


Pilotní ověření návrhu nové metodiky hodnocení výzkumných organizací

Pilot Test of New Evaluation Methodology of Research Organisations

**Samostatný doplňující dokument 8:
Komentáře členů hlavních a oborových panelů
k metodice hodnocení a pilotnímu ověření**

***Background document 8:
Main and Subject Panel Members'
Feedback on Methodology
and Pilot Test***

The document summarizes feedback to the Methodology and pilot test received from main and subject panel members. The document was prepared by Hana Bartková, Andrea Weinbergerová, and Vlastimil Růžička from the received responses to twelve submitted questions.



Tento dokument byl zpracován v rámci Individuálního projektu národního pro oblast terciárního vzdělávání, výzkumu a vývoje „Efektivní systém hodnocení a financování výzkumu, vývoje a inovací, CZ.1.07/4.1.00/33.0003“. Projekt byl realizován Ministerstvem školství, mládeže a tělovýchovy a financován prostřednictvím Operačního programu Vzdělávání pro konkurenceschopnost z Evropského sociálního fondu a státního rozpočtu České republiky.

This document has been prepared as a part of the Individual National Project for the area of Tertiary Education, Research and Development „Effective System of Research Financing, Development and Innovation, CZ.1.07/4.1.00/33.0003“. The project was realised by the Ministry of Education, Youth and Sports and financed by the Operational Programme Education for Competitiveness of the European Social Fund and state budget of the Czech Republic.

Dokument „Main and Subject Panel Members' Feedback on Methodology and Pilot Test“ neprošel jazykovou korekturou.

List of content

1	Q-A. What, in your view, are the most important positive features of the proposed new R&D Evaluation Methodology and Funding Principles (“Metodika”)? Why?	3
2	Q-B. Please describe negative features of the Metodika.....	10
3	Q-C. Please provide comments and/or suggestions you would like to add regarding the future Metodika?	20
4	Q-1. Your comments on evaluation on the RU level vs. EvU level.....	28
5	Q-2. Are there five assessment criteria relevant and properly formulated? Any suggestion for change? ...	31
6	Q-3. Formulation of questions and structuring of the self-assessment report according to five criteria. Were all the questions (asked of the Research Units) useful? Would it be more advisable to limit the questions to major around ten and have an additional data-gathering section?	35
7	Q-4. Do you consider on-site visits important or even indispensable? Would an interview of RU staff members by the panel in a neutral location be an option instead?	39
8	Q-5. Do you consider grading of selected outputs by independent reviewers useful?	43
9	Q-6. Do you agree with assessment of only a small percent of research outputs or would you prefer to receive a full publication list? Would you be happy with a quality judgement based purely on bibliometrics, rather than independent review of selected publications?	46
10	Q-7. Do you consider bibliometric analysis important for your disciplinary area? What are your comments and/or suggestions to the content of the bibliometric report?	48
11	Q-8. Do you consider a calibration exercise a necessary and useful part of the Metodika?	51
12	Q-9. What are your comments to the pilot testing (overall organization, materials submitted, panel meetings logistics)?.....	53
13	Other comments	57

1 Q-A. What, in your view, are the most important positive features of the proposed new R&D Evaluation Methodology and Funding Principles (“Metodika”)? Why?

