate Participatory Research for Sustainability”.

⁴⁴ Følstad, “Living Labs for Innovation and Development of Information and Communication Technology: A Literature Review”; Joss and Bellucci, *Participatory Technology Assessment*.

⁴⁵ e.g. Rowe and Frewer, “Public Participation Methods”; Renn et al., “Public Participation in Decision Making”.

⁴⁶ Scriven, *Evaluation Thesaurus*.

of social science research methods”⁴⁷. Concerning research and engagement projects, evaluation will at a minimum focus on the “design, implementation and effectiveness” of the project⁴⁸, as well as outputs (e.g. reports, deliverables), outcomes (e.g. products, processes) and impacts (e.g. follow-on products, processes and research)⁴⁹. 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 impact⁵⁰, and provide participants with a better idea of the value of their participation by tracking influence on the process⁵¹. 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 research⁵². Criteria against which the success of the project can be assessed are typically used⁵³, and can be pre-defined or developed in-situ. Methods of data collection and analysis are often guided by prescriptive discipline-specific guidelines⁵⁴. However, despite these commonalities significant potential exists for interpretation and variation in aims, methods and normative indicators of success.

Moving away from these general concepts of evaluation, an initial distinction can be drawn in the evaluation of participatory research between process, procedural or formative evaluation which examines the quality of procedures or methods of stakeholder engagement and mutual learning while they occur, and outcome, substantive or summative evaluation which assesses the outcomes of procedures⁵⁵. The former can contribute to the refinement of research and participatory processes by identifying weaknesses or areas for improvement prior to the project’s conclusion⁵⁶, thereby acting as a feedback mechanism or ‘double loop’ to refine project activities⁵⁷ 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 its success⁵⁸, for example in meeting its stated objectives, impacts or success indicators, or against a pre-defined framework for assessing quality (see: Section 5.1.4.3). As with formative evaluation, post-project summative evaluation can inform the design and conduct of future participatory processes and research.

In determining the appropriate approach to evaluation of future participatory processes in particular MMLs, it is important to remember that evaluation can occur before, during and after a project. Before the project starts the proposed methods of inquiry can be assessed. During the project, the quality, effectiveness or other measures of the actual activities undertaken can be evaluated, whereas both during and after the project has ended, its broader impacts on participants, sponsors and broader

⁴⁷ Joss, “Evaluating Consensus Conferences: Necessity or Luxury”, 89.

⁴⁸ Tuominen et al., “Evaluating the Achievements and Impacts of EC Framework Programme Transport Projects”, 61.

⁴⁹ 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”.

⁵⁰ Research Councils UK, *Evaluation: Practical Guidelines - A Guide for Evaluating Public Engagement Activities*.

⁵¹ Rowe and Frewer, “Evaluating Public-Participation Exercises: A Research Agenda”.

⁵² O’Sullivan, “Collaborative Evaluation within a Framework of Stakeholder-Oriented Evaluation Approaches”, 518.

⁵³ e.g. Rowe and Frewer, “Public Participation Methods”; Webler, “‘Right’ Discourse in Citizen Participation: An Evaluative Yardstick”; EuropeAid, *Evaluation - Guidelines*.

⁵⁴ e.g. O’Sullivan, “Collaborative Evaluation within a Framework of Stakeholder-Oriented Evaluation Approaches”; EuropeAid, *Evaluation - Guidelines*.

⁵⁵ 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”.

⁵⁶ Chess, “Evaluating Environmental Public Participation”, 773.

⁵⁷ Abelson and Gauvin, *Assessing the Impacts of Public Participation*, 3.

⁵⁸ Ibid.

society and policy can be assessed. It is important to identify when and what type of evaluation will occur, and ideally to deal with all three types where funding and timing allow. Existing literature suggests that procedural evaluation before and during the project is most common⁵⁹, with a lack of substantive and impact evaluation both during and after the project, despite a recognised need to assess substantive outcomes⁶⁰.

5.1.2.1 Data Collection and Analysis Methods

Many methods of data collections for evaluation are possible, including interviews with project partners (or ‘experts’)⁶¹ and stakeholders, analysis of project documents and deliverables⁶², focus groups, observation of participation activities such as presentations or discussions⁶³⁶⁴, participant questionnaires and informal conversations⁶⁵, among others⁶⁶. For each of these standard data collection methods can be used⁶⁷; for example questionnaires can include a combination of pre-defined (e.g. Likert scale) and open ended questions amenable to quantitative and qualitative analysis⁶⁸. While frameworks of evaluation criteria are bespoke in the sense that they identify values against which participatory processes may be assessed, the necessity of unique forms of data collection is not immediately obvious.

The general strengths and weaknesses of the approaches mentioned above as reviewed in research methodology literature⁶⁹ apply and will undoubtedly influence the choice of methods in a particular project. The traditional divide between quantitative and qualitative methods is recognised in training materials for evaluators⁷⁰, with mixed methods approaches advocated⁷¹ to benefit from the strengths of both quantitative (generalisability, ease of analysis) and qualitative (depth of understanding) approaches, and to allow for triangulation of results⁷². Bespoke methodologies from both ends of the qualitative-quantitative spectrum have been developed for evaluation of participatory processes⁷³,

⁵⁹ Rowe and Frewer, “Public Participation Methods”, 10.

⁶⁰ cf. Abelson and Gauvin, *Assessing the Impacts of Public Participation*; Arnold, “Understanding Long-Term Impacts of R&D Funding”; Bornmann and Marx, “How Should the Societal Impact of Research Be Generated and Measured?”; Rowe and Frewer, “Public Participation Methods”.

⁶¹ Aichholzer and Westholm, “Evaluating eParticipation Projects”; EuropeAid, *Evaluation - Guidelines*; Goldschmidt and Renn, *Meeting of Minds - European Citizens’ Deliberation on Brain Sciences*.

⁶² EuropeAid, *Evaluation - Guidelines*.

⁶³ Aichholzer and Westholm, “Evaluating eParticipation Projects”; Goldschmidt and Renn, *Meeting of Minds - European Citizens’ Deliberation on Brain Sciences*; Research Councils UK, *Evaluation: Practical Guidelines - A Guide for Evaluating Public Engagement Activities*.

⁶⁴ For a sample structured observation sheet and generic interview questions, refer to Goldschmidt and Renn, *Meeting of Minds - European Citizens’ Deliberation on Brain Sciences*..

⁶⁵ Walls, Rowe, and Frewer, “Stakeholder Engagement in Food Risk Management Evaluation of an Iterated Workshop Approach”.

⁶⁶ For an overview of potential methods of data collection and sampling in research evaluation, see: Research Councils UK, *Evaluation: Practical Guidelines - A Guide for Evaluating Public Engagement Activities*..

⁶⁷ Denzin and Lincoln, “Introduction: The Discipline and Practice of Qualitative Research”; Guba and Lincoln, *Naturalistic Inquiry*; Research Councils UK, *Evaluation: Practical Guidelines - A Guide for Evaluating Public Engagement Activities*.

⁶⁸ Walls, Rowe, and Frewer, “Stakeholder Engagement in Food Risk Management Evaluation of an Iterated Workshop Approach”.

⁶⁹ e.g. Denzin and Lincoln, “Introduction: The Discipline and Practice of Qualitative Research”; Lincoln and Guba, “Competing Paradigms in Qualitative Research”; Oates, *Researching Information Systems and Computing*.

⁷⁰ e.g. EuropeAid, *Evaluation - Guidelines*; Research Councils UK, *Evaluation: Practical Guidelines - A Guide for Evaluating Public Engagement Activities*.

⁷¹ Chess, “Evaluating Environmental Public Participation”, 778.

⁷² Goldschmidt and Renn, *Meeting of Minds - European Citizens’ Deliberation on Brain Sciences*.

⁷³ e.g. Bornmann and Marx, “How Should the Societal Impact of Research Be Generated and Measured?”; Bornmann, “What Is Societal Impact of Research and How Can It Be Assessed?”; Chess and Johnson, “Organizational Learning

although a lack of empirically validated methods for assessing impact of such research has been noted⁷⁴. Relevant methods which imply a particular approach to data collection are noted throughout the review, particularly in relation to tracking learning over time (see: Section 6.2.5.4.1).

5.1.2.2 Who Should Serve as Evaluators?

Beyond the choice of data collection and analysis methods for evaluation, the choice of evaluators may also have a significant impact on the perceived validity of an evaluation mechanism. A distinction can be drawn between independent and participatory approaches. As argued by Scriven⁷⁵, the inclusion of stakeholders in the design and conduct of evaluation potentially biases methods and outputs towards stakeholder goals and expectations, undermining the validity and objectivity of the evaluation. From this objectivist position it would appear external evaluators without a stake in the issue addressed are a necessity for valid evaluation. Scriven⁷⁶ suggests that evaluators should avoid interaction with participants wherever possible, which limits the range of appropriate data collection methods.

In contrast to this position, advocates of participatory evaluation suggest stakeholders should be involved in the design and implementation of evaluation⁷⁷. Rather than evaluators being seen as judges of ‘right and wrong’, in participatory approaches evaluators take up a ‘learning’ based role, encouraging evaluative dialogue rather than issuing one-way judgments on the quality of the process. Perceived benefits of stakeholder participation include an increase in the perceived credibility of the evaluation among those affected by the issue discussed⁷⁸, as well as mutual learning among different stakeholders (including project partners and evaluators), which may be conceived of as an ‘impact’ of the research (see: Section 5.1.3.2). One potential way to include participants as evaluators is to conduct a questionnaire during participatory activities. Through this method of data collection the performance of the activity against pre-defined criteria can be assessed, while also gaining insight into the aims and criteria the participant would recommend the activity be evaluated against⁷⁹.

The inclusion of stakeholders in evaluation would appear to be the position preferred by a majority of evaluators of participatory processes⁸⁰. With that said, different authors advocate varying degrees and purposes of participation⁸¹; for example, O’Sullivan (2012) mentions four general approaches to participatory evaluation: ‘Collaborative Evaluation’, ‘Participatory Evaluation’, ‘Empowerment Evaluation’ and ‘Utilization-Focused Evaluation’. Approaches to participatory evaluation can be distinguished along at least three criteria: (1) which stakeholders are selected to participate; (2) the level of control stakeholders exhibit over the technical approach to evaluation; and (3) depth of participation⁸². For example, ‘Collaborative Evaluation’ operates with a sliding scale of to

about Public Participation”; Chess, “Evaluating Environmental Public Participation”; Walter et al., “Measuring Societal Effects of Transdisciplinary Research Projects”.

⁷⁴ Bornmann, “What Is Societal Impact of Research and How Can It Be Assessed?”.

⁷⁵ Scriven, “Truth and Objectivity in Evaluation”.

⁷⁶ Ibid.

⁷⁷ e.g. Bryson, Patton, and Bowman, “Working with Evaluation Stakeholders”; O’Sullivan, “Collaborative Evaluation within a Framework of Stakeholder-Oriented Evaluation Approaches”.

⁷⁸ Chess, “Evaluating Environmental Public Participation”, 777; O’Sullivan, “Collaborative Evaluation within a Framework of Stakeholder-Oriented Evaluation Approaches”, 518.

⁷⁹ Walls, Rowe, and Frewer, “Stakeholder Engagement in Food Risk Management Evaluation of an Iterated Workshop Approach”, 249.

⁸⁰ Bryson, Patton, and Bowman, “Working with Evaluation Stakeholders”, 2.

⁸¹ e.g. Blackstock, Kelly, and Horsey, “Developing and Applying a Framework to Evaluate Participatory Research for Sustainability”; O’Sullivan, “Collaborative Evaluation within a Framework of Stakeholder-Oriented Evaluation Approaches”; Alkin, *Evaluation Roots: Tracing Theorists’ Views and Influences*.

⁸² Cousins and Whitmore, “Framing Participatory Evaluation”.

distinguish ‘depth’ of participation which highlights the stage of the evaluation at which stakeholder participation first occurs—from planning the evaluation, to collecting and assessing data to reviewing the project’s impacts and outputs after it has ended. The aim of this approach is to engage stakeholders in the evaluation as far as they are willing⁸³. Other approaches advocate stakeholder participation as a means to ‘empower’ stakeholders in research⁸⁴, revealing influence of critical theory⁸⁵ on evaluators, best seen in the influence of the works of Jürgen Habermas on contemporary evaluative frameworks (see: Section 5.1.4).

Following from this, if stakeholders are to be included as evaluators the representativeness of the participants achieves new importance, as groups not participating lack a voice in the process itself and evaluation (see: Section 4). 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 research⁸⁶. While a particular degree of stakeholder involvement according to the above scale cannot be advocated outside of a particular project context due to lacking information about the willingness of the stakeholders and discipline-specific expertise required for evaluation, a participatory approach can be advocated due to its conduciveness to two-way learning compared to ‘objective’ approaches to evaluation.

5.1.3 Theories of Learning

The ‘mutual learning’ sought in MMLs can be conceptualised as part of broader movements towards incorporating ‘science in society’ as called for in recent EC funding bids. A similar emphasis on two-way learning involving in the public can be seen in the recent movements such as ‘Public Engagement in Science’ or the ‘Public Participation in Scientific Research’ (PPSR) approach, the latter of which seeks to facilitate direct collaboration between ‘expert’ scientists and members of the public in scientific project addressing large-scale societal challenges (e.g. environmental concerns)⁸⁷. These types of research move away from one-way conceptions of learning to a form of collaboration in which citizens are empowered to varying degrees to influence research:

“Whereas the competence approach is primarily concerned with oneway translation and dissemination from science to passive citizens, the participatory approach is experimenting with new formats that involve direct interaction, dialogue, and participation in a two-way communication, where citizens can in fact “speak back” to science (Gibbons, 1999), make their concerns heard to scientists as well as science policy makers, and contribute to setting the agenda for research”⁸⁸.

Such two-way models of scientific learning are meant to overcome social “science deficits” by aligning research with societal values while ensuring a multitude of interests are heard across a variety of stakeholder groups⁸⁹. However, the development of coherent approaches to two-way learning is not as simple as the discussion thus far suggests. Significant epistemic, methodological and theoretical challenges exist which constrain the justification of evaluative criteria or indicators

⁸³ O’Sullivan, “Collaborative Evaluation within a Framework of Stakeholder-Oriented Evaluation Approaches”.

⁸⁴ Fetterman and Wandersman, *Empowerment Evaluation Principles in Practice*; Cousins and Whitmore, “Framing Participatory Evaluation”.

⁸⁵ e.g. Stahl, “Emancipation in Cross-Cultural IS Research: The Fine Line between Relativism and Dictatorship of the Intellectual”.

⁸⁶ cf. Renn et al., “Public Participation in Decision Making”, 190–1.

⁸⁷ Haywood and Besley, “Education, Outreach, and Inclusive Engagement”.

⁸⁸ Mejlgaard and Stares, “Participation and Competence as Joint Components in a Cross-National Analysis of Scientific Citizenship”, 546.

⁸⁹ Haywood and Besley, “Education, Outreach, and Inclusive Engagement”, 94–5.

which indicate successful mutual learning—for example, how should uncertain or contested ‘scientific knowledge’ be presented? Do stakeholders possess the necessary background knowledge and conception of how research ‘creates’ knowledge to understand such uncertainties, or the relative validity of information presented to them? When and how should participatory discourses consider wider policy debates and societal impacts⁹⁰? How and according to which criteria can learning be deemed successful?

In the context of such difficulties faced in integrating mutual learning in research, it is important to know which theories of learning and development inform which approaches to evaluation to better understand how these questions can be answered. In general, a reflexive account of the theoretical basis is not provided in evaluation studies⁹¹. However, Chilvers⁹² suggests that in their place studies exist which focus on the relationship between participation and learning. Such studies “have tended to focus on individual and collective learning in discrete engagement events,” presenting learning as a dichotomy between ‘instrumental learning’ (or ‘knowledge acquisition’) concerned with the acquisition of knowledge and skills by participants, and ‘communicative learning’ oriented towards cooperation in discourse and learning about how others perceive a situation⁹³, the latter of which can be seen in evaluative frameworks based on Habermasian thought⁹⁴.

The conception of learning adopted by evaluators may have significant influence on the choice of criteria to evaluate the quality of learning which occurs in a participatory process, 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). In the reviewed literature four theories of learning were identified, each of which conceptualises learning, and therefore ‘successful’ learning, differently: (1) Organisational Learning; (2) Expansive Learning Theory; (3) Transformative and Reflective Learning; and (4) Social Learning.

5.1.3.1 Organisational Learning

Organisational learning describes the process through which organisations ‘learn’ and adapt over time. Such learning may “manifest in changes in policy, rules, or practices,” and can be seen in “collective understanding of history” or “story lines” which are influential in the organisation⁹⁵. This is not to suggest that an organisation has a single set of stories or histories shared across it; rather, organisational learning can be diverse and vary between departmental units⁹⁶. Despite the moniker, organisational learning is a diverse field of academic study, with at least three fundamentally different conceptions of learning. Learning may alternatively be described as (1) acquisition of knowledge, skills and competencies, wherein knowledge is “a commodity that can be moved around and transferred from person to person” (see for example⁹⁷; (2) a process of participation in cultural practices, meaning “people learn through engaging in interaction with others”; and (3) knowledge creation, meaning the “innovative and explorative processes of co-creating something that does not

⁹⁰ Chilvers, “Deliberating Competence Theoretical and Practitioner Perspectives on Effective Participatory Appraisal Practice”, 157.

⁹¹ Saari and Kallio, “Developmental Impact Evaluation for Facilitating Learning in Innovation Networks”, 228.

⁹² Chilvers, “Reflexive Engagement?”.

⁹³ *Ibid.*, 294–5.

⁹⁴ e.g. Wabler, “‘Right’ Discourse in Citizen Participation: An Evaluative Yardstick”.

⁹⁵ Chess and Johnson, “Organizational Learning about Public Participation”, 183.

⁹⁶ *Ibid.*; Senge, “The Fifth Discipline”.

⁹⁷ Stagl, “Multicriteria Evaluation and Public Participation”.

yet exist.” These three approaches are known as layman’s, sociocultural and knowledge creation approaches to learning, respectively⁹⁸. These three conceptions of knowledge are implicit in other theories of learning described here.

Organisational learning can occur in a critical and uncritical fashion, traditionally referred to as ‘double loop’ and ‘single loop’ learning. In single loop learning outside perspectives, values and experiences are not considered in making a decision; rather, the decision is made only on the basis of knowledge already accepted by the organisation. In contrast, double loop learning involves the consideration of outside perspectives which are potentially critical of or conflict with the accepted wisdom, stories, values and histories of the organisation⁹⁹, leading to a change in the ‘frame of reference’ that informs decision-making. In participatory processes ‘double loop’ learning can be seen when evaluation leads to changes in project activities, for example when evaluators make specific recommendations to address weaknesses of prior participatory processes. In this case the second loop is formed by the feedback provided by evaluators to project partners, although partners themselves may engage in reflexive analysis of progress and provide such feedback themselves (see: Section 5.1.3.3.1). However, a potential deficiency of the double loop model is that it implies learning is one-way (see: Figure 1); the organisation or learner’s perspective is expanded by information received from an outside source, but influence, dialogue or information travelling in the other direction is not present. It is therefore questionable whether organisational learning can represent two-way mutual learning, and not merely one-way feedback loops.

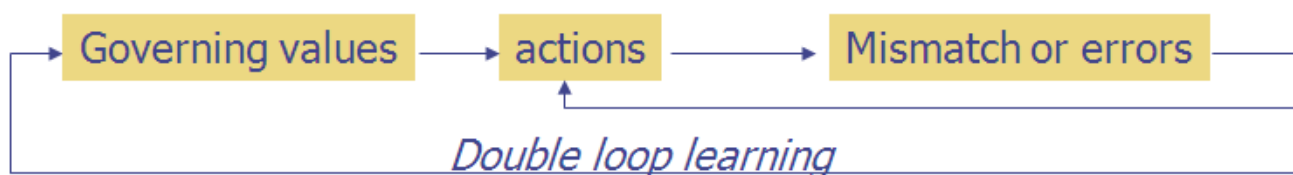


Figure 1 – Double Loop Learning

5.1.3.2 Expansive Learning

Expansive learning theory 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 creation¹⁰⁰. This view of learning is equivalent to ‘double-loop learning’ described in theories of organisational learning, 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 refined¹⁰¹.

Expansive learning theory is 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 such dialogues between stakeholders whose behaviours, attitudes or mindset have been

⁹⁸ Saari and Kallio, “Developmental Impact Evaluation for Facilitating Learning in Innovation Networks”, 230.

⁹⁹ Senge, “The Fifth Discipline”.

¹⁰⁰ Saari and Kallio, “Developmental Impact Evaluation for Facilitating Learning in Innovation Networks”, 228.

¹⁰¹ Ibid., 231.

affected by a completed piece of research through which impacts can be seen after the project has ended¹⁰². The quality of a participatory process could therefore be evaluated according to the quality of outputs 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.

5.1.3.3 Transformative and Reflective 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 speech¹⁰³, is changed¹⁰⁴. 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 interpreted¹⁰⁵.

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.

Pre-analytic assumptions can also be called a 'frame of reference' in the hermeneutic tradition which consists of a "coherent body of experience—associations, concepts, values, feelings, conditioned responses" which shape a person's interpretation of the lifeworld¹⁰⁶. As a person encounters unfamiliar experiences and ideas his frame of reference is 'tested', in the sense that his preconceptions have not equipped him to interpret and understand the unfamiliar phenomenon¹⁰⁷. While there may be a tendency to dismiss unfamiliar ideas and experiences as in some way incorrect or irrelevant to one's experiences, transformative learning occurs whenever the interpreter chooses to expand his frame of reference and find new meaning for the unfamiliar phenomena, rather than rejecting it. Transformative learners can therefore be said to "move toward a frame of reference that is more inclusive, discriminating, self-reflective, and integrative of experience"¹⁰⁸.

Four sub-types of transformative learning can be identified according to the Theory of Communicative Action¹⁰⁹:

- (1) Instrumental – learning to manipulate or control the environment or other people to enhance efficacy in improving performance;
- (2) Impressionistic – learning to enhance one's impression on others, to present oneself;
- (3) Normative – learning oriented to common values and a normative sense of entitlement...;
- (4) Communicative – learning to understand the meaning of what is being communicated"¹¹⁰.

Each type of learning is based on interpretation—learning is possible when the ability to interpret a phenomenon fails or is incomplete due to a lack of relevant experiences, meanings, evidence or

¹⁰² For such a methodology, see: Saari and Kallio, "Developmental Impact Evaluation for Facilitating Learning in Innovation Networks".

¹⁰³ cf. Gadamer, *Truth and Method*; Heidegger, *Being and Time*.

¹⁰⁴ Mezirow, "Transformative Learning".

¹⁰⁵ Chilvers, "Reflexive Engagement?".

¹⁰⁶ Mezirow, "Transformative Learning", 5.

¹⁰⁷ Heidegger, *Being and Time*.

¹⁰⁸ Mezirow, "Transformative Learning", 5.

¹⁰⁹ Habermas, *The Theory of Communicative Action: Volume 1: Reason and the Rationalization of Society*.

¹¹⁰ Mezirow, "Transformative Learning", 5.

beliefs in one's frame of reference. Learning occurs when the frame of reference is changed to accommodate new phenomena.

A critical distinction to explain how one's frame of reference changes is made between 'points of view' and 'habits of mind'. While both are products of experience and 'cultural assimilation', the former are more transient than the latter¹¹¹. 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 interpretation¹¹². 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.

5.1.3.3.1 Reflection and Reflexivity

Transformative learning becomes reflective when the "legitimacy of other sources of knowledge and the perspectives of other actors" is acknowledged¹¹³, meaning the learner's frame of reference or assumptions are opened to modification. To understand this idea a 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"¹¹⁴. 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 operationalised 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 occurs to subject decisions made across the consortium 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 irrelevant¹¹⁵, 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 process.

5.1.3.4 Social Learning

Participatory processes in which participants enter a deliberation to problem solve or discuss potential solutions implies that "co-production of new knowledge through sharing perspectives and

¹¹¹ Mezirow, "Transformative Learning".

¹¹² Ibid.; Gadamer, *Truth and Method*; Patterson and Williams, *Collecting and Analyzing Qualitative Data: Hermeneutic Principles, Methods and Case Examples*.

¹¹³ Chilvers, "Reflexive Engagement?".

¹¹⁴ Stirling, "9. Precaution, Foresight and Sustainability: Reflection and Reflexivity in the Governance of Science and Technology", 226.

¹¹⁵ Chilvers, "Reflexive Engagement?", 300.

experiences” will occur. Social learning is a transformative type of learning¹¹⁶, meaning the co-production of knowledge involves changing the frames of reference of participants in a discourse. It occurs when an individual comes to understand the “interests, values, experiences, beliefs and feelings” of others and acts in pursuit of the collective good¹¹⁷. Such learning may occur merely through interaction with information materials and other participants in participatory processes¹¹⁸. Social learning can be distinguished from others as a type of reflexive learning which requires not only understanding the perspectives of others, but sympathising in as much as the learner comes to pursue a collective rather than individual good. This type of learning has proven very influential in developing frameworks for the evaluation of participatory processes as described below (see: Section 5.1.4.3.2).

5.1.3.5 Criteria for Evaluating Mutual Learning

The theories of learning seen in academic discourse imply particular quality criteria for participatory processes in which mutual learning is sought, based upon how each conceives of learning as occurring (see: Table 5). 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-arguments¹¹⁹. 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 discourse¹²⁰ conducive to transformative learning.

¹¹⁶ Webler, “‘Right’ Discourse in Citizen Participation: An Evaluative Yardstick”.

¹¹⁷ Blackstock, Kelly, and Horsey, “Developing and Applying a Framework to Evaluate Participatory Research for Sustainability”, 728.

¹¹⁸ e.g. Stagl, “Multicriteria Evaluation and Public Participation”.

¹¹⁹ e.g. Ibid.

¹²⁰ Ibid.

