

Comments on the First Interim Report of Key Activities (KA) of IPN Metodika

Content

KA1 comments to the First Interim Report	3
Comments and questions on the content of the report	5
Comments on the text formatting	5
Final suggestions	5
Conclusion	7
KA2 comments to the First Interim Report	8
KA4 comments to the First Interim Report	23
R&D Evaluation Methodology and Funding Principles. Background report: Tools for the Evaluation Exercise Implementation	23
R&D Evaluation Methodology and Funding Principles.....	24
Formal comments.....	25
KA6 comments to the First Interim Report	29
KA7 comments to the First Interim Report	31
General comments and requirements	31
The specific comments and recommendation	33

KA1 comments to the First Interim Report

The R&D Evaluation Methodology and Funding Principles

First Interim Report is well elaborated, the following notes are written for further discussion and possible improvements.

- 1) Note that in chemical sciences, a large proportion is represented by organic, macromolecular chemistry or biochemistry, which is nearer to biological sciences than to physics.
- 2) Publishing profile – mean number and median number of authors – participation in large collaborations (ATLAS, CMS, ALICE etc) should be evaluated separately – these publications should be excluded from bibliometric reports or analysed separately. The number of outputs in large collaborations and standard publications (up to cca 30 authors) should be given.
- 3) Publishing profile, point 6 – what does „counted from the top by the number of articles in the field,“ mean? Journals publish very different numbers of articles, therefore, the distribution according to IF between quartiles just dividing the total number of journals by 4 is not appropriate (quartiles will have a very different total number of publications; in extreme cases nearly all publications might be in the first quartile and zero in the others). Has that been taken into account?
- 4) Eligible research outputs – is it correct that an EvU or RU can register any number of eligible outputs they have? If so, it should be written clearly - we have not found it.
- 5) The criteria for the calculation of the number of outputs (a book is 4) – are they taken into account in evaluation or is it only for the determination of the threshold? We recommend a strictly field-specific evaluation.
- 6) Field Normalized Citation Impact – for RU or EvU? Only selected outputs will be evaluated – does FNCI make sense? Is there not a possibility for gaming? How will it be calculated for a multidisciplinary RU? What limits for recent publications, fields with few numbers of publications per year, low citation rates or low number of publications (probably 50 is enough). Statistical stability is important.
- 7) Key publications of the RU should be distinguished (with a substantial contribution of the RU). Perhaps it is ensured that only key publications will be registered for evaluation or it will be registered as excellence? How? This is connected to co-publications – if any output can be counted twice, it can be a reason for gaming (in fact some groups are based mostly on collaborations – is that OK?).

- 8) Number and percentage of publications among the top 10%, and 25% most cited publications (world, EU28) – it should be evaluated for the given field of science, year and type of output. Percentage of what – of all publications presented for evaluation by the evaluated unit? Percentage of all Czech most cited publications would be also informative. It should be described in more detail in the second part of the report.
- 9) RUs not only do not correspond to the organization structure but as defined (above 50 outputs, one field of science non-overlapping with other RU), it will mean in many cases $EvU=RU$ at least for AS institutes (with an average number of 500 outputs per RU). Consequently, the results of evaluation will not be informative for management on the institutional level. It is not sufficient to determine and divide a RU based just on 36 fields. A solution might be the possibility to have more RUs in the frame of one RO assigned to one field.
Some of the bibliometric indicators described in section 4.5.5 are relevant only if they are determined for a more detailed classification of fields than the second OECD level, with 36 fields. We propose to calculate with bibliometric indicators on the third level of the OECD classification that has approximately 190 sub-fields.
- 10) Where will all the data for indicators and their values be collected (mainly for the indicator such as licence income, external funding, international mobility)? In the R&D IS or in some new dynamic IS?
- 11) Although individual researcher evaluation is not recommended, an author identifier such as ORCID should be implemented (it is not mentioned in the report). How will the results of teams be identified for the purpose of the bibliometric report?
- 12) What will happen with publications that are outside the RU's main field - will they not be included in bibliometrics at all? The evaluation will not consider publications that will not have an existing RU in the particular RO (in fields that do not reach the threshold for the number of results in a given period)?
- 13) The report considers only Thomson Reuters tools - it should be more general.
- 14) It is unclear how the Czech R&D IS (RIV) will be used for evaluation (will it be e.g. widened by research impact (citations, reviews)).
- 15) It is necessary to create a guide for evaluators describing how they should consider the bibliometrics in particular fields.

Comments and questions on the content of the report

The introduction of the report and the descriptions might be more explanatory. It is important to explain the role of the "submission guidelines" and "submission forms" in the evaluation methodology. We recommend moving an example of the submission form to the end of the report (or include it as an attachment). Some parts of the chapter "Data report on research outputs" are lacking recommendations and the interpretations of statistical data (e.g. Citation Impact). We are missing an overview of the bibliometric indicators. Graphs should be more explanatory. What exactly are they showing? The chapter "Comprehensive data report" is not completed.

Comments on the text formatting

The text of report is hard to read. Paging is poorly described. Tables are badly formatted and some of them are divided and continue on the next page. Text references are missing in the report ("Error! Reference source not found"). We miss a list of tables and a list of abbreviations. The tables also need more description. The report is unfinished!

Final suggestions

- Make the report more explanatory.
- Fix all flaws in the formatted text, references, tables and graphs.
- Give an example of the Submission form as an attachment to the report.
- Create special documentation. Create an overview of bibliometric indicators and metrics.
- Fill in the missing text.

The appendix to these comments is the report with the analysis, which has been performed from 98% of the Czech outputs in cooperation with more than 30 authors. The Large collaborators (CMS, ATLAS, etc.) are not involved in the analyzed outputs.

Highly cited outputs correlate with the low proportion of Czech authors

WoS data taken for cu=Czech, py=2010-2014. The numbers of authors of Czech and foreign organizations have been analyzed and plotted against TC (number of citations). The following graphs and tables have been obtained.

