Review of a BIS/Sciencewise-ERC Project

Science, Trust and Public Engagement: Exploring Future Pathways to Good Governance (Phase 1)

Dr Richard Watermeyer

Dr Gene Rowe

Dr Jessica Paddock

Professor Anne Murcott

Professor Tom Horlick-Jones

Cardiff University School of Social Sciences

November 2012

Contents

Executive Summary

This document presents our review of the *Science, Trust and Public Engagement: Exploring Future Pathways to Good Governance* (henceforth STAPE) project. The work was funded by the Sciencewise Expert Resource Centre (Sciencewise-ERC)[1]. The report begins by describing the STAPE project objectives and our involvement as evaluators, noting how the project changed from the original brief. We then describe our approach to the evaluation, and how we adapted our approach from a conventional evaluation to processes closer to peer review to address the revised project brief.

Our approach was based upon the meta-criterion of *translation quality*, which is concerned with the efficiency of information/knowledge gathering, recording, transmission and interpretation between the various stages of the activity, and involving various parties (including the sponsors and stakeholders/participants), and its comprehensiveness and appropriateness. Use of the translation criterion by necessity requires the consideration of the sponsor objectives (in this case, both project objectives and Sciencewise-ERC principles of good practice in public dialogue), as these specify the initial information/knowledge targets for the project to achieve.

We used various data-gathering methods (document analysis, event observation, questionnaires) to capture the translation efficiency of each stage of the project, as part of our overall (summative) analysis. These project stages comprised: a) a literature review, b) interviews, c) in depth reviews of two organisations, d) a stakeholder workshop, e) a final report. We also provided observations during the course of the project at various stages (formative assessment), with the aim of informing possible 'mid-course corrections' in the implementation of the project. To that end, and to ease communications generally, a member of our team was appointed to the Steering Group for the project.

Regarding the literature review, we took a peer review approach. We recognise that the review was not expected to be comprehensive. However, we suggest that, although the review did generally provide answers to most of the questions posed, there were some important limitations. For example, our view was that there could have been a broader consideration of questions concerning the organisational and policy-making context in which engagement policy is formulated and enacted. Our analysis of the evaluation reports relating to the project reports on which the review was based also suggested that there were a number of relevant engagement problems identified in past evaluations that were not captured by the review. We conclude that the review was a valuable contribution to the project, although potentially with some reduced sensitivity to some practical features of the implementation of engagement practice in real-world settings.

Regarding the interviews, these had begun before the literature review was finalised, although the interview schedules were developed to take account of emerging findings from the review. We did not have access to the interviewees, which we accepted, recognising difficulties regarding confidentiality. Our analysis has therefore been directed at the extent to

[1] The Sciencewise Expert Resource Centre (ERC), funded by the Department for Business, Innovation and Skills (BIS), is the UK's national centre for public dialogue in policy making involving science and technology issues. www.sciencewise-erc.org.uk

which appropriate and comprehensive information was recorded by these processes (what questions were asked; how data was collected; how it was recorded; how it was summarised; whether it appeared to speak to the project's central questions or not). As Steering Group members, we reviewed drafts of the interview questions and lists of potential interviewees, and received updates on the progress of the interview process. Information about the resulting data, and the means by which it was analysed, was contained in the final project report.

It is important to note that the project team was successful in obtaining significant access to many high level people in a large number of relevant science-related organisations.

In terms of the reviews of organisations (initially described as 'ethnography'), we did not have access as observers and had little first hand evidence of what was done. The final project report describes this activity as two organisation visits involving four or five semi-structured interviews each (at which field notes were taken). The report also explains that these reviews were not reported or analysed separately from the interviews, so it was not possible to track the data obtained or the role it played in the project conclusions. We cannot therefore comment on the translation quality associated with this element.

Regarding the workshop we first note that new project objectives agreed by the project sponsor replaced a series of opportunities for stakeholder dialogue with a much stronger focus on research, primarily though interviews and with a single workshop. Our analysis was hampered by only being allowed a single observer at the workshop, and by being prohibited from distributing a participant questionnaire immediately after the event (to avoid adversely impacting the event). We were allowed to distribute a questionnaire by email subsequently, although this had the impact of drastically reducing response rate, and hence of useful material.

In the workshop itself, there were a number of translational inefficiencies. For example, the participants' tasks were not clearly outlined; there was no invitation for participants to introduce themselves; there was little use of material for aiding the various discussions (e.g. questions written on whiteboards or handouts to help focus discussion); there was no recording of conversations (and so information collection was reliant upon note-taking by the project team, and the efforts of participant-rapporteurs); there was insufficient time for in depth discussions; and in some instances project team facilitators dominated proceedings during small group discussions.

As an opportunity for key representatives of different organisations to share experiences and views, the workshop was a success. However, we do not have sufficient evidence to determine the wider contribution of the workshop to capturing organisational perspectives and responses to the project team's findings ('validation'), or as a method of collective analysis.

The first draft of the project report was circulated in March 2011, and the final report was published in November 2011, following extensive revisions. The final project report provides some methodological details, as well as a comprehensive discussion of the results from the various post-literature review activities. However, given our lack of access to interview transcripts and data analysis, and no clear distinction between results coming from the different data-gathering processes (such as the organisation reviews and the workshop), we are not able to comment upon the validity of corresponding results. One possible interpretation is that similar messages were emerging from the different processes, and therefore no significant distinction emerged between datasets. If this was indeed the case then the outcome could be read as a positive one, with the different data-gathering activities essentially validating each other. Overall, we note that the results are interesting, and not unsurprising, and have a good degree of face validity.

Turning to the overall status of our own report, it is important to note that the findings are limited in various ways, due in large part to a number of barriers to our collection of evidence. Some were understandable, though others, we suggest, were not. We appreciate that the project involved discussions with senior people engaged in high-level work, bringing issues of confidentiality to the fore. However, transparency is an important concept for Sciencewise-ERC, and mechanisms such as Chatham House rules could have been invoked to assuage concerns in certain quarters. Efforts to interview the project team were not entirely successful - which was unfortunate, as they were the only possible sources of information concerning some project stages.

We were originally commissioned to provide an evaluation of the STAPE project at the same time as the project itself was commissioned, to meet the Sciencewise-ERC principle of starting evaluation early. However, as noted above, the project itself quickly changed significantly and the final project brief and objectives, as agreed with the project sponsor and Steering Group, included less stakeholder dialogue and more research and review. As a result, the requirement for a formal project evaluation changed, and the potential for conventional evaluation research diminished. We therefore developed an approach that reflected these changes, focusing more on peer review and methodological analysis, still using the meta-criterion of translation quality but taking account of the limited access to project activities that we were able to obtain. Our findings from the various forms of review and assessment we were able to complete are provided in the full report.

Regarding the project as a whole, we conclude that, in terms of the translation criterion, there were clearly some deficiencies at each stage that undermined the project’s ability to fully provide its intended deliverables. However, the main objectives appear to have been delivered (and indeed, the quantity of high profile individuals interviewed was impressive), and the conclusions do make sense (they have a certain face validity).

Finally, although we recognise the difficulties in reviewing what essentially became a research rather than a dialogue project, we do suggest that a general lesson can be drawn from the experience of this project in terms of access for reviewers or evaluators in future. We suggest that it is always important to ensure project team buy-in to evaluation, or other agreed forms of review, in order to identify good practice and lessons for the future - whatever form and methods are required and used for either the project or any review.

1. Introduction: the STAPE project and its assessment

At the end of September 2010, Sciencewise-ERC[2] and the UK Government Department for Business, Innovation and Skills (BIS) issued an invitation to tender for the task of evaluating a project entitled *Science, Trust and Public Engagement: Exploring Future Pathways to Good Governance* (henceforth STAPE). Originally, STAPE was to be a learning project, which, in its Phase one, would seek to identify examples of good, effective practice across government concerning mechanisms for governance, transparency, trust-building and public engagement in policy areas involving science and emerging technologies. At that stage, a second phase of STAPE was planned, initially conceived as a series of action research interventions, building on what was learned during Phase one, in order to develop and validate new and improved engagement mechanisms.

We responded to the tender invitation, and were subsequently appointed to evaluate Phase one of STAPE. In our proposal to do the work, we recognised that the task of evaluation would be centrally concerned with assessing the credibility, effectiveness and success of STAPE, measured against its deliverables and objectives. We noted that there were additional objectives concerning the provision of formative evaluation-based support during the course of STAPE, and the provision of evidence of good practice and impacts to support Sciencewise-ERC's wider mission in creating excellence and capacity-building for engagement practice. The evaluation was required to be line with Sciencewise-ERC's evaluation principles – namely: *starting early*, *clarity*, *rigour*, *appropriate level of participation*, *proportionality*, *transparency*, *practicality*, *utility*, *independence* and *credibility* - an approach we were happy to support.