Main panels

1. The most important feature of the new proposal is the combined approach in which panels of internationally acknowledged experts evaluate research units according to 5 well defined evaluation criteria. This approach combines bibliometric and other data with expert opinions; the combination seems to be well balanced and suitable as a basis for decisions on the further development of scientific institutions in the Czech Republic. It must be added that performing a self-evaluation to be read by experts in itself can be a highly useful exercise for an institution as well as for the individual researchers, especially if they have a constructive attitude towards the exercise.
2. In my opinion the most important feature of the new proposal is combined approach using both bibliometric method and expert evaluation.
3. Important features
 - That it makes a clear effort to be demonstrably fair to all research units.
 - That it takes account of the different types of RU, not assuming that all are engaged in the same types of activity
 - That it takes account of all of the types of activity in which the RUs are engaged – fundamental and applied research and PhD training.
4. The most positive is an approach of complex evaluation of the RU or EvU based on detailed analysis of data provided by high-level experts with international experience having chance to compare national level with international performance. The positive feature is also qualitative assessment excluding very demanding quantitative approaches.
5. Qualitative assessment of RU and/or EvU and national/international comparison of their R&D performance play important role for governmental authorities which are responsible for tuning of finance distribution supporting sciences, research and development in the country. International experiences of panellists and evaluation methodology are very important feature without any doubt.
6. The Idea to implement a new evaluation methodology and funding principles is important for several reasons: First, it creates a new peer-review based evaluation system for the whole country. Second, it provides for the first time the tool for the evaluation on various levels (disciplines, research institutes, universities, departments etc.).
7. We evaluate highly positively the fact that the methodology has diverted from the mechanical “coffee grinder” quantitative method and focused on key publications of the evaluated units. A further positive feature is that the funding is not based on purely one criterion (evaluated publication outputs), put attempts to take into account other aspects important when determining (awarding) the scientific potential of the evaluated institutions.

Subject panels

1. Chemical Sciences Panel has evaluated 9 RUs in 3-4 real days. Each RU was ascribed to two designated experts (a main expert (Expert 1) and a secondary (Expert 2) a few weeks before the meeting. Documents were available before the meetings on a dedicated website. Experts 1 were in charge of preparing a draft report with the help of Expert 2. They suggest notes in the pre-review to start the discussion. All scores and texts were fully discussed during the meeting. Debates were sometime intense and long, finally converged to a common view. For some RUs, it results in different appreciations than the original suggestions showing the interest of a common evaluation during the meeting. After debates, scores were not the scores of a single expert but became the ones of the panel based allowing a similar standard for all RUs. Expert panels working remote to prepare the meeting is a good method for doing this evaluation in a cost-effective way. However, sessions dedicated to the meeting and discussions were obviously too short (see below).
 - On the whole, the general evaluation procedure as considered in “Metodika” project, follows the international standards as it could be found in several European countries. It remains that the procedure should be adapted to the specificities of the research and organization in the Czech Republic. The five main assessment criteria are well chosen but contains some overlapping depending on their interpretation by the evaluated units. This has led to some misunderstandings when the units have filled the forms.
 - At the beginning of the meeting, experts had a discussion concerning the “A” rating but exchange with the other panels show that this point remains unclear. In this respect, a better calibration exercise is needed to ensure a similar judgment between panels.
 - On-site visits were found as a very important source of information. On-site visits are strongly recommended for the reviewing process of the RU, which is assigned to a referee. It should follow the examination of the documents to be efficient in a short time. Duration of the visit should be adapted to the size of the RU.
 - A strong positive impact of the method is the self-evaluation that allows careful self-reflection of the RUs under evaluation. The procedure provides a mirror for controlled self-reflection.
 - As a general statement, top experts in a particular field can only evaluate the scientific excellence. A panel of 6 experts could hardly contain a specialist of all domains. Thus, the assistance of external experts complementing the expertise of the panel experts is a useful component for a fair evaluation. However, a general procedure of grading of selected outputs by independent reviewers (as done in the project) is less efficient and does not solve the difficulty of finding the right experts. This procedure was often seen as a simple reevaluation of a work that has already been reviewed by the journal before publication and advices were probably based on the journal ranking. In case of cost limitation, the majority of panel members suggests that visits be privileged to the “selected output procedure”. Panel experts have a sufficient knowledge to judge the quality of a publication based on the journal name.
2. The methodology encourages institutions to assess their own performance against excellence defined on a global scale. A result of this could be a better understanding by them of their competitive positioning and could positively influence their goal setting. The inclusion of a SWOT analysis in the self assessment is also valuable. The collection of the financial statistics regarding the main types of funding is also useful and the changes over time give reviewers a good insight into the trends. This coupled with the staff numbers and output of trained persons gives a good common basis on which to carry out comparisons. The use of

bibliometric information is an important element although this has the potential to be overcomplicated and possibly misleading. Use of simple 5 point scale in 5 categories is good and temptations to make a finer scale should be resisted.