THEORY	LEARNING AS	CHARACTERISTICS
Organisational	<ul style="list-style-type: none"> ▪ Knowledge acquisition (layman's) ▪ Participation in cultural practices (sociocultural) ▪ Knowledge co-creation 	<ul style="list-style-type: none"> ▪ One-way ▪ Group learning ▪ Single/double loop ▪ Presenting/receiving knowledge
Expansive	<ul style="list-style-type: none"> ▪ Knowledge co-creation 	<ul style="list-style-type: none"> ▪ Two-way ▪ Double loop ▪ Stakeholder dialogue ▪ Encounters with unfamiliar perspectives
Transformative or Reflective	<ul style="list-style-type: none"> ▪ Knowledge co-creation ▪ Challenging existing understanding and assumptions 	<ul style="list-style-type: none"> ▪ Two-way ▪ Stakeholder dialogue ▪ Encounters with unfamiliar perspectives ▪ Challenging assumptions ▪ Transformation of frame of reference used in interpretation ▪ Critical reflexivity
Social	<ul style="list-style-type: none"> ▪ Knowledge co-creation ▪ Challenging existing understanding and assumptions ▪ Adoption of collective interests 	<ul style="list-style-type: none"> ▪ Two-way ▪ Stakeholder dialogue ▪ Encounters with unfamiliar perspectives ▪ Challenging assumptions ▪ Affirmation of importance of interests of other stakeholders ▪ Pursuit of collective good ▪ Identification of (effects of) power relationships in dialogue

Table 5 – Theories of Learning

5.1.4 Evaluation Criteria

A central issue in project evaluation is how to assess the quality of a research activities and participatory processes. Discourse in this area tends to focus on establishing and justifying evaluative criteria to qualify effectiveness based on the quality of the process (or procedure) and its outcomes¹²¹. Many potential criteria of effectiveness 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 evaluated¹²². All of the many frameworks and criteria used for project evaluation should be understood as helping assess the quality (or as is sometimes said, effectiveness or success) of the activity¹²³; while the criteria vary across disciplines, frameworks and projects, the central purpose of evaluative criteria as facilitating assessment of quality does not change.

¹²¹ Rowe and Frewer, "Evaluating Public-Participation Exercises: A Research Agenda".

¹²² Rowe and Frewer, "Public Participation Methods", 3; Smith, Nell, and Prystupa, "FORUM".

¹²³ Rowe and Frewer, "Evaluating Public-Participation Exercises: A Research Agenda", 517.

5.1.4.1 Inter- and Transdisciplinarity in Choosing Criteria

A universally accepted framework to define effectiveness does not exist in evaluation literature¹²⁴, due perhaps to participatory processes occurring across numerous disciplines with different epistemic frameworks and standards of quality¹²⁵. MMLs can therefore be thought of as interdisciplinary or transdisciplinary activities, although definitions of interdisciplinarity and transdisciplinarity are contentious. For example, Huutoniemi¹²⁶ defines interdisciplinarity as “a genus of integrative research activities that combine more than one discipline, field, or body of knowledge. In contrast, the ‘transdisciplinarity’ is more focused, referring to “trans-sector problem solving where various stakeholders in society are actively involved in knowledge production.” According to Walter, Helgenberger, Wiek, & Scholz (2007), transdisciplinary research involves a “process of collaboration between scientists and non-scientists on a specific real-world problem...the research process is opened up to the stakeholders, aiming at a mutual learning process.” Initially, then, it would appear MMLs are transdisciplinary, while potentially lacking a specific (single) real-world problem focus (see: Section 5.1.1).

With this said, the particular definitions adopted are of little consequence to the discussion here, so long as MMLs are conceived of as potentially interdisciplinary or transdisciplinary, meaning they therefore involve multiple disciplines with varying epistemic and evaluative frameworks. It appears necessary to prescribe evaluative principles according to which specific criteria are chosen which reflect good practice in the specific disciplines involved in the project¹²⁷. This situation arises due to the various definitions of ‘quality’ adopted in different disciplines, in which the concept is defined according to the “codified rules, beliefs, perceptions and procedures with regard to producing and evaluating knowledge” accepted in a particular discipline¹²⁸. While this does not preclude recommending particular criteria for the evaluation of participatory processes, it may limit the degree of detail and prescription that is appropriate in identifying principles of good practice for MML evaluation. This difficulty can be understood as part of a broader debate on the possibility of a universal definition or set of criteria to evaluate the effectiveness of participatory processes, or whether the best that can be hoped for it ‘local sets’ that are specific to certain types of projects or disciplines¹²⁹. Without taking a position in this debate, in principle it should be possible to identify relatively general principles of good practice in MML evaluation on the basis that all MMLs will share certain characteristics—namely, an emphasis on mutual learning and coordinating dialogue and collaboration between stakeholders in the sciences, government and civil society.

5.1.4.2 Whose Criteria?

A related question to be answered in planning an evaluation is *whose* criteria will be used. Stakeholders tend to bring a diversity of goals, values and evaluative criteria to participatory processes¹³⁰; indeed, engaging with a representative range of stakeholders may be necessary for the process to be considered ‘fair’ or ‘rational’¹³¹ (see: Section 5.1.4.3.2). The diversity of goals, values and evaluative criteria encountered inevitably leads to conflicts, meaning a universally accepted set

¹²⁴ Rowe and Frewer, “Evaluating Public-Participation Exercises: A Research Agenda”.

¹²⁵ 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?”.

¹²⁶ Huutoniemi, “Evaluating Interdisciplinary Research”, 309.

¹²⁷ cf. Huutoniemi, “Evaluating Interdisciplinary Research”.

¹²⁸ *Ibid.*, 310.

¹²⁹ Rowe and Frewer, “Evaluating Public-Participation Exercises: A Research Agenda”.

¹³⁰ *Ibid.*

¹³¹ Renn, “Risk Communication: Towards a Rational Discourse with the Public”, 497.

of objectives to be met by the process which satisfies all stakeholders is unlikely to emerge. Criteria can therefore show preference to the interests of a particular group of stakeholders¹³², be they members of civil society or partners in the project sponsoring the participatory process. Choosing evaluation criteria should therefore be understood as a normative process which can show preference to particular stakeholders and values, potentially undermining the ‘fairness’ of the process (see: Section 5.1.4.3.2).

Three general approaches can be taken in deciding upon criteria: user-based, theory-based or goal-free evaluation¹³³. User-based evaluation accepts the existence of a diversity of potentially contradictory goals and expectations across stakeholders, and develops evaluation mechanisms (e.g. questionnaires) and criteria that either reflect the goals and expectations of all involved, or any areas of consensus across participants¹³⁴. In a user-based approach criteria are contextually developed or specified from pre-defined lists, rather than following a pre-existing framework of criteria as would occur with theory-based evaluation. Influential frameworks which are adopted in taking a theory-based approach, such as those based on ‘fairness’ and ‘competence’ criteria¹³⁵, are reviewed below (see: Section 5.1.4.3.3 and 5.1.4.3.2). Goal-free evaluation seeks to avoid ‘biased’ criteria created from the goals and expectations of stakeholders, and instead focuses more narrowly on “program effects and efficiency”¹³⁶. Several examples of user and theory-based frameworks were found in the literature.

5.1.4.3 Frameworks of Evaluation Criteria

Evaluation criteria can be conceived of as ‘tools’ which allow for the quality of activities to be compared across projects. Criteria represent the actualisation of different epistemic frameworks to identify particular aspects of an activity as representative of its quality or epistemic credibility. For example, choosing to evaluate the availability of information to participating stakeholders suggests that projects which fail to provide sufficient information materials in an easily accessible way lack quality. Recognising this, the criteria reviewed in this section can be treated as individual recommendations for good practice in evaluation of participatory research. As they represent value judgments, some disagreement should be expected in the sense that authors of evaluation frameworks will not recommend evaluating identical aspects.

Many frameworks for evaluation of participatory research exist, although adoption of pre-defined frameworks and criteria has not been widely seen in evaluation of public participation¹³⁷. An overview of historically influential frameworks provided by Sewell & Philips¹³⁸ is helpful in tracing the development of criteria which are influential in contemporary evaluation. The earliest framework identified by Sewell, the Vindasius framework, was primarily concerned with the *effectiveness* (degree to which objectives are met) and *efficiency* (economic cost of meeting objectives) of public participation, understood as requiring a two-way exchange of information

¹³² Rowe and Frewer, “Evaluating Public-Participation Exercises: A Research Agenda”.

¹³³ Abelson and Gauvin, *Assessing the Impacts of Public Participation*.

¹³⁴ Chess, “Evaluating Environmental Public Participation”, 775; Abelson and Gauvin, *Assessing the Impacts of Public Participation*.

¹³⁵ e.g. Rowe and Frewer, “Public Participation Methods”; Webler, Kastenholz, and Renn, “Public Participation in Impact Assessment”.

¹³⁶ Chess, “Evaluating Environmental Public Participation”, 776; Abelson and Gauvin, *Assessing the Impacts of Public Participation*.

¹³⁷ Rowe and Frewer, “Evaluating Public-Participation Exercises: A Research Agenda”; Sewell and Phillips, “Models for Evaluation of Public Participation Programmes”.

¹³⁸ Sewell and Phillips, “Models for Evaluation of Public Participation Programmes”.

between the public and researchers to influence planning processes and outcomes¹³⁹. Vindasius' concerns with information exchange were expanded upon by the Hampton framework, which gave attention to the types of stakeholders involved, in particular whether they are members of majorities, minorities or elite groups¹⁴⁰. The Hampton framework begins to incorporate concerns with power relationships and *representativeness* of stakeholders into participatory processes, although without explicitly creating such criteria for evaluation. Elements of this framework can be seen in modern evaluation, for example in concerns that participatory discourses may be dominated by senior individuals or 'experts' who drown out the voices of 'non-expert' stakeholders¹⁴¹. Another framework reviewed, the Farrell model, shows the beginnings of a now classic split in evaluation between process and outcome evaluation¹⁴², seen in modern discourse in the distinction between formative and summative evaluation, as well as procedural and impact evaluation. The influence of these frameworks can be seen clearly in three influential contemporary frameworks: those of EuropeAid/OECD, Weblar and Rowe & Frewer (see: Table 6).

¹³⁹ Ibid., 339.

¹⁴⁰ Ibid., 340.

¹⁴¹ e.g. Walls, Rowe, and Frewer, "Stakeholder Engagement in Food Risk Management Evaluation of an Iterated Workshop Approach", 254.

¹⁴² Sewell and Phillips, "Models for Evaluation of Public Participation Programmes", 341.

FRAMEWORK	TYPE OF CRITERIA	EVALUATION CRITERIA
EuropeAid/OECD	Procedural and Outcome	<ul style="list-style-type: none"> ▪ Relevance ▪ Effectiveness ▪ Efficiency ▪ Sustainability ▪ Impact ▪ Coherence/Complementarity ▪ Community Value Added
Webler	Procedural and Outcome (based on Social Learning and Habermas' concept of 'Rational Discourse')	<ul style="list-style-type: none"> ▪ Fairness <ul style="list-style-type: none"> ○ Equity ○ Openness ○ Ease of Attendance ▪ Competence <ul style="list-style-type: none"> ○ Stakeholder technical expertise ○ Quality of information ○ Accessibility of information
Rowe & Frewer	Procedural (based on Webler's 'Fairness' and 'Competence' meta-criteria)	<ul style="list-style-type: none"> ▪ Acceptance <ul style="list-style-type: none"> ○ Representativeness ○ Independence ○ Early Involvement ○ Influence ○ Transparency ▪ Process <ul style="list-style-type: none"> ○ Resource Accessibility ○ Task Definition ○ Structured Decision Making ○ Cost-Effectiveness

Table 6 – Evaluation Frameworks

5.1.4.3.1 European Commission's EuropeAid Framework

A generic approach to evaluation evident in the evaluation currently carried out by MMLs and similar EC research and coordination projects (see: Section 6.1) is the framework described by EuropeAid¹⁴³, the EC's Directorate-General responsible for development policy and aid delivery. The framework prescribes a structured approach to evaluation consisting of three distinct phases: Preparatory, Field and Synthesis. Specific practical steps to be undertaken by external evaluators are also described. Without reviewing the methodology as a whole, the most relevant segments of the framework are its seven evaluation criteria and methods of evaluating the quality of participatory processes.

The criteria are heavily based on criteria developed by the Organisation for Economic Co-operation and Development's (OECD) Development Co-operation Directorate (DAC), which are intended for application in research programmes in development assistance¹⁴⁴. Despite its initially 'limited' scope, the OECD's criteria have achieved influence across the European Commission and in evaluation discourses. The seven criteria, as adapted by EuropeAid, include:

¹⁴³ EuropeAid, *Evaluation - Guidelines*.

¹⁴⁴ OECD, "Evaluation of Development Programmes".

- **“Relevance:** The extent to which the objectives of a development intervention are consistent with beneficiaries' requirements, country needs, global priorities and partners' and donors' policies.
- **Effectiveness:** The extent to which the development intervention's objectives were achieved, or are expected to be achieved, taking into account their relative importance.
- **Efficiency:** A measure of how economically resources/inputs (funds, expertise, time, etc.) are converted to results.
- **Sustainability:** The continuation of benefits from a development intervention after major development assistance has been completed. The probability of continued long-term benefits. The resilience to risk of the net benefit flows over time.
- **Impact:** Positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended.
- **Coherence/complementarity:** This criterion may have several dimensions:
 - 1) Coherence within the Commission's development programme
 - 2) Coherence/complementarity with the partner country's policies and with other donors' interventions
 - 3) Coherence/complementarity with the other Community policies
- **Community value added:** The extent to which the project/programme adds benefits to what would have resulted from Member States' interventions in the same context”¹⁴⁵.

In terms of applying the EuropeAid framework to MMLs, the criteria can be easily adapted to other types of projects without distorting the meaning of the criteria, for example by replacing the phrase ‘development intervention’ with ‘engagement project’ or ‘participatory process’. To evaluate projects against these criteria evaluators are directed to develop a list of ‘Objectively Verifiable Indicators’ of success, or characteristics which must be possessed by project outputs to label a particular activity or deliverable as successful. Such indicators should be identified, along with ‘Sources of Information’, prior to delivery of the deliverable and in cooperation with responsible stakeholders (e.g. project partners)¹⁴⁶.

Procedural and outcome criteria are found in this framework, along with the influence of *representativeness* (or relevance/community value added) and *coherence* as concepts indicative of the quality of participatory processes. The framework is particularly helpful in identifying pragmatic quality criteria applicable to funded projects regardless of discipline, such as efficiency conceived of as a cost-benefit calculation or sustainability concerning the perceived lifetime of the process or intervention and its outcomes/impacts. These criteria provide a useful basis for understanding how projects may be generically evaluated according to the basic components shared by any type of funded research, such as costs, processes, impacts, and relevance to the target audience.

5.1.4.3.2 Webler’s Framework

Moving into theory-based frameworks, one of the most influential centres on the concepts of ‘fairness’ and ‘competence’ is that developed by Webler¹⁴⁷. Based around Habermas’ concept of rational discourse and the ‘ideal speech situation’¹⁴⁸, the framework is described as a set of ‘normative criteria’ for evaluation of public participation processes based on the concept of ‘social

¹⁴⁵ EuropeAid, *Evaluation - Guidelines*.

¹⁴⁶ Ibid.

¹⁴⁷ Webler, Kastenholz, and Renn, “Public Participation in Impact Assessment”; Webler, “‘Right’ Discourse in Citizen Participation: An Evaluative Yardstick”.

¹⁴⁸ Abelson and Gauvin, *Assessing the Impacts of Public Participation*, 5; Chilvers, “Deliberating Competence Theoretical and Practitioner Perspectives on Effective Participatory Appraisal Practice”, 160.

learning', which translate "uncoordinated individual actions into collective actions that support and reflect collective needs and understandings"¹⁴⁹. In the context of participatory processes, one way to define a rational discourse is as a process in which "affected parties resolve a conflict by:

- (1) reaching a consensus on the procedure that the participants want to employ in order to derive the final decision or compromise, such as majority vote or the involvement of a mediator;
- (2) basing their factual claims on the "state of the art" of scientific knowledge and other forms of legitimate knowledge; in the case of scientific dissent all relevant camps should be represented;
- (3) interpreting factual evidence in accordance with the laws of formal logic and argumentative reasoning;
- (4) disclosing the values and preferences of each party, thus avoiding hidden agendas and strategic game playing;
- (5) attempting to find a fair solution whenever conflicting values or preferences occur, including compensation or other forms of benefit exchange"¹⁵⁰.

Two components of this definition do not necessarily apply to participatory processes in MMLs. First, a conflict requiring resolution need not be the topic of discourse. Second, consensus among all affected parties in the sense that a final decision or compromise is reached need not occur, as it is not clear that consensus is required for social, or mutual, learning to occur. Social learning is described as learning that occurs from social interaction, related but not limited to learning by imitation: "social learning means more than merely individuals learning in a social situation. We envision a community of people with diverse personal interests, but also common interests, who must come together to reach agreement on collective action to solve a mutual problem"¹⁵¹. Here, public participation is clearly linked with problem solving, meaning 'social learning' may only be an appropriate evaluative criteria for MMLs in which stakeholder participation in policy-making, framework building or other processes addressing a clear area of conflict, and not those that seek awareness raising or mutual learning as an end in itself.

Social learning consists of two components: cognitive enhancement, or knowledge acquisition and technical development, and moral development, or the willingness and ability to engage with others with different values in a discourse. Both are said to occur in participatory processes in which some degree of consensus is sought among stakeholders, so the quality of such processes can be evaluated in terms of the degree to which stakeholders were supported in learning about technical knowledge and the views of others on the topic of discourse, and their willingness to take the values or moral viewpoint of others seriously. 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 participants¹⁵², and the degree to which this structure contributed to a high quality discourse in which social or mutual learning occurred between stakeholders (and researchers). Through discourse participants co-produce knowledge, meaning both researchers and other stakeholders come to understand each other's values, beliefs and 'local' knowledge¹⁵³; the quality of such a discourse can therefore be assessed according to the mutual understanding (or learning) achieved by all

¹⁴⁹ Webler, Kastenholz, and Renn, "Public Participation in Impact Assessment", 460.

¹⁵⁰ Renn, "Risk Communication: Towards a Rational Discourse with the Public", 494.

¹⁵¹ Webler, Kastenholz, and Renn, "Public Participation in Impact Assessment", 445.

¹⁵² *Ibid.*, 456.

¹⁵³ Blackstock, Kelly, and Horsey, "Developing and Applying a Framework to Evaluate Participatory Research for Sustainability", 729.

participants, not only the ‘experts’ who initially planned the process through which the discourse occurred.

Social learning is described as learning that occurs from social interaction, related but not limited to learning by imitation: “social learning means more than merely individuals learning in a social situation. We envision a community of people with diverse personal interests, but also common interests, who must come together to reach agreement on collective action to solve a mutual problem”¹⁵⁴. Here, public participation is clearly linked with problem solving, meaning ‘social learning’ may only be an appropriate evaluative criteria for MMLs in which stakeholder participation in policy-making, framework building or other processes addressing a clear area of conflict, and not those that seek awareness raising or mutual learning as an end in itself.

This type of learning, which can be described as a transformative type of mutual learning¹⁵⁵ (see: Section 5.1.3.3), as well as the criteria derived from it, are based upon Habermas’ theory of communicative action and the cooperative discourses through which social action and learning emerge¹⁵⁶. Social learning occurs through such participatory processes. Criteria based upon socio-psychological dynamics of group interactions are said to transcend the “the limitations of evaluative approaches that focus solely on empowerment, degree of influence, or subjective satisfaction with the results,” meaning process evaluation will often focus only on these aspects without recognising the importance of ‘group learning’ which occurs in participatory processes¹⁵⁷. To correct this deficiency the authors conclude that evaluation of participatory processes should include ‘social learning’ alongside ‘fairness’ and ‘competence’ as normative meta-principles of effectiveness.

‘Fairness’ refers to the equity, openness¹⁵⁸ and ease of attendance¹⁵⁹ for participatory processes, and can be evaluated by assessing stakeholder attitudes towards the equity, sense of inclusion and importance of their participation in the process; for example, do they feel their opinions will actually influence outcomes or decision-making, or in terms of mutual learning, do they feel the researchers are learning from their perspectives? Fairness means opportunities to “act meaningfully” are shared among all participants, who should have influence over “agenda setting, establishing procedural rules, selecting the information and expertise to inform the process and assessing the validity of claims”¹⁶⁰. A potential difficulty created by such an approach is that it may grant control over the scope of information included in a discourse to stakeholders lacking the expertise necessary to identify and understand relevant information—a balance must therefore be struck between granting these opportunities to all stakeholders while also ensuring stakeholders with relevant (scientific) expertise are included in the process.

This latter point relates to the ‘competence’ meta-principle, which defines a ‘competent process’ as one in which “appropriate knowledge and understanding of the issue is achieved” in a discourse through access to and interpretation of relevant information¹⁶¹. ‘Competence’ here refers to the

¹⁵⁴ Webler, Kastenholz, and Renn, “Public Participation in Impact Assessment”, 445.

¹⁵⁵ Blackstock, Kelly, and Horsey, “Developing and Applying a Framework to Evaluate Participatory Research for Sustainability”, 728.

¹⁵⁶ cf. Habermas, *The Theory of Communicative Action: Volume 1: Reason and the Rationalization of Society*; Habermas, *The Theory of Communicative Action: Volume 2: Lifeworld and System: A Critique of Functionalist Reason*.

¹⁵⁷ Webler, Kastenholz, and Renn, “Public Participation in Impact Assessment”, 443.

¹⁵⁸ Webler, Kastenholz, and Renn, “Public Participation in Impact Assessment”; Webler, “‘Right’ Discourse in Citizen Participation: An Evaluative Yardstick”.

¹⁵⁹ Rowe and Frewer, “Public Participation Methods”, 13.

¹⁶⁰ Abelson et al., “Deliberations about Deliberative Methods”, 244.

¹⁶¹ Ibid.

technical competence or expertise of stakeholders necessary to understand and evaluate the issue around which participatory processes are organised¹⁶². Competency can ground both substantive and procedural evaluations. Substantively, 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?

According to Chilvers¹⁶³, when considering these criteria against the Habermasian background of rational discourse, the quality of participatory discourses can be evaluated from a process perspective according to two primary requirements: “(1) access to information (knowledge) and its interpretations; and (2) the use of the best available procedures for knowledge selection to resolve disputes.” The second criteria hints at broader requirements of Habermas’ Theory of Communicative Action¹⁶⁴, which according to Genus & Coles¹⁶⁵ means that discourses in participatory research and stakeholder engagement need to support mutual understanding among participants, ensure participants recognise the legitimacy of the perspectives and claims of other participants (even when such claims are not accepted), and allow all actors equal opportunity for discussion “free from any form of domination whether arising from ‘strategic behaviour’” or claims to power or authority. These requirements help explain potential barriers to mutual learning where stakeholders participate in MMLs and related research, both in terms of capacities potentially lacked by stakeholders as well as ‘strategic’ structures or actions taken by participants to diminish the voice or influence of others involved.

A potential criticism of a Habermasian approach is that an overemphasis on the equality and fairness of dialogue can present a false reality in which consensus is always possible, meaning intractable epistemic differences are not a possible outcome of discourse¹⁶⁶. The importance of such a criticism may be overstated so long as participatory discourses are seen as searching for areas of agreement rather than absolute consensus; if, indeed, discourse is seen as an ideal to strive towards in facilitating ‘cooperative life’¹⁶⁷ rather than a realistic end point to be expected in participatory processes searching for consensus without tolerance for divergence. It would therefore seem that consensus is not required for mutual learning to have occurred.

5.1.4.3.3 Rowe and Frewer’s Framework

Webler’s framework has been extremely influential in evaluation of participatory processes. Perhaps the most obvious influence appears in another widely cited procedural framework¹⁶⁸ describing procedural evaluation criteria to assess the effectiveness of public participation. The framework, developed by Rowe and Frewer¹⁶⁹, addresses “public acceptance and good process in participation

¹⁶² Webler, Kastenholz, and Renn, “Public Participation in Impact Assessment”, 453.

¹⁶³ Chilvers, “Deliberating Competence Theoretical and Practitioner Perspectives on Effective Participatory Appraisal Practice”, 160.

¹⁶⁴ Habermas, *The Theory of Communicative Action: Volume 1: Reason and the Rationalization of Society*; Habermas, *The Theory of Communicative Action: Volume 2: Lifeworld and System: A Critique of Functionalist Reason*.

¹⁶⁵ Genus and Coles, “On Constructive Technology Assessment and Limitations on Public Participation in Technology Assessment”, 438.

¹⁶⁶ Chilvers, “Deliberating Competence Theoretical and Practitioner Perspectives on Effective Participatory Appraisal Practice”.

¹⁶⁷ cf. Bauman, *Hermeneutics and Social Science: Approaches to Understanding*; Gadamer, *Truth and Method*.

¹⁶⁸ Abelson and Gauvin, *Assessing the Impacts of Public Participation*, 12.

¹⁶⁹ Rowe and Frewer, “Public Participation Methods”, 11.

exercises” by distinguishing between acceptance criteria, or those “related to the effective construction and implementation of a procedure,” and process criteria, or those “related to the potential public acceptance of a procedure.” The former type of criteria are derived from Webler’s principle of ‘fairness’, while the latter are based upon the principle of ‘competence’. The following acceptance criteria are proposed:

- **“Criterion of representativeness:** The public participants should comprise a broadly representative sample of the population of the affected public.
- **Criterion of independence:** The participation process should be conducted in an independent, unbiased way.
- **Criterion of early involvement:** The public should be involved as early as possible in the process as soon as value judgments become salient.
- **Criterion of influence:** The output of the procedure should have a genuine impact on policy.
- **Criterion of transparency:** The process should be transparent so that the public can see what is going on and how decisions are being made”¹⁷⁰.

Similarly, the following process criteria are proposed:

- **“Criterion of resource accessibility:** Public participants should have access to the appropriate resources to enable them to successfully fulfil their brief.
- **Criterion of task definition:** The nature and scope of the participation task should be clearly defined.
- **Criterion of structured decision making:** The participation exercise should use/provide appropriate mechanisms for structuring and displaying the decision-making process.
- **Criterion of cost-effectiveness:** The procedure should in some sense be cost-effective”¹⁷¹.

Each criterion can be used to create indicators of success against which participatory processes can be procedurally evaluated; for example, the criterion of independence would rely upon discourses being chaired by individuals willing to allow perspectives with which they disagree fair exposure in the discourse¹⁷². Following on from Webler, the 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*, 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, 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 participants¹⁷³, perhaps due to a mismatch between the language or concepts in the information and the background expertise of the participants. According to the framework, ‘learning’ occurs when participants and sponsors have *effectively processed information* according to

¹⁷⁰ Rowe and Frewer, “Public Participation Methods”.

¹⁷¹ Ibid.

¹⁷² cf. Walls, Rowe, and Frewer, “Stakeholder Engagement in Food Risk Management Evaluation of an Iterated Workshop Approach”, 255.

¹⁷³ Rowe and Frewer, “A Typology of Public Engagement Mechanisms”, 262–3.

the above criteria, implying a knowledge-acquisition theory of learning despite the emphasis on two-way flow of information between sponsor and participant in public participation (see: Section 5.1.1).

This framework for procedural evaluation is pragmatic in the sense that, for each of the criteria, the authors provide a set of practical recommendations for activities, tools or processes to be undertaken by projects. Specific guidelines for implementation of these recommendations are not provided, as strategies will need to be built in response to the needs and focus of individual projects and the stakeholders engaged, meaning the framework straddles the distinction between user- and theory-based frameworks. However, by providing not only evaluative criteria but examples of how the criteria can be realised in practice, Rowe and Frewer's framework creates a clear vision of what evaluators should look for in evaluating participatory research; in other words, the framework provides a set of recommendations for good practice in evaluation. The framework is also unique in the sense that evaluation instruments have been developed from it and tested to a limited extent¹⁷⁴, which is rare among frameworks for evaluation of participatory processes¹⁷⁵.