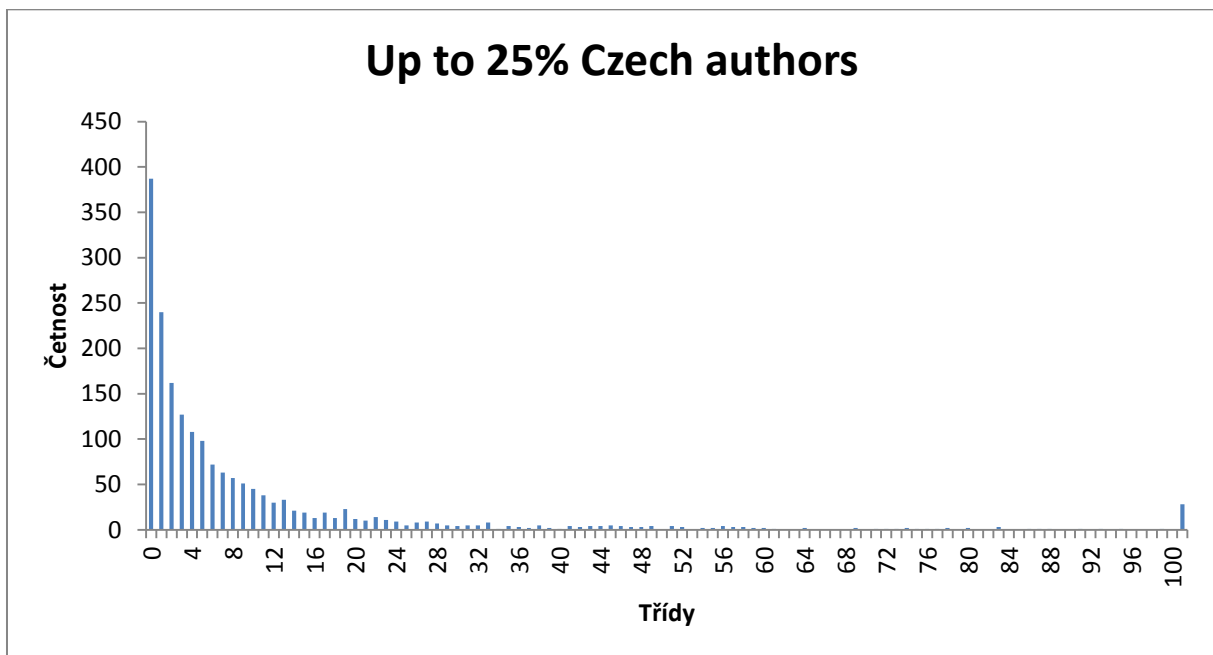
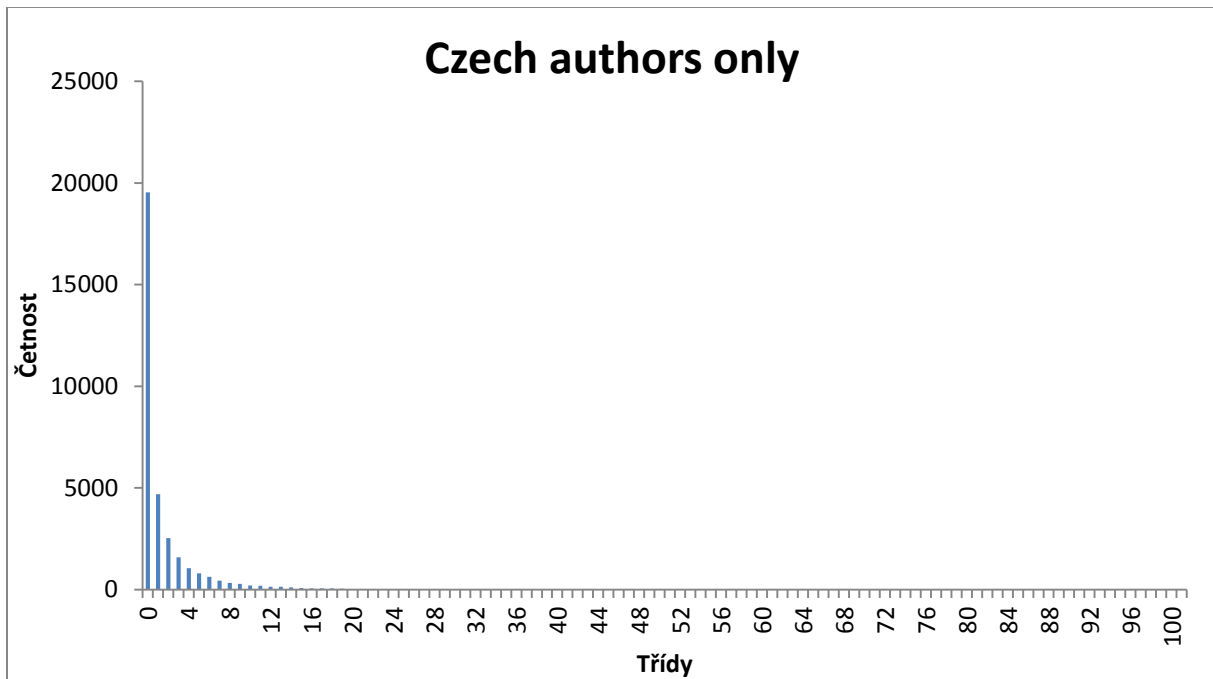


Table 1. Different citations categories in comparison to the proportion of Czech authors among all authors of the output.

Authors CR/ALL	0 citations	<=5 citations	>30 citations	>100 citations
Only czech authors	60%	90%	0,4%	0,01%
Less than 10% authors from CR	20%	20%	6,5%	1,5%

Similar proportions can be observed for various categories. In addition, among the 100 most cited outputs are various categories (not all): physics, chemistry, medicine, biology, etc.

Conclusion

Highly cited papers of Czech Republic in various fields of science are in correlation with low proportion of Czech co-authors among the authors of the output. The result of the bibliometric evaluation of the outputs for RU or EvU will be, therefore, strongly dependent on the contribution of outputs obtained in collaboration with foreign partners. This phenomenon has not observed in advanced countries such as Switzerland. Therefore, the construction of indicators for CR may differ from that of advanced countries such as GB. Probably, one possible solution could be the construction of indicators separately for individual key outputs (with a substantial contribution of RU to the output) and other outputs or the construction of indicators for key outputs only.

KA2 comments to the First Interim Report

- 1) Why should RUs report outputs if they are already available in the R&D IS? It would be sufficient if EvU marks them in the IS.
- 2) Pp. 32 Draft refers to the San Francisco Declaration “Do not use journal-based metrics, such as Journal Impact Factors, as a surrogate measure of the quality of **individual** research articles, to assess an **individual** scientist’s contributions, or in hiring, promotion, or funding decisions.” But the EM is NOT about evaluating individual articles and individual researchers! EM is to evaluate institutions! Why is this misleading reference made here and its very partial relevance explained?
- 3) Pp. 34Draft lists possible levels of evaluation but it does not include research teams and departments.
- 4) How is it that outputs to be submitted for the evaluation of excellence should be limited to publications and other than publications types are not to be included? [pp.68Draft].
- 5) The key principle of voluntary participation in evaluation has funding implications? [pp. 59Draft] Does this mean that it will be the management of the HEI deciding which of its EvUs will be subject to evaluation?
- 6) The threshold criteria for participation in the evaluation are somewhat confusing pp.62 (allow for various interpretations). This should be defined more rigorously.
- 7) The eligibility of RU for interdisciplinary evaluation uses the term **research activities**. This needs to be defined. If you have in mind the number of publications, why are other than published outputs not considered?
- 8) Why do research performance criteria contain *Ability to attract PhD students*? This indicator is easy to game because there is no effectively binding ceiling on the number of PhD enrolled (and funded by the government). What matters more is the number of PhD grads and their placement.