3. The most important features of the “Metodika” are 1) the reviewer opinion as main basic principle; 2) the quality of scientific activity (output) is proposed as the main principle; and 3) application of some numerical parameters to characterize specific scientific activity of EvU/RU as supporting material for panellists.
4. International compatibility and selection of criteria
 - Criteria and methodology follow international examples and upcoming standards. In this respect, the evaluation is to some extent pioneering, particularly taking into account that an evaluation raster designed for the research organizations in UK (and adopted by some other countries) is imposed to the research organizations of a country where universities and institutes of the Academy of Science follow a completely different tradition and history. I think that the strongest impact of this method based on self-evaluation is the careful self-reflection of the RUs under evaluation. The procedure provides a mirror for controlled self-reflection. This is good.
 - The five main assessment criteria are well chosen. They are to some extent orthogonal. However, if a final over-all scoring will be done in one or the other way (e.g. by impacting funding), these criteria should not attain the same statistical weight. Good scores in “membership of the national and global research community” cannot compensate a lack in “scientific research excellence”, etc.
 - Scientific excellence can only be evaluated by top experts in the particular field considered. The assistance of external reviewers complementing the expertise of the panel experts (which cannot be as near to the research topics in all cases as external reviewer experts can be) are an essential component of a fair excellence evaluation.
 - A field-specific and RO-type-specific interpretation of the assessment criteria is targeted. In this respect, however, a better calibration exercise is still needed.
- 2. Cost effectivity Expert panels working remote is a good method for doing this evaluation in a cost-effective way. Computer based bibliometric data are fine and time- as well as cost effective for supporting expert judgement.
5. First time evaluation of Czech science, which can be used for future benchmarking and alteration of the current system to prepare Czech science for the future. It is very clear that the groups are not used to an evaluation and there is also little awareness about the need for such assessment. In the future when there is a serious assessment it should be clear upfront what the purpose is of such evaluation.
6. Important features:
 - First, I would say, the will to find an objective way to evaluate university and research centers in a country, that has undergone deep changes for two large decades, in particular with the recent creations of so many formation and research centers, often outside from Prague or large, well-known cities.
 - The sincere will to make the country “progress” toward other European countries (but everything is not pink and white in “western” Europe...)
 - The criteria selected for evaluation, even though a slight interference between criteria 3 and 4 really exists.

7. A.1 The Research Unit Evaluation Report:

- Completeness: the five quality criteria cover the full range of research activities and outputs pretty well.
- Internationally applicable and compatible scales: the quality levels at each criterion span the entire spectrum from an underperforming national research player to globally leading top level research groups.
- Comparability: Key words are offered per research quality level to anchor qualitative assessment of research quality level. These key words enable a more structured discussion and debate between evaluation experts and make things better comparable. This is true even if one accounts for some different personal interpretation of words because of different cultural and linguistic backgrounds of reviewers.

A.2 The Self-Assessment report & Registration form:

- Aspects of managing human resources in research and career development are properly addressed. The form gives room to express a research strategy, to offer an overview of current research activities as well as the funding and infrastructure aspects. Though it is noticed that Research Units vary in their adeptness to present such topics. The form features research (output) excellence and performance as well as societal relevance. These features offer a wide enough space to position research for a diversity of institutions, in line with their missions. Institutions differ by their mission, which may range from studying fundamentals in some branch of science to development and pilot testing of engineering applications.
- The inclusion of HC (head count) and FTE (full time equivalent) is experienced as useful and should be contained. In spite of the suggestions given for the criteria, the descriptions as given for each criterion were clear and largely not liable to interpretation. Again, in spite of the suggestion for minor additions, the bibliometric information given is useful and should be contained.