The tender invitation also required the evaluation to address six key questions including whether the dialogue met standards of good practice for dialogue[3]. At that early stage, Sciencewise-ERC recognised that its principles of good practice for public dialogue might not, on their own, provide a sufficient basis to encompass the totality of activities entailed in implementing STAPE, in particular the project's focus on working with stakeholders rather than the public. Evaluators were asked to suggest other standards of good engagement practice that might be used to complement the Sciencewise-ERC principles in this case. In our response, we argued that our approach to the work would provide a suitable basis for evaluating a wider range of information-gathering and analysis, stakeholder engagement, and action research interventions, in addition to citizen engagement.

In the remainder of this report we begin by discussing our approach to evaluating the project, and the shift from a conventional evaluation of a dialogue project to a focus on peer review and methodological critique. Central to our approach overall has been the meta-criterion of *translation*, and we note how this incorporates many normative criteria from the academic literature (discussed in e.g. Rowe and Frewer, 2000), as well as addressing many of the Sciencewise-ERC criteria for good public dialogue practice.

We also set out our methodological approach, by way of a discussion of agreed changes in the project specification / objectives, which had practical implications for the assessment task. Following the methodological discussion, the report considers in detail the various steps in the project, attempting to record what was done during the project and to what end,

---

[2] The Sciencewise Expert Resource Centre (ERC), funded by the Department for Business, Innovation and Skills (BIS), is the UK's national centre for public dialogue in policy making involving science and technology issues. www.sciencewise-erc.org.uk

[3] Sciencewise Expert Resource Centre (2008). *The Government's approach to public dialogue on science and technology*. http://www.sciencewise-erc.org.uk/cms/publications/

and to what extent each step represented 'good translation' and/or matched Sciencewise-ERC's principles. The steps in the project involved, in order: the literature review; the interview of key stakeholders and organisation reviews; the workshop; and the drawing up of a project report. In the final section, we discuss a number of practical matters that placed constraints upon our capacity to evaluate the project as effectively as we would have wished, prior to setting out our conclusions.

2. Our approach to reviewing the project

The Invitation to Tender (ITT) for the evaluation of the STAPE project specified a number of objectives for the evaluation, together with a set of 'key questions' that should be addressed by the activity. The five objectives were to:

- Provide an independent assessment of the project's credibility, effectiveness and success against its deliverables and objectives, throughout and at the end of the project.
- Contribute to the overall Sciencewise-ERC aim of creating excellence in public dialogue to inspire and inform better policy-making in science and technology.
- Contribute to the development of mechanisms throughout the project to aid reflection and learning in relation to the project's own engagement processes.
- Gather and present objective and robust evidence of activities, achievements and impacts to support Sciencewise-ERC work in increasing wider understanding and awareness of the value of this work.
- Identify lessons from the project to support Sciencewise-ERC work in capacity-building across Government, and the development of future good practice.

The key questions were:

- Has the dialogue met its objectives?
- Has the dialogue met (Sciencewise-ERC) standards of good practice[4]?
- Have those involved been satisfied with the dialogue (of value to them)?
- What difference/impact has been made by the dialogue?
- What was the overall balance of costs and benefits for the dialogue?
- What are the lessons for the future? What worked well and less well, and more widely?

The ITT for the evaluation was distributed at the same time as the ITT for delivering the project, which is normal practice for these projects to meet the Sciencewise-ERC principle of starting evaluation early. At this stage, the project was still conceived as essentially a *dialogue* project.

In addition to adhering to Sciencewise-ERC evaluation principles, our approach to conducting the evaluation was strongly shaped by our commitment to the following additional six principles:

- <u>Evaluation mode</u> – our approach combined summative and formative approaches. In other words, we aimed to provide an overall assessment of credibility, effectiveness and success, as well as ongoing feedback during the course of the project on what was working well, and what might need a little fine-tuning, so supporting the possibility of 'mid-term corrections' in the implementation of the project.
- <u>The role of criteria</u> - we recognised the need for our assessment to be based upon appropriate (transparent and theoretically supported) criteria, whilst adopting a style of analysis that allowed us to learn, and therefore not to be unhelpfully constrained by the criteria should new information come to light.
- <u>The central role of knowledge translation</u> – our overarching framework (or meta-criterion) was that of translation quality, a criterion that was identified and

[4] Sciencewise Expert Resource Centre (2008). *The Government's approach to public dialogue on science and technology*. http://www.sciencewise-erc.org.uk/cms/publications/

successfully used in evaluating the *GM Nation?* public debate (Horlick-Jones *et al*, 2007a; 2007b). Translation quality is concerned with the efficiency by which information/knowledge is captured and utilised by engagement processes. More precisely, it concerns an assessment of the extent to which the project is based on an adequate understanding of the brief; designs and implements a process that is capable of achieving its set objectives and delivers the agreed outcomes; and successfully collects, analyses and reports information on that process to the project sponsors. There is potential information/knowledge loss at each project stage that may undermine the overall process. The central task of the review was to describe these stages and, through multiple data collection methods, assess their 'translation efficiency'.

- Methodological indifference - we have been committed to complete neutrality with respect to the rights and wrongs of policy issues addressed by STAPE. In methodological terms, we have also been committed to not seeking to adjudicate between knowledge claims by different actors in STAPE, and to not regarding any given source of knowledge as creating a benchmark of what is right or wrong (cf. Tom Horlick-Jones' preface to Horlick-Jones *et al*, 2007b).
- Plural use of methods - we were committed to gathering evidence using a range of qualitative and quantitative methods, and have been sensitive to the methodological difficulties sometimes inherent in seeking to triangulate between accounts generated by different research instruments.
- Publication - we have a serious commitment to peer-reviewed publications on the detail and results of this review. As part of our commitment to transparency (an important Sciencewise-ERC evaluation principle), and in recognition of the need for the review itself to be assessed (by external sources), we believe it is fundamentally important that our work is exposed to intense scrutiny by expert observers – not only the sponsors but to other significant stakeholders (notably, the academic community).

At the heart of STAPE lay a series of information/knowledge elicitation, gathering, integration, analysis and validation processes. As such, the meta-criterion of translation quality lent itself to providing a basis to assess the efficiency and effectiveness of these processes. Some of this information/knowledge was accessible by means of documentary sources, however, STAPE also necessitated an engagement with practitioners/stakeholders who possessed practical experience and tacit/experiential knowledges that, crucially, needed to be accessed and integrated with more formally-codified knowledges. We expected that the evaluation of STAPE would entail 'evaluation of evaluation' in the sense that we would attempt to acquire evidence of the project's success in evaluating the quality of evidence assembled to support the findings.

Having been appointed as evaluators for STAPE, an inception meeting took place on 15 November 2010, at which we met with Sciencewise-ERC representatives (the Sciencewise-ERC Evaluation Manager and a Dialogue and Engagement Specialist responsible for the project were both present). At this point, changes in the nature of the project were discussed. The continuing uncertainty over the Phase two part of the project was also noted (this was not covered by the contract for this evaluation).

The key change to Phase one of the project was a shift in the balance of activities towards more of a research-type process, with dialogue/engagement activities having a less significant role than originally conceived. A number of workshops were still included in the overall project design, but it was not clear at that stage how many of such events would take place, or their nature or likely participants.

On 26 November 2010, the evaluators were invited to join the Steering Group for the project in order to ease communications. That was agreed immediately. The evaluators produced a note summarising their revised approach to the evaluation to take account of the changes to the project design. This revised approach was finalised and agreed in December 2010.

The final STAPE project objectives, as agreed by the project Steering Group on 11 November 2010 and provided to the evaluators on 3 December 2010, were as follows:

- Synthesise public concerns and values on science and technology across a range of domain areas, the drivers that underpin and structure concerns.
- Synthesise insights around the substantive governance issues raised through public dialogues in recent years.
- Understand how governance is framed by scientific policy-makers.
- Understand the drivers that will impact on governance in the next 5-10 years.
- Understand barriers and enablers to taking account of the governance issues raised by the public.
- Develop an account of these findings, together with specific interventions that can be used to overcome barriers and promote enablers.
- Develop recommendations to galvanise action in this area amongst science-based organisations.

At this stage, the project methodology was described as comprising five stages:

- Stage 1. A review of findings from public dialogues, building on existing studies, and focusing on a) "synthesis of public concerns and values on science and technology across a range of domain areas"[5], and b) "formulation of a series of core governance issues that evolve from the above".
- Stage 2. Depth interviews with policy-makers "to consider how governance is framed and the drivers shaping this; and clarify the barriers and enablers to taking account of the governance challenges posed by public concerns and values towards science and technology".
- Stage 3. Ethnographic work "to observe and explore in depth how governance issues play out within two organisations".
- Stage 4. Workshops with "influential figures in scientific and policy communities to consider the findings from the interviews and help develop recommendations that can galvanise future institutional response".
- Stage 5. The development of a report, authored by the Royal Society, "highlighting and disseminating findings, and stimulating a wider policy review of science governance".