- 9) As proposed, the publishing profile of a RU will bring only information on the number of publications and no proxy information on their potential quality. This will lead to more weight to quantity than quality. [pp.74 Draft]
- 10) It is explained [pp 76 Draft] that in fields with insufficient coverage of publications in WoS/Scopus, it is normal practice to give more weight to qualitative judgment by peers. But no approach is proposed for the EM. We would expect that, for example, in fields with poor coverage, the percentage of publication outputs devoted to peer-review evaluation (the segment of excellence evaluation) is higher. If not resolved, a panellist will have almost no information on publications such as books and Jrec/Jneimp articles in the given field.
- 11) Information on outputs requested applies mostly to HEI and AS CR Institutes. Not so for business sector ROs, as well as for most state research organizations (museums, galleries, libraries) and organizational parts of the state, i.e. about half of all ROs.
- 12) The threshold size for RUs. The interval of 1-2% of research outputs that RUs submit for peer review is not justified. Why do books count as 4 outputs. Controversial quantitative indicators have been left out, with the assessment moved into the hands of the panel chairs. This is especially true for SSH, where more weight is given to qualitative judgment by peers due to the low-coverage and hence, partly invalid nature of quantitative indicators. However, it is hard to understand why the above thresholds still apply even for SSH.
- 13) Ad categorization of EvU's, saying that "In the EM2015 it is also foreseen that this categorisation may feed into a different weighting system for the allocation of institutional funding." only opens up undesired gaming effects.
- 14) Why are the provided forms and their explanations tailored for the Small Pilot Evaluation, without any comments on how they apply to a future large-scale evaluation?
- 15) What does Section 1.1.3 and Table 1 in Q007 referred to?
- 16) Fix "*Error! Reference source not found.*"

- 17) No definition of "institutional funding" other than stating, "The word 'institutional' means that the funding is not given to carry out specific 'projects'."
- 18) The "Use of the Data" explanations do not clarify sufficiently what EvU and RU should specifically enter (or not), but they rather give them additional motivation to manipulate the data supplied.
- 19) Stating sentences such as "These data will be normalised taking into account the size of the RU and will inform the evaluation panels on the level of research productivity." without any further explanation is not a good idea. Either explain or do not state.
- 20) The volume and detail of information required by EM from all ROs seems large and not so relevant for most of them. In fact, the required information makes sense only for HEI and institutes of the Academy of Sciences and a few Public ROs. For business sector RO, as well as most State Research organizations (museums, galleries, libraries) and organizational parts of the state (for instance Institut ochrany obyvatelstva Lázně Bohdaneč, former Civil Defense, part of the Ministry of Interior), which make up half of the currently acknowledged ROs and for which the income from R&D makes only a small part of the whole budget, the information of research staff, research training, career development, international research presence and collaboration, PhD students, competitive funding, research management makes no sense and has little relation to their mission and operation. It should be made clear how the EM provides such ROs with information on the following RO the criteria.
- Institutional management and development potential
 - Membership in the national and global research community
 - Research excellence
 - Research performance
- 21) Even for HEI and institutes of Academy of Sciences, because of the definition of RU, the required input data will not contain much relevant information. For instance, the information on competitive funding will, in many cases, concern the whole institutes, whereas the grants and projects concern mostly small units. For instance, The Institute of Physics of the Academy is spending billions of Crowns from the EU Structural funds to build the ELI-Beamlines laser facility, but this project concerns only a very small part of the Institute and thus the huge sum of money coming from competitive funding tells us nothing about the level of funding of the whole Institute.

- 22) The information required for the assessment of Societal relevance is so detailed and difficult to document and process that the resulting score will inevitably be of low reliability.
- 23) According to the document First Interim Report: the R&D Evaluation Methodology, there will be 6 main panels, organized at the level of disciplinary areas, and approximately 24 subject panels... Panels need to be small and high-level. The recommendation is to keep the number up to 5 max 6 members per panel. Will such small panels be able to competently process the large volume and complexity of information required in the submissions forms? Large Fields like Physical, Chemical or Biological sciences have many different subfields that would require many more panel members to organize the evaluation competently.
- 24) The definition of **Research Fields** will, in many Evaluation Units, lead to **huge RUs** including hundreds of researchers. For such large RUs the crude characterization of each of the five assessment criteria with one numerical value (4 to unclassified) will make such an assessment **useless** in terms of a formative function and **will not provide** “strategic information to actors at all levels”.
- 25) **Note inconsistent statements:**
- Page 4:** *It was expected to include **all research organizations** (ROs) while taking into account the differences among types of research organizations and disciplinary cultures.*
 - Page 7:** *The evaluation covers all research **organizations of a minimum size, on a voluntary basis.***
 - Page 59:** *The research organizations will participate to the evaluation on a **voluntary basis.***
- 26) Participation in an evaluation can be voluntary, but only if it will automatically imply resignation on institutional funding.
- Page 21:** *A requirement is that the system should provide **some** institutional research funding for all parts of the RD&I system that do research.* It ignores the fact that **only** Research Organizations are eligible for institutional funding.
- 27) In several places, the text shows poor knowledge of the reality of the Czech R&D environment and governance. Examples:

- a. Page 4: *The research community in the Czech Republic is composed of researchers employed in universities, public research institutes (including those 'grouped' in the Academy of Sciences and 'sectoral' public research institutes), private research institutions, and industry.*
- b. The list is incomplete as there is also a large number of **State Research Organizations** (Statní příspěvkové organizace, **SRO**), established by various Ministries, as well as a number of **Organizational parts of the State (OPS)** that do research and are mostly considered to be Research Organizations. The lack of knowledge of this large part of RO is similarly evident from the statement on page 27: *In the official statistics, research-performing organisations are grouped into free main categories: Higher Education Institutes (HEIs), Public Research Institutes (constituting the 'government' sector), and private research organisations.*
- 28) Besides PROs, the government sector also includes SRO and OPS. Neglecting the existence of **State Research Organizations** and how they differ from Public ROs, established by Ministries also leads to a wrong statement on page 16: *Public research organisations accounted for the majority of the registered research organisations in 2014 (163 on a total of 219).*
- 29) The number of PRO also includes the SROs and OPSs. The true number of Public Research Organizations is actually only 76. The difference between a PRO and SRO is very important and concerns not only the legal status, but also the overall mission and the way of funding. If the EM does not capture this difference, it will not work properly.
- 30) The Report fails to mention the main task of the Council (RVVI), which is the preparation of a proposal of a very detailed budget for research and development.
- 31) Page 13 of the Report mistakenly claims that the **Academy of Sciences is an organizational part of the State** and in this function, apart from Ministries and grant agencies, is also responsible for the "implementation of RD&I policy".
- 32) The claim on page 15: *The Academy of Sciences has historically a special position in the Czech R&D system, answering directly to the Prime Minister rather than a Ministry. It is wrong.*

33) Definition of research fields and panels

- a. Subject panels. According to the statement on page 83: *“There will be 6 main panels, organised at the level of disciplinary area, and approximately 24 subject panels, organised at the level of field.”*, but there is no specification how these 24 subject panels will be formed from 42 Fields in Exhibit 28.
- b. Meaning of the statement on page 78: *“The current categorisation is shown in Exhibit 28, below. A transition period of approximately 5 years will be needed before a full-fledged use of this new classification system will be possible.”* Does this really mean that the EM could be used in earnest only after 5 years?

