



Horizon Europe

Observer Report

25 July 2022

**CONFIDENTIAL****OBSERVER REPORT**

CALL	
Call:	HORIZON-ER-JU-2022-01
Topics:	HORIZON-ER-JU-2022-FA1-TT-01 HORIZON-ER-JU-2022-FA2-01 HORIZON-ER-JU-2022-FA5-01 HORIZON-ER-JU-2022-FA6-01 HORIZON-ER-JU-2022-FA3-01 HORIZON-ER-JU-2022-FA4-01
Types of actions	IA (Innovation Actions)
Type of Grant	Horizon Lump Sum Grants
Service:	EUROPE's RAIL Joint Undertaking (ER-JU)
Opening / deadline dates:	10 March 2022 / 23 June 2022


EVALUATION	
Type:	Single stage
Panel(s):	Panel 1 (topic HORIZON-ER-JU-2022-FA1-TT-01) Panel 2 (topics HORIZON-ER-JU-2022-FA2-01 and HORIZON-ER-JU-2022-FA6-01) Panel 3 (topic HORIZON-ER-JU-2022-FA3-01) Panel 4 (topics HORIZON-ER-JU-2022-FA4-01 and HORIZON-ER-JU-2022-FA5-01)
Observer:	Jo Prieur 

TABLE OF CONTENTS

1. BACKGROUND AND SCOPE	3
2. OBSERVER ASSESSMENT	4
Methodology	4
Assessment	5
Recommendations	13

1. BACKGROUND AND SCOPE

Background and scope

This report describes the observer's assessment of the evaluation of the following call:

Call for proposals: HORIZON-ER-JU-2022-01

Opening: 10.03.2022

Deadline: 23.06.2022

Budget: € 234 millions

This call covers the following topics /types of actions: HORIZON-ER-JU-2022-FA1-TT-01 (IA)

HORIZON-ER-JU-2022-FA2-01 (IA)

HORIZON-ER-JU-2022-FA5-01 (IA)

HORIZON-ER-JU-2022-FA6-01 (IA)

HORIZON-ER-JU-2022-FA3-01 (IA)

HORIZON-ER-JU-2022-FA4-01 (IA)

The observer report analyses the efficiency of the procedures, usability of the instruments (including IT tools), conduct and fairness of the evaluation sessions, and compliance with the applicable rules.

The objective is to give independent advice for improving the evaluation processes for EU funding.

The observer is NOT an evaluator. He does not express any views on the proposals under evaluation, nor on the opinion or the qualification of the experts (the evaluators).

The independent observer Jo Prieur has been retired from ONERA, the French national aerospace lab, since June 2011. He has no connexion whatsoever with the call under evaluation, nor with the proposal(s) being evaluated, nor with any of the applicants. He has a wide experience of EU research framework programmes and proposal evaluations, first as END/SNE in DG RTD and DG GROW (then DG ENTR) from 2001 to 2005, and, from 2006, as an independent expert (evaluator, rapporteur, and observer), on several themes (Aeronautics, Surface Transport, Security, Space, Clean Sky JU, S2R JU, BBI JU, NMP, EIT, Metrology), in various places (primarily Brussels, but also Luxembourg, Budapest, Rotterdam, Monaco), for FP6, FP7, Horizon 2020 and Horizon Europe programmes.

The call covers the following topic/type of action:

Topic code & title	Type of action and funding	Number of proposals (eligible/received)	Budget (EU Contribution)	Expected EU contribution per project (EUR million)	Number of projects expected to be funded
HORIZON-ER-JU-2022-FA1-TT-01 Network management planning and control & Mobility Management in a multimodal	Innovation Action (IA) Lump Sum	1/1	€ 38 Millions	€ 38 Millions	1

environment and Digital Enablers 01					
HORIZON-ER-JU-2022-FA2-01 Digital & Automatic up to Automated Train Operations	Innovation Action (IA) Lump Sum	1/1	€ 54.3 Millions	€ 54.3 Millions	1
HORIZON-ER-JU-2022-FA3-01 Intelligent & Integrated asset management	Innovation Action (IA) Lump Sum	1/1	€ 46.3 Millions	€ 46.3 Millions	1
HORIZON-ER-JU-2022-FA4-01 A sustainable and green rail system	Innovation Action (IA) Lump Sum	2/2	€ 38.3 Millions	€ 38.3 Millions	1
HORIZON-ER-JU-2022-FA5-01 Sustainable Competitive Digital Green Rail Freight Services	Innovation Action (IA) Lump Sum	1/1	€ 40.6 Millions	€ 40.6 Millions	1
HORIZON-ER-JU-2022-FA6-01 Regional rail services / Innovative rail services to revitalise capillary lines	Innovation Action (IA) Lump Sum	1/1	€ 16.5 Millions	€ 16.5 Millions	1
Total		7/7	€ 234 Millions		6

All topics generated 1 single proposal with the exception of topic HORIZON-ERJU-2022-FA4-01 which generated 2 proposals. All 7 proposals received in response to the call were considered eligible.

Each of the 7 proposals was evaluated by a group (panel) of 5 or 6 evaluators. Each panel had its own dedicated recorder (who is NOT and evaluator). Internal quality checkers (from within EU-Rail JU team) assisted the panels for iterative CR quality checks (several exchanges between the QC and the panels, via the moderators)

There were 4 expert panels (Panel 1 for topic HORIZON-ER-JU-2022-FA1-TT-01, Panel 2 for topics HORIZON-ER-JU-2022-FA6-01 and HORIZON-ER-JU-2022-FA2-01, Panel 3 for topic HORIZON-ER-JU-2022-FA3-01, and Panel 4 for topics HORIZON-ER-JU-2022-FA4-01 and HORIZON-ER-JU-2022-FA5-01)

The consensus and panel phases of the evaluation were performed 100% remotely (virtual central meetings), in contrast with local consensus/panel phases generally held in Brussels before the COVID outbreak.