We proposed to evaluate these stages as follows:

Stage 1. Evaluation of the review of engagement experience, later described as a 'literature review', was to take the form of, in essence, a peer review.

Stage 2. Interviews. We recognised that the intimate style of stakeholder interview anticipated in this project raised important questions of sensitivity and confidentiality. In view of these considerations, we felt that it was probably unworkable for us to seek access to e.g. audio recordings of interviews, or to speak with the respondents. We therefore proposed to focus on the extent to which these activities demonstrated good practice, and the ways in

[5] All quotes in this section taken from the *TNS-BMRB Proposal*, dated 8.10.10.

which the interviewing fitted into the overall flow of activities. We hoped to be able to conduct informal interviews with the project team about the interview work.

Stage 3. Ethnographic work. At this stage, it was not entirely clear what this activity would comprise. Again, a focus on seeking to ascertain the nature of the data gathering and analysis, and an assessment of its quality, seemed the only means to evaluate this activity.

Stage 4. Workshops. Whilst there was still uncertainty about the nature and number of workshop-based stakeholder engagements, we anticipated using an observational protocol focused on translation quality, informal interviews with participants with possible follow-up telephone interviews, and the use of a short participant questionnaire administered at the events with postage-paid envelopes. We also wished to examine stimulus materials produced for the workshop(s). The observation checklist is included in Annex A, and the questionnaire in Annex B.

Stage 5. Drawing conclusions. We recognised that the production of project outputs would play a key role in the overall quality of translation processes associated with the project. We proposed to examine carefully the outputs along the lines of peer review, and also consider the quality of the flows of work that led from project activities to these outputs.

In practice, the project activities, as agreed with the Steering Group (including the project sponsors), included less stakeholder dialogue (one workshop) and more research and review than originally envisaged. As a result, the requirement for a formal project evaluation changed, and the potential for conventional evaluation research diminished. We therefore developed an approach that reflected these changes, focusing more on peer review and methodological critique, still using the meta-criterion of translation quality but taking account of the limited access to project activities that we were able to obtain. A summary of the chronology of the project events as they actually transpired is provided in the next section of this report.

3. The implementation of the project

In Table 1 we provide a chronology of key events in the implementation of the STAPE (Phase one) project, and our corresponding review and assessment activity.

Date	Project	Review and assessment
November 2010	15th - outline literature review produced	15th – Inception meeting 16th – circulated comments on outline literature review. 26th - THJ appointed to Steering Group, to facilitate communications
December 2010	Setting up interviews	8th – THJ telephone conversation with the project team to discuss implementation of the project and practicalities of the evaluation work. 9th – RW conversation with project team, possible meeting discussed 20th – RW conversation with project team
January 2011	Start of January – interviews began 6th – interview guide circulated 10th full literature review circulated Interviews in progress	19th - circulated comments on the full literature review 19th – RW telephone conversation with project team 28th – RW telephone conversation with project team arranged – postponed until 1st February.
February 2011	39 of the 40 interviews complete before workshop 9th – workshop at the Royal Society 23rd – meeting of the Steering Group	1st – RW telephone conversation with project team 2nd – RW email contact with project team 9th – RW observed workshop 14th - questionnaire sent by email to workshop participants 23rd – RW observed meeting of the Steering Group and contributed to discussion of the evaluation.
March 2011	10th – draft project report delivered	18th – draft executive summary of evaluation report delivered 22nd – draft evaluation report delivered
Post-March 2011	Final report completed	Final review report revised and delivered

Table 1: Timeline for the project and the review and assessment

4. Literature review

4.1 Early stages

The literature review set out to provide a review of findings from public dialogues, building on existing studies, and focusing on a) a synthesis of public concerns and values on science and technology across a range of domain areas, and b) formulation of a series of core governance issues that evolve from the above.

This corresponded to the first two of the agreed objectives:

> 'Synthesise public concerns and values on science and technology across a range of domain areas, the drivers that underpin and structure concerns.'
>
> 'Synthesise insights around the substantive governance issues raised through public dialogues in recent years.'

At the inception meeting, which signalled the start of the evaluation work, we were provided with a short outline of the forthcoming literature review. We were informed of the short deadline but were invited to provide comments by the end of the following day. Our comments, which constituted our first formative contribution to the project, were as follows:

- Although we recognised the important role of ambivalence in lay perceptions of science and technology, we noted that significant numbers of people *are* antagonistic to some technological innovations. Often this resistance is not associated in any simple way to e.g. membership of pressure groups, but it can be very strong. Psychologists try to understand such resistance in terms of concepts like 'core values'. Others have discussed the 'brand' of certain technologies and powerful negative associations they engender. Of course, others are enthusiastically in favour of certain technologies (e.g. mobile telephones, i-technologies).
- We noted that differences *do* exist across gender and social class dimensions, though perhaps age is more ambiguous. More important, there are clear associations with what one might call 'attitudinal clusters' or perhaps 'forms of life' that may not map onto classical variables in any simple way. Of course, DEFRA and other bodies have used such associations as the basis for segmentation models of behaviours. We also suggested that the degree to which the technology is in some sense 'unknown' should perhaps be considered.
- We felt it was important to look closely at the ways in which different technologies have their own specific characteristics, and indeed, one of the evaluation team has written about the 'signature' of different technologies (Horlick-Jones 2007).
- We wondered about the extent to which the four technology areas mentioned were, in some sense, 'representative'. We thought that a useful exercise might have been to try to characterise them according to some fairly uncontroversial dimensions, and then to map them to see if they cluster. Of course, it may be that the four areas identified needed to be the ones studied for all sorts of practical reasons, but it would have been useful to have a sense of the degree to which they 'span' the entire domain of technologies, or omit certain areas.

4.2 The literature review as a basis for the interview guide

We then circulated comments on the full literature review, which was completed on 10th January 2011. The timing of the production of the review meant that the programme of interviews had started by the time the report of the review was finalised. The literature review was intended to inform the guide drawn up for the interviews and indeed the findings from the review were known in advance of the completion of the final review report and enabled the interview schedule to be informed before the interviews were carried out. Nevertheless, the timing did mean that it was not possible for the various comments (from us and the Steering Group) on the full literature review to impact that interview schedule design.

Our interim comments on the literature review, circulated on 19th January, focused on any ways in which the review could have the potential to create imperfections in translation to the interviews. In these comments we:

- Expressed concern with the way the interview schedule drew upon the literature review, and the extent to which any possible imperfections in the review might be reflected in the design of the schedule. Ideally, of course, there would have been more time for discussion between the review being delivered and the interview schedule finalised.
- Suggested that the lack of attention to the practical, real-world, implementation of engagement and other governance initiatives might be important. In particular, we had hoped to find some consideration of the various unofficial 'uses' of engagement (i.e. the various organisational agendas that are served by such initiatives); the unanticipated impacts of implementing initiatives of this kind; and the ways in which such initiatives intersect (sometimes in positive and sometimes negative ways) with other policy objectives. We also wondered whether ethical issues rather got lost in the process of drawing up the interview schedule.
- Recommended that the interview schedule was reviewed and possibly amended to open the possibility of the interviews eliciting accounts that provided some insight into the informal aspects of implementation of engagement etc. initiatives. These were important practical considerations in seeking to capture the full range of constraints and opportunities associated with policy development in this area.

4.3 The substance of the review

Overall, we found that the review provided a reasonably thorough synthesis of the recent scholarly literature concerned with engagement practice, and represented a valuable contribution to the project. Our view was that there could have been a broader consideration of questions concerning the organisational and policy-making context in which engagement policy is formulated and enacted, as the review focused on the ways in which science and policy institutions were responding to governance challenges in three areas - genomics, nanotechnology and climate science. In addition, we might quibble slightly over the focus on certain specific tendencies in the literature that were dominant in the review. However, we acknowledge the project team's prerogative to take this particular view of things.

We welcomed the attempt to ground the findings in the practical experience of 17 Sciencewise-ERC dialogue projects (the review covered the 17 project reports). We focused our analysis on the effectiveness of this process, and therefore examined the 10 available *evaluation* reports corresponding to a subset of these Sciencewise-ERC projects. We conducted a thematic analysis of these evaluation reports, focusing on trends concerning practical aspects of the dialogues that went well, and things that did not go so well.

The literature review identifies six ‘potential’ criticisms of engagement, namely:

1) Group discussions are often organised on the terms of the host decision making institution.
2) Small group discussions tend to reinforce consensus and homogenise views.
3) Stimulus materials and expert presentations obstruct the deliberative process by leaving insufficient time for discussion on participants’ own terms.
4) Selection techniques produce a stable image of “the public”.