8. The proposed new R&D Evaluation Methodology and Funding Principles represent a comprehensive and well-structured guideline for the RUs and EvUs. The relevant criteria are well defined. It is also the impression that these criteria can be addressed by the RU and the EvU. If these guidelines are followed carefully then it should be possible to obtain a clear picture of the organisation of the RU(or EvU), the performance of the RU(or EvU) as research unit, the place of the RU (or EvU) in the national and international research community and the impact of the research. The current set up of the proposed new R&D Evaluation Methodology and Funding Principles has a structure that makes it possible to include future improvements without changing the basic structure
9. The methodology enables and underpins comparison and ranking of RU's and EvU's in the Czech Republic, one of the main goals of the Evaluation
10. The attempt to introduce in the CR clear rules (Metodika, following EU-like models) for an evaluation system for the whole scientific landscape of the country is a great step forward – a great deal of the IPN and the Ministry! It will foster the research activities via a more quality driven mode of funding by state authorities who might enable institutional flexibility needed to react on new trends. International evaluators will have three sources for their work: self-assessments of the RUs, peer reviewed outstanding products (mainly books) plus complete book lists, and the talks with the management and the staff during site visits. This worked well for the pilot phase, but even here only for half of the RUs.

11. The new R&D Evaluation Methodology promises to replace the purely quantitative criterion used so far with a qualitative assessment. The evaluation is carried out by independent scholars who should not have interests in common with the institutions evaluated. This kind of evaluation can be used for all research centres (universities, academy of sciences, archives, etc.). Once implemented, the system can lead to a gradual improvement of individual research centres, which can more easily identify their strengths and weaknesses.
12. A good opportunity to support aspiring faculties, helping to identify problems in their strategies and to discuss options for future action. A valid and reliable assessment seems possible even if there are some open questions. The most important part is the self-assessment report combined with the list of books. It is unclear how to attribute weaknesses "of the RUs" (but not truly their weaknesses) such as the absolutely insufficient funding of doctoral students.
13. The possibility to provide an independent evaluation, based on foreign participation in order to overcome the difficulties given by the size of the country and the inconvenient of a narrow Czech community of researchers. The opportunity to share experiences and points of view from different countries in the process of evaluating. The effort to search criteria and patterns of evaluation that allows to approach indifferently natural sciences and engineering, social sciences and humanities. The possibility to include in the procedure institutions that are very different in nature, size, mission etc. That may bring surprising picture. (Why for instance some very problematic topics from political and societal point of view are studied in some place rather than elsewhere. That could help also to rethink the relations between Academy of Sciences and Universities. The effort to mix quantitative criteria and qualitative criteria, a very crucial objective to support and foster research in "soft sciences". (As we can see during the panel, a lot of questions stay opened).
14. The principle of evaluation on national and international level is the key for improving the quality of research, as well as making it relevant nationally (providing argument on societal level) and internationally (regarding the level of national research). Adaptation of EU(in this case perhaps mainly UK) -inspired practices is undoubtedly a big step forward. Internal evaluation systems have existed/exist on both EvU and RU levels, but they are not very functional or reliable (lack of comparison, personal relations/nepotism, insufficient funding, not considered as important by the staff, etc.) thus evaluation on a national level makes the evaluation system transparent, impartial (hopefully) and relevant. Indeed, the evaluation system itself is a part of the process of defining what is good /excellent research, which is an important development even in humanities, where the issues of "quality", "impact", "competitiveness", or even "profitability" are much more difficult (if not impossible) to apply. A positive experience of the Metodika Pilot testing project is the existing awareness of the specific situation in humanities in comparison with hard sciences (natural sciences, medicine, etc.), not least the difficulty to use bibliometric data when assessing the quality and competitiveness of the research, and so the necessity to introduce other quality measures in humanities. Naturally, this makes the evaluation in humanities more time consuming (and so resource requiring), but it is an absolute necessity. Another aspect regarding humanities, that should be reflected more in the future evolution processes, is the specific situation of humanities in small countries (small languages). For example, scoring as Global Leader (A) or Strong International Player (B) are sometimes meaningless, if the research field is specific part of Czech history or literature; there is nothing to compare with internationally. But, and it is important to stress that, all history research is not "only" national relevant and specific, and even Czech historians (and humanist researchers) can play a global leader roles if the research

tasks are defined in international context (examples from global, environmental, or economic history show this clearly); even humanist research is internationally comparable and results are transferable.