2. OBSERVER ASSESSMENT

Methodology

Methodology

The approach followed by the observer was in line with the requirements of the ER-JU evaluation team and with the observer contract and its annex. All relevant information was made available to the observer in a timely manner.

The observer was involved in the evaluation process after the entry into force of his contract CT-EX2002B070194-158 on 17 June 2022 and was given access from that date to all relevant documents and to the SEP system. The observer work consisted in the following activities:

Review general and specific Horizon Europe documents and relevant evaluation rules such as:

1. The Horizon Europe programme and rules of participation (<https://eur-lex.europa.eu/eli/reg/2021/695/oj>)
2. The Horizon Europe Programme guide (https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf)
3. The Europe's Rail Master Plan https://rail-research.europa.eu/wp-content/uploads/2022/03/EURAIL_Master-Plan.pdf
4. The Europe's Rail Joint Undertaking Multi-Annual Work Programme dated 1 March 2022 https://rail-research.europa.eu/wp-content/uploads/2022/03/EURAIL_MAWP_final.pdf

5.	The Europe's Rail Work Programme 2021-2022 adopted by the EURO Rail Governing Board on 1 March 2022 https://rail-research.europa.eu/wp-content/uploads/2022/03/AWP_2022_2024_Final_Published.pdf
6.	The Questions and Answers (Releases 1 to 8) published during the proposal preparation period from the call opening date (10 March 2022) to the deadline date (23 June 2022)
7.	The Work Programme General Annexes (https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/wp-call/2021-2022/wp-13-general-annexes_horizon-2021-2022_en.pdf) especially those related to the award criteria (Annex D) and to the evaluation procedure and ranking (Annex F), complemented by Annex 8 of the Europe's Rail Work Programme 2021-2022 providing additional award sub-criteria
8.	The EU Funding & Tenders online Manual (https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/common/guidance/om_en.pdf) especially section 3.2.2 dealing with Evaluation
9.	The "Decision authorising the use of lump sum contributions under the Horizon Europe Programme" https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/lis-decision_he_en.pdf
10.	The standard application form https://rail-research.europa.eu/wp-content/uploads/2022/03/Application-From-Part-B-HE-ER-IA.pdf and evaluation form https://rail-research.europa.eu/wp-content/uploads/2022/03/Evaluation-Form-IER-CR-ESR-HE-.pdf utilised by EU-Rail JU for IAs
11.	The standard observer report template to prepare the present document

Review the call text for the call under consideration (topic description, destination, expected impact for each of the 6 topics) and applicable call conditions

Attend a specific Observer briefing delivered by the EU-Rail evaluation team on 7 July 2022

Review other briefing material provided to the evaluators and recorders by the EU-Rail evaluation team at the start of the individual evaluation phase on 27 and 29 June 2022 (Ethics briefing, Evaluator Briefing, Recorder briefing)

During the Individual Evaluation Phase, check on a regular basis (approximately daily or every second day) the progress of the evaluation for individual experts (percentage of IER completion)

During the initial drafting of CRs by recorders, perform random checks of progress on SEP

During the remote consensus and panel phases (13 to 19 July 2022), attend and observe the initial consensus briefing presentation, the consensus meetings and the elaboration of the consensus reports (from the initial draft CRs prepared by the recorders until the finalized version) and panel meetings for the different expert panels and different topics. As several meetings were taking place in parallel, the observer had to make choices and tried to cover partly or fully all 4 moderators

Attend a specific "lessons learned" debriefing meeting on 18 July 2022

Review additional relevant material provided to the experts such as Person Month statistical data, for each topic, to assist the experts in making a financial assessment of the costs estimated by the applicants (specific for lump sum grants)

During the whole evaluation period have various exchanges with the call coordinator, to raise questions, seek clarifications and obtain clear and prompt answers

Prepare and submit the present report to the EU-Rail Team

Assessment

Assessment
Scale of complexity of the evaluation task
<p>With a total of 7 proposals received in response to the call and to be evaluated, the present evaluation may sound relatively simple in terms of magnitude and workload. However, the evaluation required a lot of effort and special attention for the several reasons:</p> <ul style="list-style-type: none"> The proposals are submitted in response to topics which cover 6 of the 7 so called "Flagship Areas" (FAs) of the EU-Rail Innovation pillar + Transverse Topic (TT). These FAs are covering at large key subsystems of the rail system, multidisciplinary, and call for significant expected impacts, thus requiring an integrated sector systemic approach. The proposals submitted in response to these integrated subsystems' topics are necessarily complex and large in terms of durations (48 months), requested EU budgets (16 to 54 million), range of activities (large number of work packages), number of participants (in the order of hundred partners and/or affiliates) and deliverables. An illustration of the recognized complexity and wide range of activities to be covered in these proposals is their allowable size which was increased to 120 pages, in contrast to the more standard

and simpler Innovation Action proposals which are normally limited to 45 pages under Horizon Europe. As such, each of these proposals may be seen as an integrated sub-programme rather than a narrow and focussed project. This is also one of the reasons why only one proposal per topic is expected to be funded. This encouraged stakeholders of the rail sector and beyond to join forces to put together and submit a comprehensive proposal with the main objective to deliver impact in accordance with the SBA.