34) Categorization of research organizations according to their missions

- a. **Page 7:** The Evaluation Methodology distinguishes between
 - i. Scientific Research Institutions,
 - ii. Research and Technology Organisations (RTOs),
 - iii. Public service Research Organisations, and
 - iv. Infrastructures, providing service to research.
- b. This categorization **does not reflect the real differences** between missions of various ROs, which are primarily given in their statutes. Many RTOs act at the same time as Scientific Research Institutions and vice versa. Public service Research Organization is an artificially created category which has no legal substance. This even more true for "Infrastructures, providing service to research", because in the new GBER, research infrastructures are defined as a separate concept, different from Research organizations: *'research infrastructure' means facilities, resources and related services that are used by the scientific community to conduct research in their respective fields and covers scientific equipment or sets of instruments, knowledge-based resources such as collections, archives or structured scientific information, enabling information and communication technology-based infrastructures such as grid, computing, software and communication, or any other entity of a unique nature essential to conduct research. Such infrastructures may be 'single-sited' or 'distributed' (an organised network of resources) in accordance with Article 2(a) of Council Regulation (EC) N°723/2009 of 25.6.2009 on the Community legal framework for a European Research Infrastructure Consortium (ERIC); The form of their financing is set in Article 26 Investment aid for research*

infrastructures. Consequently, infrastructures CANNOT be treated and financed as RO!

- c. The only meaningful categorization of genuine RO is as follows:
 - i. HEI, primarily Universities
 - ii. Public RO of the Academy of Sciences
 - iii. Public RO and State RO (Statní příspěvkové organizace) established by Ministries
 - iv. Business sector RO
- d. This appropriate categorization of RO is crucial in order to meet the goal of EM expressed on page 8: *The Evaluation Methodology allows for a reasonable level of field- and RO typology-specific variations to the common generic indicators and assessment criteria. While the scores against the different assessment criteria will therefore be 'field-neutral', the different missions of the research organizations will be taken into account in the second stage of the PRFS process, i.e. for the allocation of the funding.*

35) Remote peer-review.

- a. Peer-review without on-site visits is a shortcoming as far as the assessments of Research excellence and Research performance are concerned. The problem is that peer-review of publications cannot easily determine the extent and importance of the contribution of an assessed team to the published results. This is particularly pressing in the era of globalized research and concerns not only large international collaborations counting thousands of authors, like in particle physics and astronomy, but also small teams made up of authors from several countries. Not determining this contribution, a mere assessment of publications cannot give a true enough picture about the quality and importance of research in an assessed workplace. In the prepared evaluation of institutes of the Academy of Sciences, this problem will be tackled in Phase II, which will involve on-site visits of panel Commissions.

36) Assessment criteria.

- a. *Institutional management and development potential* makes no sense for RU, only for EvUs.
- b. *Quality of the managements and the research infrastructure is outstanding and beyond international norms* is too vague and very subjective.
- c. *Research performance is world-leading or internationally excellent in terms of outputs, productivity and overall quality*: Guidance is missing on what will be

the measure of performance, how will the panels assess the “productivity” and “overall quality” and how will they establish the “international excellence” in these aspects?

- d. Much the same can be said for **Societal relevance** and the characterization of the top level. What is the definition of “*outstanding impact*” compared to “*very considerable impact*”?
- e. In order to make these characterizations meaningful, they should be accompanied by a specification of reasons guiding the panel to particular classification.

37) With a maximum of 2% of outputs submitted for Research excellence assessment, this criterion does not have to be informative, as it corresponds (assuming one output per researchers per year) typically to 0.1 output per researchers. In UK RAE, one researcher submits 4 outputs and in the prepared evaluation of the institutes of the Academy the corresponding number will be 2.

38) Weighting of main indicators of a given criterion. The statement on page 77 above: *At the level of main indicators, the subject panels will describe their specific approach for the weighting of the main indicators* implies mixing apples with oranges and makes the final score difficult to interpret. How will you weight, for instance, adding together “reputation and esteem” with “national co-publications”? Maybe, the Evaluation report should assess the RU for each criterion from all the main indicators without combining them together with ad hoc weights.

39) Evaluation of results

- a. If the evaluation of results should also directly inform public institutional funding for research, as stated on page 59, it will have to be formulated with links to the basic principles of the institutional funding (2nd IR).
- b. Inconsistency in the specification of what will be evaluated. In Section 4.8.1 *Structure and content of the panel reports* on page 79 one reads: “*The evaluation results will consist in a panel report covering each Research Unit and each field.*”, whereas the statement on page 87: “*Developing the assessment and providing scores. For each of the EvUs scores are provided on each of the five criteria.*”, seems to indicate that scores would be defined for the whole EvUs, which would require weighting the scores of different RU within one Evaluation unit, typically the whole institute of the Academy or the whole faculty of HEI. For this reason, and in order to see the form of results of the proposed EM clearly, the Report should include an example of a complete Evaluation Protocol.

- c. RU must get a chance to express opinion on the preliminary version of the Evaluation Protocol and this opinion must be part of the final version of the Evaluation Protocol.

- 40) Does the 30% threshold (pp. 62) suggest identifying evidence-based interdisciplinary, i.e. based on analysis indicating that there is, in fact, a significant drop of around 30%, or is it rather a kind of rule-of-thumb? Since gaming is likely around this threshold, it would be preferable to consider in more detail the available evidence in order to specify the exact value of this threshold. In other words, it should be supported by computation based data from the Czech system (as the data for this is there, isn't it?).
- 41) How is cross-affiliation (of authors of publications) tackled, i.e. the same person affiliated at the same time to an institute of the Academy of Science and a university (or possibly more) and claiming these multiple affiliations in a top journal article? Does "co-publications of researchers active in different EvUs will be counted as one each" (pg. 73) also apply to the same person active in different EvUs? Note that this is particularly important to clarify in a system with no upper limits on the sum of part-time positions, i.e. no upper limit of say 120% is in place here. Again, if there is a threshold used here, is there any evidence-based support to use say a 50% contract as the threshold or does this number come from thin air? Note, for example, that there is 50% upper limit for the so-called "DPP and DPČ" contracts in the Czech system, so if a 50% threshold is used, there is a window for gaming with these types of contracts right there. Also note that the wage paid in these contracts depends on negotiations, hence there is necessarily no direct connection between the wage paid and the size of the contract. It will be impossible to detect gaming along these lines, unless wage data is reported (and checked) during the course of the evaluation. For how long or at which given point in time does the contract exist, i.e. only on 31.12. or through the whole evaluation period? How will cases of people moving between organizations be tackled, i.e. having multiple contracts in different parts of the system over the evaluation period (fractionally accounted for or what?)
- 42) "Co-publications of researchers active in the same EvU but registered in different RUs will be counted as one each, on condition that they are of a clear interdisciplinary nature" (pg. 73). How and who will determine whether the publication is of "a clear interdisciplinary nature"? How many publications registered in the Czech system would classify as "clearly interdisciplinary" over 2008-2012?