15. As I could not take part in the first calibration meeting in May in Prague, I cannot see the exact evolution between the frames or models; nevertheless, I think I will address the issue below (answers to 2). But above all, I would like to specify that my appreciations could be influenced (“distorted”) by my (rather long) experience in the French Assessment system (AERES-HCERES). That said, I would try not to be “ethnocentric” because the systems are different. For instance, the French RU are (scientifically) independent from the EvU (Faculties or Universities) where, nevertheless, the RUs are located. So, the on-site visits above all consist on meetings with the RU, but also (this issue will be addressed one more time below), on (separate) meetings with which is called “tutelles” (“supervisory authorities”, given that the RU are not independent from the EvU from an institutional, i.e. financial, juridical, point of view)
16. The most important positive feature of the new evaluation methodology is the detailed questionnaire for the EvU. The universities and the RUs will receive now not just a questionnaire, but also some principles, how to develop their organization as well their research activities. The universities and other research units need a stable and calculable funding system, and the questionnaire has to be in concordance with the principles of the funding system. I think, this methodology and the questionnaire is clear and logic, so you can use it. However I have some suggestion, to make this questionnaire better: I would like to give points, and not just 5 categories to score the results. A, B ... to E are just 5 categories. The Czech Republic has a lot of universities so you can’t make a difference among universities which are relatively similar. For each category I would introduce 10 up to 50 points, so can the universities accumulate for example 5×50 points = 250 points. In this case the evaluators have the possibility to make a better difference. On other example: In the Romanian system each employee from one university has to complete a self-assessment. Each possible activity has a score, so the Romanian colleagues accumulate points. The RU, in the Romanian case the faculties have the score the sum of the point from the employees. I attached this questionnaire and my own score. This method is pretty complicate, because the template is not enough for long list, which had to put in the template. However, the Metodika team can try, to make de evaluation more precise.
17. The will of promoting institutional self-awareness and the hint given to research beyond teaching duties; the encouragement of a non-self-referential institutional attitude, rather of a systemic one concerning the whole academic organization, on a national as well as international level.
18. The abandonment of a methodology based on mainly quantitative data about publications, citations and impact factor indicators in favour of a much more qualitative approach based on peer reviewing is clearly positive. A strong emphasis on quantity does not yield an adequate basis for the evaluation of research quality in general and may have detrimental effects on publication practice and research ethics. It also suffers from not being sufficiently field specific: the definition of quality criteria needs to be sensitive to the fine differences between the humanities, the social sciences and the natural sciences with regard to the nature and societal function of research. The invitation to write the self-assessment reports has the potential to strengthen the self-awareness of the institutions concerned. It allows them to think about their past, present and future in ways that stimulate them to improve their research. However, to successfully do so, the questions and format of the self-assessment

reports become very important. In this sense, the intention behind the new evaluation methodology is commendable, given the situation in the Czech Republic so far. At the same time, this is a unique opportunity to explore productive and reasonable ways of evaluating scientific research, researchers, and institutions.