5) Sampling strategies along standard demographic lines produce groups with weak ties and hence shallow conversation.
6) Qualitative methodology is not well equipped to deliver information with regards to attitudinal segmentation.

Our analysis of the 10 Sciencewise-ERC project evaluation reports, which correspond to a subset of the 17 Sciencewise-ERC project reports considered by the literature review, suggests that what the review describes as 'potential' criticisms do not fully correspond very closely with our analysis of these reports.

We found that whilst these six criticisms do indeed resonate with the evaluation of one of the Sciencewise-ERC projects, namely the one concerned with stem cells (2008), they do not apply more widely across the rest of the dialogue evaluation reports. Three of these 'potential criticisms' in particular are not supported by evidence offered by the real world experiences of previous dialogues.

Consider the second 'potential criticism' for example. Here, the review suggests that 'small group discussions tend to reinforce consensus and homogenise views'. It is clear from five of the dialogue evaluation reports that this was not the case. Rather, the small group discussion method used in these projects resulted in the following evaluation conclusions:

- That, in overwhelming terms, participants rarely reached a consensus of opinion (Hybrid and Chimera Embryos p31).
- That facilitators gave a one sided view of the groups' discussions, reporting a consensus when there was none (Animals Containing Human Material p17).
- Efforts were made to defuse tensions arising from conflicts of opinion in a number of the groups (Synthetic Biology p 16). This suggests that there were indeed heated debates, and thus little consensus amongst the group members. In fact there are many examples of participants of small group discussion reporting 'lively' and 'passionate' discussions (Drugs Futures). Moreover, a facilitator's need to defuse such tensions may be suggestive of a desire on behalf of facilitators to create consensus amongst group members. Such a trend is noted by the evaluators when suggesting that experts commented upon the impact of facilitation on the groups whereby groups were seen to develop a view accepted by the group even if these views were not representative of the diversity of the opinions held (Synthetic Biology p32).
- Similarly, the evaluators cite an example where participants felt under some pressure from the process to come up with an answer (Big Energy Shift p 19). This could be suggestive of a means by which consensus may not naturally be reached by participants, but is the result of facilitation.

The third potential criticism of engagement identified in the review suggests that stimulus materials and expert presentations obstruct the deliberative process by leaving insufficient time for discussion on participants' own terms. Our analysis of the evaluation reports suggests that the stimulus materials and the involvement of experts in fact provided an indispensable mechanism in engaging the participants. Indeed, it was often suggested that *more* time should be allocated for the dissemination and absorption of stimulus materials. It seems here that the problem may be more of a failure to allocate sufficient time for the dialogue process as a whole, rather than being something to do with the nature of stimulus materials.

We could find little evidence to indicate stimulus materials being seen as obstructing deliberative dialogue. Rather, our analysis found the following:

- Stimulus materials were seen to contribute to the *quality* of the discussion (e.g. Science Horizons pp.58-9) but that the issue of time poverty was often a barrier to the *depth* of discussion (*ibid* p.88) (i.e. reduced time for dialogue due to the extent of information).
- An obstruction to deliberative discussion related to room size rather than a focus on stimulus material (Stem Cells p 42).
- Stimulus materials and expert involvement added to helping participants experience and in helping to inform their opinions (Animals Containing Human Material p 20).

In relation to the fifth potential criticism of engagement, about sampling strategies, our analysis suggests that sampling strategies that have depended on standard demographic characteristics did *not* generate weak group ties amongst participants resulting in shallow conversation as the review suggests. The following examples from a number of evaluation reports suggest the opposite:

- If shallow discussions were reported, these were seen by the evaluators to result from the structure of the day being repetitive (Animals Containing Human Material p15). Moreover, participants felt they were getting a greater depth of discussion the more they spoke with people different to themselves (p 32).
- A diverse mix of people was seen to lead to 'good discussions' (Drugs Futures p 47) about 'challenging topics' (p 22).
- Even amongst those that represented a deliberately homogenous cross section of the public, shallow conversation was reported (DNA p14).
- Where the quality of the discussion was identified by the evaluators as an aspect of the dialogue that 'worked well' this was attributed to the diverse sample with a good mix of people (Drugs Futures p 68).

The literature review did not aim to cover engagement practice in detail, and did not consider project evaluation reports. However, the STAPE project was conceived as one that would draw upon the wisdom of past engagement experience, and the literature review does suggest that the set of Sciencewise-ERC projects were examined as part of the process of review. Our analysis suggests this may have been more limited than might have been expected.

4.4 Conclusions

The review is well-written and condenses a significant literature. However, our view is that the translation quality of the literature review process was possibly limited in two ways. First, the overall scope of the review, in terms of the scholarly literature, could have included more on the broader literatures on organisations and on the practicalities of policy-making. Second, our analysis of evaluations of earlier Sciencewise-ERC dialogue projects suggests that the review was more limited than might have been expected and that therefore the project may have had too little sensitivity to practical features of the implementation of engagement practice in real-world settings.

5. Interviews and organisation reviews

5.1 Preamble

Here we consider the two component parts of the project to which we were unable to gain direct access. Early discussions with the project team and Steering Group suggested that the interviews would take an intimate and confidential form, so precluding us having access to audio recordings or to the interviewees themselves – which prohibition we accepted, recognising that there might have been valid difficulties regarding confidentiality. As noted above, we concluded that we would need to obtain information about the interviews and what was done with the data by interviewing the project team.

5.2 Interviews

Our analysis has been based upon the description of the interview and data analysis process given in the final report, and directed at the extent to which appropriate and comprehensive information was recorded by these processes (what questions were asked; how data was collected; how it was recorded; how it was summarised; whether it appeared to speak to the project's central questions or not).

The project team conducted telephone interviews with around forty senior decision-makers in 'science and technology based organisations'. Most of these organisations were UK central government department and agencies; relatively few of them were from the private sector or from NGOs. Two people from a local authority were interviewed. The project was successful in recruiting as interviewees a range of senior people from public sector bodies.

In relation to the analysis of data, the project used what was described in the final project report as 'core matrix mapping'. It was not entirely clear whether this referred to a variant of standard analytic processes or a customised technique that might have been regarded as commercially confidential. Both thematic and content analyses were implied. The final project report contains information on the analytical process but few additional details are provided.

5.3 Organisation reviews

At the time of our 8th December interview with the project team, the ethnography activity was yet to be arranged.

The final project report describes how four / five people in each of two organisations were involved in what were effectively semi-structured interviews at which field notes were taken. The lack of recording may have led to imperfect data collection, although in the methodological context this seems acceptable and understandable. The final report does not, however, identify what extra insights emerged from this part of the process: the report appears to combine insights from all of the methods (ex-review) into one narrative. Further detail on the unique contributions of this piece of work would have been useful. The absence of differentiation in the report may, however, stem from the fact that insights across methods were similar and hint at cross-method validation.

6. Workshop

6.1 Preamble

One workshop took place, providing an opportunity for people from the organisations from which the interviewees were drawn to discuss provisional findings based upon the earlier stages of the project. The event was designed to provide 'a safe space' where issues could be discussed in detail and without inhibition under Chatham House rules.

6.2 The event

The workshop took place in the conference room of the Royal Society (RS). Four tables were set out to each comfortably accommodate between 4-5 participants. A total of roughly twenty-five participants faced a raised stage with lectern, board table and a large power-point presentation, which read, 'Science and Trust: Some Strategic Issues Facing Science Governance'. A policy officer at the RS, a member of the project team, chaired the event. He introduced the President of the RS, Sir Paul Nurse, who introduced proceedings, covering:

- The need for improved governance of science through increased investment in public engagement
- How public engagement provides a justification of science and public funding - improving life, enhancing culture, and contributing to economic growth
- Why scientists need to engage with the public, to 'earn our licence to operate'
- How democracy deals with complex questions, which demand transparency, an open ear and determination of the kinds of questions that need to be asked
- How the RS is committed to address these issues
- His personal enthusiasm for the workshop, and the need for critical reflections on 'how we work together'

The chair provided a brief overview of the day's proceedings, although with little in the way of 'housekeeping' information. Participants were assured of the confidentiality of the workshop, and their own anonymity in any published documentation, but were not invited to introduce themselves. The member of the evaluation team present was pointed out by the chair. The Chair then handed over to the project director.