5.1.4.3.4 Related Frameworks

Beyond projects in which these frameworks have been applied or adapted¹⁷⁶, elements of the frameworks of Webler and Rowe & Frewer can be seen in many other evaluation frameworks and methodologies reported in the reviewed literature¹⁷⁷, in particular those which adapt 'fairness' and 'competence' principles and criteria¹⁷⁸¹⁷⁹. For example, Abelson et al.¹⁸⁰ specify the 'fairness' and 'competence' meta-principles into a series of principles for evaluation of four aspects of a participatory process: (1) representation of different stakeholder groups; (2) location of control and deliberation as enacted in procedural rules; (3) quality of information; and (4) quality of outcomes/decisions (see: Table 4).

¹⁷⁴ e.g. Rowe and Frewer, "Evaluating Public-Participation Exercises: A Research Agenda"; Horlick-Jones, Rowe, and Walls, "Citizen Engagement Processes as Information Systems"; Walls, Rowe, and Frewer, "Stakeholder Engagement in Food Risk Management Evaluation of an Iterated Workshop Approach".

¹⁷⁵ Rowe and Frewer, "Evaluating Public-Participation Exercises: A Research Agenda".

¹⁷⁶ e.g. Walls, Rowe, and Frewer, "Stakeholder Engagement in Food Risk Management Evaluation of an Iterated Workshop Approach"; Horlick-Jones, Rowe, and Walls, "Citizen Engagement Processes as Information Systems".

¹⁷⁷ e.g. Research Councils UK, *Evaluation: Practical Guidelines - A Guide for Evaluating Public Engagement Activities*.

¹⁷⁸ e.g. Abelson and Gauvin, *Assessing the Impacts of Public Participation*, 5; Renn, "The Challenge of Integrating Deliberation and Expertise".

¹⁷⁹ For a comprehensive review of evaluation frameworks, many of which show influence from these two frameworks, see: Rowe and Frewer, "Evaluating Public-Participation Exercises: A Research Agenda".

¹⁸⁰ Abelson et al., "Deliberations about Deliberative Methods", 245.

Representation	Procedural rules	Information	Outcomes/decisions
Legitimacy and fairness of selection process	Degree of citizen control/input into agenda setting, establishing rules, selecting experts, information	Characteristics	Legitimacy and accountability of:
Is there a representative sample?	Deliberation	Accessibility Readability Digestibility Selection and presentation	Decision-making Communication of decisions Responses to decision or input More informed citizenry
Geographic Demographic	Amount of time Emphasis on challenging experts, information Mutual respect	Who chooses the information Who chooses the experts	
Political Community			
Participant selection vs. Self-selection	Credibility/legitimacy of process	Interpretation	Achievement of consensus over the decision
Inclusiveness (broad) vs. Exclusiveness (narrow)	What point in the decision-making process is input being sought? Who is listening? (e.g., Influential decision-makers or junior staff)	Adequacy of time provided to consider, discuss and challenge the information	(I.e., Broad-based understanding and acceptance of final decision) Better (or different) decisions

Table 4 – Principles for the Design and Evaluation of Public Participation Processes

The influence of Webler’s framework and Habermas’ philosophy can be seen, for example, in the emphasis on deliberation, stakeholder control and early input into decision-making processes. If stakeholders are excluded from early stages of the process, or not allowed to openly deliberate and assess the validity of claims, then two-way learning between stakeholder groups and project partners cannot be said to occur throughout the participatory process¹⁸¹.

Another framework demonstrating a clear influence from Webler and Habermas is that of Renn (2004), which specifies four evaluation criteria: fairness, competence, transparency and efficiency. Fairness is adapted directly from Webler’s framework and concerns the fairness of participation understood in terms of access to the participatory process and the quality of the discourse in Habermasian terms. Competence is adapted from Habermas and refers to the participant’s capacities to deliberate, understand alternative views and appreciate the consequences of claims forwarded in the discourse. Participants should have sufficient levels of communicative proficiencies to participate in the discourse. Transparency requires that all methods, proceedings, rules, information and results from a participatory process be openly communicated to all participants, with an appropriate balance between the quantity/quality of information provided and the mutual understanding of alternative viewpoints in the discourse. Finally, efficiency refers to the balance between resource expenditure and results of the process, in particular concerning the satisfaction of participants with the process and its outcomes.

A framework developed by Aichholzer and Westholm identifies the ‘Project Perspective’ as the centrepiece of most participatory evaluation. Overlap can be seen between their ‘relevant dimensions’ to be assessed and the frameworks described above, particularly between ‘engaging with a wider audience’ with fairness, and ‘feedback behaviour’ with competence:

- “Project management
 - Goal clarity; resource planning; responsibilities
 - Quality of tool selection and implementation; resource efficiency

¹⁸¹ cf. Ibid., 246.

- Coordination of online and offline processes
- Engaging with a wider audience
 - Promotion measures; outreach
 - Incorporation of (multiple) target group perspectives in service design
 - Accessibility; inclusiveness; barriers to participation
- Community development
 - Participation and networking patterns
- Obtaining better-informed opinions
 - Relevance and quality of information
 - Learning effects over the participation process
- Process quality
 - Gap analysis against standards and good practice
 - User and stakeholder identified areas for enhancement
 - Integration of online and offline processes
 - Harmonisation of work-practices of authority and eParticipation processes
- Scope of deliberation
 - Extent of interaction amongst participants; level of involvement
 - Extent of rationality and use of arguments
- Effectiveness
 - Cost/time effectiveness of processes and structures (e.g. cost savings/time savings in providing, aggregating and evaluating input)
- Feedback behaviour
 - Response measures set by project organisers; rates and timeliness of response
 - Feedback content and quality; participants' satisfaction with feedback
- Sustainability
 - Level of key stakeholder support; provision of resources and maintenance
 - Stakeholder perception of continuity barriers
 - Level of institutionalisation of education and training for government officials¹⁸².

This framework shows clear elements the frameworks of Webler, Rowe & Frewer and the OECD by combining 'generic' project evaluation criteria with others specific to evaluation of participatory processes, which can be connected to fairness and competence.

5.1.4.3.5 Other Frameworks

Other frameworks exist which are not explicitly related to Webler or Rowe & Frewer's frameworks. A recent review of literature concerning public engagement in research and science in society¹⁸³ identified several sets of 'indicators of success' concerning the quality of learning (Table 1), participatory processes (Table 2) and societal impacts (Table 3)¹⁸⁴. While these indicators incorporate views from both frameworks, they extend beyond the frameworks in a number of ways. First, the indicators concern not only the processes of participatory engagement, but also outcomes and societal impacts. While core 'fairness' and 'competence' criteria are evident in Table 2, the indicators extend beyond these concepts by emphasising the creation of new social networks and relationships through which exchange of perspectives and information may occur, even after the conclusion of the project. These extensions are important because they allow for evaluation of

¹⁸² Aichholzer and Westholm, "Evaluating eParticipation Projects", 11–2.

¹⁸³ Haywood and Besley, "Education, Outreach, and Inclusive Engagement".

¹⁸⁴ Tables adapted from Haywood & Besley, 2014.

indicators of mutual learning; for example, the establishment of relationships and perhaps community between participants is to be expected in mutual learning conceived of as social learning (see: Section 5.1.3.4). Overlap can be seen with Aichholzer and Westholm's framework in this emphasis on community building.

Table 1. Segment one indicators for science education and learning (micro-level).

Indicator	Description (participants refers to all project members)
Science Concepts, Theories, and Phenomena	Degree to which participants interact with, analyze, and assimilate information about scientific concepts, theories, or phenomena into existing knowledge
Scientific Process and Skills	Degree to which participants increase awareness and understanding about science processes (i.e. framing measurable questions, designing research protocols, collecting and analyzing data)
Career Connections	Degree to which participants expand awareness and understanding of careers in science, associated contributions to society, and the relevance of science to other careers
Transferable Skills	Degree to which participants are able to cultivate transferable skills throughout the project (i.e. writing, technology use, oral presentations)
Values, Perspectives, Opinions, and Attitudes (VPOA) about Concepts, Theories, and Phenomena; the Scientific Process; and Science and Society	Degree to which participants are encouraged and challenged to negotiate VPOA about science concepts, theories, and phenomena; the scientific process and the knowledge produced; and the relevance and applicability of science to society
Attitudes about the Environment	Degree to which participants are engaged in reflection and discussion about VPOA on the environment to include how social, economic, and environmental priorities are set
Lifestyle Changes	Degree to which participation in the project influences changes in behaviors or lifestyle choices (i.e. pro-environmental behaviors, time engaged in similar projects, time engaged outdoors).
Citizenship	Degree to which participation in the project influences changes in behavior related to citizen-action activities, community involvement, or general participation in decision-making processes
Engagement in Science	Degree to which participation in the project influences changes in behaviors related to participation in science-related activities, discussions, and policy making

Table 2. Segment two process indicators for participatory engagement

Indicator	Description
Extent of Involvement	Degree to which citizen participants are integrated into the research project and process (e.g. research design, data collection, data analysis and interpretation, reporting)
Accessibility (Information, Human, Material, Time)	Degree to which citizen participants have appropriate information, materials, and time to contribute to the research project or process (e.g. are participants able to review research abstracts, literature reviews, or mandates compelling the research?)
Role Definition/ Instruction/ Organization	Degree to which citizen participants are involved in the creation and defining of group and individual tasks, the clarity and structure of this process, and the availability of sufficient instruction where needed
Ownership of Outcomes and Control	Degree to which citizen participants are engaged in data analysis, involvement in the compilation and dissemination of results and application of research
Representativeness/ Inclusion	Degree to which project participants are representative of the population that is/ may be affected by the research. This may be measured by assessing the degree to which project participants represent the range of VPOA relating to the topic
Transparency, Accountability, and Fairness	Degree to which the project, research process, and decision-making procedures are open and responsive (e.g. all interested parties can participate, procedures are clear and communicated openly, participants are treated fairly in the process of engagement)
Quality and Rigor	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
Social Networks and Relationships	Degree to which the project facilitates new networks and relationships among project members or reinforces existing bonds

Table 3. Segment three PPSR science in society indicators (meso-level).

Indicator	Description (participants refers to all project members)
Needs Met	Degree to which the products generated (intellectual or material) meet the legitimate needs and expectations of participants (e.g. early warning systems, relevant information by which to make health decisions)
Scope and Influence	Degree to which products generated (intellectual or material) impact broader social, economic, or environmental systems and relevant policy (e.g. local laws and procedures, national standards, corporate practices)
Community/ Social Capacity	Degree to which the project influences the capacity of communities/social groups to respond to social or ecological challenges, negotiate conflicts, and develop solutions
Trust, Confidence, and Respect	Degree to which the project fosters general trust, confidence, and respect among project participants and in science

5.1.4.4 Emerging Consensus

Influence from each of these frameworks can best be seen in emerging consensus in the evaluation of participatory processes. For example, a widely cited comprehensive review of public participation evaluation literature¹⁸⁵ showed that the majority of evaluation studies with pre-defined criteria employ outcome criteria, with roughly half also making use of process criteria. Of the reviewed studies, only two use only process or procedural criteria in evaluating the effectiveness public participation. Outcome criteria concerning the representativeness of the process and its sample were extremely common, as were criteria concerning the impact of the process on policy or sponsor decision-making, and on the knowledge or awareness of the segment of the public involved. For process criteria concerns over the fairness of the process were common, conceived of as affecting the two-way communication and group interactions central to public participation.

Expanding on these initial themes, Chilvers¹⁸⁶ suggests “considerable consensus exists” among evaluators of participatory research on certain criteria, according to which participatory processes should:

- “be representative of all those interested and affected by a decision or action and remove unnecessary barriers to participation (*representativeness and inclusivity*);
- allow all those involved to enter the discourse and put forward their views in interactive deliberation that develops mutual understanding between participants (*fair deliberation*);
- provide sufficient resources (information, expertise, time) for effective participation (*access to resources*);
- be transparent to all those inside and outside of the process about objectives, boundaries, and how participation relates to decision making (*transparency and accountability*);
- enhance social learning of all those involved, including participants, specialists, decision makers, and wider institutions (*learning*);
- be conducted (managed and facilitated) in an independent and unbiased way (*independence*); and
- be cost-effective and timely (*efficiency*).”

To build on these seven areas of consensus, Chilvers conducted a series of interviews and workshops with evaluators to identify criteria that all public participation evaluation should operationalise. A total of eight criteria were identified demonstrating significant overlap with both frameworks: representativeness, inclusivity, clarity, transparency, legitimacy, adaptability, learning and efficiency. The overlap between these empirically developed criteria and those seen in the literature suggests that, at least to a limited degree, consensus is emerging in terms of principles of evaluation, even if these ‘criteria’ need to be specified and operationalised¹⁸⁷ within specific disciplines or projects before being practically useful. Chilvers’ sample started to do so by specifying several principles which participatory processes should fulfil, clearly showing the influence of the Webler and Rowe & Frewer frameworks:

- “Scientific analysis should support deliberation and be accessible, relevant, and usable to participants within the process.

¹⁸⁵ Rowe and Frewer, “Evaluating Public-Participation Exercises: A Research Agenda”.

¹⁸⁶ Chilvers, “Deliberating Competence Theoretical and Practitioner Perspectives on Effective Participatory Appraisal Practice”, 159.

¹⁸⁷ Rowe and Frewer, “Evaluating Public-Participation Exercises: A Research Agenda”.

- Scientific analysis conducted within the process should be responsive to the needs, issues, and concerns expressed by participants in an iterative way.
- Scientific analysis should be transparent to participants within the process and make underlying uncertainties and assumptions explicit.
- Information provided should be appropriate, meaningful and understandable from the perspective of those participating.
- Information provided within the process should faithfully represent the range/diversity of views that exist on the issue being considered.
- Information provided within the process should be responsive to the needs of participants.
- The process should have access to specialist expertise and control over who provides this assistance.
- The deliberative process should ensure a highly interactive, symmetrical, and critical relationship between participants and specialists.
- While recognizing the role of consensus, the deliberative process should emphasize diversity and difference through representing alternative viewpoints, exploring uncertainties, and exposing underlying assumptions.
- The deliberative process should allow enough time for participants to become informed and develop competent understandings.
- Those facilitating the deliberative process should have adequate substantive understanding of the issues being discussed while remaining independent and impartial as to the outcomes of the process”¹⁸⁸.

Some of these principles may be useful in identifying principles of good practice in evaluating MML. Chilvers notes however that the practitioners consulted were dubious of the possibility of operationalising the principles in evaluation due to practical institutional, cultural, political and economic constraints. It is therefore suggested that to overcome these constraints future evaluation may need to be increasingly critical¹⁸⁹, identifying limitations not only with the process itself but the context in which it and its evaluation occur.

5.1.5 Impact Evaluation

Whereas procedural evaluation will assess the internal procedures of a project, impact evaluation will address the intentional and unintentional outcomes and influence of a project, including indirect outcomes that could not have been predicted¹⁹⁰. Initial distinctions between anticipated and unanticipated impacts and those within and outside the area or field intended to be influenced by the project may be helpful in categorising and evaluating different types of impacts. 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 influenced¹⁹¹. Another distinction can also be made between beneficial and detrimental impacts¹⁹², although the determination in these terms will be dependent upon the stakeholder interests considered; one impact can be beneficial to some stakeholders, and detrimental to others.

¹⁸⁸ Chilvers, “Deliberating Competence Theoretical and Practitioner Perspectives on Effective Participatory Appraisal Practice”.

¹⁸⁹ *Ibid.*, 180.

¹⁹⁰ Tuominen et al., “Evaluating the Achievements and Impacts of EC Framework Programme Transport Projects”, 61.

¹⁹¹ Penfield et al., “Assessment, Evaluations, and Definitions of Research Impact”, 21.

¹⁹² Mickwitz, “A Framework for Evaluating Environmental Policy Instruments Context and Key Concepts”.

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 returns¹⁹³. Impacts can be said to “add to the social, economic, natural and cultural capital of a nation”¹⁹⁴. For participatory processes, impact can also be seen through change in stakeholder behaviours, attitudes or knowledge resulting from participation¹⁹⁵. 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 deliverables¹⁹⁶¹⁹⁷. Outputs and outcomes should be readily identifiable to a consortium at the conclusion of a project, whereas impact is cannot be predicted with certainty because it is defined by *future* uptake and influence of project outputs. Defining a clear endpoint for a research activity, at which point ‘impacts’ begin to materialise, is a difficult normative choice in designing evaluation¹⁹⁸.

5.1.5.1 Types of Impact in MMLs

Numerous approaches to categorise impacts exist which vary according to discipline and intended domain of impact (e.g. political, social, industrial). A general dichotomy can be seen in approaches to impact assessment, according to which studies tend to focus either exclusively on the policy domain as the “locus for identifying impacts,” or on a wider range of domains including civil society and industry, which consider other indicators of impact such as media coverage, social media activity or changes in R&D, among many others¹⁹⁹. Impact assessments limited to influence on policy have been seen in recent years as insufficient measurements of impact, due in part to the relatively small chance of a particular participatory process overly influencing a political decision²⁰⁰. In contrast, searching for impact through media coverage or ‘learning’ among participants may reveal impacts which are more difficult to track, but equally important due to the shifts in social discourse and knowledge they imply. These latter impacts have been referred to as ‘resonance’, or the possibility of raising awareness, influencing social attitudes and initialising stakeholder actions²⁰¹

As MMLs focus on mutual learning and encouraging collaboration and communication among a variety of social, industrial and intergovernmental stakeholders, focusing on both social and political impacts seems appropriate. One political typology suggests there are three different types of impacts regardless of domain: (1) raising awareness among policy makers or in public debate; (2) forming opinions of policy makers or actors in such debates; and (3) initialising actions undertaken by these groups²⁰². While this typology was created in reference to Technology Assessment and is therefore very policy-oriented, its distinction between three types of impact holds for non-political impacts as well in terms of describing the effect of the ‘impact’ on a stakeholder’s perceptions, beliefs or actions.

¹⁹³ Bornmann and Marx, “How Should the Societal Impact of Research Be Generated and Measured?”, 211.

¹⁹⁴ Saari and Kallio, “Developmental Impact Evaluation for Facilitating Learning in Innovation Networks”, 229.

¹⁹⁵ Walter et al., “Measuring Societal Effects of Transdisciplinary Research Projects”.

¹⁹⁶ Abelson et al., “Deliberations about Deliberative Methods”; Walter et al., “Measuring Societal Effects of Transdisciplinary Research Projects”.

¹⁹⁷ The use of vocabulary in the literature is not consistent (see: Section 5.1.2). 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.

¹⁹⁸ Rowe and Frewer, “Evaluating Public-Participation Exercises: A Research Agenda”, 520.

¹⁹⁹ Loeber, Versteeg, and Griessler, “Stop Looking up the Ladder”, 599–600.

²⁰⁰ Hennen et al., “Towards a Framework for Assessing the Impact of Technology Assessment”; Loeber, Versteeg, and Griessler, “Stop Looking up the Ladder”, 601.

²⁰¹ Hennen, “Impacts of Participatory Technology Assessment on Its Societal Environment”.

²⁰² Hennen et al., “Towards a Framework for Assessing the Impact of Technology Assessment”, 61–2.

For the purposes of identifying principles of good practice for MML evaluation, the ongoing need to assess impact on policy should be acknowledged as important, but in no way unique to MMLs. Beyond this, MMLs focus on the mobilisation of stakeholders and mutual learning between societal actors and researchers. This emphasis suggests the ‘learning outcomes’ of MMLs are key to evaluating the success of a MML. Furthermore, 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’²⁰³ or future perspective, beliefs and actions, it follows that learning outcomes may influence MML participants in the long-term. Therefore, it can be recommended that principles or criteria of MML evaluation should be established which evaluate the long-term impact of mutual learning on MML participants; such a focus distinguishes MMLs from other types of research in evaluating impact.

Focusing on mutual learning, Walter et al.²⁰⁴ identify 8 different types of societal impact which can be interpreted as indicative of different types of mutual learning (according to different theories of learning): (1) network building; (2) trust in others; (3) understanding of others; (4) community identification; (5) distribution of knowledge; (6) system knowledge; (7) goal knowledge; and (8) transformation knowledge. Networking building refers to the extent to which a participatory process allows the participant to establish connections with new individuals and institutions, and can be measured according to the number and strength of new contacts. Trust in others refers to an “individual’s behavioural reliance on another person under a condition of risk”²⁰⁵. Understanding of others refers to the participant’s ability to understand the perspectives of other participants on the topic of discourse. Community identification refers to the participant’s ‘sense of belonging’ which influences him to become involved in a local community. Distribution of knowledge refers to the number of times the project in which the process occurred was discussed by or in the presence of the participant. System, goal and transformation knowledge each refer to the “capacity and knowledge that the participants of the project gained during the project,” which in turn concern the “current state of the problem (*system knowledge*), the knowledge about the goals aimed at (*goal knowledge*), and the knowledge about how to transition from the current to the target situation (*transformation knowledge*)”²⁰⁶. Each type of societal impact mentioned concerns learning outcomes of the participatory process, where learning is understood in terms of knowledge acquisition rather than co-generation (see: Section 5.1.3.1).

5.1.5.2 Difficulties of Evaluating Impact

While the evaluation of impact needs to be considered, it is not unique to MMLs—indeed, the need for better methods of evaluating the impacts of publicly funded research and its implications for future expenditures is widely acknowledged²⁰⁷. Frameworks such as the UK’s Research Excellence Framework and ongoing evaluations of the EU’s Framework Programmes²⁰⁸ demonstrate the broad importance of this topic. A lack of empirically validated and broadly accepted methods has been

²⁰³ cf. Gadamer, *Truth and Method*; Heidegger, *Being and Time*.

²⁰⁴ Walter et al., “Measuring Societal Effects of Transdisciplinary Research Projects”.

²⁰⁵ *Ibid.*, 332.

²⁰⁶ *Ibid.*, 333.

²⁰⁷ cf. Arnold, “Understanding Long-Term Impacts of R&D Funding”; Bornmann and Marx, “How Should the Societal Impact of Research Be Generated and Measured?”.

²⁰⁸ cf. Arnold, “Understanding Long-Term Impacts of R&D Funding”; Tuominen et al., “Evaluating the Achievements and Impacts of EC Framework Programme Transport Projects”.

widely noted, meaning tools comparable to those for evaluating academic impact (e.g. H-indexes, impact factors) do not exist for societal impact²⁰⁹ or policy impact²¹⁰.

Evaluating societal impact faces numerous methodological and epistemic challenges, which often lead to evaluation being limited to the quality of participatory procedures and mechanisms as surrogates for effectiveness and impact²¹¹. Put simply, impact is difficult to evaluate because the influence and effects of a particular project or output will often not materialise until several years after the end of the project, at which point assessment of impact has ended. Additionally, subtle impacts such as changes to the behaviours and attitudes of individual participants are practically difficult to measure. Four primary difficulties are identified by Martin²¹²: (1) establishing causality of impacts; (2) attributing complex impacts to a particular project or input—for example, calculating the proportion a new environmental policy can be attributed to a particular project; (3) identifying impact across international borders, as research and development are inherent global activities; (4) focusing on too short of a timescale when impacts may occur many years after a project has ended, meaning research with immediate benefits appear more impacting. The difficulties associated with evaluating long-term impacts of a project can be connected to the “cost, a need for commitment over an extended period of time and problems showing that results are caused by a single programme or activity, as opposed to many other variables”²¹³. Another difficulty unique to assessing impact in terms of learning is that evaluation relies upon the participant’s memory and capability to identify attitudinal change²¹⁴, which is dependent on the variable capacities of individual participants and the lapse of time between participation and evaluation.

Despite these difficulties, a small number of systematic methods for the evaluation of societal impacts of research (such as mutual learning) during a project’s duration have been created. One such method attempts to establish and quantitatively measure a causal chain along stakeholder involvement, impacts and effects on decision-making capacity via multiple participant questionnaires, interviews and social impact assessments²¹⁵. A basic method to evaluate learning that has occurred as the result of a participatory process is to conduct an identical questionnaire of participant attitudes at the beginning and end of the process²¹⁶, and perhaps several days or weeks afterwards as well to ‘measure’ further reflection potentially stemming from the process. Another (Q-Sort) examines changes over time in participant’s perceptions and attitudes by examining how they sort statements encountered in engagement discourses (Q-sort)²¹⁷. Changes in mindset are allegedly reflected in the sorting and provide evidence of learning, allowing for direct ‘objective’ comparison of viewpoints alongside the participants’ reasoning behind the categorisations²¹⁸. Other approaches rely on data collected after the project’s end which creates practical difficulties with access to stakeholders and funding, although it is worth noting that societal impact need not be conceived of as a static set of objectives or measures to be taken during and after a piece of research;

²⁰⁹ Bornmann and Marx, “How Should the Societal Impact of Research Be Generated and Measured?”, 212; Bornmann, “What Is Societal Impact of Research and How Can It Be Assessed?”; de Jong et al., “Evaluation of Research in Context”.

²¹⁰ Emery, Mulder, and Frewer, “Maximising the Policy Impacts of Public Engagement: A European Study”.

²¹¹ Ibid.; Rowe and Frewer, “Evaluating Public-Participation Exercises: A Research Agenda”.

²¹² Martin, “Assessing the Impact of Basic Research on Society and the Economy”.

²¹³ Chess, “Evaluating Environmental Public Participation”, 773.

²¹⁴ Walter et al., “Measuring Societal Effects of Transdisciplinary Research Projects”.

²¹⁵ Ibid.

²¹⁶ e.g. Stagl, “Multicriteria Evaluation and Public Participation”.

²¹⁷ Brown, “A Primer on Q Methodology”.

²¹⁸ Chess and Johnson, “Organizational Learning about Public Participation”.

for example, expansive learning theory conceives of impact as an ongoing “qualitative learning challenge”²¹⁹, meaning impact is equated with the learning outcomes of a piece of research. Such an approach requires researchers to engage in critical ‘double-loop’ dialogue with stakeholders wherever possible through which learning occurs (see: Section 5.1.3.2).

While assessing mutual learning and thus the influence of MMLs on participant perceptions, beliefs and actions may be methodologically difficult without extensive post-project research, this does not justify ignoring impact altogether. As suggested above, it may instead possible to evaluate the likelihood of future impacts according to the quality of the procedures through which mutual learning occurs (see: Section 5.1.5). If a rational discourse approach is adopted the quality of the discourse is directly equivalent to changes in mindset among participants. If it is assumed attitudinal change leads to behavioural change over time, then it can be said that rational discourses marked by open minded participation are likely to have lasting impacts on participants compared to other types of discourses.