2 Q-B. Please describe negative features of the Metodika.

Main panels

1. There are no obvious negative features with the main general setup, although some may emerge when the system is put into practice. The presence of a human factor must be acknowledged and should be simply taken into account by all parties concerned. It may be considered to give the RUs a right of appeal against decisions of panels to an independent controlling board. However, this should be designed so that appeals do not take overhand. Site visits seem to be a necessity in several cases. In cases where no site visits are made, the panels should be given a chance to clarify outstanding questions in a meeting with an RU leader. The evaluation is rather complicated and expensive in comparison with the present system. It is important to define the principal goals of the evaluation. If it is only a tool for distribution of a small part of the funding among institutions, it may be considered too expensive. Another problem with such a system may be that the grades of RUs within an EvU may vary considerably, but the funding will not go to the individual RUs, but to the EvU. The expected interval between evaluations is 5 years. The preparation of the evaluations requires much work, possibly up to one year for the RUs and EvUs as well as for those responsible for the evaluation. If all evaluations are done the same year, the resulting variations in the work load between evaluation years and other years may be inconvenient, especially for the latter (evaluation) group. Some RUs and EvUs may worry about both the amount of time they have to spend on the evaluation and, in particular, what the evaluation will be used for. They are likely to wonder if a simple relationship exists between the grades and future funding? It may be possible to further minimize the bother the RUs and EvUs have with filling questionnaires and at the same time improve the information provided, based on the experiences reported by the RU panels. If it is planned to let future funding automatically follow the grades (although this is hardly a perfect solution) it will further increase the worries about the grades. It may be better to say that the grades are part of the background for funding decisions, and that no automatic relation exists. The roles of the members of the main panel should be more precisely defined. If their task, in addition to assist the subject panels, is to take care of the evaluation of EvUs, this should have been more clearly specified from the start.
2. I do not see any strongly negative features at this time although some may emerge when the system is put into practice. There is always the human factor present but since it cannot be removed it should be simply taken into account by all parties concerned.
3. That it allows RUs to submit an arbitrary number of outputs, not linked clearly to staff (I would link the outputs to specific staff more strongly, so that the number of outputs relates to the number of staff, OR ask for a full publication list.) Too much information is asked for in the self-assessments. I liked the idea suggested that there should be two sections: data-gathering (automated as much as possible) and a shorter textual part.
4. The main negative feature is unavailability of subjectivity of the evaluators. From pilot testing seems that VSCHT is evaluated too positively compared to other RU/EvU having big respect to this University.
5. There are evident differences between individual RUs in one EvU. But, research finance transfer is addressed to EvU. Its top-management plays crucial role because strongly influences the balance between RUs in future perspective

6. A problem of the testing of Metodika was the time pressure (short term evaluation in a limited period) that affected the self-evaluation reports as well as the evaluation of the publications.
7. The system of evaluation as presented is **clumsy** and **erratic**. Evaluation of key outputs is compensated by additional, in their defined format objective, but without direct examination of the evaluated unit *in situ*, hardly verifiable evaluation parameters. This concerns both the equipment and the operations of the institution, as well as its public (e.g. media) activities. However, when this form of evaluation methodology is applied in practice, it will not be possible for the committees (groups of panel members) to visit all the units evaluated. As a result, a substantial proportion of the so far considered “hard data” criteria is going to turn into unverifiable conclusions. Thus, a “feeling” or “perception” that the committee will have from the written source material, provided by the evaluated subjects, will play a key role. Their quality (including how well the given institution will predict what the evaluation committee wants to hear, or read) and, even more likely, “presentation appeal”, is going to play more of a role in the end-result of the evaluation than real productivity, quality and the situation of the given workplace.

From the point of view of Czech participants, i.e. experts of the main panel in the disciplinary area of Humanities, a number of aspects that could potentially negatively impact the objectivity of evaluation appears: Unfamiliarity with the decision-making process regarding the projects in terms of very limited funding of projects through GAČR, i.e. the current limited possibility of Czech research to do conceptual, and, especially long-term research plays a key role, as well as minimum insight of mainly foreign referees into a largely complicated system of funding of research and publicity (popularising, i.e. particularly with “civic involvement”, thus not only “generally educational”) activities of faculties, institutes and, for example, archives.

This also concerns evaluating scientific outputs. As long as institutions do not know what the panels mean by “key” publications, they will submit the “scientifically best” they have produced, not only thinking that “it is necessary to introduce all segments” of the institute. When an institution already has a new evaluation experience and time to apply it, it is going to prepare a tasty “sandwich” of the key publications that the panel wishes to see. This means that in the perspective of three (perhaps even better five) years the faculty or institute is going to reduce funds to all its researchers and workplaces that do not structurally have the chance to produce the outputs required, and, without much consideration of scientific relevance, will primarily support those, that are going to “appeal” to the committee.