The project director began by referring to the 'safe umbrella of the workshop' as a space in which to 'critically' discuss interview findings. His subsequent Powerpoint-based presentation addressed:

- Public engagement activities over the last ten years, identifying which substantive governance issues had been raised by these activities
- Five key issues identified by the literature review
- The project interviews – how interviews were designed and conducted, and the headline findings

The presentation was in many ways detailed and comprehensive, although we identified several problems that could have caused difficulties for later workshop discussions:

- A lack of detail on methodology and research design (who were the individuals that were interviewed, why were they interviewed, who interviewed them, for how long, how was this information coded etc.)
- Numerous text-heavy slides, containing far too much information for participants to reasonably assimilate
- A tentative tone, which served to rather undermine the credibility of the findings

This twenty-five minute presentation was followed by two short question and answer sessions about the presentation, managed by the workshop Chair. These were tightly controlled, offering little opportunity for the critical elaboration of ideas or for participants posing questions to enter into any sort of dialogue with the project director. Without introductions having taken place, it was unclear who was asking the questions, with just a few participants choosing to state their identity and/or organisational affiliation. Following the first round of questions, the workshop Chair asked if the issues raised during the session 'resonated', which failed to elicit any noticeable response from the audience. In total, nine questions were posed, most of which were general comments, although one specifically concerned the 'vacuum' created by personnel moving between mainstream activities and public engagement work.

After an extended break, a representative of Sciencewise-ERC made a short, informal, presentation, largely directed at making the case for the virtues of engagement. It was not entirely clear how this contribution was related to what had gone before, or what was to follow.

A break-out group activity followed. Participants were asked by the Chair to join groups with people having different perspectives on public engagement to them. Given the lack of introductions or general discussion in which these perspectives might have become apparent, it was not clear how this was to work. Unsurprisingly, no-one moved from the groups they had arrived in at random, save one individual who was encouraged to join a group to make up the numbers. At this stage, three participants had already left the workshop. This left four small groups, each comprising four-five people sitting around a table. A member of the project team was allocated to each group.

Each group was asked to appoint a rapporteur. They were then asked to discuss 'threats, opportunities, and risks of doing nothing', and were allocated twenty-five minutes for this discussion. They were provided with no prompts, Powerpoint slides, nor hand-outs to support this activity. There was no evidence of introductions taking place within the groups. Whilst some of these groups produced copious notes, others relied purely on the appointed *rapporteur* to record conversations.

Given that the room was relatively small, and tables closely spaced, it was relatively easy for the member of the evaluation team present to monitor parts of all the various group-based conversations. These conversations were slow to start, with participants apparently unsure of how best to proceed. Some members of the project team responded to this hesitancy by being proactive in seeking to promote exchanges, and in so doing perhaps became a little overly-prominent in the dialogue. At the end of this activity, rapporteurs were asked to provide a short 'one-minute' summary of their discussion to be fed back to a plenary gathering. We felt that this process was a little hurried, with little time allocated for questions or discussion about what the rapporteurs had to say.

Participants were then asked to return to their break-out groups and to appoint new rapporteurs. In this second discussion period, they were asked to address the question 'how to take things forward: what do you need to do as a group or as organisations?' Once again, the groups were not provided with prompts or other materials to support the dialogue. The group conversations were a little more lively this time, but we noted that a number were increasingly characterised by one-to-one exchanges between individual participants, rather than dialogue involving the whole group. We also found that the topics of conversation had digressed and wandered away from the key questions posed by the Chair.

The rapporteurs were once again asked to provide quick feedback to the plenary group. The Chair informed participants that they would be provided with a summary of workshop discussions for approval, and then brought the event to an end.

6.3 Findings from observation

a) Information comprehensiveness

We observed that the purpose of the workshop and the agenda for the event were not clearly defined. Reference was made to science governance, and issues of science, trust and public engagement, but the relevance of these topics to the project objectives were not fully articulated. Rather, we found that these themes were, rather loosely, ‘up for discussion’. It was unclear who owned the event, and what, precisely, was the relationship between the project, BIS and Sciencewise-ERC. The composition of the project team was not revealed, leading to some ambiguity about whether the Royal Society was merely a host, part of the project team, or was endorsing the event with its imprimatur. The different component parts of the workshop were not well coordinated, and appeared fragmented and disconnected with, at times, jerky transitions.

b) Information appropriateness/fairness

The project team provided an account of public engagement, science, and trust issues, and stressed that all interpretations and ideas presented were open to challenge and debate. The opportunity for disputing ideas was, however, rather limited. As the primary space for such activity, the break-out groups occasionally suffered from the over-involvement of project team members. On the whole, however, participants were able to interact and put forward ideas in a relatively unrestrained manner. On some occasions, individual participants were able to dominate and overshadow the contribution of others and, in these cases, the task of moderating was made increasingly onerous and was not always feasible.

c) Process limitations to effective translation

We observed that the workshop process was noticeably constrained by a shortage of time, with too many tasks crammed into too short a time. The volume and distribution of information provided to participants was erratic and inconsistent. In contrast to the information-heavy initial Powerpoint, there was a dearth of more basic information about the workshop process. The use of elected or volunteered participant *rapporteurs* allowed the collection of data from each break-out group to be largely impartial or external to the views of the project team, but it was not clear how these views were to be integrated in the whole project report. Given the time constraints, *rapporteur* reports tended to be little more than brief, headline summaries.

d) *Information synthesis*

Our observation found limited opportunities for interaction between project team members, rapporteurs and participants, and between participants themselves. A lack of stimulus materials and facilitating technologies, such as flip-charts, whiteboard, Powerpoint, video or audio, meant that participants relied solely on one-off instructions by the workshop Chair to prompt discussion. Project team members were observed to be making notes, but these were not used to support the workshop process. Participants were not provided with any workshop packs, stimulus materials, or orienting literatures, or printed agendas or delegate lists. Copies of the project team's initial presentation were left on tables, but we saw few participants picking up or reading these materials.

6.4 Findings from questionnaire data

a) Preamble

We had planned to issue a short questionnaire to the workshop participants at the end of the event, together with a postage-paid envelope. This questionnaire, which is reproduced at Annex B, is related to other evaluation questionnaires we have used in the past (see e.g. Horlick-Jones *et al*, 2007b), so using a tried and tested instrument to not only generate valuable data, but also to create the possibility of comparability with other engagement exercises. However, following discussion, we were prohibited from issuing the questionnaire on the day of the workshop as it was suggested this might be disruptive to the intimacy of the occasion and possibly inconsistent with the anticipated high-level decision-maker participants. (In the event, a number of the high-level people who had been interviewed sent substitute participants on their behalf.) We were, however, permitted to circulate the questionnaire in electronic form a few days after the event - a delay that inevitably eroded the response rate to just four replies. A follow-up request, sent a week later, elicited a single additional completed questionnaire. In total, therefore, we gained just five responses from around 25 participants.

Clearly, there are significant limits on what we can say on the basis of so poor a response rate. We have no way of determining how atypical the small sample of five respondents was with respect to the group of workshop participants as a whole. Nevertheless, taking into account both descriptive statistics and contributions in free text boxes, a number of interesting agreements as well as differences of opinion were in evidence. Needless to say, these findings should be regarded with an appropriate degree of caution.

b) Clarity of purpose

Four out of five respondents agreed that they thought that the workshop was *well run*. However, when asked if it was clear from the information that they were sent prior to the event what the workshop was about, two out of five respondents were unsure.

c) Appropriate attendees

Two out of five respondents thought the audience was appropriate for the event, two thought it was *not*, whilst one remained unsure. Two were unsure why *they* were invited. Three out of five provided further comment as to whom they believe was missing from this event. One respondent said:

> *"It would have been better if there were more high level people, i.e. CEOs but I know that the RS tried! Maybe next time don't put 'public engagement' in the title as CEOs automatically pass it to their PE person…"*

d) Influence on public engagement policy

Two out of five respondents thought this event *would* have influence on engagement policy, although three gave less positive feedback. One respondent thought the event would have 'no' influence on engagement policy, explaining:

> *"Largely this event spoke to the converted with insufficient attention to barriers or weaknesses in PE narrative."*

Two others indicated that they were unsure as to whether the workshop would have any influence:

"I wasn't really sure what the next steps were i.e. how the outputs of the interviews and workshop would be used."

"… I believe there was valuable dialogue but not sure what difference it will make (…) what are the success measures? More case studies of success might help."

In addition to these comments, participants were invited to reflect upon the extent to which they believed the event *should* have influence on engagement policy. Comments included:

"I think the discussions made a lot of sense. The attendees were intelligent and thoughtful and it would be nice for them to be 'heard'."

"As head of (science strategy) at (organisation) I will certainly expect the report to be valuable in my work to develop (the organisation's) approach to science governance, and colleagues here are also positive about the project. More generally, I would expect that the status of the report, coming from the RS, and the quality of the people involved, to produce well-grounded and useful outcomes".

"In many circumstances PE can be a valuable tool. But sometimes not and often not when key players/institutions not bought in."

"Particularly valuable if a consensus/common understanding of shared goals, and if approach can be achieved to encompass policymakers, academic researchers/research funders and industry, rather than fragmented approaches."

"We need to consider the challenge of public engagement rather better and this will help raise the issue and perhaps get some momentum - especially from those involved with the commissioning of research and the end users in the translation."