- Compared to more standard Innovation Actions (IAs) under Horizon Europe programme, the EU-Rail IAs of this call have additional award sub-criteria for each main criterion, as a result of a Commission's Internal Audit Service recommendation. One particularly important example is the quality of the management structure under criterion 3, which has been re-introduced as a sub-criterion. Indeed, the management of such big projects with consortium size in the order of 100 participants (partners and affiliates) has to be carefully organized.
- In contrast with more conventional types of grants with payments made as reimbursements of actual costs, the lump sum approach pushes the applicants to breakdown their work into a larger number of packages and activities with well identified tasks, milestones and deliverables whose completion will be the basis for payments, privileging impact and delivery rather than focusing on the administration of costs.
- The applicable funding scheme and related grant type (lump sum grant LSG) have been utilised previously, as pilot, under Horizon 2020, including also for S2R projects, but never as widely as they appear to be utilised now under Horizon Europe, nor for such very large projects. The lump sum scheme is deemed to provide some administrative and financial simplification in the management of the funded projects, but requires additional attention to the coherence between resources and activities, and to financial aspects at the proposal evaluation and grant preparation stages. The present call is the first EU-Rail call of the lump sum type launched under Horizon Europe and the approach to deal with this aspect at the evaluation stage is in itself a novelty. It has also to be noted that the LSG approach is implemented in the context of delivery of in-kind contributions by the members of the JU participating to the actions; in this respect, the members have obligations coming from the SBA that complement the LSG and address key assurance aspects.
- As an exception to the "standard" evaluation process, in which experts are requested to refrain from formulating recommendations, the lump sum funding scheme evaluation process invites experts, when they judge it necessary, to recommend (and hence justify) budget modifications and/or reallocations, possibly leading to a decrease (and not an increase) or a redistribution of the total requested budget between work packages and / or partners. Therefore, the lump sum scheme requires that evaluators scrutinize in more detail the relation between resources (financials and others) and results to be achieved, including the rather detailed budget table requested from the applicants (which is not normally required for the more conventional funding schemes) and many other data under the several hundred pages of part A of the proposals.
- To some extent the lump sum approach departs from the conventional "No Negotiation, therefore No Recommendation" motto of the previous programme Horizon 2020, by creating a new approach in which the motto becomes "Recommendations OK, but still No Negotiation" ...which makes the whole process more difficult

Therefore, the combined specificities of the EU-Rail programme and of the lump sum funding scheme contributed to make this evaluation a delicate and complex exercise, requiring a lot of effort and detailed attention.

Transparency of the procedures

The procedures for the evaluation of the proposals were transparent to all participants (evaluators, recorders, and observer) who had access to all publicly available information, primarily on the EU Funding and Tenders Portal and on the Europe's Rail Joint Undertaking (EU-Rail) web site.

Each type of participant was delivered its own specific briefing in due time for the evaluation (Experts' briefing, Recorders' briefing, Observer's briefing) and another comprehensive Consensus briefing on the first day of the consensus phase.

In addition to the general information on the call itself and its content, these briefings contained all relevant information for the evaluation: planning of all phases (individual evaluation, CR drafting, consensus and panel phases), evaluation criteria and sub-criteria, scoring system and score interpretation table, tips for writing high quality reports (IERs and CRs)

In addition, the specificities of the lump sum funding scheme were clearly explained to all participants

All external participants (experts, recorders, observer) were also made aware about transparency requirements and absence of conflicts of interest (Col) from the moment they sign their contract. By signing their contracts (and its annex about the code of conduct they have to abide by), they do recognize that they are aware of these requirements and that, to the best of their knowledge, they have no Col. Observers coming from other European institutions do not sign a specific contract for this evaluation but are covered by Staff regulations in terms of Confidentiality.

The presence of an independent observer, and his freedom to raise any questions to the evaluation staff is also a clear indication of the willingness to run the evaluation exercise in full transparency. All questions from the observer always received prompt and clear answers, without any restrictions whatsoever. In addition, for transparency and

openness reasons, the experts were clearly informed that they could raise their concerns, if any, with the observer via e-mail (due to the entirely remote character of the evaluation), should they wish to do so.

The observer is fully convinced that all applicable procedures and rules, either general Horizon Europe rules, or related specifically to this EU-Rail call, including those related to the lump sum funding scheme, were clear and transparent to all experts and the observer, and that the evaluation process was conducted to high standards of diligence, fairness and transparency throughout the entire evaluation exercise

Throughput time of the evaluation and the efficiency of the procedures

The overall evaluation process took place from 30 June 2022 until 19 July 2022. This was quite a sufficient period to cover all phases of the evaluation process (IER preparation, CR drafting, Consensus meetings, quality checks, and panel meetings) for the 7 proposals being evaluated, and for experts to complete their work. And indeed, the evaluation exercise was shorter than the time allocation originally foreseen.

Each evaluator had 1, or 2, or maximum 3 IERs to prepare over the individual evaluation phase (over a period of 6 working days), which is, in principle, a very reasonable timeframe, even for those experts who do have a job. However, this work load may be somehow on the high side for those who were given 3 proposals to be evaluated considering the large size of the proposals (120 pages for part B, a huge part A, a detailed budget table), their complexity, and the consideration of the lump sum scheme specificities.

The recorders had 1, or 2, or maximum 3 CRs to draft within 3 working days, again quite reasonable on average, but possibly on the high side for the recorder who had 3 CRs to prepare considering that the number of IERs to be taken into account for each proposal was high (5, and even 6 in one case).

The work of the Quality Checkers was not directly visible to the observer, but considering the number of iterations needed between the quality checkers and the expert panel (via the moderator) and the waiting time, in some cases, before QC comments were made available to the expert panels, it appears that the QC process represented a high workload.

As indicated above, with only 7 proposals, the overall consensus phase was shorter than originally planned. Considering the specificities of this EU-Rail call with large and complex proposals, and the specificities of the lump sum scheme requiring a particular attention to the financial dimension of the proposals at the evaluation stage, the observer considers that the efficiency and the "productivity" of the evaluation process were reasonable, but attention should be paid in the future that the QC step, with multiple iterations, does not become a bottleneck in the evaluation process.

On the basis of other Horizon Europe call evaluations attended recently by the observer, EU-Rail may want to consider, for the future, the option of having quality checkers participating directly into the consensus meetings. This may contribute to save some iterations between, QC and evaluators. However, the experience shows that this requires a clear understanding by all participants that quality checkers are NOT evaluators (just like recorders). It should be noted however that in these other evaluations, quality checkers were external experts selected primarily for their writing skills and therefore their QC work was limited to the quality of the wording

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

The SEP system appears generally operational and reliable, and the experts involved in the present evaluation appeared familiar with the system and did not seem to have any problem using it.