Already during the discussions of the pilot panel this showed clearly throughout the debates regarding current and absent “excellence” of outputs, which do not have the features of a research study, i.e. specifically for example historical **series**, and the relevance of whether they should be prepared by scientists at universities or AS CR. Though, generally, it is understandable that Egyptology series on material, papyrus and the Dead Sea or even illegible medieval tractates that can hardly be interpreted do have the features of scientific work, the production of the medieval charters series at the AS CR institute was strongly doubted by the foreign experts, given the fact that exactly the same can be done by the archive itself, which stores the documents. Furthermore, presentations of editions of modern e.g. diplomatic material were very clearly rejected as a form of significant scientific output.

This debate is the consequence of the fact that the foreign experts that were addressed preferably, considering their ability to read scientific work in Czech, approached the

evaluation very subjectively and according to their own narrow research background and specialisation (mostly modern historians or contemporary historians), the tradition of their country of origin and with an a priori idea of the “ideal” type of scientific output.

It was interesting to observe how a whole range of problems was gradually clarified during on-site visits of the evaluated workplaces and during discussions with their representatives and workers on specific aspects of scientific work in Czech conditions and the specific forms of its outputs.

Subject panels

1. Important features:

- Time dedicated for the evaluation was obviously too short. Chemical Sciences panel had 9 RUs to evaluate in 3-4 days and it is difficult to treat more than 2-3 units per day. This made difficult to finalize the reports especially for the final step dedicated to the homogenization of the ranking. Notice that it is almost impossible to make standardization later when the meeting is over.
- Several times during our panel sessions, panel member came to the conclusion that a final assessment of the evaluation methodology is essential. Unfortunately, we had not enough time to collect all the comments given during the panel discussions and bring them in a final form during the meeting. This was done afterward, but it would have been better to do it during the panel meeting.
- Common analysis of EvUs – Institutes of Academy of sci., private institutions and Universities by “one common approach”, creates difficulties for the evaluation on proposed scheme. Classification of fields was often not adequate as not sufficiently precise especially for large RUs.
- Differentiation by OECD classification of the disciplinary areas and scientific fields is not correctly applicable for all EvU without accounting character of organizations. It also creates heterogeneity of size of the RU under investigation as the report template and methodology is unique.
- Some of the RUs were large (some small) and could include researches in different domains of expertise combining basic and applied works. Heterogeneity between the RU size makes more difficult to provide a fair and equal evaluation for all of them. For the largest unit, it limits the evaluation to a superficial view that may be not even useful for the unit itself.
- Universities have as the first target the training of masters and PhDs, have possibility “to use” students for scientific works as well as some staff members. In contrast, Academy and private institutions, especially the last ones, have other missions, limited students (PhD) and “project concentrated” money. Pure science for private institute occupies “secondary” place. These differences do not appear in the questions/answers or reports templates.
- Some other difficulties have arisen from the evaluation of specific RU that were more “service dedicated” units than a research department, this is notably the case for analytical laboratories or services. If such units are separated for evaluation, it’s desirable to present data about “work-hours” on service and “work-hours” on “own science”. It’s

clear that the service function is very important also for whole EvU and other users as so called “instrument shared access center”.