5.2 DISCUSSION

Much of the literature addresses the quality of participatory processes in particular, meaning evaluation primarily addresses the quality of project activities through which stakeholders are engaged. However, this should not be taken to mean that only participatory activities need to be evaluated; rather, the quality of such activities should also be assessed through analysis of project documents including reports, communications between partners and other deliverables²²⁰. The lack of attention given to such pragmatic evaluation should not be taken as a lack of importance, but rather that project documents are typically taken for granted as a source of data in the literature and related training materials²²¹. This finding may be a result of the search terminology used, which focused on learning and evaluation of participatory processes in particular, as these were the characteristics of MMLs identified as unique and in need of further consideration (see: Section 4). In doing so, the importance of ‘generic’ evaluation of project activities and deliverables (e.g. EuropeAid framework) should not be forgotten; to do so would be to eliminate a key source of information concerning the quality of project activities. As explained below (see: Section 6.1.1), evaluation of project documents and communications is common in current MMLs.

5.2.1 Stakeholder Inclusion in Evaluation

Overlap is evident among the frameworks found in the reviewed literature. Chilvers (2008) empirically identified areas of consensus in evaluation literature concerning participatory processes, which creates an impromptu framework of generic criteria to be employed in the evaluation of participatory processes. These areas of overlap implicitly promote a view of participatory processes as rational discourse in which the fairness and competence of the dialogue are crucial to assessing its quality, and thus the quality of the learning which occurs through such dialogues. Fairness, competence and learning all imply that the quality of a dialogue cannot be assessed entirely by a third-party, but instead requires input from participants in the dialogue on the perceived fairness and competence of the dialogue, its participants, and the information/perspectives considered. This area

²¹⁹ Saari and Kallio, “Developmental Impact Evaluation for Facilitating Learning in Innovation Networks”, 240.

²²⁰ cf. Rowe and Frewer, “Public Participation Methods”.

²²¹ e.g. Bergmann et al., *Quality Criteria of Transdisciplinary Research: A Guide for the Formative Evaluation of Research Projects*; EuropeAid, *Evaluation - Guidelines*; OECD, “Evaluation of Development Programmes”; Petts and Leach, *Evaluating Methods for Public Participation: Literature Review*; Research Councils UK, *Evaluation: Practical Guidelines - A Guide for Evaluating Public Engagement Activities*; Scriven, *Evaluation Thesaurus*; Spaapen, Wamelink, and Roberts, *The Evaluation of University Research*.

of consensus in the literature suggests that MML evaluation should include stakeholders in assessing the quality of dialogue facilitated by the project because assessments of ‘fairness’, ‘competence’ and ‘learning’ will vary according to individual perspectives of participants in the discourse. An objectivist approach to evaluation in which the evaluator makes judgments of ‘right’ and ‘wrong’ independently of the perspectives of participants would therefore be methodologically and epistemologically inappropriate.

5.2.2 Evaluative Framework for MMLs

The framework developed by Haywood & Besley (see: Section 5.1.4.3.5) integrates contributions from the ‘fairness’ and ‘competence’ oriented frameworks seen throughout the reviewed literature, different theories of learning through which mutual learning can be conceptualised, and traditional evaluative criteria such as those specified in the EuropeAid framework (see: Section 5.1.4.3.1). In the context of MML evaluation, indicators of education and mutual learning are particularly helpful. Among the reviewed literature, Haywood & Besley’s framework can be considered the closest to an evaluative framework for MMLs. Without being overly prescriptive and disrespectful of the significant discipline and context-specific challenges faced in interdisciplinary and transdisciplinary research²²², Haywood & Besley’s framework provides a clear structure for identifying prescriptive principles of good practice in MML evaluation which respects the distinctions between process and outcome, learning and education, procedural and substantive evaluation.

This is not to say all MMLs should adopt Haywood & Besley’s evaluative framework—despite its clear potential in evaluating the quality of mutual learning through criteria concerning capacity building, communication between stakeholders, acquisition of skills and knowledge, among others, it is not yet empirically proven as an effective framework for MMLs. The areas of overlap identified by Chilvers (2008) may provide a more conservative, less prescriptive starting point, as they reflect consensus on criteria seen throughout the reviewed literature rather than representing a particular framework, such as Haywood and Besley’s. A combination of the two may be the way forward for MMLs; both provide (meta)criteria or areas requiring evaluation, which can provide a starting point for evaluators in MMLs to specify a set of criteria appropriate to the discipline, stakeholder and deliverable-specific needs of a particular MML. A principle of good practice in defining appropriate evaluative criteria can therefore be identified: 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).

5.2.3 The Need for Reflexivity in MMLs

If truly reflexive two-way learning is sought in MMLs (which can be contested, see: Section 5.1.3.3.1), then theories of transformative and reflective learning suggest that particular perspectives and activities need to be adopted by MML consortium. As suggested by Chilvers²²³, participants and partners alike will need to acknowledge “their underlying assumptions, motives, and commitments

²²² cf. Bergmann et al., *Quality Criteria of Transdisciplinary Research: A Guide for the Formative Evaluation of Research Projects*; Huutoniemi, “Evaluating Interdisciplinary Research”; Klein, “Evaluation of Interdisciplinary and Transdisciplinary Research”.

²²³ Chilvers, “Reflexive Engagement?”, 300.

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. Partners will need to show respect for alternative views and trust in the integrity of other partners if such events are to progress beyond mere (dis)agreements on proposed actions²²⁴. Although these activities and attitudes will involve all members of a MML consortium, adoption can be facilitated by evaluators and thus expressed as a principle of good practice.

5.3 INITIAL PRINCIPLES OF GOOD PRACTICE

The following is a list of initial principles of good practice for carrying out evaluation and reflection in MMLs. The principles have been developed according to the author’s interpretation of the literature (see: Section 5.1). Wherever possible principles are kept as general as possible to avoid specifying methods or criteria to be adapted which may conflict with discipline-specific requirements of quality and evaluation due to the inherent inter- or transdisciplinarity of MMLs (see: Section 5.1.4.1). The principles will be further refined in Section 5 after considering findings from an empirical study with existing MML coordinators, evaluators and project partners.

Criteria Principles

- Evaluative criteria should be specified according to the context of the particular MML, including potentially engaging the consortium to identify appropriate discipline-specific criteria for particular activities and deliverables (see: Section 3.1.4.1).
- MMLs should have clearly defined indicators of success concerning the quality of processes and outcomes prior to the start of evaluation (see: Section 5.1.4.3).
- Evaluation should address the ‘generic’ qualities of participatory processes such as those areas of consensus in evaluation literature identified by Chilvers (2008). 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 & Besley (2013) and Walter et al. (2007).

Methodology Principles

- 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 impact²²⁵, and provide stakeholders with a better idea of the value of their participation by tracking influence on the process²²⁶. MML evaluation should, at a minimum, seek to meet these three generic aims.
- A clear ‘endpoint’ should be specified at which point project impacts can start to be identified and evaluated (see: Section 5.1.5).
- 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) (see: Section 5.1.5.2).

²²⁴ e.g. Stagl, “Multicriteria Evaluation and Public Participation”.

²²⁵ Research Councils UK, *Evaluation: Practical Guidelines - A Guide for Evaluating Public Engagement Activities*.

²²⁶ Rowe and Frewer, “Evaluating Public-Participation Exercises: A Research Agenda”.

- Evaluation should include stakeholders in assessing the quality of dialogue facilitated by the project, as fairness, competence and learning all have an implicit component of subjectivity, requiring the perspectives of participants (or ‘learners’) to be collected and assessed (see: Section 5.1.4.3).
- Evaluation should occur before, during and after the project to ensure all processes and impacts are evaluated to some degree (see: Section 5.1.2).

Mutual Learning Principles

- Mutual learning outcomes among project participants should be assessed (see: Section 5.1.3), for example by monitoring changes in participant perspectives, beliefs and actions over time.
- In evaluating the quality of mutual learning that has occurred, the possibility of mutual learning without absolute consensus should be recognised (see: Section 5.1.4.3.2).
- Data collection and analysis methods conducive to evaluating learning or attitudinal change over time should be employed in evaluation (see: Section 5.1.2.1)
- 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 (see: Section 5.1.2.2).
- 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 (see: Section 5.1.3).

Reflection Principles

- Reflexive questioning of project progress, indicators of success and alterations to planned activities should occur across the MML consortium, potentially through workshops or meetings scheduled at set intervals, to ensure feedback on project progress leads to corrections and improvements to project activities (see: Section 5.1.3).

Each of the principles specified will be re-considered in light of the responses of existing MML coordinators, evaluators and project partners in the empirical study. The next section describes the results of this study.

6 STUDY OF MMLS

In parallel to the completion of the literature survey, a two-part empirical study was conducted with consortia partners from existing MMLs. The study involved reviewing MML publications and interviewing partners to better understand how current MMLs are being evaluated. The study was carried out to contribute to the set of initial principles of good practice in MML evaluation and reflection identified above. After reviewing project documents describing the strategies and findings of evaluation, MML partners were interviewed concerning their experiences with evaluation and reflection, and asked for recommendations for future MMLs.

6.1 MML PUBLICATION SURVEY

MML publications including evaluation reports, journal publications, terms of reference and other deliverables concerning evaluation were reviewed prior to interviewing partners to identify methods, results, recommendations and principles of evaluation and reflection. The survey of MML publications was systematic, with sources identified through MML project web sites as well as contact with project coordinators and evaluators. Cooperation was sought from project partners (e.g. evaluators or coordinators) all MMLs to assist in identifying relevant publications. Of the 17 other MMLs currently in existence, publications were obtained from six. A total of fifteen documents were reviewed: 5 deliverables describing evaluation methods, tools or criteria, 5 evaluation reports, 2 responses to evaluation reports, 1 terms of reference for evaluation, 1 presentation summary and 1 journal article. The relatively small sample size can be explained to some extent by the recent start dates of many MMLs for which these types of documents have not yet been produced or made publicly available (see: Appendix 3 – MML Criteria).

6.1.1 Results

Overlap was identified in many areas between the documents and literature survey. The results of the document analysis, which were analysed in the same manner as the literature survey (see: Section 5), are arranged by project to ensure data collection and analysis methods, evaluative criteria, indicators of success and recommendations are considered as part of a whole evaluation methodology rather than distinct from the project context. For the sake of anonymity of participants in the interview study specific documents and project names or areas of investigation are not mentioned in this section; doing otherwise may facilitate reverse engineering of the identities of interview respondents.

6.1.1.1 Project A

One project stood out from others by providing a set of reports clearly explaining the approach, justification and results of evaluation. Reviewing the methodology of this project in detail reveals significant overlap with the findings of the literature survey, and suggests methods for putting the initial set of principles specified above into practice (see: Section 5.3).

A mixed methods approach was taken in explicit recognition of the transdisciplinary nature of the project, the variety of stakeholders engaged and complex organisational structures among partners. The approach draws on several methods for evaluating transdisciplinary research from the Evaluation Network for Transdisciplinary Research²²⁷, the Research Embedment and Performance

²²⁷ Bergmann et al., *Quality Criteria of Transdisciplinary Research: A Guide for the Formative Evaluation of Research Projects*.

Profile (REPP)²²⁸, and the Context, Input, Process and Product (CIPP) model²²⁹. According to an evaluation report these methods were chosen due to their emphasis on evaluating the success of research according to how well societal interests are evident in (non)academic processes and outputs, as opposed to purely bibliometric measures.

Evaluation began with a pre-evaluation questionnaire distributed to consortium partners to gather their expectations of participation and evaluation, including types of outputs to track and indicators of success for project activities. Specific outputs were identified, which were ranked by partners according to importance. Societal impact ranked highest among expected outcomes. Academic partners tended to rank academic outputs (e.g. journal articles) as key, whereas activists wanted attention to be given to the quality of training, networking opportunities and impacts on social awareness.

Evaluation reports with recommendations are set to be produced every 12 to 20 months throughout the life of the project. In the first report a series of recommendations were made, including the need for external review of project reports by an expert in a relevant discipline to assess their academic quality, where the evaluator may lack the necessary expertise. Evaluation data was collected through a number of methods, including interviews with partners and participants, surveys and observations of participants at engagement events, project reports and policy briefs, training documents, partner reports on dissemination, and evaluation surveys distributed to partners and WP leaders after each workshop. For several of the data types specified, specific methods of analysis were not identified in initial strategy documents.

Evaluation consisted of formative evaluation of stakeholder input into the project including the quality of stakeholder engagement processes and consideration of the project's wider context, as well as a summative evaluation of the quality of 'outputs' and 'impacts' including impact on the wider context. Formative evaluation specifically evaluated the representativeness of stakeholders involved, the extent of networking and learning between activist groups and with researchers, the building of capacities among stakeholders through training (in particular evidence of integration of scientific ideas into behaviour), 'contextual factors' consisting of possibilities of future actions according to the input of stakeholders into the project, partner progress against milestones and deliverables specified in the description of work, and the quality of communication and collaboration between the consortium. Summative evaluation looked at the quality of research findings in terms of understanding the causes of problems in the research area and how to use stakeholder conflicts to create solutions, quality and quantity of outputs including academic articles, project deliverables and media, and finally societal impact conceptualised as societal learning and knowledge transfer, public awareness of the research area, policy impacts and accountability among groups seen as the cause of problems in the research area. Influence of the project's outputs on future research and activism was also specified, albeit without a clear method for evaluating such impacts which are traditionally difficult to track²³⁰. The combination of data collection methods specified above means the evaluation was based on a mix of self-reported partner and participant perspectives, as well as independent document analysis and observation by the evaluator.

²²⁸ Spaapen, Wamelink, and Roberts, *The Evaluation of University Research*.

²²⁹ Stufflebeam, "The CIPP Model for Evaluation".

²³⁰ e.g. Bornmann and Marx, "How Should the Societal Impact of Research Be Generated and Measured?"; Bornmann, "What Is Societal Impact of Research and How Can It Be Assessed?".

Indicators of success focused on implementation (stakeholder inputs, processes and contexts), societal impact (effects on policies and public awareness) and scientific quality and quantity (research findings and academic outputs). Non-bibliometric indicators included positive evaluation of participatory events by stakeholders, an increase in linkages between involved organisations, media coverage of project activities, enhanced training among participants, and enhanced influence among traditionally underrepresented stakeholders in public discourse.

6.1.1.2 Project B

The second project used a four-step evaluation method, starting with all task leaders providing a list of evaluation indicators (e.g. indicators of success) specific to individual tasks. This approach involves partner input at a more detailed level than Project A, which gathered general rather than task-specific indicators. These indicators were then reviewed by the evaluators in terms of whether they would actually lead to the operationalisation of objectives to be measured in the evaluation, and thus whether they will lead to specific deliverables and outcomes (e.g. impacts). If necessary, indicators were tweaked or added to ensure deliverables, and thus impact, was high quality and met the project's objectives. This method created a specific set of indicators for each sub-task, allowing for a highly detailed evaluation. Thirdly, evaluation data was collected through a variety of methods. Finally, the data was analysed, conclusions drawn and recommendations made to improve the quality of ongoing processes or the next step of a particular task. Three evaluation reports will be produced in which an analysis, conclusions and recommendations are provided for each sub-task.

Data collection and evaluation primarily occurred around a series of conferences organised by the project to debate a particular assessment method in their area of research. Success criteria identified for the conference included the representativeness of participants, coverage of a broad spectrum of relevant subjects and policies, and increases in the awareness of the research area among countries not currently involved. These criteria were summarised as three determinants of success: the conference's fitness in mobilising suitable actors, ability to safeguard the quality of its content, and ability to practically facilitate a debate. To evaluate the conferences according to these criteria data was gathered and assessed via desk work, analysis of developed materials (leaflets, papers, invitations), interviews with conference organisers, policy makers and participants, expert appraisals, online participant surveys, examination of participant lists and observations by evaluators.

6.1.1.3 Project C

In comparison to the previous two projects, Project C clearly distinguished four stages of evaluation with particular aims at each. In 'early-stage' evaluation a checklist was distributed to the project partners before substantive interventions occurred to ensure the project's aims, deliverables, primary questions, societal and research aims, methods, sources of data collection and methods of data analysis were all clearly defined and appropriate resources allocated for each. Questionnaires were then distributed during a mid-point, end-point and post-project evaluation. Mid-point represented a phase in the project at which its aims and methods could still be modified without damaging the validity of its outcomes. The end-point questionnaire is intended to be distributed when the final project report is submitted to measure levels of partner satisfaction with outputs and processes. Finally, and uniquely, the post-project questionnaire will be sent to partners 12 months after the end of the project in an attempt to formally assess longer-term impacts by asking partners to report any published results and reflect on the experiences of participants and new connections formed with partner organisations. At each stage evaluation reports are submitted containing conclusions and recommendations for improving ongoing and future project activities.

In outlining the evaluation methodology, the project made six recommendations for planning project evaluation in general: (1) clearly identify who is responsible for overseeing evaluation; (2) identify partners and stakeholders to be involved in the evaluation; (3) discuss the purpose and methods of evaluation with participants and set out its scope and aims; (4) prepare partners for the possibility of negative or critical results; (5) clarify any differences in relation to the objectives of evaluation, and deal with the differences openly; and (6) have all forms prepared to be handed out at events rather than sending them at a later date to ensure a higher completion rate. In considering these recommendations, it should be noted that the methodology undertaken in Project C is conducted by all consortium partners rather than a single evaluator, which may necessitate greater clarity in defining the partner(s) responsible for overseeing the evaluation.

6.1.1.4 Other Projects

Three of the projects provided only one or two short documents lacking an overview of the evaluation methodology. One of the projects provided a formative evaluation report and an initial set of questions for evaluating the project's impact. The report revealed that evaluation was designed to occur annually and move from formative to summative over the life of the project. Problems and recommendations were meant to be identified based on annual interviews with project partners and other stakeholders and anonymous e-surveys distributed to conference attendees. Participatory 'reflection sessions' convened at project conferences were also to be observed, which (according to the report) were intended to create 'double loop' learning among the consortium. However, this design was not employed throughout the life of the project due to a change in evaluator. Later evaluation is set to focus on impact, with 'impact questions' looking for mutual learning as shown in changes to partners' mindsets as a result of the project. Mindset is seen to consist of understanding, beliefs, aims, questions asked and ways to address and answer those questions. Mindset change is thought to be evident when partners reassess the feasibility of objectives to achieve during the project, as well as the overall aims of their involvement. Questions also focused on the perceived sustainable achievements of the project in terms of creating lasting change in the project's research area.

Another project employs self-evaluation questionnaires and brief snapshot interviews with participants in engagement events. The questionnaire focuses on the quality of the event itself, with participants asked about the demands placed on them by the event, suggestions for improvements to future events, and to rate the event against a set of specific indicators of success unique to the particular event or culture in which it occurs. In contrast, the snapshot interviews assess the learning that has occurred as a result of the event by asking participants about their knowledge of the topic prior to the event, the clarity of the information provided (see: Section 5.1.4.3.3), changes to opinions or knowledge as a result of the event, and interest in future learning on the topic. The interview topics implicitly support a reflective approach to learning (see: Section 5.1.3.3), as the participant is asked about challenges to his opinions during the event and how these opinions changed as a result. Both the questionnaire and interviews provide data for case studies of each culture in which engagement events occur, which are then evaluated by an external evaluator.

The final project was unique in its planned use of a formal quantitative method to compare viewpoints of participants and track learning over time. The project used Q-sort methodology²³¹ to track changes in perceptions and attitudes over participants in project cases over time, reflecting learning resulting from participation. In conducting the Q-sort analysis difficulties were faced in

²³¹ See: Brown, "A Primer on Q Methodology".

identifying a ‘start point’ of learning due to interactions between stakeholders and partners building upon existing collaborations, making it difficult to attribute a particular belief or behaviour to involvement in the project. Recognising these difficulties, rather than conducting a second Q-sort towards the end of the project, in-depth qualitative interviews, institutional analysis and case study writing events will be conducted instead.

6.1.1.5 Discussion

The documents reviewed show evidence of support for some of the principles identified above (see: Section 5.3). For example, the focus on evaluating stakeholder mutual learning through interviews in one project supports the need to evaluate MML success in terms of mutual learning within and between all partners and participant groups. Beyond the principles already identified, the indicators of success mentioned in Project A point towards generic criteria for MML evaluation. Positive evaluation of participatory events by stakeholders suggests that the quality of the event is directly attributable to stakeholder reaction. This indicator potentially reduces quality merely to stakeholder satisfaction, which should be resisted as quality can conceivably be connected to factors beyond the stakeholders’ awareness such as representativeness or transparency of decision-making. Equating success 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. Connecting training among participants with quality is also helpful, although Project A failed to suggest an approach to categorise (e.g. quantify or qualify) training. The emphasis on training in evaluation is helpful in the sense that MMLs are intended to build capacities among involved stakeholders, but the contribution of Project A requires the support of a typology or framework of training, or an account of evidence which indicates training has built capacities. Basic tests of skills or knowledge pre- and post-involvement may provide such evidence.

Finally, enhancing the ‘voice’ of underrepresented stakeholders 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. This indicator 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 others²³². A similar concern was reflected in Project B’s indicator concerning increasing the awareness of the research area among countries not currently involved in the societal discourse(s) addressed by the MML.

The approach taken in Project B to set indicators of success has certain advantages which may justify the extra effort on part of the consortium to assist in evaluation. The approach has the advantage of deriving evaluation from the person(s) most familiar with a particular deliverable or task, while still allowing for it to be evaluated by someone with a holistic view of the project assuming the evaluators retain the right to review the indicators prior to conducting the evaluation. The approach allows members of different disciplines to bring discipline-specific quality criteria to bear on their work, even when the evaluators lack expertise in their discipline(s). The approach is therefore amenable to inter- and transdisciplinary projects in particular, including MMLs.

²³² cf. Foucault, *Discipline & Punish*; Habermas, *The Theory of Communicative Action: Volume 1: Reason and the Rationalization of Society*.

Considered together, the indicators identified in the documents review suggest an additional ‘Criteria Principle’ of good practice in MML evaluation:

- 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.

Methodological suggestions can also be found in the approaches mentioned reviewed documents, for instance the snapshot interviews mentioned in one project can be seen as a way to evaluate learning from participation, conceived of as behavioural and attitudinal change. Concerning Project C, the surveys distributed in the mid and end phases of the project appeared to have collected data of limited utility because partners are prompted to agree or disagree with a series of generic statements concerning progress, with two open ended questions concerning the ‘most valuable’ and ‘most difficult’ aspect of the project. However, the survey may have value in requiring partners to reflect on objectives/aims which have been forgotten or delayed. The post-project survey is also of questionable value. Statements helpfully conceive of impact in terms of the project’s impact on participant knowledge, relationships between stakeholders, influence on subsequent research and funding, raising awareness and publications; however, it is not clear that partners merely agreeing or disagreeing with simple statements is of much value in understanding the content or meaning of the project’s impact.

Project C also provided six recommendations for planning project evaluation in MMLs. These recommendations point towards an additional ‘Methodological Principle’:

- To ensure the consortium understands the process and has an opportunity to express concerns and expectations, engage in a dialogue concerning the scope, aims, methods and (where pre-defined) indicators of success used in the evaluation before it begins.

The reviewed documents can be seen as the start of a process of identifying principles of good practice in MML evaluation from existing MMLs. The discussion of these documents should therefore be considered in parallel to the data collected from interviews with MML coordinators, evaluators and other partners.

6.2 INTERVIEW STUDY

Members of consortia responsible for existing MMLs were approached for interviews concerning their experiences with evaluation and reflection. Of the 17 other MMLs, interviews were conducted with partners from 14 of 17, six of which also provided publications. One of the three MMLs for which information was not obtained had not yet started, while another was in ‘Very Early’ stages did not yet have publications or an evaluation strategy which they were willing to discuss with the author. The third was a ‘Middle’ stage MML (see: Appendix 3 – MML Criteria) for whom publicly available publications could not be located via website searching or by contacting the consortia, despite repeated requests for interviews and information concerning evaluation.

6.2.1 Methodology

A primary challenge faced in the study was to learn from the experiences of a relatively small group of people who have experience with evaluation and reflection in MMLs. As mentioned above, only 18 MMLs have been funded, with many starting so recently that evaluation has not yet been carried

out (see: Appendix 3 – MML Criteria). Qualitative methods were chosen on the basis that relatively few individuals possess experiences and insight stemming from MML evaluation. The interview study can therefore be conceived of as an exploratory study looking at a type of project evaluation and reflection which is not yet widely reported on in academic discourse. A quantitative study would therefore be difficult because predetermined quantities amenable to measurement cannot be established prior to understanding the experiences of the targeted sample with MML evaluation.

Supporting this, qualitative research is often useful in forming knowledge about phenomena and beliefs about which little is known²³³. This usefulness is derived from the capacity of qualitative methods to capture meanings participants assign to phenomena²³⁴. Meaning is captured through interviews that focus on the context and beliefs of practitioners²³⁵. Qualitative methods therefore contrast the tendency of quantitative research to strip data of secondary variables or context that may ‘corrupt’ results in the pursuit of objective understanding. In this case, experience with MML evaluation is the context targeted.

6.2.2 Recruitment and Sample

Qualitative research typically focuses on a small number of participants to develop an in-depth understanding of the phenomenon under study²³⁶. To this end purposive sampling²³⁷ was used to recruit MML coordinators, evaluators and stakeholders/participants to the study. While the sampling was purposive in that individuals with particular experiences were recruited, it did not have to be limited to ‘representative types’ typically used to define a desired sample of individuals that are representative of a larger population²³⁸. Rather, recruitment was relatively comprehensive in that coordinators, evaluators and other project partners from all existing MMLs were invited to participate, with 82% of MMLs represented in the sample (14 of 17) including multiple participants in some cases.

Twenty three individuals were interviewed: 9 coordinators, 9 evaluators, 4 project partners responsible for areas related to evaluation (see: Table 4), and one project officer from the EC. Evaluators were sampled to provide first-hand experiences and reflections on evaluation. Coordinators and other partners were sampled to provide experiences and reflections based on participation in the evaluation and reflection activities carried out by project evaluators. The chosen sample thus allowed for an equal balance between the ‘external’ views of evaluators who actually coordinated the evaluation and the ‘internal’ experiences of coordinators and partners who participated in it.

Potential participants were identified by reviewing project websites and the CORDIS database. Coordinators were typically used as a point of first contact, unless contact details were publicly available for evaluators. Participants were invited to the study via e-mail and provided with an Information Sheet and Consent Sheet at least 24 hours prior to the interview (see: Appendix 2 – Human Research Ethics Documentation). All individuals that indicated interest in participating were

²³³ van Hooren et al., “Providing Good Care in the Context of Restrictive Measures: The Case of Prevention of Obesity in Youngsters with Prader-Willi Syndrome”, 167.