The MS Team platform used in the remote consensus phase appeared also to be familiar to all participants, and there was no major technical issue with the use of the system. Occasionally, there were some minor connexion problems for some participants but these small issues did not have any adverse impact at all on the evaluation process. Most of the participants did switch off their microphones and cameras whenever they did not want to speak, in order not to deteriorate the quality of the exchanges.

Overall, the observer believes the procedures used in the present EU-Rail evaluation are reliable and robust and the implementation of these procedures was good.

Impartiality, fairness and confidentiality of the evaluation

From his attendance to the consensus and panel phases, the observer is convinced that the evaluation was impartial, fair and conducted in full compliance with all general Horizon Europe rules, taking also into account the specificities of this EU-Rail call (for example, additional evaluation sub criteria), and specific lump sum aspects of the call, i.e. the assessment of the detailed budget table requested from the applicants and possibly recommendations for budget modifications (overall reduction and/or redistribution over the different work packages or partners). These rules and procedures were known to all experts and were reminded to them as part of their briefing.

The early detection and treatment of potential conflicts of interests contributes to achieve impartial and fair evaluations. All experts, and the observer, sign a declaration of absence of conflict of interest as part of their contracts. The importance of avoiding conflicts of interest is also reminded to experts during briefings. Conflicts of Interest are detected in general before the start of the evaluation, but can be uncovered during the evaluation. Experts with a conflict of interest with one proposal are excluded from participating to the evaluation of the proposal

in question. They are not necessarily replaced as long as the minimum number of evaluators (>3) are present. Other observers from EU institutions are bound by Staff regulations concerning Confidentiality.

The overall evaluation process was fully compliant with the principles established by the Commission: independence, impartiality, objectivity, accuracy and consistency.

The involvement of several participants to the evaluation process for each proposal, having well defined roles (5 or 6 experts/evaluators, 1 recorder, 1 independent observer, the moderator, other observers) is believed to contribute significantly to achieve impartial and fair evaluations. With all these participants, often participating together in consensus and panel meetings, it is almost impossible that any deviation from the applicable rules would go unnoticed by at least one of these participants. In other words, the simultaneous participation of all of them reinforces the fairness of each of them.

Conformity of the evaluation with the applicable rules (including guidance documents)

The evaluation process was fully in line with the applicable rules given in the general annexes of the Work Programme, especially those of particular relevance such as annex D (award criteria) and F (evaluation procedure) and complied with the principles established by the Commission and reminded to experts during their briefing (independence, impartiality, absence of conflicts of interest, objectivity, consistency, accuracy).

In addition to the above-mentioned general procedures, the EU-Rail JU evaluation team prepared and delivered to the experts and observer detailed and comprehensive briefing material concerning the context and the content of the call and, within the call, of each of the 6 topics. The attention of the evaluators was also drawn to the Annex 8 of the Annual Work Programme detailing the specific award sub-criteria, additional to those of the general Annex D. A specific briefing part dedicated to the lump sum scheme was also delivered to the experts

Quality of the evaluation process in comparison with similar national/international evaluation procedures

The observer has never taken part in an evaluation process other than EU research ones. However, on the basis of earlier discussions with several experienced evaluators over different calls of the previous H2020 programme, it is generally recognized that the EU evaluation process is viewed as robust and of high quality and can be generally considered better than or as good as other national and/or other international research funding schemes. Several countries are known for being inspired by the EU evaluation methodology to improve their own national process.

At international level, the main evaluation principles are: objectivity and impartiality, independence of evaluators, participation of all parties concerned in the entire process, transparency and focus, reliability, consistency, completeness and clarity of reports, fairness and protection of the interests of the parties involved.

The observer believes that all the above-mentioned principles are well embedded and efficiently implemented in the EU evaluation process.

Overall quality of the evaluation

The observer believes the evaluation was conducted in a transparent and open way and all proposals received very detailed attention and fair treatment, in compliance with the applicable rules and procedures.

The observation of the IER phase was rather limited, and reduced almost to the follow up of progress on SEP at regular intervals (approximately on a daily basis or every second day) and as such is not very different from the situation prevailing during pre-covid evaluations.

For the consensus and panel phases, the remote character of the evaluation is indeed a constraint that all participants have to live with. Nevertheless, from discussions, telephone conversations, exchange of mails with the EU-Rail call coordinator and some moderators, attendance to briefings, reading of relevant documentation, and full-time observation of the "virtual-central" consensus and panel phases, the observer is convinced that the quality of the evaluation was excellent

Throughout the entire process, the EU-Rail evaluation staff demonstrated their commitment to transparency and adherence to all applicable rules and procedures. The professionalism and availability of all participants contributed greatly to the smooth and efficient running of the evaluation process.

The observer considers that the whole evaluation process was conducted by expert evaluators and EU-Rail staff with the highest professional and quality standards and in accordance with the rules and guidelines for the HORIZON Europe programme, as well as specific EU-Rail JU rules whenever applicable

Other remarks

- *quality of the documentation provided to experts beforehand*

The independent observer had his own observer briefing delivered to him on 7 July 2022. On the same day he was also given access to the briefings delivered to the other external participants (expert briefing, recorder briefing, ethics briefing) on 27 and 29 June 2022 just before the start of the individual evaluation phase (30 June 2022). These high quality and comprehensive briefings covered general information about the EU-Rail JU and the EU-Rail call for proposals 2022, explaining in detail the hierarchy of the key documents (Regulation, Master Plan, Multi Annual Work Programme, Annual Work Programme), the overall EU-Rail Programme and major expected Impacts, the call 2022 context, content and conditions, and the evaluation process with the role of each type of participants and the sequence of the process (individual, consensus, and panel phases). In addition, it contained very relevant and useful information and guidance about the reports (IERs and CRs) and the required quality of these documents as well as some writing tips, the evaluation criteria and sub-criteria, the scoring system and score interpretation table, etc.

