



H2020 INDEPENDENT OBSERVER'S REPORT

Call ID: H2020-S2RJU-2019

Date of evaluation: 20 June – 26 July 2019

Number of pages in this report (title page included): 13

Name of the observer: Dragana Avramov

Present at the evaluation: 22-26 July 2019

Date: 30 July 2019

TABLE OF CONTENTS

1. Introduction and the approach taken by the observer	3
2. Overall impression	4
2.1. Scale of complexity of the evaluation task	4
2.2. The allocation of external experts to proposals: balance (gender, geographic, sector), relevance and balance of expertise	4
2.3. Transparency of the procedures	5
2.4. Through output time of the process and the efficiency of the processes	6
2.5. The understanding by experts of the call (context, topics), of the evaluation process and their role and of the award criteria and scoring schemes	6
2.6. Efficiency, reliability and usability of the implementation of the procedures, including the IT-tools	7
2.7. Impartiality, fairness and confidentiality	7
2.8. Overall conduct of staff: responsiveness, hospitality, competence	7
2.9. Quality of the S2R JU evaluation process overall	8
3. Any other remarks	8
3.1. Step 1: Briefing of experts for remote individual assessment	9
3.2. Step 2: Remote consolidation of individual assessment forms by the recorder	9
3.3. Step 3: Consensus phase	9
3.3.1. Central briefing of experts	9
3.3.2. The process of the consensus meetings and the actors involved	10
3.4. Step 4: The review panel	11
4. Summary of recommendations	11

1. Introduction and the approach taken by the observer

S2R JU appointed an independent expert to observe the evaluation process from the point of view of its working and execution. Dr Dragana Avramov (Ms) monitored the process including the remote evaluations and central consensus phase. Her mandate was defined in the Expert Contract in view of informing S2R JU of the conduct and fairness of all evaluation phases of the evaluation of proposals submitted under the HORIZON 2020 Shift2Rail Joint Undertaking (S2R JU) 2019 call. She was also mandated to monitor the way in which the external experts acting as evaluators apply the evaluation criteria, and how the evaluation procedures could be improved.

H2020-S2RJU-2019 call for S2R JU members and non-S2R JU members (open call) was published on 15 January 2019 and the submission deadline was 18 June 2019. For the S2R JU members the number of topics to be addressed was 6 and the estimated S2RJU funding was € 57.5 million. The non-S2R JU members call included 12 topics and estimated S2R JU funding € 19.3 million. The evaluation process included two phases: remote individual evaluation from 20 June 2019 to 12 July 2019 and central consensus meetings from 22 July to 26 July 2019.

The independent observer was given access to all the relevant documents and could follow all the activities throughout the evaluation process. The S2R JU staff were very cooperative and acted in a fully professional way. The independent observer was given access to the online IT tool "Submission and Evaluation of proposals" (SEP) in order to be able to monitor the progress on the completion of the reports online and also examine the quality of the individual evaluation reports (IER), which were completed remotely. In addition, at the launch of the remote evaluation the observer was provided the same briefing materials as the external experts evaluators. The independent observer was present at evaluation premises during four and a half days of central evaluations. She attended the central briefing and consensus meetings for the six panels. Sometimes the independent observer attended the discussion for all three criteria (Excellence, Impact, and Implementation), and sometimes was present during only part of the consensus meeting. The purpose was to get a broadest possible insight into the way consensus meetings were moderated and how engaged were experts in discussions. This allowed for assessing also how the principles for evaluation process were applied - namely the award criteria, equal treatment of proposals, coherence and transparency. The latter were instrumental for ensuring that comments and scores were well aligned and that adequate feedback to applicants was provided.

The independent observer also had individual talks with the S2R staff regarding the different aspects of the evaluation process, and with experts involved in the evaluations. The meetings with experts took place either as individual interviews or conversation with a group of experts from the same panel. She also interviewed several independent experts who contributed as recorders/dedicated rapporteurs. The independent observer also had informal talks with several DG MOVE, European Railway Agency (ERA) and DG RTD staff that were present as observers.

This report provides S2R JU with an independent assessment of all the review phases and outcomes, and suggests recommendations on possible improvements that could be put into practice in the future. It is based on desk review of background documents, reading of randomly selected independent evaluation reports (IER), attendance of briefing meetings, attendance of

consensus meetings, and talks with S2R JU staff, independent experts, and observers. All inputs of the stakeholders are integrated in this report.