²³⁴ Casterlé et al., “Researching Lived Experience in Health Care: Significance for Care Ethics”, 234.

²³⁵ Lincoln and Guba, “Competing Paradigms in Qualitative Research”, 106.

²³⁶ Miles and Huberman, *An Expanded Source Book: Qualitative Data Analysis*.

²³⁷ e.g. Mays and Pope, “Rigour and Qualitative Research”, 109–10; Reed, “A Sampling Strategy for Qualitative Research”, 54; Tuckett, “Qualitative Research Sampling: The Very Real Complexities”, 2–3.

²³⁸ Patterson and Williams, *Collecting and Analyzing Qualitative Data: Hermeneutic Principles, Methods and Case Examples*, 9:41.

interviewed, with the exception of one coordinator and one evaluator, both of whom cancelled the interview due to a scheduling conflict. Consent was taken prior to each interview, with respondents given sufficient time to ask any questions about the study and their role. The interviewer ensured the participant had read and understood the information provided prior to taking consent.

6.2.3 Data Collection

Data collection occurred through semi-structured interviews typically lasting between 15 and 40 minutes, with one exceptional interview lasting 3 hours. As the goal of the study was to develop principles of good practice in MML evaluation according to participant insights and experiences, an approach to data collection was required which allowed for revision of the interview tool as the study progressed to reflect the topics identified by participants as important. This approach was necessary to reflect the fact that participants possess unique and privileged experiences from MML evaluation.

6.2.3.1 Interview Tool

The interview tool initially consisted of a set of broad topics and questions informed by the results of the literature survey (see: Appendix 4 – Interview Schedule). Iterative development of the tool occurred throughout the study by DMU and partners at Trilateral responsible for Task 2.1. The tool was initially piloted with partners at DMU before being used with MML participants.

Data collection, analysis (see: Section 6.2.4) and tool revision occurred in parallel. With that said, a particular stance towards good practice in project evaluation was not taken in refining the interview tool to avoid limiting the interviews to a particular strategic or academic approach. In this sense the determination of good practice in the empirical study was participant-led.

As the interviews are semi-structured, the tool consisted of a list of potential interview topics and questions rather than a pre-defined list of questions to be asked in the same order. For each interview, initial questions focused on general thoughts and experiences with evaluation/reflection in the MML, including the participant's role in the project in general and evaluation in particular. Responses to these introductory questions were then used to identify topics of relevance in the interview. All topics mentioned on the schedule were not covered in each interview; rather, the participant's area of expertise and responses determined which topics were covered.

6.2.4 Data Analysis

Interview data underwent thematic analysis using the NVivo 10 software package. Key terms were identified, interpreted and combined into themes present across multiple. Words and passages were highlighted in interview transcripts which appear to indicate principles or examples of good practice in evaluation and reflection. Highlighted segments were then coded. Similar codes were assigned to themes, which were informed by but not limited to the results of the literature survey.

Thematic analysis as described here is based in Gadamerian hermeneutics²³⁹, which emphasises iteration in data collection and analysis. This paradigm is well-suited to research approach described above because it allows for data collection and analysis to occur in parallel. The interview tool was therefore able to be iteratively developed according to developing results from analysis of the literature and interviews, meaning multiple versions of the interview tool were produced as analysis progressed. An increasingly specific interview tool focusing on principles of good practice identified at earlier stages of the study was produced in this way.

²³⁹ Gadamer, *The Historicity of Understanding*; Patterson and Williams, *Collecting and Analyzing Qualitative Data: Hermeneutic Principles, Methods and Case Examples*; Gadamer, *Truth and Method*.

6.2.4.1 Method

The analysis of qualitative data typically involves reducing, categorising or otherwise organising the data in such a way that links or themes can be identified. Coding or “identifying meaning units” is perhaps the most common way to start analysis²⁴⁰, during which the researcher reads and re-reads the texts and begins to abstract and summarise the data. The process of analysis is often iterative, opaque and without a clear methodology, but this ‘inherent messiness’ is not necessarily a weakness²⁴¹. Attempts have been made to ‘clean up’ qualitative research to meet positivistic ideals of rigour²⁴², yet under the hermeneutic paradigm such attempts are misguided and harmful to the outcomes of qualitative research²⁴³.

6.2.4.1.1 The Organising System

The central structure of hermeneutic analysis is the organising system²⁴⁴, which provides a framework for the organisation, interpretation and presentation of the interviews²⁴⁵. The creation of an organising system, with categories, themes and relationships, is analogous to the process of data analysis. The final organising system is the product of data analysis, which should provide a “thick description” of the themes found in the interview data²⁴⁶. At its most basic, the organising system should structure the phenomenon under study in such a way that new insights are revealed. Analysis of the literature and empirical data will create an organising system for the identification of principles and criteria of good practice in project evaluation and reflection.

6.2.4.1.2 The Steps of Data Analysis

The data analysis method to employed here is inspired by the approach outlined by Patterson and Williams (2002, p.46). Data analysis was carried out systematically to ensure all data was given equal consideration. This is not to say all the data was weighted equally in the researcher’s final interpretation, but rather that the steps of data analysis which led to that interpretation (coding, categorising, identification of themes and relationships) was carried out systematically and rigorously, involving iteration and revisiting texts as the researcher’s understanding of the data developed over time. As such, the findings should be taken as a fair representation of the range of data collected and themes identified. Variations on themes are presented when found in the data, and outlying cases are mentioned.

Thematic analysis can be considered a form of textual analysis. Interviews were transcribed prior to analysis with the exception of one interview (C04) for which the audio quality was too poor to allow for transcription. In line with ‘good practice’ in qualitative research²⁴⁷, all interviews were recorded on an audio device and transcribed as soon as possible after they occur.

²⁴⁰ Denzin and Lincoln, “Introduction: The Discipline and Practice of Qualitative Research”; Patterson and Williams, *Collecting and Analyzing Qualitative Data: Hermeneutic Principles, Methods and Case Examples*.

²⁴¹ cf. Marshall and Rossman, *Designing Qualitative Research*, 153.

²⁴² Urquhart and Fernández, “Grounded Theory Method: The Researcher as Blank Slate and Other Myths”; Jones and Alony, “Guiding the Use of Grounded Theory in Doctoral Studies – an Example from the Australian Film Industry”; Strauss and Corbin, “Grounded Theory Methodology”; Mays and Pope, “Rigour and Qualitative Research”.

²⁴³ For example, Grounded Theory attempts to eliminate bias by standardising the steps of data analysis to allow themes to ‘emerge’ from the data and not the researcher’s interpretation.

²⁴⁴ Tesch, *Qualitative Research*.

²⁴⁵ Patterson and Williams, *Collecting and Analyzing Qualitative Data: Hermeneutic Principles, Methods and Case Examples*, 9:45.

²⁴⁶ *Ibid.*, 9:45–46.

²⁴⁷ cf. Crabtree and Miller, *Doing Qualitative Research*; Marshall and Rossman, *Designing Qualitative Research*.

Once the transcripts were prepared, each text was read once to provide an initial understanding. ‘Meaning units’, or statements that, according to the researcher’s interpretation, provide insight into the phenomenon being studied, were then identified. Sentences were treated as meaning units and subsequently coded. Sentences were chosen because they represent complete thoughts or claims, and are often grouped together to reflect more complex claims. After an initial reading, sentences of interest for further analysis were marked in the text. The meaning given to sentences was located within a holistic view of the text²⁴⁸.

6.2.4.1.2.1 Organisational Coding

The next step involved sorting or labelling the meaning units based on the researcher’s interpretation of the text, which has developed from reading the text and identifying meaning units. Labelling occurs by applying codes to sentences. While meaning units are the empirical statements or ‘raw’ data to which codes are applied, codes are the researcher’s interpretation of “what the meaning units reveal regarding the phenomenon being studied”²⁴⁹. This type of coding is a common step in many qualitative research methodologies²⁵⁰. Codes allow the researcher to group meaning units by common themes, content or meaning for further analysis.

‘Organisational Codes’ (OC) were applied to the meaning units to summarise the content of each unit in the words of the participant. OC served as labels for statements in need of further analysis, or those containing contextual information relevant to understanding the meaning of the participant’s claims²⁵¹. These codes emerge directly from the data, in the sense that the participant’s words were used to give each code a name. The purpose of these codes is to create a shorthand way of viewing the data in the words of the participant, allowing for comparison between the participant’s literal statements and the researcher’s interpretation of the meaning of the statement. These codes were not revised as analysis progressed, so as to not lose the ‘original’ meaning expressed in the participant’s words. OC ensured the participant’s voice was retained in presenting the findings of the study, while recognising that the meaning given to a statement by a participant cannot be perfectly recreated in interpreting and coding the data.

OC were divided into two types: Biographical and Evaluative. Biographical OC contained content about the background of the participant including details of the MML and their role in the project. Biographical OC are referred to in finding connections between the texts of multiple participants sharing a similar experience or characteristic. Evaluative OC will contain content related to good practice in evaluation and reflection, including relevant experiences in project activities.

6.2.4.1.3 Substantive Coding

Substantive coding involved the researcher interpreting the meaning of the sentences labelled with an Evaluative OC where the participant’s statement was contentious, confusing or otherwise open to multiple reason interpretations. Substantive Codes (SC) make claims about the data, or represent interpretations of the data, and can thus be proven wrong in a way that OC generally are not²⁵² because the latter are intended to be a shorthand summary of the participant’s or author’s words rather than a meaningful interpretation by the researcher. Substantive codes are the researcher’s

²⁴⁸ Patterson and Williams, *Collecting and Analyzing Qualitative Data: Hermeneutic Principles, Methods and Case Examples*, 9:47.

²⁴⁹ Ibid., 9:48.

²⁵⁰ cf. Strauss and Corbin, “Grounded Theory Methodology”; Patterson and Williams, *Collecting and Analyzing Qualitative Data: Hermeneutic Principles, Methods and Case Examples*; Tesch, *Qualitative Research*.

²⁵¹ Maxwell, “Designing a Qualitative Study”, 237.

²⁵² Ibid.

interpretation of the data and are iteratively revised as analysis proceeds by revisiting texts throughout the analysis process. Ideally, codes become increasingly specific as analysis progresses, showing that the researcher's understanding of the data is developing. The researcher's interpretation is based on his preconceptions²⁵³ or frame of reference consisting in part of his familiarity with each participant or text. SC go beyond the words of the participant or author and bring in themes from the researcher's conceptual understanding of principles of good practice, which develops through the literature surveys and empirical study.

SC were assigned by reviewing each meaning unit assigned an Evaluative OC and adding a SC when the meaning of the original statement was not obvious. In this way the participant's or author's words influence the researcher's interpretation of the participant's statements. This aspect of coding prevents the researcher from 'forcing' units into his framework of prejudices. Coding was iterative, meaning codes were shared across the interview texts. Codes were revised, further specified and grouped as analysis progressed (see: Figure 4.1).

The concept of meaning units as presented in Patterson and Williams (2002) is somewhat problematic, as sentences are described as "complete thoughts," implying a sentence has a single clearly defined or objective meaning. Recognising this, multiple SC will be often assigned to a single meaning unit to show different possible interpretations, reflecting the fact that meaning units are not subject to a single 'correct' interpretation.

6.2.4.1.3.1 Grouping

During and immediately following the substantive coding of a text, initial groupings of SC were created and iteratively revised, starting with headings derived from the researcher's background familiarity from the literature survey. Grouping can be understood as the researcher applying his interpretation to both Evaluative OC and SC, removing the analysis one-step from the participant's actual utterances. The purpose of grouping is to iteratively identify relationships between the codes as the researcher moves between texts to assist in further analysis and the identification of new insights in the next step of analysis (see: Figure 1). Groupings were reviewed after coding or revisiting each text. Once all transcripts and literature were thematically analysed, a final review took place during which SC were placed into final groupings that emerged from the literature and transcripts.

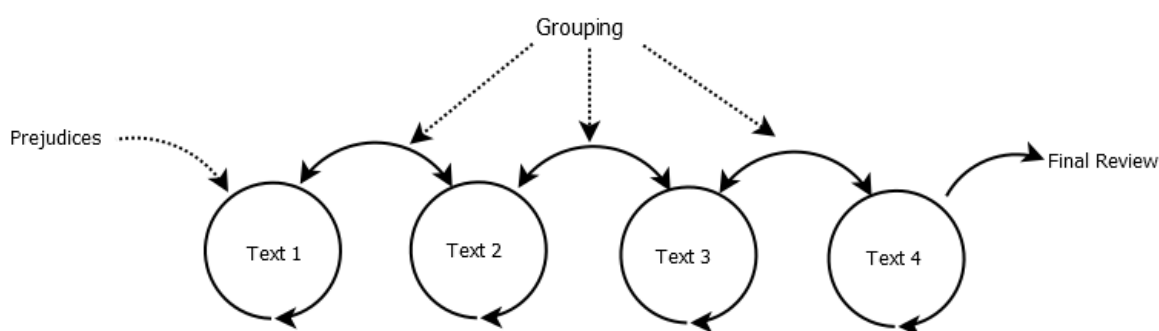


Figure 1 - Iterative Analysis of Multiple Texts

²⁵³ Ibid.

6.2.4.1.4 Discussion of Analysis Method

As hinted at in the figures above, the method of data analysis can be understood as a hermeneutic circle²⁵⁴, meaning data collection and analysis were iterative. The hermeneutic circle describes the structure of data collection and analysis, in which the researcher's preconceptions and the phenomenon studied interact²⁵⁵. The structure of analysis as presented in Figure 4.1 shows a type of hermeneutic circle, in which the researcher's understanding of good practice in MML evaluation develops iteratively through analysing and revisiting the texts. In this sense the hermeneutic circle is a name given to hermeneutic data analysis, which proceeds through a simultaneously holistic and deconstructive reading of a text. A holistic view of the text provides an initial understanding of the phenomenon to the researcher, and informs the interpretation of separate parts of the text²⁵⁶. The process becomes a circle when the interpretation of separate parts of the text leads to re-interpreting the text as a whole, and the interpretation of other texts leads to reinterpreting the original text.

6.2.4.1.4.1 Reviewing Hermeneutic Interpretation

Although interpretive research typically recognises some degree of inherent subjectivity of explanations of the world²⁵⁷, validation, or the search for common ground between these explanations that convince us of their credibility, cannot be abandoned entirely if pragmatic evaluations of the relative credibility of research are to be possible. Particular interpretations cannot be seen as absolutely or objectively true—hence the picture of understanding as a never-ending circle in which the understanding of a phenomenon improves through endless openness to new interpretations and evidence. Despite never arriving at static conclusions, hermeneutic understanding (and research) can facilitate 'cooperative life' by establishing mutual understanding.

Recognising this, some form of review was required if the analysis of the empirical data is to be seen as reliable and persuasive. Interpretations of claims can be more or less credible according to how close the researcher's interpretation comes to the 'original' meaning of the participant, supported by the text itself and the participant's other claims. The 'original' meaning was reviewed by re-reading of texts and checking the author's interpretation represented in SC and groupings against Evaluative OC which represent the participant's actual words. The review process was intended to ensure the author's interpretation of the data is credible.

Credibility as used here is not a synonym for 'true'. Rather, it means that the researcher's interpretation matches the words of the participant in some way. As an example, the statement "I am a very private person" uttered in response to a question about being 'watched' by a camera could be reasonably interpreted as meaning "I value being left alone," or "I value my independence from others," or "I don't want to be watched" if it was uttered in response to a question about surveillance. Interpreting this same statement as "I dislike the company of others" or "Privacy is my most important value" is less reasonable because the statement does not compare the relative importance of particular values or refer to social attitudes beyond the immediate topic of dialogue (surveillance). Reasonable interpretations rely upon an understanding of the text as a whole and the participant as a socially embodied person with a particular history and set of values. Supporting arguments are required when interpretations vary significantly from the actual words uttered by the participant.

²⁵⁴ Bauman, *Hermeneutics and Social Science: Approaches to Understanding*.

²⁵⁵ Patterson and Williams, *Collecting and Analyzing Qualitative Data: Hermeneutic Principles, Methods and Case Examples*, 9:27.

²⁵⁶ Ibid.

²⁵⁷ Lincoln and Guba, "Competing Paradigms in Qualitative Research".

6.2.4.1.4.2 Theoretical Outliers in Coding

Coding with two types of codes to separate the participant's voice from the researcher's voice facilitates reflexivity in the presentation of results. OC and SC can be understood as the difference between the researcher describing the participant's world as accurately as possible for the reader (imperfect as this account necessarily must be), and the subsequent interpretation of that world within a specific framework of prejudices. This is not to suggest the former is objective and the latter subjective; the researcher's prejudices necessarily influence any encounter with and interpretation of the participant's lifeworld. Instead, the difference is that SC attempt to fit the participant's statements into the researcher's framework of understanding so as to identify outliers or unfamiliar experiences and interpretations which can subsequently expand the conceptual framework beyond its current limitations.

6.2.4.2 Conclusions

The data analysis method described here facilitates reflexive assessment of interviews with MML partners to 'extract' principles and criteria of good practice stemming from their unique experiences with MML evaluation and reflection. The outcome of this method of data analysis is identification of principles and criteria that can inform DMU's approach to evaluation in SATORI WP12 (see: Section 7).

6.2.5 Results

As with the literature survey results of the interviews are presented as a thematic overview. Themes were organised according to the topic areas covered in the Interview Schedule (see: Appendix 4 – Interview Schedule). Table 5 shows the breakdown of the sample according to project role, stage and evaluation type. Representatives of MMLs in a 'Late' were more responsive to invitations to the study. The categories reflected in the table refer to the follow: for 'Stakeholder ID', 'C' refers to Coordinator, 'E' refers to Evaluator, 'P' refers to Project Partner, and 'O' refers to Project Officer (from the European Commission). For 'Project Stage', 'Late' means the project has been running for 2.5+ years (at the time of interview), 'Middle' means it has been running for 1 to 2.5 years, and 'Early' means it started less than a year ago. 'Type of Evaluation' is discussed below (see: Section 6.2.5.3).

Participant ID	Project Stage	Type of Evaluation
C01	Late	Independent Internal
C02	Middle	External
C03	Middle	Internal/External
C04	Early	Internal
C05	Early	Internal
C06	Late	Internal/External
C07	Late	Internal/External
C08	Middle	Other
C09	Late	External
E01	Early	Internal/External
E02	Late	Internal/External
E03	Late	External
E04	Late	Internal/External
E05	Late	Internal/External
E06	Late	External
E07	Early	Independent Internal
E08	Early	Internal
E09	Early	Internal
P01	Late	Internal/External
P02	Late	Internal/External
P03	Middle	Internal/External
P04	Late	External
O01	N/A	N/A

Table 5 – Interview Study Participants

6.2.5.1 Defining MMLs

One of the challenges facing researchers carrying out MMLs is in understanding the expectations and objectives attached by the EC to this relatively new project type. Participants were asked to identify defining characteristics of MMLs, the purpose of which is to match evaluation methods and criteria to the types of activities which define MMLs.

An EC project officer (O01) responsible for three MMLs was asked about the Commission’s perception of how MMLs compare with other types of EC-funded projects. According to O01, MMLs are not projects, but rather a mechanism to “setup dialogue between four types of actors: researchers, policy makers, industry and civil society organisations.” In engaging these stakeholder actors from as many EU countries as possible should be engaged, showing that MMLs have a clear concern with European representativeness.

According to O01 MMLs have a variety of potential aims, and can be considered a type of public engagement project intended to involve the public in R&D, to raise awareness about research among actors typically not involved, and to embed Responsible Research & Innovation (RRI) in these R&D and engagement processes. Awareness raising may be part of a broader programme to address a specific societal issue through stakeholder outreach and mutual learning. Whereas in projects dissemination of findings is typically an unofficial endpoint, for MMLs stakeholder engagement should occur throughout the entire project cycle and ideally build networks and channels through which communication and collaboration between stakeholders will continue after the project’s lifecycle. Rather than information being ‘disseminated’, stakeholders are rather ‘engaged’ in a two-way dialogue in which both parties learn from one another. In this way the public’s involvement

feeds into and shapes R&D. Beyond the networks and channels to be created, MMLs are also meant to provide input for research policy useable by the EC to write future work programmes and calls.

These characteristics were mirrored to some extent by other respondents. For example, E08 and E09 suggested the key actions in MMLs are participatory processes involving stakeholders and partners. Concerning stakeholders, C01 identified a variety of institutional and social stakeholders broadly fitting the categories identified by O01, including members of civil society typically underrepresented in R&D. C08 claims the main purpose of his MML is to get civil society involved in government R&D processes. C05 saw MMLs as centrally focused on engaging such stakeholders as members of the public beyond government R&D processes:

“It is very much focused on public engagement. It mobilises not only experts or policy makers, but also a lot of methods are, and approaches are, employed by the different projects to mobilise the society at large which means, you know, the ordinary citizens...the challenging thing in an MML is to bring and to make the dialogue between experts and lay people work.”

C07 described MMLs as a ‘philosophy’ rather than specific approach to research and engagement, saying, “it's about how to engage in research, and the type of activities, and how they're conducted.” In terms of actions, C07 saw nothing different between a MML and CSA. P01 claimed MMLs are a higher risk form of CSA due to the larger consortia, increased flexibility and transparency, and the call to address larger social problems.

It could be argued that MMLs are distinct from CSAs in their emphasis on mutual learning. According to C01, “it's not only mutual learning for me between, for instance, the different MMLs, but also mutual learning between experts and non experts, stake holders and citizens.” E06 agreed, saying that mutual learning means two-way learning, for example between activists and researchers who learn to use each other's language and concepts to “understand their problems” through each other's frames of reference. According to P03 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.”

C02 saw stakeholder participation as a central ‘phase’ of her project which, in her case, consisted of a series of mutual learning exercises encouraging societal debate. These events operate on certain basic ideas, including that the events “should be to collect the opinions of the public, should be bottom-up not top-down...should be debating widely within the society and getting their views and not trying to impose or to channel the discussion in any direction.” This approach suggests that the partners are not fully engaging in a discourse (and mutual learning) with stakeholders, but rather only encouraging these sorts of activities and mutual learning among stakeholder groups beyond the consortia. C02 saw her MML as “encouraging debate” between different social, governmental and institutional stakeholders, without requiring consortia partners to become involved in the content of the debate itself.

This hands off approach is not necessarily shared by other consortia, suggesting room for interpretation in translating the broad aims and characteristics specified by the EC (stakeholder engagement, mutual learning, network and channel building) into specific objectives for a particular MML. This may in part be due to the size and disciplinary variety of consortia as suggested by C05. Following from this, C06 reported significant difficulty in the first 18 months of the project to unify the aims or overall purpose perceived by consortium partners, although they eventually came to

understand that the project's primary aims were to encourage mutual learning and create connections between stakeholders that would otherwise not interact, rather than to conduct research as an end in itself. The coordinator also feared that other MMLs would face a similar difficulty due to the project type having been created by the EC only recently, meaning prior examples from which consortia may learn regarding appropriate aims and methods do not yet exist. E06 reflected these fears to some degree, noting that MMLs are "so complex in terms of the kinds of partners and the kinds of outputs." Similarly, E03 finds MMLs a confusing concept without a clear purpose. However, he suggests key characteristics include stakeholder engagement, communication, diversity, dissemination of ideas from different perspectives, network building and mutual learning, which are broadly similar to those specified by the EC.

Analysis of MML documents provides hope that C06's fears are unfounded, at least among certain MMLs. The evaluation report of Project A, reviewed above (see: Section Project A), identified two-way learning, communication and collaborative research between activists and researchers as primary aims against which the project's success will be evaluated. Awareness raising regarding the problems faced by particular stakeholders, as well as their potential in addressing similar issues on a global scale was also identified as a key objective. C06 suggests a way forward is to clearly distinguish MMLs from other types of research and engagement projects by emphasising learning as a valuable outcome in itself, as opposed to traditional 'outputs' (e.g. policy briefs, academic publications) sought in the EC's approach to impact evaluation²⁵⁸.

6.2.5.2 Evaluation Methods and Purpose

Project evaluation can be broadly described as "looking at how well [the consortium] delivered against the milestones and objectives" set out in the DoW and beyond (E02), for example against indicators of success identified by task leaders or a set of evaluative criteria defined by the evaluator. No matter the framework used a basic distinction in approaches to evaluation adopted by MMLs can be seen here, between evaluation as a 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.

The distinction may not always be recognised, but can be seen in whether a particular methods attempts a normative assessment of the quality of activities or outputs, or merely an administrative assessment to facilitate project activities. For example, the management side of evaluation requires ensuring objectives, milestones and deliverables described in the DoW are both delivered on time and of acceptable quality. The coordinator may be responsible for this type of 'management' evaluation (C02), or a partner can be classed as an internal observer (C05). The EC may also fulfil this role, for example through a project officer who performs "monitoring and tracking of milestones, and deliverables, and submission of reports, and, you know, the periodic reports... that's a project management and implementation monitoring and evaluation. So, is the project doing what it said it would do? Are the partners performing their roles as they said they would? Do the outcomes match the spend?" (C07).

6.2.5.2.1 Types of Data Considered in Evaluation

Many potential methods of data collection for evaluation are viable (see: Section 5.1.2.1). Those described by the participants include questionnaires distributed to participants in stakeholder

²⁵⁸ cf. EuropeAid, *Evaluation - Guidelines*.

engagement and mutual learning events (C02, C09, P04), observations of stakeholder engagement events by consortia partners and external evaluators (C02, C05, C09), and interviews with participants and partners to gather perspectives on project progress, quality of mutual learning events and influence of participation (P03, see: Section 6.1.1). Online materials, such as information or training materials, may also be relevant (P04).

Project deliverables are typically considered in evaluation, for example as suggested by C02 whose project uses a scientific committee as an alternative to having an evaluator. The committee's responsibility is to "analyse all the documents, all the deliverables, and have the big picture about the project." Similarly, C07 described a "steering group" made up of members of intended recipient stakeholder groups who were responsible for "fit for purpose" evaluation of project activities by reviewing deliverables for evidence of impact on intended recipients and stakeholders. In other words, the group assessed whether the outputs of the project are valuable to its intended audience.

6.2.5.2.2 Consortium Involvement in Evaluation

Transparency and engagement of the entire consortium in evaluation were identified as two important aspects of MML evaluation by C05. In his project templates were created to evaluate the quality of deliverables and stakeholder engagement events. The templates have two parts: "One is focused on the overall logistics, facilities, the whole technicalities around having an event and the second part is focused on the content of the event. So, all this will be summarized and, you know, information will be sent to the partners and this will be discussed and these summaries will be discussed during the evaluation meeting." The method aims to be as transparent as possible for all participants and stakeholders, with regular reports made by the internal observer based on assessment of partner and stakeholder feedback from the templates. The observer provides recommendations for improving how activities are implemented, or to adjust methodology where required. Also, partners review each other's deliverables, along with three external experts with expertise in engagement methods/processes, sustainable innovation and the concept of MMLs.