A specific part of the briefing was also dedicated to the lump sum funding scheme applicable to this call and what was expected from the experts in that respect (are the resources and lump sum shares adequate to cover the activities and expected outputs? Are the budget estimates accurate and in line with proposed resources? Are applicants going to use their own accounting practices? Is the budget correct on the basis of statistical financial data?) Specific information (such as statistical financial information on personnel costs per partner and per country) was provided to the experts to help them make their financial Assessment

All experts and the observer had also access to the usual documents that applicants had been using to prepare their proposals, available on the "Funding & tender opportunities portal" (such as the applicable work programme and the relevant general annexes, the standard application forms, the standard evaluation forms, usable for both IERs and CRs).

The observer is convinced that all relevant information was made available to all those who "need to know" and that the information was clear, comprehensive and of high quality

- *quality of the on-site briefing sessions*

In addition to the above-mentioned initial briefings, virtual on-site briefings were delivered to the experts on the morning of the very first day of the consensus phase (13 July) with all necessary information for the evaluators and recorders to perform their work during the consensus and panel phases. A representative of DG MOVE participated in the briefing to give an overview of EU priorities in the rail sector and therefore the context of the call. Again, all relevant aspects (already mentioned above) were addressed during these briefings, with a particular emphasis on the confidentiality requirements, the Conflict-of-Interest situations, and all major principles of the process (impartiality, objectivity, accuracy, independence, consistency). The briefing covered again all relevant aspects related to the lump sum scheme utilised for this call, with indications as to where should these aspects be covered in the evaluation forms (under criterion 3, or under overall comments or in the minutes of the consensus meetings).

This was also an opportunity for experts to ask any remaining questions before starting the consensus phase. For example, some experts outlined the difficulty to assess properly the role of the affiliates on the basis of the limited information available in part B of the proposal, even with 120 pages, while others were unclear as to how to use the provided statistical financial data to assess the correctness of lump sum budgets.

Specific information related to expert payment was also delivered

The briefing was again of high quality, complete, clear and comprehensive

- *the understanding by experts of the call (context, topics), of the evaluation process and their role and of the award criteria and scoring scheme*

All experts were perfectly aware of the call and its 6 IA topics and of the specific nature of the applicable funding scheme (lump sum). They understood clearly the award criteria and the scoring scheme. They all were fully aware of what was expected from them, how important their work was in relation to the decision of funding or not the proposed projects, the budget available for each topic, the recommended requested budget per project and therefore the expected number of projects to be funded for each topic (1 per topic)

The need to evaluate the proposal as written and to avoid recommending any substantial change or improvement to the proposal was well understood by evaluators, with the special exception of possible recommended changes to the budget amount and distribution, due to the nature of the funding scheme (lump sum).

Nevertheless, some experts did not feel very at ease with the task of assessing how the allocated resources were in line with the proposed activities and whether the required budget per participant and per work package was reasonable and in line with the allocated person months. A few experts expressed the concern that the additional allowed number of pages in the proposal (up to 120), as well as the vast amount of data contained in Part A of the proposal, in the budget details, and in the statistical financial data provided to them, did not allow them to formulate easily an informed judgement about these resources and

financial issues. In addition, not all experts seemed to have the same understanding as to where in the CR (under criterion 3, or in the minutes of consensus meeting, or in “overall comments”) should some of the lump sum issues be addressed. As these elements (already explained to experts in their initial and pre-consensus briefings) concerned only a small number of experts, it didn't prevent each panel to reach successfully the necessary consensus with the expected assurance. Nevertheless, those points could be taken into consideration by the JU as lesson learned in order to improve the process for the experts and the future LS call evaluation.

Other important aspects well understood by evaluators were the need to comment before scoring, avoid penalizing twice (or rewarding twice!) the proposal for the same reason under two different criteria or sub criteria, pay attention to the quality of the consensus report in order to (i) provide clear feedback to the applicants and (ii) avoid wording which may generate claims, and ensure consistency of the comments with the scores.

One aspect which would deserve maybe to be more strongly emphasized in briefings is that, although each of the many participants has his/her own specific role (evaluators, recorders, quality checkers, moderators, independent observer, other observers), the quality of the CR, eventually, is everybody's business and NOT the exclusive job of the rapporteurs and quality checkers. The presence of quality checkers should not be considered as an excuse for evaluators for not formulating properly their comments nor considering that quality is entirely somebody else's business!

- *the allocation of experts to proposals: balance (gender, geographic, sector), relevance and balance of expertise, ...*

For the evaluation of the 7 proposals of this call, 54 experts (evaluators and recorders) were selected. These experts were selected primarily on the basis of their expertise and their coverage of the various facets of the call and all the aspects to be taken into account in the various criteria and sub-criteria, while trying to maintain a proper gender balance, and balanced geographical coverage and type of organisations

1/3 of the experts were women. 83% were first time experts (never contracted before under Horizon Europe), and 17% were experienced experts already involved in Horizon Europe evaluations. Experts were coming from about 20 different countries, most of them being EU Member States. The vast majority of experts (about 75%) were originating from ES, NL, IT, GR, BE, UK and IR.

About 1/3 of these experts were from private for-profit organisations, 1/3 from higher or secondary education organisations and the last 1/3 was shared between public organisations, research organisations or others.

The expertise range covered innovation (11%), SSH (24%), business (20%), gender (13%), and other aspects (32%).

Some experts expressed the view that formulating well informed judgements about the cost estimates and their coherence with the statistical data provided to them was somehow beyond their knowledge and competence; nevertheless, the experts were able to complete successfully the CR and express an overall assessment, and asked the JU to conduct some verification during the GAP phase.