All five respondents can therefore be seen to be enthusiastic with regards to the future of public engagement. The workshop clearly served to 'raise the issues', perhaps paving the way for future activities involving key players and organisations.

e) Sufficient time to discuss everything

Here we found contradictory evidence. Four out of five respondents said they had said all that they had wanted to say during the event; one did not answer this question. Similarly, four out of five agreed that there was sufficient time to discuss all that needed to be discussed, and one was unsure. However, a lack of time was identified as a barrier to having discussions in sufficient depth.

"Given this had to provide an introduction, and get us all to the same place etc., the answer….is probably yes. However the breadth of topics raised made it impossible to give most, if not all, the detail they merited."

"I think we covered a lot of ground but inevitably were not able in the time available to explore these in much depth."

Four of the five respondents thought that there were several issues not raised that they felt should have been discussed. Two suggested the following specific issues:

"Context in which any PE project takes place"

"More cases (case studies) to illustrate where public engagement had helped achieve a better outcome. These seemed rather thin and one example was repeatedly mentioned in discussion. Are there others?"

f) Overall perceived success of the event

On balance, the workshop was perceived as a success in that respondents on the whole found it to be a pleasurable experience, where they enjoyed conversations with what one respondent described as a ‘mix of people’ (seen as a positive feature of the event).

The questionnaire data provides (albeit very limited) evidence of both positive and negative views concerning the success of the workshop event. Negative points related to the absence of some key players and the lack of time given in order to explore issues in depth. A perceived success of the event was that it was seen to possibly pave the way for future events to explore pathways to effective public engagement in the governance of science and technology policy. While we should not read too much into these questionnaire results, given the low response rate, the results are still interesting for having been provided freely by a number of relatively senior participants, and for supplementing the previously described observations of the event.

6.5 Some conclusions

In summary, the workshop provided a valuable space for senior representatives from a number of significant stakeholders to come together to hear the initial findings from this project, to meet other relevant parties, and to have some discussion about the central issues of engagement and governance. However, there were a number of weaknesses in the design and delivery of the event that lessened the value of the event, including a lack of clarity of purpose and of time, a lack of materials to aid in assessing the questions for discussion, and limited recording of the content of the dialogue that took place.

7. Project reports

We received the first draft of the project report on 10th March 2011, so, at the time of finalising our draft report, we had had just a few working days to examine its text.

A substantially revised project report was completed at the end of July 2011. Following the summer break, a further series of sponsor and Steering Group comments prompted a final phase of fine-tuning, and the final report was delivered in September and published in November 2011.

We welcome the much more comprehensive nature of the final report. From an assessment and review perspective, however, we regret the lack of detail linking the collection of evidence to the conclusions (an audit trail of some sort). We do recognise that such methodological detail was not requested by the project sponsors or the Steering Group, and is rarely provided in project reports of this type. Nevertheless, it may be useful to consider and develop some ways of tracking the 'translation' of data through the various stages of the project that are both proportionate (not too onerous for participants and others) and valid. This is something that might be considered in future Sciencewise-ERC funded projects.

8. Conclusions

8.1 Preamble

Here we set out a synthesis of our findings. Earlier in this report we have documented details of the project that we set out to evaluate, our revised approach focusing on peer review and methodological critique, and the nature of the evidence that we were able to gather in performing this work. For each stage of the project process, we have examined the evidence available on the extent to which the project achieved translation quality and adherence to Sciencewise-ERC principles.

In Figure 1 we have presented a schematic representation of the STAPE project. This diagram highlights the processes of information/knowledge flow, or translation, that we understand took place.

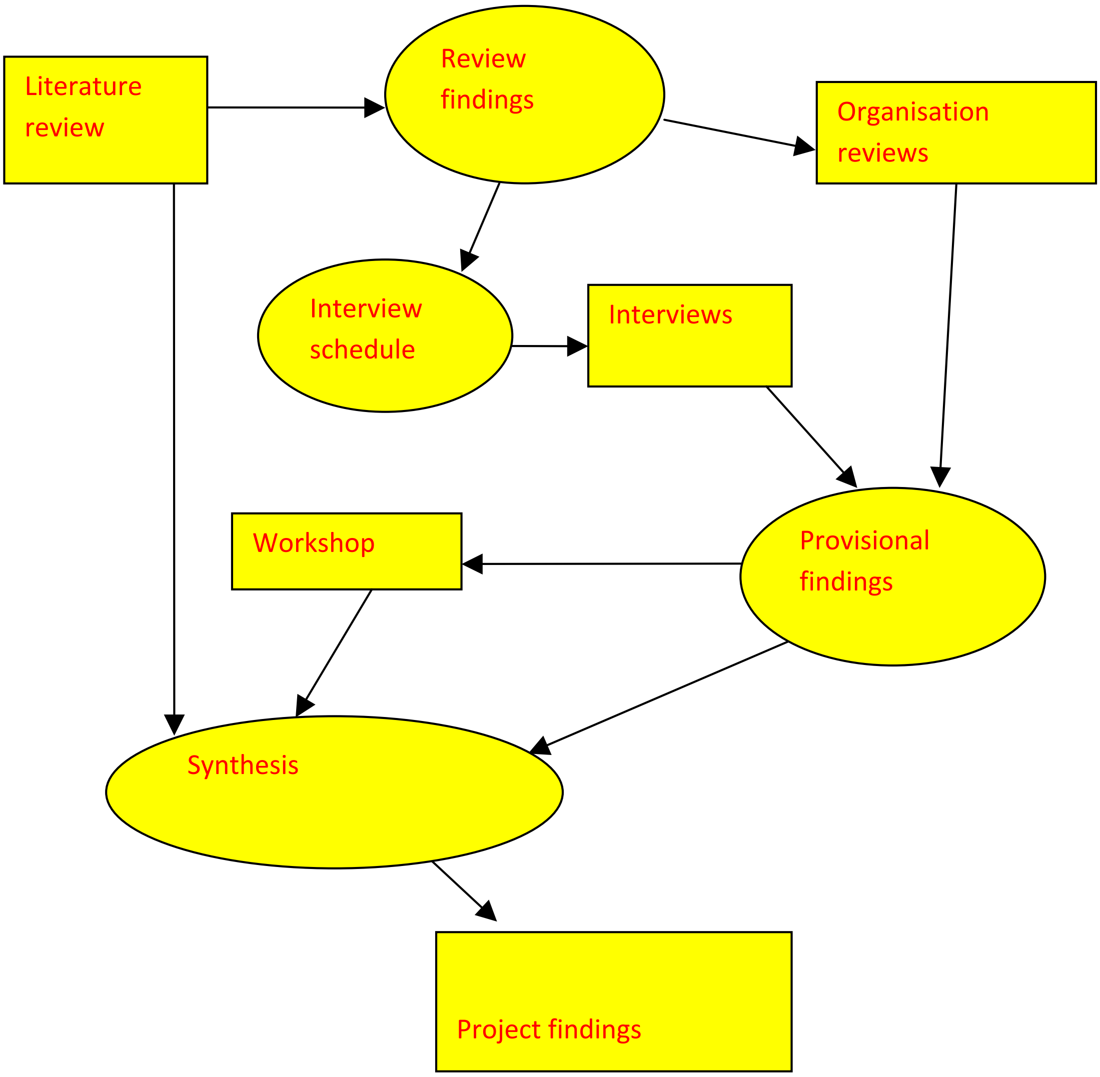

Figure 1: A schematic representation of the project process

8.2 Translation quality

In Table 2 we have summarised our understanding of the effectiveness of various stages of the project against our meta-criterion of translation quality.

Stage	**Source**	**Outputs**	**Translation quality**
Literature review	Scholarly literature, Sciencewise-ERC evaluation and project reports	Review report	Generally good review although, in the spirit of peer review, we suggest it could have been more widely framed and had more focus on sensitivity to practical implementation issues, with less reliance on particular aspects of the scholarly literature.
Drawing up an interview schedule	Review findings	Interview schedule/ topic guide	Although the emerging findings from the review informed the development of the interview schedule, translation quality was reduced by time pressures limiting the opportunities for feedback based on discussion of the full and final report of the review (by the Steering Group and others).
Interviews	Key stakeholders	Findings presented in final report	Difficult to assess as we have little information on the precise nature of the analysis of interview data, or access to interview transcripts (which we accept as procedurally correct).
Organisation reviews	Everyday life in two key stakeholder organisations	Discussed briefly in final report	Difficult to assess as limited information in final report, and no information on the specific results from this process.
Workshop	Provisional findings explored with a group of key stakeholders	Rapporteur feedback and notes by project team.	The workshop provided a valuable opportunity for senior representatives from significant stakeholders to discuss the implications of emerging findings from the project. However, the translation quality of the workshop was reduced by lack of time and clarity of purpose, and lack of effective means for recording output. Limited evaluation questionnaire data received.
Synthesis	Workshop outputs, interview findings, literature review, and organisation review data	Final project report	Little data in report on how the various forms of evidence were analysed and synthesised, and how quality control was achieved. In particular, it is unclear how the interview/ organisation review/ workshop processes separately and distinctly fed into the final conclusions in the report.