The independent observer took in particular the following aspects into consideration in this report:

- Checking the functioning and execution of the whole process involving external experts;
- Verifying compliance with the procedures;
- Advising on the conduct and fairness of the evaluation sessions, how the experts apply the criteria, and how the efficiency of the process could be improved.

2. Overall impression

2.1. Scale of complexity of the evaluation task

The evaluation process was complex as it included the call for members (CFM) and open calls for 18 different topics. All phases required a great commitment of the S2R JU staff in all the evaluation phases - from the availability request to experts, eligibility check, allocation of proposals, panel preparation and organisation, to managing consensus meetings. All these steps required an accurate timetable, coordination, clear definition of concepts and efficient execution.

Based on monitoring and observation, the independent observer confirms that the complex evaluation tasks were managed very well in full accordance with the H2020 rules and guiding principles and were completed in a timely manner.

2.2. The allocation of external experts to proposals: balance (gender, geographic, sector), relevance and balance of expertise

For the evaluation purposes proposals were grouped into 6 panels covering different IPs and cross-cutting themes considering complementarity between members proposals and open-call proposals. Six panel moderators from the S2R JU staff supported the external experts by monitoring the remote evaluation phase, managing consensus meetings and guiding the discussions during the central consolidation phase. They also provided relevant information to external experts whenever necessary.

27 external experts evaluators were selected from the EC expert database. In addition, three independent financial experts contributed to the assessment of the call for members (CFM). Six independent experts performing the task of recorders supported evaluators. These experts, also selected from the EC database, facilitated the drafting of consensus reports as dedicated rapporteurs but did not assess proposals.

The roles of experts performing the tasks of evaluators, recorders, and financial experts for CFM, were clearly defined and explained in the briefing for experts.

A broad geographic representation was ensured, as experts with a sufficiently broad range of nationalities were involved. A significant proportion had no previous experience with S2R JU

evaluations. Maximum effort was made to ensure a balanced gender representation, as 3 out of every 10 evaluators were woman. Experts came from all relevant organization types.

Each proposal has been evaluated by at least three experts. For the CFM six external experts were involved. The number of evaluators in each panel was adequate as it matched the complexity of the evaluation and the topics addressed in the call.

The level of competences and expertise of new and experienced experts that were selected was generally high and in line with the requirements of the calls. Experts involved in the panels had complementary knowledge and skills necessary for assessing each proposal.

2.3. Transparency of the procedures

Each step of the evaluation process has been carried out in a very transparent way at all levels. Experts were given access to all relevant documents prior to the launch of remote evaluations including the key reference documents, work plans, action plan, master plan and the S2R regulation. Experts were well informed about the call texts and the guide for applicants. At all stages the different guidelines and rules have been duly considered in order to get a fair and consistent treatment and evaluation of the proposals.

All the justifications and relevant elements of the individual judgement by external experts were identified and recorded online in the individual evaluation report (IER). The IERs were recorded in the online SEP IT tool.

Experts worked effectively together in the consensus phase and were fully committed to reaching agreement. The staff from S2R JU moderated the consensus meetings very effectively. They secured that the views from all experts were taken into consideration. The similarities and the differences in the assessment by the individual experts for all criteria were discussed in order to draw joint conclusions. The scores and comments agreed by the external experts were drafted by the recorder and checked and agreed upon by all the experts ensuring consistency, and reflecting the consensus reached. All the consensus reports were proof read by the S2R JU staff. This was important for achieving consistency of assigned scores with the comments, and ensuring that they were of high quality and suitable for feedback to the applicants.

The evaluation criteria were consistently applied and experts provided justification of the scores applied. All the S2R JU staff operated in a neutral way during the consensus meetings and moderated the discussions towards clear and substantiated statements securing an adequate evaluation of each proposal under each criterion.

Representatives from relevant European institutions (e.g. European Commission, the European Union Agency for Railways, the European GNSS Agency) were invited to participate to the consensus week as observers. Confidentiality was ensured as these representatives are covered by the staff regulation. Chairperson (or Vice Chairperson) of IPs/CCA Steering Committees were also invited to participate to the consensus week, exclusively in panels covering open call topics under the condition of signing the non disclosure agreement (NDA) before the start of the consensus week.

The observers from the relevant DGs and ERA provided useful information when relevant.

The transparency of the evaluation procedure and outcomes was recognized by the experts and is confirmed by the independent observer. The consensus meetings involving experts and moderators were always carried out in an open and transparent atmosphere, with a comprehensive presentation of different and specific points of view, generally clearly stated, noted by the recorder and extensively motivated in the consensus report.