- *the process of the individual evaluations and the actors involved*

The individual evaluation took place after the initial briefings to the evaluators and recorders on 27 and 29 June 2022, starting on 30 June 2022, until 7 July 2022. This was followed from 8 to 12 July 2022 by the period when IERs were reviewed by EU-Rail staff and initial CRs were drafted by dedicated rapporteurs. Over that period the observer noted on SEP at regular intervals the smooth progress towards finalizing the IERs.

As already mentioned, in view of the small number of proposals, the duration of this phase was adequate, but for those experts who had 3 IERs to prepare the workload was significant considering the size and complexity of the proposals (120 pages in part B, large amount of information to be reviewed in part A, budget, financial information)

- *the process of the consensus meetings and the actors involved*

The consensus phase was initially scheduled from 13 to 19 July 2022 but ended up being shorter than planned, in view of the small number of proposals to be evaluated. Yet some of the meetings were very long not only due to the extensive discussions between evaluators but also in view of the number of Quality Check iterations between quality checkers and the evaluator panels. Indeed, there were cases when meetings were somehow disrupted while waiting for QC comments to come back to the evaluators, which may be somehow detrimental to the fluidity of the process. Some experts expressed the view that spending 2 days for one proposal was excessive.

It was noted that several evaluators were not familiar with all the implications of the lump sum funding scheme on the evaluation process and therefore needed guidance throughout the whole consensus meeting. This guidance was provided by the moderator but rather frequently also by the quality checkers through their comments on the CRs, to the extent that several evaluators had the feeling that the quality check process was much deeper than what they had expected and was dealing with the content, the substance of the comments and was not only limited to a quality check “stricto sensu”.

A few details might be worth considering also for the future. For example, there was a case when the instruction “commenting before scoring” seemed to be interpreted (by at least one evaluator) as if all 3 criteria had to be commented and agreed before addressing scores. Clear guidance should be given to experts that the instruction “commenting before scoring” should be understood for each criterion individually and not globally for all 3 criteria. Clearly the sequence should be comment X, score X, comment Y, score Y, comment Z, score Z... and NOT comment X, comment Y, comment Z, score X, score Y, score Z. Failing that would bring the risks of (i) mixing up criteria and (ii) compensating an over-generous score on one criterion by an over-severe score on another criterion!

Likewise, there was at least one situation when a recorder, after agreement on comments on a particular criterion, wanted to initiate the score discussion by disclosing openly the individual scores of each evaluator. The observer would like to suggest that this should be avoided. Evaluators should not be influenced in any way by the individual scores of their co-evaluators. At the consensus stage, individual scores are almost irrelevant; they have become “past history”: what is to be scored are the agreed comments and nothing else! The observer was pleased to see one moderator making it very clear to his panel by asking explicitly the evaluators to “forget about your own individual scores” at the consensus stage and “ignore criteria A and B when you are scoring criterion C”

The observer noted also that negative comments were ALL qualified as “shortcomings” (at least in the observed meetings) as if there were absolutely no “weaknesses” in the proposals. Or as if experts were reluctant to use the full severity scale or the full score range, even though they were encouraged, during briefings, to not hesitate to use the full score scale. In addition, evaluators have sometime a tendency to deviate from the standard wording of the score interpretation table by inventing new qualifiers such as “small shortcoming”, “important shortcoming” which makes it more difficult to give a score. For a significant number of negative points, evaluators agreed that they should not be qualified at all, arguing that these negative points were mentioned “only for information”, which sounds somehow strange because these negative points (sometimes numerous) do not seem to be taken into account at all when proposing scores. Likewise, comments which are merely descriptive but do not express a judgement should be avoided. This was noted also several times as part of the comments of Quality Checkers who had to intervene to ask the evaluators to **assess** the proposals and NOT to **describe** the proposals.

One more aspect to be considered from the start of the consensus meetings is the limitation in the number of characters (4000) for comments on each criterion. There were situations where it was only at the very end of the process, after the final iteration of the quality check process, that it was discovered that finalized comments were too long by a significant amount. At that late stage, the only way to bring the comment length down within allowable limits was to replace many words by acronyms (Master Plan by MP, Project Management by PM, System Pillar by SP by, Work Stream by WS, Data Management Plan by DMP, etc...): the end result was that a perfectly written and clear comment, understandable by almost anybody, was replaced by another one, supposedly identical, but full of acronyms and understandable only by insiders!

The observer was pleased to note that draft CRs shared by the recorder to all evaluators did not contain, at the start of the consensus meetings, any proposed score. Indeed, evaluators should not be, even unconsciously, influenced by a proposed initial score (either the average of IERs or any other suggestion from the recorder or from any other origin). Scores should only be discussed and entered in the evaluation form only after finalizing the comments, for each criterion, and the score suggestion should not be initiated always by the same evaluator, in order to indicate clearly that all evaluators are equal, and to avoid giving the impression that one particular evaluator has more “weight” than his/her co-evaluators. Likewise, when the consensus meetings start with a short tour de table and experts give an overall judgement on the proposal (either globally or for each criterion), it is advisable not to start always with the same evaluator: all evaluators are equal and none of them is “more equal” than the others! The observer was pleased to note that this was generally done in this evaluation

- *criteria and scoring scheme: appropriateness, completeness, relevance, clarity, consistency in application, ...*

The applicable 3 main criteria and sub-criteria for the particular action type (IA) were explained in detail to the experts during the briefings. Although the evaluation methodology does not depart significantly from the previous programme Horizon 2020, there were some specific aspects which had to be reminded repeatedly to the experts (who were newcomers, for most of them) especially those related to the lump sum scheme and the additional sub-criteria listed in Annex 8 of the Work Programme 2022-2024. Due to the relative novelty of the lump sum approach, there were several situations when quality checkers, as part of their review work, had to remind the experts about some specific aspects which seemed to have been ignored or forgotten by evaluators and needed to be considered...thus giving some evaluators the feeling that quality checkers were getting involved directly into the substance (or missing substance) of the comments, beyond strict quality aspects.