- 43) How is the disciplinary classification of a publication determined, i.e. is it the disciplines classification of journals in the WoS or something else? Note that many journals in WoS are assigned to multiple disciplines. Are all of these going to be taken into account? How is this going to be tackled in order to avoid gaming (of bibliometric evidence), supporting claims for doing interdisciplinary research?
- 44) It is not clear what is supposed to be the final output of the Czech valuation, i.e. how the EvU will be classified? There is an example of a Likert scale used in the UK (pg. 52), but it is not explained in Section 4, whether the same or similar scale will be used in the Czech Republic. There are 1 to 4 scales outlined for the sub-criteria but not for the overarching score.
- 45) I would classify the Czech system in the lower left corner, i.e. Research & Outputs, in Exhibit 14.
- 46) How will the decision-making in the panels be organized, i.e. voting based on a plain majority, etc?
- 47) There is “XXX” on pg. 54 and “Error!” on pg. 62 and 63.
- 48) A great deal of the evaluation criteria should be known to the research community IN ADVANCE. The Interim Report expects each panel to largely agree on how it will evaluate the respective field, taking into account its specificity. However, given the Czech experience with changing criteria of submissions and evaluations of research outputs, it is essential for the academic community to state at least a basic set of criteria that will be the basis for evaluation IN THEIR FIELD well in advance when this knowledge can still have an impact on their research and management strategy (e.g. it is not sufficient to mention that the WoS will not be the only or main criterion for research in the history of arts, but also what other criteria will be taken into account instead of these data).
- 49) The number of evaluated outputs (for the criterion of Research Excellence) - 1-2% - seems to problematize the inclusion of smaller research units (like the Oriental Institute of the Academy of Sciences with under 15 researchers). This may be inevitable, but it might be a good idea to address it (especially regarding the possible funding implications).

- 50) It seems that the section on Conflict of interests has been adopted from material on grants, as it consistently speaks of "proposals" and "principal investigators".
- 51) Sec. 4.1 pp. 59, among the "key principles" of evaluation methodology, the fourth bullet point states that "Research Unit includes all individual researchers in an EvU (across the organisation structure) that conduct research in a single scientific field" (according to OECD field classification as listed in exhibit 28: structure of disciplinary fields). The problem is that within the given EvU (say a faculty), the scientific field under evaluation may be composed of several departments of vastly diverse quality. For example, in any Philosophical faculty (EvU), the evaluation of RU "art" would include departments of art history, aesthetics, film studies, music and drama studies; in reality there are substantial qualitative differences among these departments in all major philosophical faculties (Prague, Brno, Olomouc), RU "art" would thus be composed of excellent as well as very mediocre departments. Evaluating the aggregate output and impact of such RUs by a single grade may yield relevant information at the level of comparing evaluation units, but at the same time it will obscure the real differences among departments within research units of any given EvU. This is particularly worrisome if the process of evaluation is directly linked to management. Authors of the text implicitly assume that the consequences of the evaluation process of the given RU do need to be addressed in the document, as they are fully in the competence of management of any EvU. However, unless the evaluation process is finalized at the level of the given EvU by highlighting the internal differences within RUs, the whole process will not promote excellence and may be de-motivating for people in excellent departments.
- 52) Sec. 4.4, p. 63 Is it necessary to stipulate that peer review will be remote with no on-site visits? There surely could be some variability in this regard, allowing for an inclusion of on-site visits whenever they would make sense in any given case.
- 53) Sec. 4.5, p. 63-64: Assessment criteria in response to the policy objectives. This part has not been changed, formulation of social relevance in terms of collaboration with the needs of an industry is extremely narrow and leaves out other desirable aims - such as public education.

- 54) Sec. 4.5.4, p. 73 Eligible research outputs. The stipulation that *each Research Unit will submit for review a number of research outputs that accounts for minimum 1% and maximum 2%* is too restrictive (particularly for humanities, where the main output takes form of a book-monograph). For example, 5% minimum and 10% maximum would be more appropriate.
- 55) Sec. 4.5.6., p. 75: Partial coverage creating tensions among fields. Overall, the document covers the problem of disciplinary differences and how they should be reflected in the EM fairly well.
- 56) On p. 75 it states that: "To avoid creating tensions among fields in the SSH, it is important not to regard coverage in the commercial databases as an indication of research quality in a given field." This is correct and to the point. However, the following stipulation that - to quote - "Instead, less covered fields should be respected for their specific publishing pattern that are insufficiently covered" is problematic. It opens the door to introducing double standards, as the notion of "specific publishing pattern" is indeed wide and vague and prone to be used as a handy explanation for second-rate publication output.
- 57) R&D Evaluation Methodology pp. 69: Ability to attract PhD students: One of the attraction effects could be to have easy, low level PhD theses, as provided by some faculties. Therefore, additionally to the statistical numbers, the panellists should check at a few randomly selected theses and their level and quality.
- 58) R&D Evaluation Methodology p 85: Subject panel members: Due to the practice of the election of the faculty deans in the CR they do not, in general, represent a privileged category of specialists suitable for panels.
- 59) R&D Evaluation Methodology p 74: 4.5.5 Bibliographic indicators: Statement "The selection has been made after discussion with an(d) input (comments in response to the input appear on the last page below)". No such comments can be found.

- 60) Follow comments on the First Interim Report: the R&D Evaluation Methodology
- 61) To avoid ambiguity, the term “Evaluation Methodology” (abbreviated EM) should be consistently used for the new evaluation methodology proposed by Technopolis, the Czech word “Metodika” should be used for the existing evaluation methodology in the Czech Republic, and the term “evaluation methodology” should be used in a general, unspecified context.
- 62) The abbreviations should be used in a consistent form (e.g. R&D&I / R&DI, R&D IS / RD&I IS / R&D&D IS / R&DIS / IS etc.).
- 63) Page 4, 4th paragraph: The meaning of the expression “the most publication-active disciplines” is not quite clear. Does it mean “disciplines with the highest absolute number of publications”? If so, then the comparison of disciplines makes no sense because the reason for the described differences may be due to the nature of the disciplines.
- 64) Page 7, 6th paragraph: It is not clear what evaluation period is in fact being proposed. Here (and also on p. 62), it is vaguely said that it will be 5 to 6 years while the Background report: Tools for the Evaluation Exercise Implementation claims on p. 8 that the evaluation will take place every 6 years.
- 65) Page 8, 4th paragraph: Correct the typo – “extent” instead of “extend”.
- 66) Page 13, last paragraph: The sentence “The two agencies have a unique status, with their governing bodies is nominated by the Government...” should be correctly reformulated.
- 67) Page 15, 7th paragraph: The correct of the number of the act is 211/2009.
- 68) Page 17, exhibits 3 and 5: The abbreviations “No” and “Nr” for “number” should be unified.
- 69) Page 21, 1st paragraph: The sentence “A requirement is that the system should provide some institutional research funding for all parts of the RD&I system that do research” is ambiguous.