Involvement can be encouraged even before evaluation has officially begun. For example, E06 issued a pre-evaluation questionnaire to all consortium partners to get evaluation "on their radar." She also conducted an online survey to understand what types of impacts the project participants (WP leaders, any individual closely following the project) expect, and how she could potentially measure them, which was found to be extremely useful because of the high response rate and thoughtfulness of answers.

Regular evaluation meetings held with the consortium were seen in other approaches. In C07's project a management group consisting of the internal evaluator, coordinator and work package leaders met regularly "to review the performance and outputs of the project, and look to see whether we're on line and heading in the right direction," meaning activities were evaluated in terms of delivery against the DoW as well as quality.

6.2.5.2.2.1 Methods to Facilitate Partner Reflection

Evaluation may also involve project partners reflecting on the progress and outcomes of their involvement in the project (E02). For example, C08 described informal evaluation meetings focusing on reflection on progress between partners and discussion of issues that have arisen in carrying out the project. The partners are asked "What are the issues you are most concerned about?" which are then discussed in the meeting, and responses drafted.

Reflection can also be embedded in peer review. An approach described by C08 involves partners reviewing each other's work reflected in reports on stakeholder events and other deliverables, which is designed as an advisory role. The purpose of the scheme is to "help the project partners at moments when they are unsure about what they're doing, or they have questions." Thus far the role has been seen as threatening or critical rather than supportive, a problem which is still being addressed by the consortium.

6.2.5.2.3 Scope of Evaluation

Beyond immediate concerns with the methods of evaluation employed an initial issue to be considered is how and when to define the scope of evaluation. C01, C04 and P01 all agree that defining the scope at an early stage is the most important aspect of planning an evaluation. In doing so the consortium needs to answer questions such as: Is the evaluation only there for risk management, which controls the reports and how money is spent? Or is it just monitoring, e.g. "this is where you started and this is where you are now, a guidance of sort"? Or is it a mixture of the two (P01)?

According to C01, the scope should clearly identify whether evaluation considers only "technical aspects" of timing and meeting DoW requirements, or if a "more holistic approach" of "evaluating the progress and feedback" is taken. C01 feels that failing to clearly define the scope for all partners, including whether the evaluator is "independent, internal, or external," creates problems because evaluation as a concept can be interpreted many ways. A "narrow interpretation of evaluation" looks only at the project's "methods, how they are executed, and if they were successful in relation to their goals." However, resources, time and expertise allowing, broader evaluation can be carried out which looks the whole project, meaning "project progress, communication inside the project, the management of the project, the management properties of the work packages leaders." In a narrow scope only the participatory processes are evaluated according to their goals, which in some cases could be pre-defined indicators of success, or at least aims specified in the DoW.

The need to evaluate the quality of mutual learning is an important aspect in defining the scope of MML evaluation which accords well with conceiving of 'mutual learning' as a defining characteristic of MMLs (see: Section 6.2.5.1). C05 described her project's evaluation of the quality of outputs and events as a standard scope for evaluating engagement projects. For MMLs, she felt that the evaluation must look at how "mutual learning among the different stake holders looks is addressed and how the public, the society at large, is engaged in the MML...not only, you know, checking if all deliverables have been published in the participant portal on time, etc., etc., but really focus on the specifics of the MMLs because MMLs include a democratic approach." Following from this, E07 feels evaluation must look at the value added by the project to existing social innovation networks, or whether the project is helping to develop networks and communication channels through which social dialogue and collaboration can occur. Mutual learning is therefore both distinct from other engagement aspects of the project, but facilitated communication and collaboration channels and networks created by the project.

6.2.5.3 Evaluator Roles

In Table 5 'Type of Evaluation' refers to the position of the evaluator in relationship to the project consortium: 'Internal' means evaluation is carried out by a consortium partner, 'External' means the evaluator is not part of the consortium, 'Independent Internal' is means the evaluator is a consortium partner without any further involvement in the project beyond evaluation, 'Internal/External' means the consortium uses internal and external evaluators with different responsibilities, and 'Other'

means the project does not have an explicit evaluator. These categories cover a range of different possible arrangements. For example, in C05's project a partner acted as an 'internal observer' analysing "the process of project implementation." The observer's purpose is to ensure "the "interdependencies between the tasks and the work strategies" are followed, and to "facilitate, to support the instruction of the partners on the progress and implementation throughout the project." In C03 a mixture of peer-review of deliverables was combined with an external 'evaluation board' with expertise in relevant fields reviewing the quality of project outputs. For C07 management of the project is being evaluated both internally and by the BC, while an external 'steering committee' reviews the quality of outputs and suggests refinements to ongoing activities. For C01, the evaluator was only involved in evaluating stakeholder events, and did not evaluate the quality of deliverables which were seen as "my business or the project officer's business."

6.2.5.3.1 Internal vs. External

According to C01, a practical problem with external evaluators for EC projects exists in writing project proposals with funding established for as-of-yet unnamed evaluators. Recognising this, a common compromise is an independent internal evaluator. Truly external evaluators, "completely free of any influence or pressure from the project...are hard to find," yet the independence they offer is seen as desirable. Regardless of whether the evaluator is internal or external, a degree of independence is required to ensure partners do not unduly influence the evaluation.

External evaluators can provide an outlet for criticisms and concerns in projects, which can in turn be fed back to the consortium as recommendations for improvement. According to E02, "an independent evaluation is an outlet...discussing things that you wouldn't necessarily write on an evaluation form." E02 sees the external as very valuable because it allows for partners to reflect on "what they've been doing" in the project. She did not rule out the possibility of an internal doing the same thing, but rather said that the internal in her MML did not evaluate reflection, but rather only learning outcomes. There was some initial hesitancy from partners, however: "There's been a semi-reluctance to engage with the external evaluator because it's been a little bit like, you're going to misquote what we're saying. Don't use this in the wrong way. Please don't misinterpret what we're saying here. There has been a little bit of reluctance on that front." She wondered if some of this may have been a knock-on effect from changing external evaluators mid-project, from another evaluator to herself: "He was actively involved in the set up meetings. I think if you build that repose and build that trust with the partners. You're not there to criticize. You're there to be a sounding board for what's happening within the project. Then, I think it can be quite valuable."

While externals have certain benefits, C06 and P01 sees a risk that external evaluators will not be able to understand the relationships between consortium partners. This is reflected in practical difficulties faced by E02 with partners speaking with her about the project, which may be attributable to a lack of trust to some degree. However, she also sees a problem if the evaluator's activities not being written into the DoW, as this is needed to grant her investigations legitimacy and practically allow for collection of relevant data: "I think also, the project's been good at making sure we have been included at every point, every key milestone, so at the consortium meetings, and conferences as well. It might be nice to be a little more included within the day to day working of the work packages. We've relied very much on the interpretation of the work package leaders to refer clients out to us. Being able to evaluate internal discussions within the work package would be quite good."

According to C01, one benefit of an internal compared to external is that the former is easier to reach or communicate with because he is part of the consortium, and therefore responsible for the project.

This could, however, potentially introduce bias if the evaluator feels pressure to present the project's outcomes in a positive light, as having an internal evaluator "runs the danger or you can bring in fear that they are too closely involved" with the project and thus not "freely evaluating." For example, when evaluating or observing an event an internal can be accessed by the partner running the event, and his perception of it potentially swayed by the partner's input. Considering the benefits and issues of trust, partner influence and bias identified here, it is therefore unsurprising that many MMLs have employed both internal and external forms of evaluation.

6.2.5.3.2 Feedback and Reflection

The concerns seen over evaluator independence and bias may reflect a broader concern over the evaluator's ability to provide valuable, potentially critical feedback to the project so as to create a 'double loop' of learning along which project activities can be refined going forward. A primary purpose of having evaluation is to create this sort of feedback loop through which partners may reflect on their actions and role in the project. C01 seemed to suggest that this sort of feedback is critical to the success of evaluation and thus the project, recommending that all MMLs "give the evaluator a broader role because he can give very important input...from an outside perspective or from a 'not involved' perspective." Specifically, he "would have liked a bit more evaluation from the early beginnings to look at the whole process," referring to dozens of stakeholder consultation workshops held throughout the project.

While this shows that evaluators are involved in 'inducing' reflection among project partners, reflection itself was seen by C01 as the responsibility of the coordinator, work package leaders and project officers, rather than the evaluator. C01 saw reflection as "a learning exercise" because "not everything works as should have been," so partners are forced to adapt their concepts and methods to meet the practical limitations 'on-the-ground'. C01 believes partners should be called upon to "give recommendations together with the evaluator from early on in the project" to ensure things are "on the right track and [identify] what needs to be changed." Evaluators can assist in identifying recommendations, but cannot undertake the reflective process in proxy of the partners themselves.

This type of formative assessment, which facilitates reflection, was mentioned by several participants. E08 and E09 mentioned that the internal evaluator's role is to support and foster the project by providing useful feedback to improve the activities and identify problem areas. C02 saw formative assessment, which provides ongoing feedback to the project, as the "most useful evaluation." In her MML they conduct formative assessment: "every time a work package finishes, we evaluate and we feed the results into the next work package." Additionally, feedback was provided by informal input from colleagues of the consortium partners. C09 saw formative evaluation in her project as useful because were asked to "make suggestions for how to improve things or how to make things better, either backwards or forwards, so that's been rather useful." In C05's project the internal observer is explicitly tasked with providing recommendations for improving project activities based on partner and participant feedback taken from workshops and observation of partner work (e.g. deliverables, communication). C07 sees evaluation as involving reflection on the changing circumstances faced by a project: "Things change, so you have to monitor and evaluate how things have changed, and how that might influence what you're going to do to keep up with those changing circumstances." Evaluation should not be done only at the end of the project, but rather influence its ongoing activities:

“They were able to do something about it. It's no good doing it at the end, that's the one thing I'd say...the purpose of that is then to write up, some of it will be things you can do, some of it will be things you can't do, within the resources of the project and the time available.”

This feedback loop allows the consortium to implement changes within the scope of their DoW and resources, and to feed these changes back to the Commission to demonstrate how the information provided by evaluators has been used by the consortium. The “whole point of evaluation and monitoring” is to facilitate this sort of change. For P01, while both formative and summative assessment are required, the former is perhaps more important because it ensures mutual learning is occurring among the partners themselves.

Evaluators in the study also reflected on their actions which ‘induced’ reflection among project partners. E02 asked WP leaders to “reflect on the process of being involved in the project” and assessed their reflections via interviews to identify formative recommendations. E07 sees evaluation as primarily about providing feedback to the project consortium as a sort of “real-time evaluation.” Evaluators identify lessons to be learned from prior project activities.

In providing feedback, evaluators may need to play a critical role. Critical reports can be a helpful form of feedback because they force the consortium to learn to overcome challenges and obstacles to meeting the project’s objectives (C06 and P01). As noted by P04 reflection on progress occurred among partners whenever a project milestone was reached; however, this reflection was not in any way critical. The input of an external evaluator was therefore found extremely helpful:

“Being so close to the project, you can't really appreciate, I suppose, some aspects of it...I found it was very useful to have an external eye appraising things, and catching some of the weaknesses, or potential points of weakness, both scientifically and as far as, you know, I mean, it's quite broad, what she was looking at. Some scientific questions, management questions, communications between the project members.”

The fact that recommendations are made does not, however, mean that changes will be accepted by partners. C03 indicated that partners will sometimes resist changes suggested by evaluators. E04 had difficulty convincing partners that evaluation was more than a report to prepare for the EC, but rather a process meant to identify problems and recommendations with their activities with the intention that the project will adapt its activities to feedback throughout its lifetime. Clearly defining the scope and purpose of the evaluation, as suggested above, may help overcome resistance to some degree. Building trust and a relationship between evaluators and consortium may also help. According to E02, for critical feedback to be accepted a degree of trust between the consortium partners and evaluator is necessary, otherwise an honest and open dialogue is not possible. Suspicion of evaluators is common, suggesting evaluators can be seen as ‘quality police’. To avoid this perception, E02 suggested keeping the same evaluator for the duration of the project.

6.2.5.4 Evaluating Mutual Learning

Formative evaluation can be conceived of as facilitating mutual learning within a project consortium. Implementing suggested changes may therefore act as evidence of mutual learning. However, other forms of mutual learning between partners and stakeholders also need to be considered in evaluating the extent to which mutual learning has occurred, where the facilitating of mutual learning is understood as one of the key objectives of MMLs.

Mutual learning was mentioned as a key issue to consider in evaluating MMLs by several participants. In P03’s project, the evaluation explicitly asks whether mutual learning actually

occurred at a mutual learning event, or if it was only one-way communication. C05 conceived of evaluation as assessing “how proactive are the project partners and the project manager to ensure that there will be mutual learning.” Without two-way learning, “the risk of these kind of projects is that the civil society organizations [for example] feel like they are being just recipients of information from academics and that’s not interesting to them.” According to E06, the connections between stakeholders from a diversity of locations, for example in exchanging concepts and ideas which they otherwise would not have encountered, are a key aspect of fostering mutual learning. E06 felt 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-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.

E06 saw evaluation of learning outcomes as an important, but far more difficult aspect of MML evaluation compared to traditional bibliometric evaluation of impact. According to C07, 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.”

Mutual learning may be conceived of as a type of impact on project partners and participants. As explained by C06 and P01, mutual learning is ‘implicit learning’ in the sense that it occurs without the explicit recognition or effort of the learner. Rather, it is a change in mindset among participants which occurs when encountering an unfamiliar perspective or idea (“at interdisciplinary meeting points,” C06), for example after speaking with someone from a different socioeconomic background or listening to a story. It is therefore more than the mere exchange of ideas between two parties, and very difficult to measure in evaluation.

6.2.5.4.1 Methods of Evaluating Mutual Learning

According to C06, two-way exchange of ideas in project activities such as stakeholder engagement events may demonstrate explicit learning, but the implicit nature of mutual learning means requiring explicit evaluation of learning outcomes may be an inappropriate approach to measure the project’s impact. Mutual learning is longer lasting than mere exchange of ideas, though, implying a longer term evaluation is required for MMLs to evaluate learning outcomes (P01). According to C06 and P02, a key question to ask to partners and participants is “How are you behaving differently since being involved with the project?” Even self-reflection and self-assessments can fail to identify learning unless explicitly asked to compare behaviours or mindset from past to present. In their MML the participants did not recognise the behavioural and attitudinal change themselves, but the coordinators of the project were able to notice it. Peer review to evaluate attitudinal and behavioural change is highly recommended from the experiences of their MML. Participants and partners may be paired with each other and monitor such changes through dialogue.

More structured approaches to evaluating mutual learning may also be possible. Several approaches require taking a ‘zero-point measurement’ of participants, and to track changes in attitudes and behaviours over time via multiple applications of the same tool (e.g. C08). The approach taken by E08 and E09 involved self-reflective surveys with stakeholders at engagement events to assess current knowledge. After one year, another survey will be conducted to see if their attitudes have changed. In C07’s project learning was evaluated using Q-Sort methodology which also involves

multiple comparisons of attitudes over time (see: Section 5.1.5.2). However, compared to subjective impressions of learning among stakeholders, the results of applying Q-sort were not seen as justifying the effort required for the Q-sort—the two produced very similar results. E07 suggests that the EC’s guidelines for evaluation (embedded in the EuropeAid framework) can be used to evaluate learning outcomes, although only with input from the consortium to define ‘learning points’:

“I think it is something we need to discuss very much with our partners, because, you know, our perception of what is useful learning points could be different to the ones they've got or want to have in mind, but I think as a starting point, we would say, well, you know, take these criteria for carrying out an evaluation of a project and then sort of just see whether there are particular actions that are undertaken that maximize the project's performance in relation to these things, and that would be a very, you know, firm foundation for the sort of evaluation of the learning outcomes.”

In relation to Q-sort methodology, C08 recognised the need to evaluate his project’s influence on learning among participants. He recommended a multi-stage attitude evaluation: “You should have a sort of a zero point measurement, interviewing people, what they know, and sort of a as-is situation measurement, and you want another one of those, maybe halfway through the project and at the end, and seeing if something has changed because of the project.” However, practical limitations on measuring such change were recognised: “That would take maybe 10 or 20% of the entire budget to properly do such an evaluation, because we want the dialogue message and the importance of dialogue in people's heads, you know, when they say, okay, when we start thinking about building a windmill park, or a solar park, or carbon capture and storage, the first thing that we want to happen in their head is saying, okay, we need a dialogue with whoever we're going to do this, because now we still have the flexibility to take local concerns into account. Our proper evaluation would see in how much did we actually achieve in getting the dialogue message in people's heads, and it would be really, really difficult to actually measure that.”

6.2.5.5 Impact

The majority of respondents recognised the need to evaluate the impact of projects funded by the EC. None of the MMLs consulted have had a chance to evaluate long-term impact, as none have ended. The difficulty of measuring impact was broadly recognised; according to P04, a period of three years at least would be required before impacts can be recognised as such. E03 felt all EC projects, not only MMLs, lack the funding structure to assess impacts which occur years after the end of the project.

Despite the methodological difficulties with evaluating impact (see: Section 5.1.5.2), several attempts to begin to investigate impact were seen across the sample. For example, E06 suggested that evaluating policy impact was difficult, but her consortium showed enthusiasm towards the perceived impacts:

“How you measure the actual impact, it’s really, really hard and actually a lot of people were honest about how they think it’s too soon to know, but others are very excited about the media coverage, they’re excited about, you know, the attention their policy briefs have gotten, interviews and videos that they’ve made, you know, and so, and how they’re being used and teaching, training, etc.”

Similarly, C08 suggests policy impacts can be assessed through analysis of forthcoming EC calls: “Once the texts for 2016 and 17 are being published, and we see that we've actually achieved getting the dialogue message into the texts, that's a point for evaluation, that would be, I mean, a really great achievement for the project.”

In terms of societal impact, C07's project conducted an evaluation via interviews and questionnaires with intended recipients to check whether the project outputs had reached them and influenced their self-reported behaviour, which can be considered part of mutual learning. P03's MML currently assesses the social and ethical role of the project in societal debate and influence on stakeholders, for example whether topics put onto the agenda for mutual learning exercises legitimises those topics and leads to change in behaviour.

6.2.5.5.1 Types of Impact

Several types of impact were recognised by the respondents, including dissemination of project results (C02, E06), societal impact including behavioural and attitude change (C03, P03), and influence on policy (C03, E06). C07 thinks of societal and policy impact as “whether or not, for the people it's supposed to have influence and relevance to, it's useful or relevant to them at all.” In accord with the previous section, E06 conceived of impact as ‘societal learning’, and felt that measuring it quantitatively was inappropriate and perhaps impossible to identify causal links between project activities and impacts:

“I'm still in the process of trying to figure out how I'm going to write about, you know, these different avenues of societal learning, but, in a sense, what the aim is not to say ‘oh, there were X media hits,’ or ‘oh, this X policy changed because of Y policy brief.’ I think that's almost impossible to draw those linkages.”

Similarly, capacity and network building among non-consortium partners, conceived of as a societal impact, was identified as particularly difficult to assess without a collaboration profile from each participation group or institution (P04). E08 and E09 felt that too much focus in impact evaluation is dedicated to academic publications, particularly in projects such as MMLs where learning, capacity and network building are ends in themselves.

6.2.5.6 Criteria and Indicators of Success

Several evaluative frameworks and criteria for assessing the success of MMLs and participatory processes were employed by the respondents to create ‘indicators of success’ which show that a particular project activity is high quality. E05 employed the Rowe & Frewer framework in evaluating stakeholder engagement events. The key benefit of the framework according to him was that it allows for comparative analysis of the merit of different events, and helps identify strengths/weaknesses of different engagement methods. Support for using pre-defined process criteria as in the Rowe & Frewer framework was implicit in E03's suggestion that sponsor or DoW objectives are often purposefully vague and unhelpful in creating evaluative criteria and indicators of success for assessing the quality of the event: for example, ‘hold an event with 25 stakeholders’ says nothing of the quality of the event itself. This suggests that relying solely on the DoW or EC prescribed objectives to create criteria and indicators of success may be inappropriate.

E03 is also using the Rowe & Frewer framework to evaluate stakeholder engagement through participant questionnaires and observations: “The evaluation framework is based upon what the participants themselves think about the event and basically they simply are a couple of open questions in my questionnaire.” In terms of criteria his main concerns are with how information

enters, moves through and is shaped in the event, conceived of as an information system. Sponsor and DoW objectives are also considered, but viewed as malleable and potentially incomplete or inappropriate to the developments which have occurred in the project over its lifetime.

The EuropeAid framework was used by E07, E08 and E09. According to E07 the framework identifies key evaluation issues, or those “one would expect to [consider]...in any program or project evaluation, relevance, efficiency, effectiveness, added value, and so on.” These issues were specified by E07 to identify specific expectations or indicators of relevance, efficiency, etc. According to E07 effectiveness refers to “what extent the project achieves its objectives.” Efficiency is a “relationship between the money put into a project and the outcomes achieved, but also non-financially, whether the focus groups were well organized or not, and that sort of thing.” There is therefore a clear overlap between effectiveness and efficiency in the sense that both address the quality of participatory processes beyond financial considerations. Added value is “what is being done by the project with EU support that couldn't have been done by the member states or the partners without this.” Sustainability refers to whether “the network or the project going to last beyond the period of funding.” According to E07 these are the key evaluation criteria for any EC project, and should therefore be included in any MML evaluation.

The use of pre-defined framework was not shared across the sample. For example, in her evaluation E06 is instead relying upon “qualitative interpretation of the data to be able to...speak confidently about the basic trends.” Quantitative measures concerning the percentage of respondents' answers to survey questions form the groundwork for qualitative analysis. E03 asserts that evaluation criteria are best created by inferring success factors the responses of participants to questions about what they found ‘good or bad’ about a particular engagement event: “Because they say it was good for this reason, clearly that reason is an important reason for them, an important basis for them to evaluate the event.” The same is true for negative assessments and reasons. E03 therefore bases his evaluation upon the evaluation schemes of “the organisers or the sponsors, the inferred criteria of the participants, whether they are public or the stakeholders, and then the normative evaluation scheme based on the information translation.” This in effect gives a project-specific approach to developing evaluative criteria.

The only ‘generic’ criterion for MMLs (beyond the frameworks discussed above in Section 5.1.4.3) was suggested by C04. According to her the quality of participatory processes can be derived from the MML's (in)ability to attract stakeholders from affected communities over the life of the project. The continued cooperation of the affected groups is taken to indicate added value, and thus quality.

6.2.5.6.1 Possibility of a Generic MML Evaluation Framework

In considering the different evaluation methods and criteria adopted by MMLs, a question is raised of whether a generic evaluation framework can or should be created for MMLs. When asked, C07 viewed such a framework as presenting a “danger” of a too “rigid structure for the evaluation process that might not be suitable or appropriate for the projects.” For him, participant-designed evaluation (e.g. criteria, indicators of success) is the best approach:

“In terms of what an evaluation might look like, the need for one is very important, but the nature of that evaluation process might be different, require different things from different projects. So, effectively, that evaluation process of the impacts and influence that we're having in the policy areas or with the people, our main users who we're aiming to influence, is best designed by them, effectively.”

The need for such a framework was also questioned by E07, who did not see MML evaluation as in any way unique, but rather a continuation of evaluation of other EC projects which can use the same abstract evaluation framework prescribed by the Commission, seen above in the EuropeAid framework (see: Section 5.1.4.3.1)²⁵⁹. However, he did recognise the need to adapt the framework to the specific requirements of MML research, for example the “very, very strong emphasis...on feeding back ideas on good practices [to the EC], and what one can learn from the different countries on setting up the networks and so on.”

These criticisms of a generic MML evaluation framework emphasise the need for project-specific evaluation methods and criteria. The need for evaluation to take into account the context of a project was mentioned by several other participants. C01 described the “special environment and ecology” of projects. C05 feels indicators of learning are very difficult to pre-emptively identify, at least those which are generic to MMLs, due to the “number of participants and events.” She feels that the indicators will be very specific to the particular aims of the project, and to reach their intended audiences and have successfully established a dialogue with and between them. These factors may be reflected in “changes in the attitudes of stakeholders and citizens as well.” Similarly, C02 described the questionnaire sent by external evaluators to be filled out by participants in her project as a “lousy questionnaire” and “too rigid,” with questions that were not well-suited to the project. Her impression was that the questionnaire was a generic tool not adapted to the needs of her project.

6.2.5.6.2 Consortium Criteria

While some respondents are dubious of the possibility of a generic framework for MML evaluation, it may be possible to address these criticisms by requiring specification of evaluative principles and criteria by projects before application. One way to achieve this was undertaken in C09 and E04’s project where task leaders helped develop indicators of success used by the external evaluator. This approach was taken largely because the project’s external evaluator is not given sufficient funds to attend and evaluate all stakeholder engagement events. The approach was positively accepted by the consortium, although it created the impression that partners would have the opportunity to respond to the criticisms and recommendations of the external evaluator. Even if the partners are not required to identify task-specific criteria, E04 feels the consortium should be given the opportunity to comment on the criteria and indicators chosen by the evaluator.

C09 was, however, critical of the quality of the indicators towards the end of the project: “I think if we could do anything better then it would probably be to be more careful when finding, for example, indicators of whether things are successful or not. Some of these indicators have been- I mean when you decide early in a project and things change along the way and maybe some of the indicators that we chose were not as relevant in the end.” Also, the consortium partners were asked to create indicators for processes with which they have a proven track record, meaning this approach may not be appropriate for MMLs in which partners are addressing new methods or topics.

A similar approach was taken by E08 and E09, who decided upon indicators for each deliverable with consortium partners. The indicators described the “perfect conditions for the objective to be met,” with an objective attached to a specific deliverable. This preparation phase helped create a holistic view of the project for the evaluators, while also clarifying partner expectations of the evaluators. The indicators are based on the objectives, and the partners are directed that they will be evaluated against these indicators, which according to the pair provides the consortium with clear

²⁵⁹ It is worth noting that E07 also mentioned that a consultation process to revise the EC’s guidelines for project evaluation is currently underway with the Secretary General.

expectations of how their involvement and outputs will be assessed. In principle this approach ensures the objectives of the project are met, or at least not ignored.

A less formal approach to specifying criteria and indicators of success was taken by E06 with a pre-evaluation questionnaire:

“I asked [partners] what they expected and what they were hoping the impact would be, so that gives me something to, in some ways, evaluate what- the data I’m getting now against what they initially thought or hoped the project would achieve, and you know, I asked them what was their motivation for being involved in the project. It’s just really really useful information going into it, it helps you to know, you know, when you do tweak your questions along the way for the midterm kind of reviews, you know why people are involved and it just helps in that sense.”