All evaluators were familiar with the scoring system, and the significance of thresholds. The score interpretation table was often referred to as a reminder, before scoring against the agreed comments, for each criterion. Negative comments were often qualified in terms of severity using the wording of the score interpretation table (minor shortcoming, shortcoming, etc.) but the use of these qualifying terms was not evenly applied by all expert panels.

Notwithstanding the above remarks, the observer considers that the criteria, the related comments and the scores were applied consistently throughout the overall process by the evaluators, with the help and the vigilance of the quality checkers.

- *final panel meeting and the actors involved*

With only 1 proposal per topic (except one topic with 2 clearly unequal proposals in terms of merits, as reflected by the scores) the panel meetings were a simple formality consisting essentially for the experts to express formally their agreement with the consensus reports, confirming their views that these final versions of the CRS indeed reflected their collective opinion and that comments and scores were aligned, and endorsing the evaluation results.

The observer noted that experts were invited to briefly express whatever views they may wish to express at the very end of the evaluation and that there was a specific section in the panel report for that purpose. In that context the observer would like to suggest that this opportunity for the experts should be mentioned to them early in the evaluation process, for example as part of the briefings, in order for them to be better prepared to make comments and suggestions. Indeed, some experts seemed surprised by this apparently unexpected invitation to express comments. In earlier evaluations attended by the observer some years ago, under Horizon 2020, there was an explicit item "Expert comments" in the final panel agenda with about 0.5-hour allocation. The observer believes this is a very good practice.

At the end of the evaluation a specific debrief / "lesson learned" meeting scheduled for the following week was organised in which experts did express some interesting views but it is unclear to the observer whether these comments were eventually incorporated in the relevant sections of the panel reports.

- *hearings (if any) and the actors involved*

Not applicable

- *occurrence and handling of specific issues (if any) such as conflicts of interest*

The absence of conflict of interest was checked at the proposal eligibility stage. The understanding of the nature and importance of real or potential conflicts of interest was clearly explained to the experts during the briefings. Col cases uncovered during the evaluation process are in general spontaneously and immediately flagged to the evaluation staff. In such situations, the concerned evaluators, if any, are simply withdrawn from all phases (individual, consensus, and panel) of the evaluation of the proposals for which a Col exist.

- *quality of evaluation summary reports*

The need to have high quality reports was clearly explained, and hopefully understood, by all experts, not only in order to avoid complaints by applicants, but also in order to provide a fair, clear and useful feed back to the applicant. Indeed, any constructive comment generated by the evaluation may be of interest for the applicant to prepare, in due time, its future applications. The presence of quality controllers, assisting the moderators and the recorders during and after the consensus and panel meetings does contribute greatly to the quality of the final ESRs.

The observer would like to say again that the ESR quality is everybody's business. It does start at the IER stage and it is not the exclusive job of the quality checkers. Properly written and completed IERs are the basis for properly drafted initial CR versions which in turn are the basis for high quality final CRs and therefore of ESRs. Too often in consensus meetings, one can hear evaluators formulating half-completed comments with half- sentences finishing by "...or something like this...you see what I mean" or, even worse, by statements like "the quality checkers will change our text any way...". This should be clearly emphasized in briefings!

- *overall conduct of staff: responsiveness, hospitality, competence, ...*

The evaluation process was conducted entirely remotely. Throughout the complete evaluation phases attended by the observer, the professionalism and support of all EU-Rail staff involved in the evaluation reflected a high degree of responsiveness and competence. All questions and remarks from the observer were answered in a prompt and efficient manner, especially from the call coordinator, thus demonstrating also a high degree of openness and transparency.

- *infrastructure and working conditions for evaluators*

During all phases (individual and consensus) of this evaluation the infrastructure and working conditions for the evaluators were their home conditions due to the 100% remote character of this evaluation.

- *workload and time given to evaluators for their work, (remotely and/or on-site, as applicable)*

The time available to the experts for their individual assessment and the consensus meetings was comfortable in terms of overall duration. For those experts who were in charge of evaluating 3 proposals, the workload was however rather high in view of the size and the complexity of the proposals (120 pages part B, huge part A, a lot of financial data).

The consensus and panel phases were well prepared and planned and did not give excessive workload to the participants. However, the iterative quality check process may generate some disruptions in the

consensus phase sometimes by alternating consensus discussion time on a proposal, with time to incorporate QC comments related to an already discussed proposal. This may be detrimental to the continuity and fluidity of the consensus phase. It is important to ensure that the quality check process does not become a bottleneck and does not generate waiting time in the consensus process.

- *remuneration of evaluators (in relation to workload)]*

Surprisingly, the now usual complaint about the daily remuneration rate (450 euros/day) which has been unchanged for many years was not heard at all by the observer. Nevertheless, some evaluators questioned the time allocation per proposal for the individual evaluation, arguing that the evaluation of a 120-page proposal requires more effort and more time than for a standard 45-page IA proposal.

The observer would also like to submit that drafting a CR from 6 IERs requires also more time and effort than drafting a CR from 3 IERs only...but the observer is not aware of how the recorder remuneration is established

All experts were provided, as part of their briefings, with specific information concerning their payment conditions and procedures to cover their attendance to the "virtual central" phases of the evaluation process.