- 70) Page 21, the next-to-last bullet: In view of the current practice in the Czech Republic, it should be emphasized that to be a research organization according to the formal definition or by the decision of the R&D&I Council is a necessary, not sufficient condition for obtaining institutional support. The research organization should be subject to evaluation but this should not automatically provide a voucher for institutional support.
- 71) Page 54, 2nd paragraph below Exhibit 20: The reference “Chapter XXX” should be specified.
- 72) Page 58: The exact meaning of the words “on a national basis at the institutional level” in the first sentence is not fully clear. Will the proposed Evaluation Methodology be conducted at national or institutional levels according to the definitions on p. 26? If at an institutional level (as it appears), then the claim “no on-site visits are foreseen” in the 3rd bullet on p. 63 does not seem in harmony with the claims on p. 26 and 31, that “these evaluations typically involve (informed) peer reviews with on-site visits”. The Background report: Tools for the Evaluation Exercise Implementation on p. 8 mentions a “national evaluation”.
- 73) Page 62, 3rd paragraph: The condition “An EvU can register for the performance assessment more than 1 Research Unit only if at least 50 research outputs have been produced in more than 1 scientific field during the evaluation period” is not suitable for many institutes in the Academy of Sciences, as well as for some faculties like the Faculty of Mathematics and Physics of the Charles University. For instance, this would mean that the Institute of Physics ASCR with its several hundreds of researchers working in different fields of physics will be formed by just one Research Unit. How can such evaluation provide strategic information to the director of this institute? In fact, this is true even for much smaller EvUs.
- 74) Page 73, the next-to-last paragraph: The limits of 1–2% seem very small. If the threshold for the participation of an EvU in the evaluation – and for the registration of a Research Unit by the EvU – is set at 50 research outputs within 1 field of research over the evaluated period (cf. p. 62, 2nd paragraph), then such a minimal EvU is going to submit just 1 research output for review!
- 75) Page 74: The references 3a, 3b, 3d in the last four bullets are not clear.

76) Follow comments on the Background report: Tools for the Evaluation Exercise Implementation.

77) Page 9, Q013: The claim “This information will provide ... a view on the positioning of the research unit in the EvU in terms of its share of that funding” is not clear. First, it seems that the proposed questionnaires do not provide such information. Second, it will be difficult or even impossible to provide such information in many EvUs.

78) Page 9, Q015: The meaning of “researcher that conducted 50% or more of their PhD research in research organization” should be well defined. It should be noted that PhD students are always officially registered at a university and receive scholarships from there. If their supervisor is in an institute of the Academy of Sciences, the students usually have (in addition to their scholarships) a part time contract in the institute which is, however, often smaller than 50%. This is not related to the real time spent on research in the institute. Thus, the extent of research a student conducts in the institute cannot be derived from the employment contract.

79) Pages 39–40, Exhibit 12: The data should be considered with caution. What is valid for Norway need not be valid universally. For instance, the following table shows numbers of mathematics papers added to arXiv, Web of Science and MathSciNet annually between 2008 and 2012:

	arXiv	Web of Science	MathSciNet
2008	14 373	20 908	86 533
2009	16 319	22 390	87 279
2010	18 765	22 079	87 162
2011	21 287	22 716	89 638
2012	24 176	23 760	92 191

It is evident that the coverage of mathematics in the Web of Science is rather poor.

KA4 comments to the First Interim Report

R&D Evaluation Methodology and Funding Principles. Background report: Tools for the Evaluation Exercise Implementation

No.	Location in the document	Comment (<i>original text from the report in Italics</i>)
1	p. 7, part 2.1.3 Q006	<i>„Each researcher affiliated to the EvU can be registered in only one Research Unit.“</i> What is the criterion – research outputs or a person’s working time?
2	p. 8, part 2.1.3 Q006	A problem may arise: several ROs will register the same researcher. Who will check the correct registration of an individual researcher (to avoid multiple occurrences)?
3	p. 8, part 2.3.1 Q008, 9, 10	<i>„Staff numbers should be provided both in headcount (number of people) as well as in Full Time Equivalent (FTE)...“</i> In Czech, HEI conversion of headcount into FTE has to be resolved systematically. A proposal: 1 faculty member = 0.5 FTE researcher.
4	p. 7, part 2.2 Q007	The term “Conference proceedings (D)” should be clearly defined. Does it mean conference proceedings or a paper in conference proceedings? Which conferences are eligible? Only those included in WoS or Scopus?
5	p. 9, Q013	The term „institutional funding for research“ must be precisely defined by using the wording from the Czech accounting act; ditto for „other funding“.
6	p. 9, Q015	What does “conduct research in the research organization” mean? Will it be judged based on a working contract or somehow else?

7	p. 11, Q027	How is sharing equipment with other research organization measured? Eg. if we provide measured data on the equipment of another institution, is it sharing equipment? If there is no clear definition of an evaluated output, there is risk of too much useless data in the self-evaluation report.
8	p. 11, section 2.4.2., Q026, Q027	Equipment that cost more than €25000 –it might be better to have two thresholds – one for social sci.and humanities, e.g. €10000 and the other for remaining disciplinary areas, e.g. €40000. The rationale is in substantially different requirements for equipment in different disciplinary areas.
9	p. 27. Q022	It should be clearly stated if V.A.T (DPH) should be included in the reported incomes from contract research.

R&D Evaluation Methodology and Funding Principles

No.	Location in the document	Comment (<i>original text from the report in Italics</i>)
1	Exec. Summary	An exec. Summary is rather a set of political statements than a brief description of the main features of the proposed methodology
2	p.7, line 7	<i>“Each individual researcher in an Evaluated Unit can register for one Research Unit only.”</i> Who is going to check the correct registration of an individual researcher (to avoid multiple occurrences)?
3	p.7, 1 st frame	<i>“evaluation covers all research organisations ..., on a voluntary basis.”.</i> <u>Same on p.21, last but one bullet;</u> <u>Same on p.59, 5th bullet: participate... on a voluntary basis”.</u> <i>“...would be unfair to oblige all research organisations to bear these costs...”</i> Comment: What happens if the RO will not take part in the evaluation? Will it still be receiving institutional funding?

4	p.15, last par.	<p><i>“A technical amendment to the Czech Act is in the process of approval”</i></p> <p>Was approved by RDI Council on October 31, 2014</p>
5	p.23, 2 nd par., 2 bullet points	In my opinion, a ‘general’ evaluation (1 st bullet) has a summative function (looking backwards) whereas ‘PRFS’ has a more pronounced formative function
6	p.62, part 4.3.2	<u>Threshold</u> : Infrastructure ROs and Public Service ROs may not be capable of reaching this threshold! Shall we exclude them from evaluation? I would say NO, they should be included.
7	p. 65 Exhibit 25	There are three main indicators - Income from competitive national research, Income from societal relevant research (competitive & contract research) and Income from commercialization activities – all of there are very close to each other. It will be very necessary to identify the differences precisely and to define data sources (verifiable and comparable).
8	p.73	What are the rules for the treatment of co-authored outputs such as patents and other IPs (see Exhibit 27)?
9	p.84-87	Terminology unclear or mismatched: Evaluation Management Board, Governing Body, Evaluation Management Committee, Management Team
10		As far as the funding is concerned, will it be possible to evaluate newly established faculties or other research organizations?