2.4. Through output time of the process and the efficiency of the processes

The evaluation process was completed in line with the foreseen time-schedule. The period between the end of the remote individual evaluations and the start of central consensus gave time to recorders to prepare the “pre-consensus draft” consolidated report as a working document for structuring the consensus discussion.

It is good practice to assign the recorder the task of preparing a “the pre-consensus draft” report based on the individual assessments of experts. Some areas for further improvement of this practice are suggested in recommendations section of this report.

Consensus planning was very good. Sufficient time was allocated to consensus discussions and drafting of consolidated reports. Planning of up to 3 consensus meeting per day is an appropriate workload for experts and moderators. Contingency slots were adequately foreseen for all the panels.

Panel 5 had the largest number of proposals to process and had as many as four consensus meeting per day. All persons involved in this panel worked very efficiently and had to stay late hours to process all the proposals.

The throughput time of the process was well planned. The implementation of the evaluation process was efficient and effective.

2.5. The understanding by experts of the call (context, topics), of the evaluation process and their role and of the award criteria and scoring schemes

In general, the understanding by experts of the call text was adequate even if the call text in some cases was rather broad. The moderators provided any needed clarification and their guidance was valuable and appreciated by the evaluators.

The observers from DG MOVE, DG RTD, and ERA gave valuable inputs for clarifying the policy framework and expected impacts mentioned in the call text. Observers remained neutral and gave useful information whenever relevant. This is good practice.

However, it is noted that the H2020 scoring scale and the description of scores presents some considerable challenges for moderators and evaluators. Namely, reaching a shared understanding of the difference between “shortcoming” and “weakness”, and granularity between shortcomings is quite demanding. Furthermore, there is insufficient clarity in H2020 about “a small number of shortcomings” (score 4) and “a number of shortcomings” (score 3). In practice also reaching a joint understanding of “minor shortcomings” (score 5) is quite engaging. Consistency in the interpretations of “a small number” and “a number” was achieved

in each panel but required some considerable time. It is suggested that lessons learnt from H2020 evaluations be taken up by the EC when defining the scores for the H2020 successor programme (see R1).

The experts were well aware about the evaluation process and their role and had received extensive briefing on the call texts, criteria and scoring schemes.

2.6. Efficiency, reliability and usability of the implementation of the procedures, including the IT-tools

The implementation procedures are very well elaborated and each step is well described with clearly identified roles and responsibilities. The evaluation procedures and roles assigned to the S2R JU staff and to experts, working first on individual evaluations followed by consensus meetings, are robust and were performed with professional rigour and confidentiality.

There was one PC available for the recorder in the meeting rooms. Experts brought their own laptop. The WiFi at the evaluation facility was stable throughout the whole on-site evaluation.

The majority of experts expressed their opinion that the evaluation IT tool SEP is user-friendly and mature. The instructions for the SEP use are clear and comprehensive.

2.7. Impartiality, fairness and confidentiality

Proper guidance was given to experts on how to deal with impartiality, fairness and confidentiality matters. All external experts had signed a no conflict of interest declaration. All the observers were bound by the no disclosure rules.

The process was impartial, all proposals being evaluated with parity and following the same evaluation criteria.

During the consensus phase, the content of the proposals was discussed exclusively in consensus meetings among experts evaluating the same proposal.

No breach of confidentiality by the external experts was observed or brought to the attention of the independent observer during the central evaluations.

2.8. Overall conduct of staff: responsiveness, hospitality, competence

All staff displayed a very high degree of skill, diplomacy and competence.

The overall conduct of the S2R JU staff involved in evaluations was exemplary.

2.9. Quality of the S2R JU evaluation process overall

The 2019 evaluation of the Shift2Rail Joint Undertaking (S2R JU) call for proposals was successfully completed, fully in line with all H2020 norms and guiding principles. The evaluation process was robust, very good experts had been selected and experienced moderators were in general effective and efficient in structuring the consensus meeting without being intrusive. Recorders supported adequately the evaluators for drafting the appropriate comments and aligning the comments and scores accordingly.

The S2R JU staff have acted in a very professional way, assuring the full respect of the fairness and impartiality required. Throughout the central evaluations they provided relevant support to experts, and staff demonstrated a high level of competence and diligence. The secretariat was highly efficient in providing technical support in the most convivial manner.