Recommendations, suggestions, remarks

Recommendations, suggestions, remarks

The evaluation was performed entirely in remote mode. During the individual evaluation phase, the observation work consisted mainly in following on a regular basis the progress of the IER preparation on SEP. During the "virtual central" phase, the actual "observation" work for the observer was concentrated primarily over the attendance to the consensus and panel meetings. For the rest, the work was based on discussions by mail or by phone with the call coordinator, attendance to several expert briefings in remote mode and reading of relevant documents

The complete evaluation process was, overall, as expected, i.e; without any issue or incident which would have required immediate corrective action. As a result of this "observation", the observer would like to submit the following list of items that EU-Rail JU may want to consider for future calls. There is no priority significance in the order in which these are presented:

- One particular comment/question concerning generally the evaluation of lump sum type of proposals is the ability (or willingness?) of several evaluators to make an informed assessment of all required aspects (do the proposed resources and split of lump sum shares allows achieving the activities and expected outputs? Are the budget estimates in line with proposed resources for each work package? How to use statistical data to assess the budget estimates?). It was apparent that some experts did not feel comfortable with this part of their work either because they felt this was beyond their competence, or because the high number of work packages and the role of the numerous partners (and, above all, affiliates which did not complete part A of the proposals) were not sufficiently described/detailed even with a proposal part B extended to 120 pages. From the reactions of several experts, it would appear that the extension from 45 to 120 pages of the part B is more than sufficient to address the Excellence and Impact criteria, but quite insufficient to address the Implementation criterion. In conclusion, as indicated above, such difficulty didn't prevent each panel of experts to reach the necessary consensus with the expected assurance. Nevertheless, those points could be taken into consideration by the JU as lesson learned to improve the process for the experts and the future LS call evaluation
- The lump sum funding scheme is deemed to make the management of the grant simpler during the implementation stage, but it does introduce some specific requirement and difficulties at the proposal evaluation and grant preparation stages. Indeed, the concept is such that potentially large budget cuts or reallocations between partners or work packages may result from the evaluator recommendations. Therefore, these recommendations are to be properly justified. This is even more important considering that the "no negotiation" principle is still in force under Horizon Europe. As a result, when faced with a non-negotiable (in principle) budget cut request during the grant preparation phase, the consortium will have in principle only one single and difficult choice: "take it or leave it". Therefore, it might be worth considering having dedicated financial experts participating to the evaluation, in addition to "normal" experts, to assess these specific lump sum aspects of the proposals and ensure that budget changes are based on very solid grounds (as this was the case 2 years ago in a previous S2R call evaluation when there was a specific financial expert involved in the evaluation). The applicable decision https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ls-decision_he_en.pdf infers that a financial knowhow is necessary for the evaluation of lump sum proposals: "... proposals shall be evaluated according to the standard Horizon Europe procedures... For each work package, experts with the necessary financial knowhow shall, in addition, check the budget estimate...). It is unclear to the observer whether the use of the lump sum scheme has really introduced new competence criteria, in the financial sector, for the selection of experts/evaluators

- There are always some variations between moderators in the way consensus meetings are run and the observer is convinced that, for the present evaluation, these differences were all within the applicable rules without any problems at all. The observer believes it is important that moderators take an active role during consensus meetings to (1) generally impose some discipline into the discussions (such as for example for the evaluators to be on time for consensus meetings), (2) ensure that all experts express their views, (3) act as time keepers, (4) encourage experts to focus on each criterion under discussion with ALL sub criteria (5) prevent the experts from discussing scores until the comments are agreed, for each criterion, and (6) press the experts to qualify their comments (minor shortcoming, shortcoming, weakness...) and refer to the score interpretation table. The observer believes that an excellent cooperation and complementarity between the moderator and the recorder are key to the success and the efficiency of consensus meetings
- The participation of the quality checkers (QC) in the consensus phase is very helpful and clearly contributes to the efficiency and quality of the process. In this particular evaluation, with evaluators recognizing themselves that they were not fully conversant with all the lump sum aspects to be considered, several QC comments were pointing towards aspects not addressed in the consensus reports. The observer believes this was an unavoidable and necessary task of the quality checkers, even though this may have appeared as exceeding the strict "quality" aspects of the CRs. Once again, the observer would suggest that evaluators should be convinced, that in the end, quality is also their own business and not the sole responsibility of the quality checker. As mentioned in a previous section of this report, EU-RAIL may want to consider, for the future, the option of having quality checkers participating directly into the consensus meetings. This was observed in other Horizon Europe evaluations and may contribute to save some iterations between QC and evaluators. However, in this case, a final quality check review should be performed internally within EU-RAIL JU. In any case, the quality check process should not become a bottleneck in the whole evaluation process
- It is important at the consensus stage that evaluators are not unduly influenced by the individual scores of their co-evaluators nor by any attempt from the recorder to introduce a priori a tentative score (for example an average score) which is totally irrelevant, in the initial draft CR. This was clearly implemented in the present evaluation. Initial suggestion of a score, for each criterion, should be made by an evaluator and NOT by the moderator nor the recorder, and only after the comments have been fully agreed by all evaluators, and the initial suggestion should not be made always by the same evaluator. It is the responsibility of the moderator to ensure that this is done properly and that individual scores are NOT coming into play at that stage. This of course does not prevent QC to comment also on scores (suggested by evaluators) and their consistency with comments
- The observer believes it is a very good practice to invite the experts, at the very end of the process (end of the panel meetings), to express their general comments about the call itself, the topics, and the evaluation process. Being experts on the topics of the call and being in the frontline of the evaluation process, they are fully legitimate in expressing their views; their ideas may indeed contribute to improve the next calls, both in terms of content of the call and the evaluation process. Maybe this practice should be better formalised as a specific item in the final panel agenda

In closing the observer would like to conclude by expressing his view that the evaluation process was conducted thoroughly and very professionally by all actors throughout all the observed phases. He is convinced, from his observations, that all proposals under scrutiny received adequate and fair treatment, and that the evaluation process resulted, for each topic, in the selection of very best proposals for funding

The observer would also like to express his thanks and gratitude to the EU-Rail JU personnel involved in this evaluation, especially the call coordinator Nadia Debza and the moderators Judit Sandor, Sebastien Denis, Gorazd Marinic and Javier Ibanez de Yrigoyen for making his observer work possible without any restriction whatsoever. The availability and permanent support of Nadia throughout all stages of the process contributed greatly to making the observer work not only a smooth but also a pleasant exercise.