Formal comments

No.	Location in the document	Comment (<i>original text from the report in Italics</i>)
1	p. 7, 2 nd par.	<i>Each individual researcher in an Evaluated Unit can register for one Research Unit only.</i>
2	p. 8 4 th par.	Recommendation: the reference for a RAE/REF definition should be added for those who are not familiar with the term

3	p.8, 4 th par.	<i>...RAE/REF methodology of relevance for this study is the extent to which panels...</i>
4	p.8, end	Please clarify what the basis for calculating the 1% limit of the public institutional funding for R&D is
5	p.32, 1 st par.	<i>This reduces the costs of the evaluation as well as the</i>
6	P.42, line 17	Undefined abbreviation „IP“ (Intellectual Property ?)
7	p.50, 4 th bullet	Meaning of the sentence is unclear
8	p.52, 4 th par	Unclear meaning of 'UoA'
9	p. 55, Exhibit 21	line „indirect cost“: without specifying the methodology of calculation of indirect costs, the figures have little value.
10	p.59, 6 th bullet	delete “draw on a” (occurs twice)
11	p.61, 4 th par.	Reference source not found
12	p.62, part 4.3.2	Reference source not found
13	p.62, last but one par.	<i>...are active only to a limited extent in the production...</i>
14	p.63, 2 nd bullet	Reference source not found
15	p.64,	<i>„The risk for bias: To a certain extent,...“</i>
16	p.72, 1 st bullet	<i>“contributions to developments in research, research leading to the...”</i>

17	p.74, numbered item 3	„Number and percentage publications in each field“: percentage of publications based on?
18	p.74, bullet points list at the bottom	„Five most frequent collaborating countries in the field, and their shares of the publications in 3a.“ where „3a“ is referred to? Similarly „3b“ and „3d“
19	p.74, 4 th par.	“Indicators are calculated and presented per OECD field“: does this mean per OECD field as defined in Exhibit 28 ?
20	p.75, 2 nd par.	„we trust that the Czech R&D Information System will cover the outputs from these fields comprehensively.“ To be verified/confirmed by Infosciences and possibly RDI Council
21	p.75, last par.	“For the citation indicators, the references in typical citing documents must mainly...” the meaning of this bullet point is unclear
22	p.76, 2 nd par.	“Thereby, one also open up for the evaluation of a larger multitude of aspects of research quality that SSH scholars are used to judge.” Meaning of the sentence is unclear to me
23	Page 77, part 4.6.2	What is the meaning of “cross-referrals” ?
24	p.84, part 5.1.3, 2 nd bullet	Evaluation Managm. Committee – was not defined. How does it differs from the Evaluation Managm. Board?
25	p.85, 3 rd bullet under Profiles	“All chairs and members of the panels will be international experts“ Do you mean <u>subject</u> panels? If yes, please add to the sentence
26	p.86, 1 st and 3 rd bullet	“Upon approval, the main panels will report the assessment to the governing board,” What is and who is the governing board? „The main panel can also suggest members for this interdisciplinary panel to the Governing Body“ What and who is the Governing Body?

27	p.86, 5 th bullet	<i>"...within the framework of the exercise, following the published procedures and criteria."</i> where are the <u>published procedures and criteria</u> ?
28	p.87, part 5.1.5, 3 rd bullet	<i>"is to design the guidelines for the evaluation process"</i> Where are the guidelines?
29	p.87, part 5.1.5, 4 th bullet	<i>"For the evaluation process several information systems are used, such as an online tool for the registration of the EvUs, submissions of research output,"</i> who will suggest and develop an <u>online tool</u> ?
30	p.88, 2 nd bullet	<i>"The main panels will be supported by a panel secretariat. The panel will have 2"</i> In the second sentence, the main panel is considered. If this is the case, please repeat the word "main" for the sake of clarity.
31	p.91, 1 st bullet	<i>„... importance of inter-disciplinary research, there is a small number of main panels (around six)"</i> at the top of page 83 it is stated that there are 6 main panels (<u>altogether!</u>)

KA6 comments to the First Interim Report

- 1) The implementation process should start with installing the Governing Body with reference to Chapter 3 (p. 87). However, in Chapter 3 there is no information on this body. Thus, it is not clear what the role of this body is and what its relation to the Management Team and to the “The research communities, represented by the RD&I Council is. Whether it is providing scientific advice as well as support for the evaluation implementation to both the Evaluation Management Board and the Evaluation Management Committee” as mentioned in Chapter 5.1.1. Furthermore, there is strong confusion between the usage of the concept of the board and the committee. The management team is sometimes called the management board and /or committee. The scheme on p. 82 does not give a clear idea of the governance structure, and especially of the relations between the aforementioned different bodies. Therefore, this part of the report urgently needs clarification of terminology.
- 2) All these bodies are presented as representative bodies – of ministries and agencies and scientific communities respectively. The only body that should provide administrative support to the panels should be “the panel secretariat”. However, it is not clear how they can guarantee “consistency in the approach among different panels” if they will not have a common administrative and operational structure. This structure should be stable and should be part of the research policy organisation in Czech Republic. While the detailed options of the establishment of this structure will be the task of feasibility study in KA6, this should be mentioned and explicitly referred to in the Report.
- 3) Regarding the implementation procedure, it is not clear who will be responsible for the bibliometric analysis which will not be a trivial task and should be delegated to the stable secretariat with expertise in this field (see note 2).
- 4) Some parts of the text need corrections and editing (eg. “Error! Reference source not found”).
- 5) The criterion “Ability to attract PhD students” can lead to an unintended outcome as it does reflect the PhD students selection process in CR. In principle, Czech universities can admit any number of PhD students. This is especially substantial in the case of part-time PhD students. In consequence, there is limited competition for highly motivated PhD students amongst Czech universities. Furthermore, this cannot be reflected by the PhD student enrolment. Out of the indicators proposed in Chapter 4.5.2. the most relevant would be the level of investment in PhD training and effectiveness of PhD education and the trend.

- 6) The role of bibliometric indicators as presented in Chapter 4.5.5. is still not sufficiently related to the assessment criteria in Chapter 4.5.2. Should the bibliometric report be a part of the bibliometric analysis mentioned in the research productivity criteria? Or research excellence criteria? Or is it a separate channel of information for panels?
- 7) There is a need to specify the notion of fraud and dishonesty, or to clearly refer to The European.
- 8) Using the Code of Conduct for Research Integrity as a reference document, not only stating that the methodology is “in line” with this document. High sensitivity to these matters in CR is underlined by several cases of fraud and misconduct that were publicly debated in mass media in the recent years.
- 9) The provision for avoidance of clientelism & nepotism are still perceived as insufficient. The declaration of unbiasedness is needed, but there should be also sanctions for misconduct in case of information provided to the panel.