Table 2: Summary of results of the assessment of translation quality

To conclude, the evidence we have collected suggests that some aspects of the project were good, in particular the impressive participation of a large number of high-profile people in diverse organisations. However, there were a number of shortcomings with respect to translation quality, as outlined in the table above.

8.3 Sciencewise-ERC criteria

We were also specifically asked to assess whether the project had answered a number of ‘key questions’ that are standard to public dialogue projects funded by Sciencewise-ERC. These questions were:

- Has the dialogue met its objectives?
- Has the dialogue met (Sciencewise-ERC) standards of good practice?
- Have those involved been satisfied with the dialogue (of value to them)?
- What difference/impact has been made by the dialogue?
- What was the overall balance of costs and benefits for the dialogue?
- What are the lessons for the future? What worked well and less well, and more widely?

Before addressing these questions it is important to recognise that this was not a public dialogue project, and we were able to gather only very limited conventional evaluation research data. We therefore focused instead on peer review and methodological critique. These 'key questions' therefore have limited relevance in reviewing this project. The following summarises the findings of our review within these constraints.

Regarding ‘objectives’ – broadly speaking we would conclude that the project *has* achieved what was specified, although the revised objectives were not drafted with the specific purpose of providing a framework for effective measurement of achievement. The final report does provide responses to the questions intended to be answered. Moreover, the project has largely enacted the processes it was intended to have done. However, the lack of methodological detail, and in particular the absence of a substantial body of evidence on data recording and synthesis, means we cannot comment on the connections between the answers provided and the processes undertaken.

The second key question asks whether the exercise reflected good public dialogue practice according to the Sciencewise-ERC principles which are, in summary:

- the conditions leading to the dialogue process are conducive to the best outcomes (Context)
- the range of issues and policy opinions covered in the dialogue reflects the participants’ interests (Scope)
- the dialogue process itself represents best practice in design and execution (Delivery)
- the outputs of dialogue can deliver the desired outcomes (Impact)
- the process is shown to be robust and contributes to learning (Evaluation)

We have already described above that the project did not include any public dialogue and it was not therefore possible to apply these principles fully. However, some of the principles were still relevant as the basis for some limited analysis, as follows. In terms of ‘context’ (as defined by Sciencewise-ERC), the project was fairly clear in its purposes and objectives from the outset and been well timed in relation to public and political concerns. Regarding ‘scope’, the project appeared to have covered that intended by the sponsors, being clear about participants' limits in influencing outcomes and being focused on informing, rather than determining policy and decisions. Regarding ‘delivery’, there was a single stakeholder workshop which was valuable but which had a number of methodological limitations. Regarding 'impact', this is still to be determined, though the sponsors are clearly in a good position to ensure the report has the best chance to be appropriately influential. And regarding 'evaluation', the sponsors have clearly made efforts to ensure a thorough assessment has taken place, though the quality of this has changed and been limited as previously discussed.

Regarding the third question, 'satisfaction', we need to consider the views of the key stakeholders who participated in the workshop, and of members of the project's Steering Group, including the BIS representative. On the basis of observational evidence, and limited questionnaire data, the few stakeholders who responded were interested in, and stimulated by, the workshop although some regretted there were not more opportunities for greater depth of discussion. We have monitored email traffic between members of the Steering Group and participated in one meeting of the Group. From that necessarily limited evidence, we conclude that the Group members have been reasonably happy with the project, and early comments on the findings have been broadly positive. We expect satisfaction to ultimately be determined by responses to both the final project report and this report, and hence, to be beyond the scope of our assessment here.

Turning to whether the project will make a 'difference' (the fourth question) it is, of course, far too early to tell. Comments by workshop participants and Steering Group members suggest the topics raised by the project have proved highly stimulating, but it is not yet clear whether concrete, practical ways forward will arise from this activity.

Regarding costs and benefits (the fifth question), we can draw no clear conclusions: it is arguably a matter for the sponsors to say whether they believe that this project has delivered value for money. Regarding 'lessons' (the sixth question), these are discussed throughout this report.

8.4 Evaluation of the evaluation activity

Finally, we turn to 'self evaluation' of the evaluation work itself. Given the significant obstacles we encountered in the gathering of evidence, we begin with some observations on this aspect of the work. We have already alluded to this difficulty on a number of occasions, and lest the reader feels that we have become a little repetitive, we should point out that in our collective experience of this sort of work, we have not previously encountered such considerable difficulty.

Conducting evaluations of public engagement projects is always potentially difficult, and it is necessary for us to elaborate on the nature of the difficulties we faced in order to assess the limitations that these have placed upon our analysis. There are essentially two classes of evaluation difficulties; those related to the theoretical question of 'what makes an engagement exercise/project good or effective', and those related to the data collection and interpretation and other practical aspects of the implementation of the evaluation. Elsewhere we have discussed these difficulties in detail (Rowe *et al*, 2005).

The STAPE evaluation proved no more conceptually difficult than others we have conducted. This feature was in part thanks to the early acceptance by the sponsors of our use of the translation meta-criterion, and the fact that our approach proved to be readily adaptable to the revised nature of the project (away from engagement towards review and research). In this case, most of the difficulties stemmed from severe constraints on our ability to acquire full, appropriate, and sufficient data to conduct our analysis and therefore to complete a full evaluation as originally planned, although we have completed as comprehensive a peer review and methodological critique as we could in the circumstances.

We have already described the nature of the project and the way it was implemented, which necessitated us seeking to gain information about some aspects of the implementation via the project team rather than directly from participants. We attempted to establish lines of communication, by means of telephone calls, that would allow us to review project activities, in terms of data collection and analysis, without creating unreasonable difficulties. In practice this task proved very difficult. We very much regretted this state of affairs, and offered to meet with the project team in order to find ways of establishing a more satisfactory working relationship. This offer was accepted, but in practice did not materialise. The upshot of this state of affairs is that there are clear gaps in the datasets that we attempted to collect, the practical implications of which are set out in this report. We have also discussed the loss of momentum introduced when we were denied the ability to distribute our questionnaires on the day of the workshop, which contributed significantly to the low response rate.

As committed professionals, it has been a frustrating experience not to be able to complete the evaluation in as thorough a manner as we had anticipated. We recommend that Sciencewise-ERC reviews its contractual and management arrangements for future such projects, with a view to ensuring that such severe obstacles to the collection of evaluation evidence are less likely to occur.

We now turn to the objectives for our evaluation work that were specified at the outset. Those objectives were to:

- Provide an independent assessment of the project's credibility, effectiveness and success against its deliverables and objectives, throughout and at the end of the project.
- Contribute to the overall Sciencewise-ERC aim of creating excellence in public dialogue to inspire and inform better policy-making in science and technology.
- Contribute to the development of mechanisms throughout the project to aid reflection and learning in relation to the project's own engagement processes.
- Gather and present objective and robust evidence of activities, achievements and impacts to support Sciencewise-ERC work in increasing wider understanding and awareness of the value of this work.
- Identify lessons for the project to support Sciencewise-ERC work in capacity-building across Government, and the development of future good practice.

This report concludes our independent review of the STAPE project's credibility, effectiveness and success. The work we conducted has been rigorous and evidence-based although limited and with a different emphasis from the original evaluation plans, for the reasons outlined above. We have nonetheless attempted to further operationalise the translation quality meta-criterion, in ways that we suggest enhance the development of good practice in engagement evaluation, and so Sciencewise-ERC's objectives more generally. Significantly, we were able to demonstrate the capacity of translation quality to form the basis for evaluating engagement initiatives comprising a wider range of activities than dialogue processes alone.

The shift in balance from engagement towards review and research, corresponding to the changed specification of the project, served to highlight a potentially significant feature of the evaluation activity. It is for these reasons that we moved from a conventional evaluation approach, designed for dialogue projects, to a focus on peer review and methodological critique. This shift emerged first and most clearly in connection with the literature review, which by necessity entailed a focus on *peer review*. In the case of the specific areas of the literature reviewed, there is considerable relevant technical expertise within the evaluation team, so this did not present a significant difficulty for us. We were, however, keen to avoid our review of this activity turning on perfectly reasonable differences between the project and evaluation teams over our reading of what is significant within the literature. As we have already noted, we might quibble slightly over the review's focus on certain specific tendencies in the literature, but we have acknowledged the project team's prerogative to take their particular view of things.