Observers from DG MOVE, DG RTD and ERA remained neutral and provided valuable information whenever needed.

The outcome of the evaluation is that the best proposals were selected after a thorough and transparent assessment. Decisions were made in a fair and impartial way. No breaches of confidentiality have been observed or brought to the attention of the independent observers.

The evaluation process for the call carried out by S2R JU was of very high quality. The quality is secured by the competence and experience of the staff involved, the selection of a sufficient number of experts, an adequate balance between experienced and new experts, high level of transparency of the process, and substantive and meaningful feedback provided to the applicants.

Some recommendations for minor improvements are given in the report. They mainly relate to maintaining good practices and making improvements for increasing efficiency of the evaluation process.

3. Any other remarks

3.1. Step 1: Briefing of experts for remote individual assessment

All background information regarding the evaluation of proposals, as well as the relevant documents and information related to the calls were accessible to the experts. Experts were asked to read carefully the evaluation-related documents before starting the evaluation of proposals.

All the interviewed experts confirmed that they had all the necessary information to start and complete their individual evaluation in remote.

In general, the individual evaluation reports were of high standard although in a few cases the individual evaluation reports were not fully comprehensive and did not include all factors to be considered under each criterion by highlighting strengths and weaknesses.

It was observed that some experts wrote too brief and generic comments. In some cases they did not cover all the aspects under criteria. Lack of comprehensive coverage of all aspects had

as consequence lengthy discussions at consensus meetings where experts provided more complete feedback.

The new to S2R JU interviewed experts reported that they would have appreciated some feedback and quality check of the first assessment uploaded in the IT tool SEP. Such feedback can set standards and provide clarity about expectations towards experts regarding coherence, comprehensiveness and quality of comments for all assessments that follow. This is relevant because the quality and effectiveness of the consensus discussions build on the quality and completeness of individual assessments (see R2).

3.2. Step 2: Remote consolidation of individual assessment forms by the recorder

The practice introduced by S2R JU of engaging recorders, independent experts with the assigned role of dedicated rapporteurs, is good EC practice that proved to be efficient and effective providing targeted briefing for recorders is sufficiently comprehensive (see R3).

The tasks of drafting a working document for each proposal based on the individual assessments prior to the central phase in Brussels was useful. This “pre-consensus draft” was a particularly helpful document in the event that there was a high degree of agreement among the three experts about the strengths and weaknesses of a proposal.

However, in the majority of cases a degree of divergence among experts did not allow for a coherent consolidated text. A “cut and paste” of conflicting views from individual reports was not useful. The “pre-consensus draft” is excellent tool for the moderator to guide the consensus discussion and ask the most relevant and pertinent questions providing that it is structured according to discussion points rather than as a list of divergent views (see R4).

Access to individual reports was given to recorders approximately a week before the central phase. This was appropriate time for all but one panel. Recorder for panel 5 having to draft 15 consolidations had a somewhat challenging task.

3.3. Step 3: Consensus phase

Consensus meetings were face-to face between the external experts. S2R JU staff moderated the meetings. In most cases an observer from DG MOVE, DG RTD and ERA were present.

3.3.1. Central briefing of experts

The central general briefing for experts was excellent. The briefing material was well tailored and was informative on all the key aspect.

3.3.2. The process of the consensus meetings and the actors involved

The consensus meetings were well organised, well scheduled and explained.

The consensus meetings were chaired by the S2R JU moderators. They ensured a fair and equal

treatment of the proposals and that the views from all experts were taken into consideration. The moderators never attempted to influence the opinion of the evaluators and did not express their views about the merits or deficiencies of any proposal. The moderators focussed on guiding the discussion. They also provided clarity on the call text and secured that the call text was taken sufficiently into consideration during the evaluation. Moderators remained neutral while providing important information to the experts.

Moderating consensus meetings is a difficult and complex task requiring topical knowledge, several skills (e.g. communication and social skills) and experience. Besides, the aspects mentioned above some important tasks for the moderators are to make sure that the time-schedule is kept, all experts are involved in the discussions and to interfere when the discussion will not advance.

The moderators were generally very well prepared and motivated for the task and the discussions were well structured.

The best practice observed was when the moderator chaired and manages the discussions throughout the consensus meeting. The floor was given to the recorder only to summarize the key points of agreement and suggest the phrasing. This practice prevented giving too much prominence to the recorder in managing the discussions.