KA7 comments to the First Interim Report

General comments and requirements

- 1) It is not clear how the structure of disciplinary areas and fields (Exhibit 28) will match with national CRIS system.
- 2) Self-reporting of EvU described in submission guidelines for the Small Pilot Evaluation includes several items that can be easily obtained from IS R&D&I, e.g. Institutional funding for research, National competitive funding, Research Outputs.
Why is this information collected again through EvU and not obtained directly from IS R&D&I.
- 3) Expert panels at the level of scientific fields are at the core of the Evaluation Methodology, but on-site visits are key to the whole process. These visits are recommended only rarely. Specificity of the Czech Republic needs more clarification and verification on-site. We recommend checking once again the means of site visits as a possible solution.
- 4) It is clear that especially allocation of institutional funding should be dependent on the type of RO but in the CR this is influenced by the existence of the Czech Academy of Science, which has the same mission in the research area as universities do. This fact should be reflected in institutional evaluation and a funding formula
- 5) Web of Science and Scopus are often recommended as basic bibliometric parameters. In other paragraphs this is Web of Science only. Is there any reason for this difference?
- 6) The representatives of Art and Musical universities have two main questions:
What about the future of the arts-based research in Czech Republic? Can we use, in particular, the experiences of e.g. Austria, UK, Australia, and New Zealand?
- 7) In EM, the „Policy objectives of the evaluation methodology and funding principles“(Exhibit 7, page 20) are the basis for Assessment criteria and indicators (Exhibit 24, page 63). The authors analyzed current policies, etc. and based on this, they proposed the Policy objectives. Bearing in mind that this breakdown is crucial for the entire assessment, it should be the result of a political consensus and approved by a government authority. It follows that verification of compliance with Policy objectives listed in EM with government priorities.
- 8) The methodology assumes assessment implementation (combination of objective indicators and international peer judgment) of Czech RO, all carried out especially with foreign evaluators-experts; requirements for experts and panel members are also described.

However, it lacks commissioning an authority responsible for the selection and appointment of individual members and then proposing a mechanism or a procedure to ensure that foreign experts are familiar with national priorities and take those into account in the evaluation.

Note: Specifying an authority responsible for selecting, appointing and overseeing the activities of panel members.

- 9) In EM, the typology of research organizations (RO) (Sec. 4.3.1., pp. 60-61) is processed and the roles of RO based on their purpose and focus are described in it.

From the above, it is clear that particular RO categories differ and have their own characteristics. Those specifics should be taken into account in the evaluation of RO as well. This fact is not included in the current version of EM.

Note: To set up the evaluation system in such a way so that it respects the specifics of individual RO categories.

- 10) In EM, 5 main Assessment criteria and a number of additional criteria resulting from them are indicated (Exhibit 24, p.63). In the evaluation draft, Assessment criteria are described in detail (Sec. 4.5.2., pp. 66 and downwards) and a rating scale Starred quality level in values 4-1 for specified criteria is shown. But EM lacks determining the weight or significance of the individual criteria.

Note: To determine the weight-significance of particular Assessment criteria.

- 11) The document contains a list of potential uses and benefits resulting from implementing an evaluation. Also, a possibility of using the methodology for the purposes of allocation of institutional funding is suggested here. However, the document lacks any instantiation of this area. Under „Tools for the Evaluation Exercise Implementation“ it is mentioned on page 7 „In the EM2015 it is also foreseen that this categorization may feed into a different weighting system for the allocation of institutional funding“. This confirms the planned use of the results of allocation assessment. However, the document lacks a description of how to use the evaluation results for the purposes of allocating institutional funding.

Note: To complete the document with a suggestion of using the evaluation results for the purposes of allocating institutional funding.

The specific comments and recommendation

Page 62, par.2, 3.

In line with international practice, we have therefore set the threshold for the participation of a EvU in the evaluation – and for the registration of a Research Unit by the EvU - on **50 research outputs** within 1 field of research over the evaluated period (i.e. 5 to 6 years).

An EvU can register for the performance assessment more than 1 Research Unit **only** if at least 50 research outputs have been produced in more than 1 scientific field during the evaluation period

Why even 50? The output should be weighted alike in the actual methodology.

(It was mentioned also in previous a review and it was not taken into account)

Page 72

The list of Research outputs (for example, Exhibit 27, p. 72) is incomplete and totally ignores such outputs as pilot plants, proven technologies, prototypes, industrial and utility designs. Failure to include those inputs significantly reduces evaluation and assessment of organizations involved in applied research, particularly technically oriented university departments and research organizations defined as RTOs (Exhibit 23, p. 61).

Note: To include pilot plants, proven technologies, prototypes, industrial and utility designs into evaluation and enumeration of possible Research outputs. (opinion of representatives of Industrial research institutes - AVO)

Page 73,

☒☒ Co-publications of researchers active in the same EvU but registered in different RUs will be counted as one each, on condition that they are of a clear *interdisciplinary* nature

☒☒ Co-publications of researchers active in different EvUs will be counted as one each, thus keeping the incentive for collaborations among the research organisations

We do not agree. It enables a gaming. We strongly recommend shared results.

(It was mentioned also in the previous review and it was not taken into account)

Page 9

The Small Pilot Evaluation, which implements the methodology in the context of 18 research organisations. The lessons learned from this exercise will regard in particular the robustness

of the processes; it will also give us a view on the potential costs for the implementation at a national level

We miss a list of their Disciplinary Area and Field

Page 62, par.2, 3.

In line with international practice, we have therefore set the threshold for the participation of a EvU in the evaluation – and for the registration of a Research Unit by the EvU - on **50 research outputs** within 1 field of research over the evaluated period (i.e. 5 to 6 years).

An EvU can register for the performance assessment more than 1 Research Unit **only** if at least 50 research outputs have been produced in more than 1 scientific field during the evaluation period

Why even 50? The output should be weighted alike in the actual methodology.

Citation impact – purpose:

??Percentage Web of Science publications exhibiting international collaboration in the addresses ...

??Percentage Web of Science publications exhibiting national collaboration among Czech institutions in the addresses

Why not also SCOPUS? (It is used below – see Chap. 4.5.6, and it is mentioned also on page 41)

The list of Research outputs (for example, Exhibit 27, p. 72) is incomplete and totally ignores those outputs as pilot plants, proven technologies, prototypes, industrial and utility designs. Failure to include those inputs significantly reduces evaluation and assessment of organizations involved in applied research, particularly technically oriented university departments and research organizations defined as RTOs (Exhibit 23, p. 61).

Note: To include pilot plants, proven technologies, prototypes, industrial and utility designs into evaluation and enumeration of possible Research outputs.