The timing of the project was another notable feature. From observation in person (at meetings) and emails trails, we found that there was too little time being allocated throughout the project's duration: the literature review arrived later than originally expected; the arrangements for the workshop were finalised at the last moment; and the draft project report arrived only a week before we were committed to delivering an early account of our own findings. Of course, we appreciate that engagement exercises in the real world often take place in difficult or time-urgent circumstances. Nevertheless, we believe that the substance of the project will have suffered from having to be implemented according to such a constrained timetable.

Finally, we turn to the topic of capacity-building and future good practice. We have already set out a range of suggestions that we hope will shape future Sciencewise-ERC practice. These include: further evidence of the significance and applicability of the translation evaluation criterion and associated perspective; the need for tightly-drafted objectives against which achievement can effectively be measured; and the need for contractual and management arrangements that seek to avoid evaluators encountering unworkable obstacles in the gathering of evidence. In the current economic circumstances it is perhaps especially relevant for us to also say something about engagement evaluation in an 'age of austerity'. It should be noted that in order to produce an evaluation report of a standard with which we are reasonably content, we have had to spend more than twice the staff time corresponding to the fixed (and rather modest) budget available. Clearly this shortfall in resources has important implications for doing high quality evaluation work in times of financial cuts, where engagement itself could easily be regarded as not being an essential priority for science and technology-related policy and practice. There is a danger, we suggest, that by starving evaluation work of resources, the resulting evaluations will be insufficiently robust to contribute constructively to the development of good engagement practice. We would therefore suggest that the limited resources available for engagement initiatives should be used to fund fewer high-quality activities and associated high-quality evaluations, rather than seeking to pursue a greater number of such activities with possibly inadequate budgets.

References

Horlick-Jones, T., Rowe, G. and Walls, J. (2007a) ‘Citizen engagement processes as information systems: the role of knowledge and the concept of translation quality’, *Public Understanding of Science*, 16(3): 259-278.

Horlick-Jones, T., Walls, J., Rowe, G., Pidgeon, N., Poortinga, W., Murdock, G. and O’Riordan, T. (2007b) *The GM Debate: Risk, Politics and Public Engagement*, Abingdon: Routledge.

Horlick-Jones, T. (2007) 'On the signature of new technologies: sociality, materiality and constructionism', in R. Flynn and P. Bellaby (eds) *Risk and the Public Acceptability of New Technologies*. Basingstoke: Palgrave.

Rowe, G. and Frewer, L. (2000) ‘Public participation methods: a framework for evaluation’, *Science, Technology & Human Values,* 25(1): 3-29.

Rowe, G., Horlick-Jones, T., Walls, J. and Pidgeon, N.F. (2005) ‘Difficulties in evaluating public engagement initiatives: reflections on an evaluation of the UK *GM Nation?* public debate’, *Public Understanding of Science*, 14(4): 331-352.

Annex A: The Observational Schedule

Observational Schedule

NB The following schedule suggests aspects to observe that are related to the 'translation' concept. The schedule is expressed in the form of various questions: the observer should seek to answer the questions and provide explanation/ evidence for their answers.

Information Comprehensiveness *(Do the sponsors provide full information to participants?)*

- Do the sponsors clearly state the aims of the event at the outset?
- Do the sponsors clearly elaborate on an agenda?
- Do the sponsors clearly explain to participants what is expected of them (defining their task)?
- Do the sponsors explain how they have selected participants/ why they are there?
- Do the sponsors explain what will follow from the event (i.e. what feedback they might expect and what will happen with the output from the event)?

Information Appropriateness/Fairness *(Do the sponsors fairly frame the problem or is there any evidence of bias in terms of information provision/ recording/ translation?)*

- At the outset, do the sponsors provide a fair summary of the subject being considered, or do they provide a particular slant, bias or frame that might lead some perspectives to be focused upon at the expense of others?
- Does the way in which information is collected suggest any particular bias (beyond, say, randomness)?
- Is the process managed in such a way that bias is introduced in terms of the information that is considered or recorded (e.g. participants with one position allowed to speak at the expense of those with another position)?
- In any summing up, is there any bias in the reporting of the output from participants?
- Is participation fair, or do some participants have much greater opportunity to speak and influence than others (whether due to facilitator bias or event logistics)?

Process Limitations to Effective Translation

- Is there sufficient time for participants to consider all the necessary information, provide all necessary information, and think about this information? Are certain debates unneccessarily cut short because of time limits?
- Are there any information resource limitations that hinder the effective consideration of the topic of debate? That is, are participants asked to discuss an issue or solve a problem on which it is clear that extra information might have been made available (report findings, academic evidence)?

- Are there sufficient resources (personnel, tape recorders etc.) to enable the full output from the event to be recorded, or do such resource/logistic deficiencies ensure that there is only a partial recording of output, or imperfect recording of information?

Information synthesis

- How is the various information outputs synthesized, and are there any apparent inefficiencies? For example, how are competing priorities compared and contrasted? How are pro and con arguments set against each other? How is such information displayed to participants – and is it in a way that may help or hinder them from synthesizing different points of view? [For example, are there whiteboard or computer screen displays of pro and con lists? Are accurate ‘minutes’ taken? Is there any form of voting process to confirm participants’ aggregate views?]

Annex B: The Questionnaire

Evaluation Questionnaire

Dear Participant,

Thank you for having taken part in the workshop. We would now like to ask you a few questions about it as part of our evaluation of this project. We would therefore be extremely grateful if you could complete this questionnaire as soon as possible, and return it to us in the FREEPOST envelope supplied with it (on the train home might be a good time!). The questionnaire should not take too long, and would be of great help to Sciencewise (and others) in terms of helping to improve events like this in future.

Finally, please be assured that your responses will be treated anonymously. Although we ask for your name and organization below, these are simply so that we can characterize those that respond to this questionnaire, and those that do not. Your name will not be cited in any evaluation report or associated with any comment you make herein.

Thanks for your cooperation.

Professor Tom Horlick-Jones, on behalf of the Evaluation Team

1. What is your name?
   ………
2. What is your affiliation?……………………………………………………………………………………………………
3. Was it clear from the information you were sent prior to the event what the workshop was about?
   Yes ☐
   No ☐
   Unsure ☐
4. At the start of the workshop, were the aims clearly specified?
   Yes ☐
   No ☐
   Unsure ☐
5. Was it clear to you from the information you were sent prior to the event why YOU were invited?
   Yes ☐
   No ☐
   Unsure ☐

6. Was it made clear to you how the participants for this event were selected?
   Yes ☐
   No ☐
   Unsure ☐

7. Do you think the audience was appropriate for this event?
   Yes ☐
   No ☐
   Unsure ☐

   If there were there any notable absentees, who were these?

   ……………………………………………………………………………………………
   ……………………………………………………………………………………………
   ……………………………………………………………………………………………
   ……………………………………………………………………………………………

8. During the event, did you have the opportunity to have your say?

I said all I wanted to say ☐

I said most of what I wanted to say ☐

I was only able to say a little of what I wanted to say ☐

I didn't get a chance to say anything ☐

9. Was there sufficient time to discuss all that needed to be discussed?
   Yes ☐
   No ☐
   Unsure ☐

10. Do you think there were any significant issues that were NOT discussed, but which should have been? What were these?

    ……………………………………………………………………………………………
    ……………………………………………………………………………………………
    ……………………………………………………………………………………………
    ……………………………………………………………………………………………

11. Were there any significant issues raised at the workshop that were not resolved? If so, what issues were these?

…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………

12. Do you think the summing-up accurately reflected what was discussed at the workshop?

Yes ☐
No ☐
Unsure ☐
There was no summing up ☐

If not, what do you think was missed or misconstrued?

…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………

13. Overall, do you think the workshop was well run?

Yes ☐
No ☐
Unsure ☐

If you said ‘no’, what was the main problem?

…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………

14. Do you expect any feedback from the event?

Yes ☐
No ☐
Unsure ☐

15. Did the event live up to your expectations?
    Yes ☐
    No ☐
    Unsure ☐
    If not, why not?

    …………………………………………………………………………………………………
    …………………………………………………………………………………………………
    …………………………………………………………………………………………………
    …………………………………………………………………………………………………

16. Do you think this event will have any influence on engagement policy?
    Yes ☐
    No ☐
    Unsure ☐

    Please explain your response.

    …………………………………………………………………………………………………
    …………………………………………………………………………………………………
    …………………………………………………………………………………………………
    ………………………………………………………………………………………………..

17. Overall, what was the best thing about the workshop?

    …………………………………………………………………………………………………
    …………………………………………………………………………………………………
    …………………………………………………………………………………………………
    ………………………………………………………………………………………………..

18. Overall, what was the worst thing about the workshop?

    …………………………………………………………………………………………………
    …………………………………………………………………………………………………
    …………………………………………………………………………………………………
    …………………………………………………………………………………………………

Once again, thank you for your time. Now please place this questionnaire in the FREEPOST envelope and post it (you do not need a stamp).