This approach may represent a middle ground between pre-defined frameworks which may fail to take into account the needs of specific projects, and partner-led identification of criteria which may lead to undue influence or pressure on the evaluator. Informal data collection of partner expectations of evaluation and success in their project activities allows for criteria to be developed specific to the project on the basis of partner views, while allowing the evaluator to retain independence in the identification process.

6.2.5.6.3 Types of Criteria Necessary to Evaluate MMLs

The methods, frameworks and criteria mentioned by participants address very different types of MML activities. E03 recognised the need for multiple types of criteria in evaluating MMLs because of the multitude of activities, aims and disciplines involved. The framework used by E03 concerns the quality of public participation and stakeholder engagement events, in particular concerning the quality of information flowing through the events. However, he views this type of evaluation of participatory processes as only part of evaluating an MML: “You would need a different set of criteria to address the other four elements or the four pillars or whatever it is, and I would suggest the same would be similar for MMLs, I mean you could, you know, you could define what an MML really is, what it involves, and all its particular elements, then, you might find that some of these criteria that I use might be relevant for some aspects of it but probably not for the whole thing.” At a minimum generic project evaluation criteria appear also necessary, such as those specified in the EuropeAid framework. Criteria to assess mutual 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.

6.2.6 Discussion

The above discussion analyses participant responses in terms of their contribution to identify principles of good practice and evaluation in MMLs. Role-specific analysis of the data, wherein themes would be identified with coordinators, evaluators and project partners treated as distinct groups, was considered but ultimately rejected following initial analysis of the interview data which did not reveal significant differences across respondent groups.

Overlap of results can be seen with the discussion of responses from a similar sample of MML partners provided by Trilateral in Task 2.1 which examines participatory processes in MMLs. One area of overlap is in identifying the unique characteristics of MMLs in terms of aims, objectives (e.g. task specific goals) and methods. According to Trilateral’s findings, MMLs are defined by the following characteristics:

- Bringing together a diverse group of actors with broad and varied expertise and experience all working towards a common goal.
- Public engagement in research.
- Engaging the participation of marginalised groups in defining a research agenda, e.g., PERARES consortium worked with travelling communities in Hungary, Spain and Ireland.
- Important role of CSOs and grassroots organisations in setting the research agenda
- Transdisciplinary approaches in some cases.
- Working at different levels, i.e., the local, regional and national levels.
- Linking participatory research and its outcomes to the policy level.
- Importance of stakeholders' views and perspectives having an impact on the research and innovation agenda.
- Concept of responsible research and innovation informing the approach, i.e., fostering discussion between providers of research and technology and other parts of society before the technology is fully formed. Can dialogue influence subsequent development?

Characteristics mentioned here not found in the discussion above include the role of civil society organisations (CSOs) in setting the research agenda, working at different levels, and the influence of responsible research and innovation on MMLs. Concerning the mobilisation and mutual learning sought in MMLs, Trilateral identified the following approaches taken by current MML practitioners:

- Across many projects, learning is linked not only to stakeholders involved in the projects but also to the various consortium partners who have very different backgrounds and perspectives
- As regards the 'mutual' aspect of learning, one SiS-CATALYST respondent emphasised societal benefit, while noting that learning occurs on a one-to-one basis. There are different learning outcomes but a mutual situation.
- The R&DIALOGUE respondent felt that the notion of "mutual learning" (with an emphasis on "mutual") is very idealistic, implying a level of consensus. The notion of "mobilisation", he felt, also implies a sort of common awareness of a certain issue.
- As regards the notion of mutual learning, the respondent from SEISMIC emphasised the consortium's ambition to introduce aspects to the research agenda that have not been well covered thus far in order to start a joint debate with stakeholders.
- Mobilisation, according to the same respondent, is about raising the level of engagement, giving people a voice and motivating them to take action.
- As regards the degree to which processes can enable mutual learning and mobilisation, the R&DIALOGUE respondent emphasised the importance of following up – after the empirical engagement – on the degree of learning that occurred, in addition to the point at which learning occurred.
- The SiS-CATALYST respondent prefers to talk about "mobilising" as opposed to "mobilisation" as this implies a more active approach. For her, mobilising includes stimulating thinking about diversity and stimulating learning.
- As regards mutual learning more generally, SYNENERGENE is a good example, as noted by the project respondent. There are 27 partners involved with broad and varied expertise. In addition, seven science museums are involved as third parties. Different perspectives are coming together in order to address the topic of synthetic biology and to share knowledge.
- According to one of the GAP2 respondents, mutual learning involves a genuine exchange between stakeholders and scientists and the creation of new knowledge. Mobilisation is about getting many people involved, also from universities that did not have engagement previously. Learning derives from the fact that people work side by side with people with very different levels of experience, i.e. those who have been working in the area of public

engagement for 25 years and those who have just started to work in the area. In addition, very established institutions and newly established grassroots organisations work together, as well as partners of different ages and with different agendas [PERARES respondent].

The general picture which emerges from the discussion here and in Trilateral's interview findings is that it is essential to evaluate the extent and quality of mutual learning outcomes for MMLs. However, empirically validated methodologies or accepted 'good practice' do not appear to exist in this area. Qualitative assessments of learning via peer review appear the most widely taken approach in current MMLs and have been met with some success according to their practitioners.

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.

6.2.6.1 Representativeness of Impacts

The comments made by C07 in Section 5.1.5.1 point towards a key issue in any MML concerning the identification of appropriate and representative set of stakeholders and intended recipients of the outputs (e.g. information and mutual learning) produced by the project. At its broadest, the term 'stakeholder' refers to any individual or group with legitimate interests in an issue or decision. Numerous methods for identifying and defining stakeholders exist²⁶⁰. Regardless of how stakeholders are defined in a particular MML, the representativeness of the groups identified as 'relevant' stakeholders or the intended audience of the project must be given consideration.

In assessing the quality of impacts the stakeholders identified as 'relevant' must be considered: projects which lack a representative set of intended recipients may be criticised for excluding certain relevant stakeholders, even where clear impact can be identified (e.g. influence on forthcoming policy), because the lack of representativeness undermines the quality of the participatory processes which led to the project's outputs and impacts. Put simply, the processes leading to the impact excluded relevant interests by not engaging all individuals or groups with an interest in the case. When the quality of impact is linked to the representativeness of stakeholders in this way, all projects can be criticised to some degree because of the practical limitations on engaging *all* affected individuals in a decision-making discourse. Assessments of the quality of impacts according to representativeness are therefore best seen as criticisms of projects against an ideal rather than realistic goal.

²⁶⁰ e.g. Bryson, "What to Do When Stakeholders Matter: Stakeholder Identification and Analysis Techniques"; Mitchell, Agle, and Wood, "Toward a Theory of Stakeholder Identification and Salience: Defining the Principle of Who and What Really Counts"; Reed, "Stakeholder Participation for Environmental Management: A Literature Review".

6.2.6.2 Influence on Evaluation

As mentioned above (see: Sections 5.1.2.2 and 6.2.5.3.1) the issue of bias in evaluation may be a worry depending upon the stance taken on objectivity and consortium influence on the evaluator. It is not universally accepted that 'bias' in evaluation results should or can be avoided. The term itself appears to refer to the possibility of an 'objective' evaluation of a project, which is free of influence from the project partners themselves. Two types of influence should be acknowledged: overt influence wherein a partner directly attempts to change the evaluator's opinion, for example in recommending revisions to an evaluation report or hiding results; and subtle influence wherein the evaluator's assessment is affected by the partner, but the specific change to the assessment or point of influence cannot be described. Overt influence appears to be of greatest concern to those wishing to avoid bias in evaluation results, whereas it may be impossible to avoid subtle influence (see, for example, issues of interpretation and understanding one's frame of reference in Section 5.1.3.3). Additionally, certain influences may be desirable, for example when mutual learning occurs from interactions between the evaluator and partners, or when the evaluator's interpretation is challenged by new perspectives or evidence that results in a higher quality 'critical' evaluation. On this basis calls to avoid 'bias' in evaluation altogether should not be accepted without consideration of the epistemic stance (e.g. objectivist, constructivist, interpretivist) of the evaluator and consortium partners. These stances may well be discipline-specific, precluding specification of a general principle favouring either stance. At best what can be called for is transparent reporting of influence on the evaluator as far as can be recognised, so as to allow individual reader's to assess the quality of evaluation within their particular epistemic framework.

7 PRINCIPLES OF GOOD PRACTICE IN MML EVALUATION

The results of the empirical study can be used to support, undermine, revise or expand the set of principles of good practice in MML evaluation and reflection specified thus far in this report (see: Sections 5.3 and 6.1.1.5). Applying the results of the empirical study in this regard can be considered a preliminary form of empirical validation of the set.

Findings from the interview study can be used to support, revise, undermine or expand the principles of good practice identified thus far. The following principles were identified in reviewing academic literature addressing evaluation and reflection in public participation, participatory research and other engagement activities related to MMLs, as well as MML documents, and were explicitly supported and/or revised by the findings of the empirical study (NB: references to supporting sections have been added, revisions in **bold**):

Criteria Principles

- 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 (see: Sections 3.1.4.1 and 6.2.5.6).
- Evaluation should address the ‘generic’ qualities of participatory processes such as those areas of consensus in evaluation literature identified by Chilvers (2008). 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 & Besley (2013) and Walter et al. (2007) (**see: Sections 6.2.5.4**).
- 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 (see: Section 6.1.1.5).

Methodology Principles

- 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 impact²⁶¹, and provide stakeholders with a better idea of the value of their participation by tracking influence on the process²⁶². MML evaluation should, at a minimum, seek to meet these three generic aims (see: Section 6.2.5.3.2).
- 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 (see: Section 5.1.4.3 **and 6.2.5.2.1**).
- 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

²⁶¹ Research Councils UK, *Evaluation: Practical Guidelines - A Guide for Evaluating Public Engagement Activities*.

²⁶² Rowe and Frewer, “Evaluating Public-Participation Exercises: A Research Agenda”.

learning (e.g. changes in stakeholder perspectives, beliefs and actions) (see: Section 5.1.5.2 and 6.2.5.5).

Mutual Learning Principles

- 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** (see: Sections 5.1.2.1 and 6.2.5.4).
- Mutual learning outcomes among project participants should be assessed (see: Section 5.1.3), 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 intended recipients or stakeholder groups (NB: self-reported data)** (see: Section 6.2.5.5.1).

The following principles were not explicitly supported or undermined by the interview study:

Methodology Principles

- A clear ‘endpoint’ should be specified at which point project impacts can start to be identified and evaluated (see: Section 5.1.5).
- Evaluation should occur before, during and after the project to ensure all processes and impacts are evaluated to some degree (see: Section 5.1.2).

Mutual Learning Principles

- 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 (see: Section 5.1.2.2).
- In evaluating the quality of mutual learning that has occurred, the possibility of mutual learning without absolute consensus should be recognised (see: Section 5.1.4.3.2).
- A reflexive account of the conception of mutual learning adapted should be provided, including its theoretical basis (where appropriate) (see: Section 5.1.3).

The following principles were deleted or replaced by similar principles developed from the interview study results:

Criteria Principles

- MMLs should have clearly defined indicators of success concerning the quality of processes and outcomes prior to the start of evaluation (see: Section 5.1.4.3).

Methodology Principles

- To ensure the consortium understands the process and has an opportunity to express concerns and expectations, engage in a dialogue concerning the scope, aims, methods and (where pre-defined) indicators of success used in the evaluation before it begins.

Reflection Principles

- Reflexive questioning of project progress, indicators of success and alterations to planned activities should occur across the MML consortium, potentially through workshops or meetings scheduled at set intervals, to ensure feedback on project progress leads to corrections and improvements to project activities (see: Section 5.1.3).

Finally, the following principles have been created on the basis of the interview study:

Criteria Principles

- 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 DoW (see: Section 6.2.5.2).
- The ability of the MML to get target stakeholder groups in attendance at engagement events may be used as an evaluative measure (see: Section 6.2.5.6).

Methodology Principles

- 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. This approach will help reduce resistance to recommendations made by the evaluators (see: Sections 6.2.5.2.2, 6.2.5.2.3 and 6.2.5.3).
- 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 (see: Section 6.2.5.2.2).
- 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 (see: Sections 6.2.5.2.2 and 6.2.5.6).

Reflection Principles

- 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 (see: Sections 6.2.5.3.2 and 6.2.6.2).
- When conducting a formative evaluation, the evaluator should provide critical feedback and recommendations to the consortium to improve ongoing research activities (see: Section 6.2.5.3.2).
- 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 (see: Sections 6.2.5.3.2 and 6.2.5.4.1).

Considered together, the following is the set of principles of good practice in MML evaluation and reflection identified through the research conducted in Task 12.1:

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 (see: Sections 3.1.4.1 and 6.2.5.6).
2. Evaluation should address the ‘generic’ qualities of participatory processes such as those areas of consensus in evaluation literature identified by Chilvers (2008). 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 & Besley (2013) and Walter et al. (2007) (see: Sections 6.2.5.4).
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 (see: Section 6.1.1.5).
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 DoW (see: Section 6.2.5.2).
5. The ability of the MML to get target stakeholder groups in attendance at engagement events may be used as an evaluative measure (see: Section 6.2.5.6).

Methodology Principles

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 impact²⁶³, and provide stakeholders with a better idea of the value of their participation by tracking influence on the process²⁶⁴. MML evaluation should, at a minimum, seek to meet these three generic aims (see: Section 6.2.5.3.2).
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 (see: Section 5.1.4.3 and 6.2.5.2.1).
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

²⁶³ Research Councils UK, *Evaluation: Practical Guidelines - A Guide for Evaluating Public Engagement Activities*.

²⁶⁴ Rowe and Frewer, “Evaluating Public-Participation Exercises: A Research Agenda”.

learning (e.g. changes in stakeholder perspectives, beliefs and actions) (see: Section 5.1.5.2 and 6.2.5.5).

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, independent) as early as possible. This approach will help reduce resistance to recommendations made by the evaluators (see: Sections 6.2.5.2.2, 6.2.5.2.3 and 6.2.5.3).
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 (see: Section 6.2.5.2.2).
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 (see: Sections 6.2.5.2.2 and 6.2.5.6).
12. A clear ‘endpoint’ should be specified at which point project impacts can start to be identified and evaluated (see: Section 5.1.5).
13. Evaluation should occur before, during and after the project to ensure all processes and impacts are evaluated to some degree (see: Section 5.1.2).

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 and participants to reflect on changes to their attitudes and behaviours caused by participating in the project and engaging with unfamiliar ideas and perspectives (see: Sections 5.1.2.1 and 6.2.5.4).
15. Mutual learning outcomes among project participants should be assessed (see: Section 5.1.3), 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 (NB: self-reported data) (see: Section 6.2.5.5.1).
16. In evaluating the quality of mutual learning that has occurred, the possibility of mutual learning without absolute consensus should be recognised (see: Section 5.1.4.3.2).
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 (see: Section 5.1.2.2).
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 (see: Section 5.1.3).

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 (see: Sections 6.2.5.3.2 and 6.2.6.2).
20. When conducting a formative evaluation, the evaluator should provide critical feedback and recommendations to the consortium to improve ongoing research activities (see: Section 6.2.5.3.2).
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 (see: Sections 6.2.5.3.2 and 6.2.5.4.1).

This set of twenty principles of good practice in MML evaluation and reflection has been created through a combination of a survey of academic literature as well as a review of project documents and interviews with coordinators, evaluators and other partners from existing MMLs. The principles are intended to be broad enough to be relevant to MMLs across a variety of topic areas and disciplines, unified by MMLAP's mission to encourage capacity building, communication, collaboration and mutual learning among a variety of societal stakeholders with complementary experiences and knowledge.

8 CONCLUSION

A set of initial principles of good practice in MML evaluation and reflection have been identified through literature and empirical research intended to support the development of a strategy for evaluating SATORI building upon good practice in the field of project evaluation. Beyond this immediate concern, these principles can be combined with evaluation methods and tools amenable to evaluating the defining characteristics, aims and activities of MMLs to begin to create a generic approach or evaluative framework for MMLs. The form of this framework is as of yet unclear, but may consist of a set of tools and principles to be specified in creating an approach to evaluation responsible to the needs and aims of specific MMLs.

The principles identified here are not meant to be final or comprehensive; the diversity of disciplines, activities and topic areas under the MML mechanism precludes any sort of conclusive framework being created from a single piece of research. Rather, the principles are intended to be specified, applied, validated and revised by existing and future MMLs. Whether all or some of these principles should be adopted is as-of-yet unclear, and undoubtedly requires empirical validation of the effectiveness of the principles at identifying and qualifying the quality of MML activities. Specification of these principles requires consideration of issues such as the theoretical approach taken to learning, which will influence what can be considered (evidence of) ‘good’ mutual learning. Still, the principles identified here can be understood as a first step towards the creation of a “common diagnostic for monitoring and evaluating projects,” which was called for in a 2012 workshop with current MML practitioners²⁶⁵.

The principles were created to be broad enough to be specified into approaches to evaluation responsive to the unique characteristics of individual MMLs. Accordingly, the next step in applying 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.

8.1 NEXT STEPS

While the principles identified here still need to be specified for SATORI and adapted into a comprehensive evaluation strategy, DMU plans to begin the evaluation in the near future. In Month 7 a pre-evaluation questionnaire will be distributed to the consortium. The questionnaire will ask about each partner’s expectations of the evaluation, including key indicators of success concerning their involvement in the project. The questionnaire will be taken into consideration in Tasks 12.2 and 12.3 to ensure the evaluation strategy is, wherever possible, in sync with the consortium’s expectations. Following this, DMU will send an observer to the consortium meeting in Rome in Month 9 or 10 (confirmation of the date of the meeting is pending). This action, which may involve a brief survey being distributed to participants following the event, will mark the beginning of DMU’s evaluation of project activities including consortium meetings, workshops and other participatory events. A final list of events subject to evaluation will be distributed by DMU in the

²⁶⁵ Healy, *Mobilisation and Mutual Learning (MML) Action Plans: Future Developments*, 14.

evaluation strategy in Deliverable 12.3 in Month 18. Prior to delivery of the strategy relevant events will be identified and observed by DMU.

9 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.
- Aichholzer, Georg, and Hilmar Westholm, "Evaluating eParticipation Projects: Practical Examples and Outline of an Evaluation Framework", 2009.
http://observgo.quebec.ca/observgo/fichiers/86749_AEPP-3.pdf.
- Alkin, Marvin C., *Evaluation Roots: Tracing Theorists' Views and Influences*, Sage, 2004.
- 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.
- Bauman, Z., *Hermeneutics and Social Science: Approaches to Understanding*, Hutchinson and Son, London, 1978.
- 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.
- Blackstock, K. L., G. J. Kelly, and B. L. Horsey, "Developing and Applying a Framework to Evaluate Participatory Research for Sustainability", *Ecological Economics*, Vol. 60, No. 4, February 1, 2007, pp. 726–742.
- Bornmann, Lutz, "What Is Societal Impact of Research and How Can It Be Assessed? A Literature Survey", *Journal of the American Society for Information Science and Technology*, Vol. 64, No. 2, February 2013, pp. 217–233.
- 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.
- Brown, Steven R., "A Primer on Q Methodology", *Operant Subjectivity*, Vol. 16, No. 3/4, 1993, pp. 91–138.
- Bryson, John M., "What to Do When Stakeholders Matter: Stakeholder Identification and Analysis Techniques", *Public Management Review*, Vol. 6, No. 1, 2004, pp. 21–53.
- Bryson, John M., Michael Quinn Patton, and Ruth A. Bowman, "Working with Evaluation Stakeholders: A Rationale, Step-Wise Approach and Toolkit", *Evaluation and Program Planning*, Vol. 34, No. 1, February 2011, pp. 1–12.
- Casterlé, Bernadette Dierckx de, Sofie Verhaeghe, Marijke Kars, Annemarie Coolbrandt, Marleen Stevens, Maaike Stubbe, Nathalie Deweirtd, Jeroen Vincke, and Maria Grypdonck, "Researching Lived Experience in Health Care: Significance for Care Ethics", *Nursing Ethics*, Vol. 18, 2011, pp. 232–242.
- Chess, Caron, "Evaluating Environmental Public Participation: Methodological Questions", *Journal of Environmental Planning and Management*, Vol. 43, No. 6, November 2000, pp. 769–784.
- Chess, Caron, and Branden B. Johnson, "Organizational Learning about Public Participation: 'Tiggers' and 'Eeyores'.", *Human Ecology Review*, Vol. 13, No. 2, 2006.
<http://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=10744827&AN=23040499&h=FpQ39ePC0tDJKCUuiKwKvpS5BKS2XZkv4qUBe%2FQrkV6FgI%2BYiGH1MT1RokKVq1mO7MPH9AnpxzUQcmBsOyM62A%3D%3D&crl=c>.

- 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.
- Cousins, J. Bradley, and Elizabeth Whitmore, “Framing Participatory Evaluation”, *New Directions for Evaluation*, Vol. 1998, No. 80, 1998, pp. 5–23.
- Crabtree, B., and W. Miller, *Doing Qualitative Research*, Sage, Thousand Oaks, CA, 1999.
- Denzin, Norman K., and Yvonna S. Lincoln, “Introduction: The Discipline and Practice of Qualitative Research”, *Handbook of Qualitative Research*, Sage, Thousand Oaks, CA, 2000, pp. 1–29.
- Emery, Steven B., Henk A. J. Mulder, and Lynn J. Frewer, “Maximising the Policy Impacts of Public Engagement: A European Study”, In Press.
- EuropeAid, *Evaluation - Guidelines*, 2005.
http://ec.europa.eu/europeaid/evaluation/methodology/guidelines/gbb_det_en.htm.
- European Commission, *Mobilisation & Mutual Learning (MML) Action Plans on Societal Challenges: Call for Proposals - Science in Society Work Programme 2013*, 2013.
http://ec.europa.eu/research/science-society/document_library/pdf_06/mml-brochure-programme-2013_en.pdf.
- , “Science in Society Home Page Research - Mobilising and Mutual Learning Action Plans”, 2014. <http://ec.europa.eu/research/science-society/index.cfm?fuseaction=public.topic&id=1226>.
- Fetterman, David M., and Abraham Wandersman, *Empowerment Evaluation Principles in Practice*, Guilford Press, 2005.
- Følstad, Asbjørn, “Living Labs for Innovation and Development of Information and Communication Technology: A Literature Review”, *eJOV: The Electronic Journal for Virtual Organization & Networks*, Vol. 10, 2008.
- Foucault, M., *Discipline & Punish: The Birth of the Prison*, Knopf Doubleday Publishing Group, 2012.
- Friedman, B, P Kahn, and A Borning, “Value Sensitive Design and Information Systems”, in K Himma and H Tavani (eds.), *The Handbook of Information and Computer Ethics*, Wiley Blackwell, 2008, pp. 69–102.
- Gadamer, H.G., *The Historicity of Understanding*, Penguin Books Ltd, Harmondsworth, UK, 1976.
- , *Truth and Method*, Continuum International Publishing Group, 2004.
- Genus, Audley, and Anne-marie Coles, “On Constructive Technology Assessment and Limitations on Public Participation in Technology Assessment”, *Technology Analysis & Strategic Management*, Vol. 17, No. 4, 2005, pp. 433–443.
- Glass, James J., “Citizen Participation in Planning: The Relationship between Objectives and Techniques”, *Journal of the American Planning Association*, Vol. 45, No. 2, 1979, pp. 180–189.
- Goldschmidt, Rüdiger, and Ortwin Renn, *Meeting of Minds - European Citizens’ Deliberation on Brain Sciences: Final Report of the External Evaluation*, Inst. für Sozialwiss., Abt. für Technik- und Umweltsoziologie, Stuttgart, 2006.
- 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.
- , *The Theory of Communicative Action: Volume 2: Lifeworld and System: A Critique of Functionalist Reason*, Beacon, Boston, 1985.

- 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.
- Healy, Hali, *Mobilisation and Mutual Learning (MML) Action Plans: Future Developments, Workshop - 17-18 April 2012*, European Commission - DG REsearch and Innovation, 2012.
- Heidegger, Martin, *Being and Time*, Blackwell, 1967.
- Hennen, Leonhard, "Impacts of Participatory Technology Assessment on Its Societal Environment", *Participatory Technology Assessment: European Perspectives*, 2002, pp. 257–275.
- Hennen, Leonhard, Sergio Bellucci, Robby Berloznik, David Cope, Laura Cruz-Castro, Theodoros Karapiperis, Miltos Ladikas, et al., "Towards a Framework for Assessing the Impact of Technology Assessment", in Dr Michael Decker, Dr Miltos Ladikas, Susanne Stephan, and Friederike Wütscher (eds.), *Bridges between Science, Society and Policy*, Vol. 22 of *Wissenschaftsethik Und Technikfolgenbeurteilung*, Springer Berlin Heidelberg, 2004, pp. 57–85. http://link.springer.com/chapter/10.1007/978-3-662-06171-8_3.
- Van Hooren, R. H., H. W. Van der Borne, L. M. G. Curfs, and G. A. M. Widdershoven, "Providing Good Care in the Context of Restrictive Measures: The Case of Prevention of Obesity in Youngsters with Prader-Willi Syndrome", *Empirical Ethics in Psychiatry*, Oxford University Press, New York, 2008, pp. 153–171.
- Horlick-Jones, Tom, Gene Rowe, and John Walls, "Citizen Engagement Processes as Information Systems: The Role of Knowledge and the Concept of Translation Quality", *Public Understanding of Science*, Vol. 16, No. 3, July 1, 2007, pp. 259–278.
- Hsieh, Hsiu-Fang, and Sarah E. Shannon, "Three Approaches to Qualitative Content Analysis", *Qualitative Health Research*, Vol. 15, No. 9, November 1, 2005, pp. 1277–1288.
- Huutoniemi, Katri, "Evaluating Interdisciplinary Research", *Oxford Handbook of Interdisciplinarity*. Oxford University Press, Oxford, 2010, pp. 309–320.
- Jones, M., and Irit Alony, "Guiding the Use of Grounded Theory in Doctoral Studies – an Example from the Australian Film Industry", *Faculty of Commerce - Papers*, January 1, 2011, pp. 95–114.
- De Jong, Stefan PL, Pleun van Arensbergen, Floortje Daemen, Barend van der Meulen, and Peter van den Besselaar, "Evaluation of Research in Context: An Approach and Two Cases", *Research Evaluation*, Vol. 20, No. 1, 2011, pp. 61–72.
- 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.
- Joss, Simon, and Sergio Bellucci, *Participatory Technology Assessment: European Perspectives*, Centre for the Study of Democracy, University of Westminster, London, 2002. <http://infoscience.epfl.ch/record/45920>.
- Klein, Julie T., "Evaluation of Interdisciplinary and Transdisciplinary Research", *American Journal of Preventive Medicine*, Vol. 35, No. 2, August 2008, pp. S116–S123.
- Lincoln, Y. S., and E. G. Guba, "Competing Paradigms in Qualitative Research", *Handbook of Qualitative Research*, Sage, London, 1994, pp. 105–117.
- Loeber, Anne, Wytske Versteeg, and Erich Griessler, "Stop Looking up the Ladder: Analyzing the Impact of Participatory Technology Assessment from a Process Perspective", *Science and Public Policy*, Vol. 38, No. 8, October 1, 2011, pp. 599–608.
- Marshall, C., and G. Rossman, *Designing Qualitative Research*, Third., Sage Publications, 1999. <http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0761913408>.
- Martin, Ben R., "Assessing the Impact of Basic Research on Society and the Economy", *FWF-ESF International Conference on Science Impact: Rethinking the Impact of Basic Research on Society and Economy, Vienna, Austria*, Vol. 11, Vol. 11, 2007.