The process for the consensus meetings was most effective when the moderator started the meeting by asking each expert for their general view of the proposal without revealing the scores. Best practice was when experts summarized in no more than two minutes the key strengths and the key shortcomings as this contributed to a joint understanding of the overall value of the proposal.

It was also good practice in some panels to give the floor to the observers from DG MOVE, DG RTD and ERA who were present to situate the discussion in the policy context of the topical area.

After this, the different criteria were discussed and scores were given. The time scheduled for each consensus meeting was limited to 1 hour and 45 minutes. The more experienced moderators allowed thorough discussions amongst the experts but did steer the meeting in such manner to avoid lengthy and unfocussed debates.

Moderators ensured fair and equal treatment of the proposals. They never attempted to influence the opinion of the evaluators as they did not express views about the merits or deficiencies of any proposal. They focused on guiding the discussion, asking questions, and seeking consensus between the external experts. Moderators remained neutral while providing information if useful. They also ensured consistency in the evaluation of different proposals.

The practice of some recorders after the consensus discussion to go back to individual scores before agreeing consensus score was not efficient. It contributed to “negotiating” based on initial reading of the proposal rather than proposing a score in the light of the new information coming from all the experts. In the end the outcome of discussions was a consensus but it caused unnecessary slowing down of the alignment of comments and scores (see R5)

Some recorders used terms “seems”, “appear”, “it is difficult to assess”, and indirect

recommendation such as “would have benefitted from” putting an unnecessary burden on the S2R staff carrying out proof reading (see R6).

Formulating the overall comment in the consensus evaluation template was time consuming and of little use. These comments were either repetitions of what was already commented under the evaluation criteria or introduced some new elements not clearly reflected in the evaluation under relevant criterion and as result were not adequately explaining and justifying the overall score (see R7).

The evaluation template includes the following sub-criterion for IMPACT: “Any substantial **impacts not mentioned in the work programme**, that would enhance innovation capacity, create new market opportunities, strengthen competitiveness and growth of companies, address issues related to climate change or the environment, or bring other important benefits for society”. This sub-criterion often creates inconsistencies in the evaluation because these impacts are not specifically required in the call text.

It has become good practice to mention them only if there is convincing evidence in the proposal that additional impacts can effectively be achieved, but not to penalize the applicants if they are not proposing to achieve something that they were not explicitly required (see R8).

3.4. Step 4. The review panel

Review panels were a straightforward exercise because there was no competition for financial resources between the topics. There was one ranking list per topic. At the review panels there was no change in scores. Only minor fine tuning of comments was implemented to ensure full coherence for all proposals under each topic. The six panel review reports provide specific details about these minor changes that have not had impact on the ranking of proposals.

4. Summary of Recommendations

Recommendation for the post H2020 programme

R1 - A clearer definition of the terms used to describe scores such as “shortcoming” and “weakness” and better granularity for scores 3, 4 is necessary

Recommendation to S2R JU

R2 - It is suggested that moderator provides feedback to experts on the quality and completeness of evaluator's first individual evaluation report uploaded in SEP. Following questions could be covered: Has the expert observed coherence between comments and score given? Do the expert comments sufficiently take into account the guiding questions of the criterion? Is the quality of the assessment satisfactory both language-wise and content-wise? If "no" S2R JU staff could comment in SEP Task comments on what needs to be improved and what needs to be taken on board in the assessments that will follow.

R3 - Specific briefing for recorders could be elaborated by S2R JU to fully define their role and method of work.

R4 - Recorders could be guided to structure well a “pre-consensus draft” based on individual evaluation reports submitted in SEP. It is best practice to first list points of agreement for which little discussion is necessary. Possible diverging views need to be summarized and formulated as discussion points.

R5 – Recorders could be instructed to focus on assisting the panel in aligning the agreed consensus comments with the proposed score without going back to the individual scores to look for an “average” score.

R6 – Recorders could be more firmly instructed to avoid using terms such as “seems”, “appear”, “it is difficult to assess”, and indirect recommendation such as “would have benefitted from” in order to diminish the work load of the S2R staff carrying out proof reading.

R7 – S2R JU could consider instructing experts not to provide the Overall comment in the evaluation template.

R8 - S2R JU could consider advising experts to comment additional impacts not listed in the call text only if there is evidence in the proposal that there is a credible potential for achieving them. If not a standard sentence “No additional impacts are convincingly presented in the proposal” could be inserted in the template box. This should not result in penalizing the proposal as only expected impacts listed in the call text are required.