- Maxwell, Joseph A., "Designing a Qualitative Study", in L. Bickman and D. J. Rog (eds.), *Handbook of Applied Social Research Methods*, Sage Publications, Inc, Thousand Oaks, CA, US, 1998, pp. 69–100.
- Mays, N., and C. Pope, "Rigour and Qualitative Research", *BMJ: British Medical Journal (International Edition)*, Vol. 311, 1995, pp. 109–112.
- Mejlgaard, Niels, and Sally Stares, "Participation and Competence as Joint Components in a Cross-National Analysis of Scientific Citizenship", *Public Understanding of Science*, Vol. 19, No. 5, 2010, pp. 545–561.
- 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.
- Mitchell, Ronald K., Bradley R. Agle, and Donna J. Wood, "Toward a Theory of Stakeholder Identification and Salience: Defining the Principle of Who and What Really Counts", *Academy of Management Review*, Vol. 22, No. 4, 1997, pp. 853–886.
- 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.
- 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.
- Oates, B., *Researching Information Systems and Computing*, Sage, London, 2006.
- OECD, "Evaluation of Development Programmes", 1991.
<http://www.oecd.org/dac/evaluation/dacriteriaforevaluatingdevelopmentassistance.htm>.
- 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.
- Petts, Judith, and Barbara Leach, *Evaluating Methods for Public Participation: Literature Review*, Environment Agency Bristol, 2000.
- Reed, J. et al., "A Sampling Strategy for Qualitative Research", *Nurse Researcher*, Vol. 3, 1996, pp. 52–68.
- Reed, Mark S., "Stakeholder Participation for Environmental Management: A Literature Review", *Biological Conservation*, Vol. 141, No. 10, 2008, pp. 2417–2431.
- Renn, Ortwin, "Risk Communication: Towards a Rational Discourse with the Public", *Journal of Hazardous Materials*, Vol. 29, No. 3, 1992, pp. 465–519.
- , "The Challenge of Integrating Deliberation and Expertise", *Risk Analysis and Society: An Interdisciplinary Characterisation of the Field*, 2004, pp. 289–366.
- Renn, Ortwin, Thomas Webler, Horst Rakel, Peter Diemel, 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.
- 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.
- Scriven, Michael, *Evaluation Thesaurus*, Sage, 1991.
- , "Truth and Objectivity in Evaluation", *Evaluation for the 21st Century: A Handbook*. Thousand Oaks, CA: Sage, 1997, pp. 477–500.
- Senge, Peter, "The Fifth Discipline", *The Art & Practice of Learning Organization*. Doubleday Currence, New York, 1990.
- Sewell, W. R., and S. D. Phillips, "Models for Evaluation of Public Participation Programmes", *Nat. Resources J.*, Vol. 19, 1979, p. 337.
- 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.
- Spaapen, Jack, Frank Wamelink, and L.L Roberts, *The Evaluation of University Research: A Method for the Incorporation of the Societal Value of Research*, National Council for Agricultural Research, The Hague, 1999.
- 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., "Emancipation in Cross-Cultural IS Research: The Fine Line between Relativism and Dictatorship of the Intellectual", *Ethics and Information Technology*, Vol. 8, No. 3, 2006, pp. 97–108.
- 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.
- Strauss, Anselm, and Juliet Corbin, "Grounded Theory Methodology: An Overview", in N. K. Denzin Y. S. Lincoln (ed.), *Handbook of Qualitative Research*, Sage Publications, Inc, Thousand Oaks, CA, US, 1994, pp. 273–285.
- Stufflebeam, Daniel L., "The CIPP Model for Evaluation", in Daniel L. Stufflebeam, George F. Madaus, and Thomas Kellaghan (eds.), *Evaluation Models*, Vol. 49 of *Evaluation in Education and Human Services*, Springer Netherlands, 2000, pp. 279–317. http://link.springer.com/chapter/10.1007/0-306-47559-6_16.
- Tesch, Renata, *Qualitative Research: Analysis Types and Software Tools*, Routledge, 1990.
- Tuckett, Anthony, "Qualitative Research Sampling: The Very Real Complexities", *Nurse Researcher*, Vol. 12, 2004, pp. 47–61.
- 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.
- Urquhart, Cathy, and Walter Fernández, "Grounded Theory Method: The Researcher as Blank Slate and Other Myths", 2006. <http://hdl.handle.net/2173/92896>.

- Walls, John, Gene Rowe, and Lynn Frewer, "Stakeholder Engagement in Food Risk Management Evaluation of an Iterated Workshop Approach", *Public Understanding of Science*, Vol. 20, No. 2, March 1, 2011, pp. 241–260.
- 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.

10 ANNEXES

APPENDIX 1 – MML DETAILS

Acronym	Title	Project Website	Project Coordinator
PERARES	Public Engagement with Research And Research Engagement with Society	http://www.livingknowledge.org/livingknowledge/perares	Dr. Henk Mulder, Science Shop, University of Groningen, The Netherlands - T. +31 (0)50.363 4436, h.a.j.mulder@rug.nl
INPROFOOD	Towards inclusive research programming for sustainable food innovations	http://www.inprofood.eu/	Klaus Hadwiger - klaus.hadwiger@uni-hohenheim.de - Contact for coordinating institution
SIS-CATALYST	SiS Catalyst: Children as Change Agents for the future of Science in Society	http://www.siscatalyst.eu/	Andrew Abrahamson (Project Coordinator for Univ. of Liverpool, PC) - a.abrahamson@liv.ac.uk
EJOLT	Environmental Justice Organisations, Liabilities and Trade	http://www.ejolt.org/	Joan Martinez-Alier - http://icta.uab.cat/icta/curriculum.jsp?id=15&nombre=Joan,Mart%EDnez%20Alier
GAP2	Bridging the gap between science, stakeholders and policy makers Phase 2: Integration of evidence-based knowledge and its application to science and management of fisheries and the marine environment	http://www.gap2.eu	Steven Mackinson, steve.mackinson@cefas.co.uk

PACITA	Parliaments and Civil Society in Technology Assessment	http://www.pacitaproject.eu/	Lars Klüver - lk@tekno.dk - Not clear if project coordinator, just contact for coordinating institution
MARLISCO	Marine Litter in European Seas - Social Awareness and Co-Responsibility	http://www.marlisco.eu	Doriana Calilli - d.calilli@provincia.teramo.it - 0039 (0)861 331407 - mobile 0039 366 5670917
R&DIALOGUE	Research and Civil Society Dialogue towards a low-carbon society	http://www.rndialogue.eu	Robert Vanderlande - robert.vanderlande@triarii.nl - +31 (0) 6 426 25 727
SFS	Sea For Society	http://seaforsociety.eu/	Nausicaa - http://www.nausicaa.co.uk/ - Contact person: Manuel Cira - manuel.cira@nausicaa.fr
NERRI	Neuro-Enhancement: Responsible Research and Innovation	http://www.nerri.eu/eng/home.aspx	Ciencia Viva (http://www.cienciaviva.pt/home/)
SiFORAGE	Social Innovation on active and healthy ageing for sustainable economic growth	http://www.siforage.eu/	GISME-UB (http://www.gisme.eu/)
BEWATER	Making society an active participant in water adaptation to global change	http://www.cyi.ac.cy/waterresearchandmanagement-ongoing/item/866-bewater-making-society-an-active-participant-in-water-adaptation-to-global-change.html (Project Site Under Construction)	CREAF - Centre for Ecological Research and Forestry Applications http://www.creaf.uab.es/eng/contact/index.htm
SYNERGENE	Engaging with New and Emerging Science and Technology in Responsible Governance of the Science and Society Relationship	http://www.synenergene.eu/	Christopher Coenen - contact@synenergene.eu

CASI	Public Participation in Developing a Common Framework for Assessment and Management of Sustainable Innovation	http://cordis.europa.eu/projects/rcn/111387_en.html	Applied Research and Communications Fund (http://cordis.europa.eu/projects/rcn/111387_en.html)
SATORI	Stakeholders Acting Together On the ethical impact assessment of Research and Innovation	http://cordis.europa.eu/projects/rcn/111019_en.html	N/A
ASSET	Action Plan on SIS Related Issues in Epidemics and Total Pandemics	http://cordis.europa.eu/projects/rcn/111213_en.html	VITAMIB SAS (http://www.vitamib.com/)
MAPPING	Managing Alternatives for Privacy, Property and INternet Governance	http://cordis.europa.eu/projects/rcn/111214_en.html	University of Groningen (http://www.rug.nl/)
SEiSMiC	Societal Engagement in Science, Mutual learning in Cities	http://jpi-urbaneurope.eu/about/why/seismic/ (Placeholder)	Austrian Institute for Technologies (AIT) (http://www.ait.ac.at/) and Platform 31 (http://www.platform31.nl/)



DE MONTFORT UNIVERSITY

APPLICATION FORM TO GAIN APPROVAL FOR ACTIVITIES INVOLVING HUMAN RESEARCH

Notice to Staff and Students

If your research involves using human tissue or fluid samples please **DO NOT** use this application form. Please approach the Research and Commercial Office, Faculty of Health and Life Sciences, 2.25L Hawthorn Building, Phone: 7891 / 7777 for the correct application form if you are using human tissue.

The University requires that approval is obtained by members of staff of the University and by students of the University who wish to engage in the type of research detailed below. Please use this form for an application if your research involves:

1. Gathering information from or/and about individual human beings (and organisations) through:
 - interviewing
 - surveying
 - questionnaires
 - observation of human behaviour
 - modifying/disturbing human behaviour
 - interfering in normal physiological and/or psychological processes
2. Using archived data in which individuals are identifiable.
3. Researching into activities which involves direct observation of or contact with those who are or who might reasonably be supposed to be engaged in or have engaged in criminal activities or activities which are related to criminal activity
4. Research which involves a risk of physical or psychological injury to the researcher or any other person involved in the research
5. Supporting innovation that might impact on human behaviour e.g. Behavioural Studies

[It should be noted that in regard to research into illegal activities there are no exclusions or blanket permissions and the University Insurance cover may not apply if the research activity has not been cleared by the University or, in certain cases with delegated authority, the appropriate Faculty Committee.]

Guidance and support will be given by your supervisor (for student research), your line manager or an appropriate designated officer/ Faculty Research Office. Queries arising out of this should be directed to:

FAILURE TO GAIN FREC APPROVAL FOR YOUR RESEARCH MEANS THAT YOUR PROJECT MAY BE FAILED OR THAT YOU ARE SUBJECT TO DISCIPLINARY ACTION.

DE MONTFORT UNIVERSITY**APPLICATION FORM FOR RESEARCH ACTIVITY REQUIRING HUMAN RESEARCH ETHICS
CONSIDERATION OR APPROVAL****Staff/Student Name****Programme (if relevant)**

Brent Mittelstadt

Title of Research Project

SATORI Task 12.1 - Best Practice in MML Evaluation

Brief description of proposed activity and its objectives:

The research study aims to interview project co-ordinators, evaluators and stakeholders of Mobilisation and Mutual Learning (MML) projects funded by the EC. Interviews will focus on evaluation activities within these projects. Twenty participants are expected in total. The project is part of Task 12.1 for the SATORI project led by the Centre for Computing and Social Responsibility.

Ethical issues identified:

Human participants are required to participate in the interview. Informed consent is required.

Participant identities should not be disclosed to protect anonymity.

Interview audio will be recorded, which may not be desired by participants.

How these will be addressed:

Participants will have the right to withdraw from the research at any time, and will be informed of this right. Informed consent must be obtained verbally and in writing at the beginning of the interview after reviewing the supporting documentation (consent and information sheets) with the participants and answering any questions. Participants will be informed of the purposes of the study and their role in it prior to giving consent. Documentation will be provided at least 24 hours prior to the interview.

Personal details which could lead to identification of the participant including contact details and role/activities within the MML will not be published or made otherwise publicly available by the research team.

Participants will be informed of recording at the beginning of the interview, and will have the option to request that the interview not be recorded, or the recording be destroyed.

To which ethical codes of conduct have you referred? These are specific to each Faculty and if you have a query please ask your supervisor or Faculty REC for advice.

BCS Code of Conduct
ESRC Research Ethics Framework
SRA Ethical Guidelines
NRES Research Ethics Guidance
GMC Consent Guidance

How have concerns regarding the safety of the researcher and/or the research subject been addressed if applicable?

No safety concerns have been identified.

Checklist for applicant:

Has the research proposal identified any of the following research procedures?

1. Gathering information from or/and about human beings through: Interviewing, Surveying, Questionnaires, Observation of human behaviour
2. Using archived data in which individuals are identifiable
3. Researching into illegal activities, activities at the margins of the law or activities that have a risk of personal injury
4. Supporting innovation that might impact on human behaviour e.g. Behavioural Studies

The following should be considered. Please tick yes as relevant:

- Providing participants with full details of the objectives of the research
- Providing information appropriate for those whose first language is not English
- Voluntary participation with informed consent
- Written description of involvement
- Freedom to withdraw
- Keeping appropriate records
- Signed acknowledgement and understanding by participants
- Relevant codes of conduct/guidelines

Are there other/additional factors that could/will give rise to ethical concerns? E.g. Communication difficulties

List of accompanying documentation to support the application:

- (1) A copy of the research proposal
- (2) The details of arrangements for participation of human subjects (including recruitment, consent and confidentiality procedures and documentation as appropriate)

- (3) A copy of all the documentation provided to the volunteer to ensure the clarity of information provided
- (4) Copies of appropriate other ethical committee permissions (internal or external) or supporting documentation
- (5) If appropriate: a list of proprietary drugs or commercial drugs to be used in the proposed investigation including formulation, dosage and route of administration and known adverse side effects
- (6) A statement of your competence to carry out this research as a student or a brief one page curriculum vitae for each applicant, including recent publications (staff only)
- (7) Other documentation as advised necessary:

There are normally four possible outcomes from reviewing the activity against the procedures in place:

1. no ethical issues
2. minor ethical issues which have been addressed and concerns resolved
3. major ethical issues which have been addressed and concerns resolved
4. ethical issues that have not been resolved/addressed

Provisional approval could be given at the discretion of the Research Ethics Committee.

Authorisation is dependent on Faculty. Please refer to your faculty guidelines for details on how outcomes are reached:

- The reviewer advises the PMB/SAB/REC of those activities in the first three outcomes.
- Activities in the fourth outcome are submitted to the Faculty REC for resolution
- The approved form must be kept with project documents, e.g. be included as an appendix in the report.

Signature of researcher / student Date

Signature of supervisor Date

Line Manager or Head of School signature (Staff only) Date

This form complies with the DMU policy statement on Human Research Ethics, a full copy of which can be found in the General Regulations and Procedures Affecting Students.

A separate form is required for each project.

ADVANCE APPROVAL OF RESEARCH ACTIVITY INVOLVING HUMAN RESEARCH ETHICS

- 1 Respondents' co-operation in a research project is entirely voluntary at all stages. They must not be misled when being asked for co-operation.
- 2 Respondents' anonymity must be strictly preserved. If the Respondent on request from the Researcher has given permission for data to be passed on in a form which allows that Respondent to be identified personally:

- (a) the Respondent must first have been told to whom the information would be supplied and the purpose for which it will be used, and also
 - (b) the Researcher must ensure that the information will not be used for any non-research purpose and that the recipient of the information has agreed to conform to the requirements of any relevant Code of Practice.
- 3 The Researcher must take all reasonable precautions to ensure that Respondents are in no way directly harmed or adversely affected as a result of their participation in a research project.
 - 4 The Researcher must take special care when interviewing children and young people. The Faculty REC will give advice on gaining consent for studies involving children or young people.
 - 5 Respondents must be told (normally at the beginning of the interview) if observation techniques or recording equipment are used, except where these are used in a public place. If a respondent so wishes, the record or relevant section of it must be destroyed or deleted. Respondents' anonymity must not be infringed by the use of such methods.
 - 6 Respondents must be enabled to check without difficulty the identity and bona fides of the Researcher.

SATORI Task 12.1 - Best Practice in MML Evaluation and Reflection

Dear Participant

We would like to ask you to participate in a study conducted by De Montfort University looking at best practice in evaluation and reflection activities in Mobilisation and Mutual Learning (MML) projects. You have been asked to participate on the basis that you are a project partner or stakeholder in a FP7 MML project.

The study is being undertaken as a first step in the SATORI (Stakeholders Acting Together On the ethical impact assessment of Research and Innovation) FP7 MML project to identify and justify an approach to evaluation and reflection based on expertise from partners and stakeholders in prior MMLs.

We hope better to understand your thoughts on best practice in the evaluation of MMLs through questions such as:

- How was your project evaluated?
- How did reflection on project progress occur?
- What was the value of evaluation and reflection to the project?
- If starting over, how would you improve evaluation and reflection in your MML?

Participation in this study is entirely voluntary. It will involve one interview of 15-30 minutes to take place on the telephone or via Skype. Before you consent to participate details of the objectives and your role in the study will be reviewed with you, and you will have the opportunity to ask any questions you may have.

You may decline to answer any of the interview questions. You may also decide to withdraw from this study at any time by advising the interviewer or by contacting the lead researcher at bmittelstadt@dmu.ac.uk or 0116 255 1551. If you notify us of your withdrawal, all identifiable data including interview recordings will be destroyed.

The information you provide is confidential, except that with your permission anonymised quotes may be used in publications in international conferences and academic journals. Your name, job title, project role and title, as well as any other personal identifying information will not appear in any publications resulting from this study and will only be available to the research team. The study may result in an academic publication on best practice in MML evaluation and reflection.

The information gained from the interview will only be used for the above objectives and will not be used for any other purpose or recorded in excess of what is required for the study.

There are no known or anticipated risks to you as a participant in this study. By participating you may benefit from reflecting on your experiences with evaluation and reflection, and by contributing to the identification of best practice for future MMLs.

The research team requests that you do not reveal information about illegal activities during the study, as such disclosures must be reported to the appropriate authorities.

The study has been approved by De Montfort University's Human Research Ethics Committee. If you have any questions or would like additional information please ask the lead researcher or interviewer at any time.

Yours faithfully,

A handwritten signature in black ink, appearing to read 'Brent Mittelstadt', with a stylized flourish at the end.

Brent Mittelstadt
Chief Investigator
De Montfort University
Centre for Computing and Social Responsibility
The Gateway, Leicester LE1 9BH, UK
Tel.: + 44 750 712 1631
E-mail: bmittelstadt@dmu.ac.uk

SATORI Task 12.1 - Best Practice in MML Evaluation and Reflection

Participant Consent Form

Issue	Respondent's initial
I have read the information presented in the information sheet (Version 1.0 – 14.02.2014 – Task 12.1 Information Sheet) about the study "SATORI Task 12.1 - Best Practice in MML Evaluation and Reflection." In addition, I am aware that the interview and all other related work of the project is and will be in English.	
I have had the opportunity to ask any questions related to this study, and received satisfactory answers to my questions, and any additional details I wanted.	
I am aware that anonymised excerpts and quotations from the interview may be included in publications to come from this research and used in future interviews.	
I give permission for the interview to be recorded using audio ¹ recording equipment.	
I understand that relevant sections of the data collected during the study may be looked at by individuals from De Montfort University and their collaborating partners. Any recordings made, or transcripts from recordings will be confidential to the research participants and their collaborating partners. Neither De Montfort University nor their collaborating partners will use the transcripts for any other purpose than the study describes. With this, I give permission for these individuals to have access to data from this study.	
I understand that I may withdraw my consent at any time by advising the researcher.	

With full knowledge of all foregoing, I agree to participate in this study.

I agree to being contacted in the future by the researchers if my responses give rise to interesting findings or require clarification:

No Yes

¹ Participants have the right to request the audio to be switched off at any point during the interview. They also have the right to withdraw at anytime during the interview process.

I wish to be notified when results of the study are made publicly available:

No Yes

If yes to either of the above, my preferred method of being contacted is:

Telephone Email

Participant Name:		Consent taken by	
Participant Signature:		Signature	
Date		Date	

APPENDIX 3 – MML CRITERIA

Acronym	Start Date	End Date	Length	Stage	Budget (EUR)	Consortium Size	Topic
PERARES	01/05/2010	31/10/2014	4.5 years	Late	3085511	28	Public engagement in research
SIS-CATALYST	01/01/2011	31/12/2014	4 years	Late	4561513	19	Engagement between children and social, cultural, political, scientific and educational institutions
EJOLT	15/03/2011	14/03/2015	4 years	Late	4078038	23	Environmental justice, trade
GAP2	01/04/2011	31/03/2015	4 years	Late	7555445	38	Fisheries management
PACITA	01/04/2011	31/03/2015	4 years	Late	5431938	15	Public engagement in Technology Assessment
INPROFOOD	01/11/2011	31/10/2014	3 years	Late	4553171	18	Research inclusion and sustainable food
MARLISCO	01/06/2012	31/05/2015	3 years	Middle	4544746	20	Marine litter
R&DIALOGUE	01/06/2012	30/11/2015	3.5 years	Middle	4482268	15	Low-carbon innovations
SFS	01/06/2012	30/11/2015	3.5 years	Middle	4893284	20	Societal issues involving the ocean
SiFORAGE	01/11/2012	31/10/2016	4 years	Middle	4098762	20	Ageing, sustainable economics
NERRI	01/03/2013	29/02/2016	3 years	Middle	3783867	18	Neurological enhancements
BEWATER	01/10/2013	31/03/2017	3.5 years	Middle	3588713	12	Public engagement in water scarcity

SYNERGENE	01/07/2013	30/06/2017	4 years	Early	4590081	27	Responsible Research & Innovation in synthetic biology
CASI	01/01/2014	30/06/2017	3.5 years	Early	4473404	19	Assessing and managing sustainable innovations
SATORI	01/01/2014	30/09/2017	3 years 10 months	Early	4723129	16	Ethical impact assessment in R&I
ASSET	01/01/2014	31/12/2017	4 years	Early	4496454	15	Epidemics, pandemics, public health
MAPPING	01/03/2014	28/02/2018	4 years	Early	4642522	13	Economic, social, legal, ethical aspects of developments in internet
SEiSMiC	N/A	N/A	N/A	Early	N/A	N/A	City development

APPENDIX 4 – INTERVIEW SCHEDULE

Notes about the Schedule: The following is a sample list of topics and questions that may be covered in interviews with participants. Topics will be chosen according to the participant's background and responses during the interview. The interviews are deliberately open and semi-structured to allow participants to draw on their experiences in MML evaluation and reflection as far as possible, identifying unforeseen areas and principles of best practice. The interviewer will probe answers and ask for/offer interpretations of responses, but the topics covered will mostly be determined by the participant.

Introductory Statement: SATORI is a recently launched MML project which aims to develop a common European framework for ethical assessment of research and innovation (R&I) (see <http://satoriproject.eu/>). The SATORI research consortium will develop an ethics assessment framework based on thorough analysis, participatory processes and engagement with stakeholders, including the public, in Europe and beyond. As MMLs are a new type of EC-funded project, one task of SATORI is to identify good practice in evaluation of and reflect on MMLs. In this interview we are trying to understand your experiences with evaluation and reflection in the MML project(s) with which you are currently involved. We're especially interested in your thoughts on best practice, including experiences with methods that you feel have or have not worked well.

Background

- Stakeholders
 - Types
 - Role in project
- Respondent
 - What was your role in the project? (If an evaluator, were you internal or external to the consortium?)
 - What's your background in terms of project evaluation?

Evaluation

- How was your project evaluated? If known:
 - Method of evaluation
 - Tools / Techniques used
 - Timing, integration into project plan
 - Procedural and/or substantive, formative and/or summative, process and/or impact?
 - Outputs
- What was evaluated?
 - Participatory processes
 - Impact
 - Types?
 - Partner progress/outputs against DoW
 - Quality of content produced by the project against project aims

- How are MMLs unique compared to other types of EC-funded projects?
 - Do these characteristics require a certain method or type of evaluation?
- How did the evaluation feed back into the project's activities?
 - Outputs
- What challenges were faced in evaluating the project?
- What was the value of evaluation to the project?
- What worked? What didn't?

Reflection

- How did reflection on project progress occur?
- Were partners encouraged to critically reflect on their progress?
- What was the value of reflection to the project?

Reflecting on Evaluation

- What worked?
- What didn't?
- Reflection vs. Evaluation
 - How do you differentiate between evaluation and reflection?

Recommendations

- Recommendations concerning evaluation for other projects
 - If starting over, how would you improve evaluation and reflection in your MML?
 - How would you recommend other MMLs evaluate themselves?
- Recommendations for policymakers
 - Should a requirement for evaluation/reflection be attached to MML funding?
 - Describe your experience of choosing a method of evaluation for the project
 - Can policymakers do anything to encourage evaluation/reflection in MMLs?
- Ethical issues
 - Did you encounter any ethical issues?
 - How did you deal with them?
 - Was there any ethical evaluation/reflection in the project?
- Anything else we forgot to ask?

APPENDIX 5 – DATABASE SEARCH RESULTS

Search String	Database	Returned	Selected
“science in society” AND evaluation	Scopus	6	3
	Web of Science	5	4
	Google Scholar	6860 (100 checked)	1
"framework program*" AND evaluation	Scopus	243	6
	Web of Science	145	6
	Google Scholar	28800 (100 checked)	2
“stakeholder engagement” AND evaluation	Scopus	125	0
	Web of Science	67	1
	Google Scholar	20500 (100 checked)	6
Journal: “Research Evaluation” Keywords: engagement OR stakeholder OR social OR learning OR society	Scopus	147	3
	Web of Science	N/A	N/A
	Google Scholar	N/A	N/A
Journal: “Evaluation and Program Planning” Keywords: “stakeholder engagement”	Scopus	15	3
	Web of Science	N/A	N/A
	Google Scholar	N/A	N/A
Journal: “Evaluation and Program Planning” Keywords: “stakeholder participation”	Scopus	77	2
	Web of Science	N/A	N/A
	Google Scholar	N/A	N/A
“public participation” AND evaluation	Scopus	759	10
	Web of Science	348	12
	Google Scholar	111000 (100 checked)	17