- Questions asked to the RUs are often too general and lead to misunderstanding of the unit. In the responses, we often found repetitions or they used a wrong criterion. As example, when the questions asked is “Do you offer courses to the public”, the correct answer is “yes”. Something like this is not useful for the referee, who is interested to learn something about the success and other specific details. Question should clearly distinguish for the evaluators as well as the evaluated RU/EvU, if they are addressed on the level of a RU or on the level of the EvU. Some confusion was observed e.g. regarding the placement of “research strategies”.
 - In many situations, it was not clear how RUs are integrated into the EvU and what connections and limitations exist between the different RUs inside the EvU. A number of quantified evidences relating in particular statistics (local, national and international) could directly included in the text of the reports of RUs and EvUs.
 - We had a problem to define the right scaling of the different notes. Discrepancies seem to exist with the notation from the other panels. In a few words, this could be translated to the definition of a world leader. The percentages of the category were missing in the template and thus lead to personal feelings.
2. The two main things struck me as a reviewer. One was that the questionnaire information is described so each of the 5 blocks of questions map exactly to the 5 evaluation categories. However I have the feeling that all of the information in the self assessment may have some relevance when assessing a particular category. This should be looked at and reviewers given some more direction on whether to assess purely on the answers to questions assigned to the category being reviewed. The second thing was that the bibliometric data was overwhelming and required a lot of study to start to understand which numbers were important. Also it seems to me that there are certain ratios based on bibliometric data such as output per researcher which reviewers instinctively calculate in order to make meaningful comparisons. Perhaps these should be pre-calculated and placed in a short summary. The selection and scoring of the Excellent Outputs was a problem partly because the grading of the outputs was late. The grades did come to assume quite high importance in the panel deliberations on the scoring, so both the selection by the Institutions of the nominated outputs and the scoring given by the evaluators of the papers becomes quite central to the results of the assessment. The integrity of both the EO selection and the expert scoring could be a weakness of the system.
 3. There are some problems in separation of one or two RU from EvU in the process of evaluation. May be, these problems arise only in the case of Academy Sci. institutes and similar research organizations that have several RUs.
 4. It is important to define what should be the final goal of the evaluation. Depending on the answer to this question, it might become evident that ROs with completely different duties (Universities, Academy of Science, Private Companies, Governmental Institutions) should not be evaluated on the basis of an identical questionnaire. It might, for instance, be useful to evaluate different ROs within a group (e.g. faculties of different universities; different institutes of the ASCR, different private institutes which are sponsored by research money from governmental/EU/public funds, etc.) or to compare single criteria (e.g. research excellence) between groups (e.g. research excellence of university based EvU vs. ASCR based

EvUs. Then, however, many other questions (regarding PhD students etc. should be adapted to allow a fair comparison.)

- Main dedication of the RU: Evaluation should be focused on the main dedication of a EvU/RU. It makes a difference, whether a RU is focused primarily on research or its main purpose lies in supporting the work of the EvU. In certain cases the supporting (research) work of the RU might lie in a quite different area of science compared to the EvU. (Eg chemists supporting the conservation of historic documents for an archive with focuses on historical research). Research excellence (e.g. in chemistry) will thus not be the most important goal for this supporting RU. It might be rather a goal to gather, attract, conserve best-practice knowledge and to implement it to the EvU according to the special needs (Example for this situation is the function of the conservation department of the National Archive).
- Computer based pre-digestion of the self-evaluation data Let the computer make what can be made best by a computer; and let the panel of scientists make what such a panel does better than a computer. The work for the computer should include the digestion of all bibliometric data. This was done already very well. However, it might be helpful to confine Tables and Figures to those numbers which are actually significant for evaluating scientific excellence and over-all research output. Typical useful indicators are: number of publications in the group of best 10%-regarding citation; over all citation normalized to world average in comparison EU- and CZ-average. Not all outputs are of equal interest: Decisive are the numbers of WoS-listed publications, book monographs and book chapters, as well as patents. Conference proceedings are less important within a usual research environment. They should not be accounted in the same way as WoS-listed articles. Head counts and FTE numbers should be handled by computer and further used for “output/FTE” data, i.e. number of papers listed in the WoS divided by the FTE-researchers. In this respect the FTE assigned to pensioned/emeriti professors should be clarified and treated in a uniform way. When evaluating the success of a RU by the money gathered by competitive funding, it has to be considered that some funding institutions (eg EU) define local areas of preferred funding. Comparing RUs inside such preferred areas with those outside on the basis of a money-acquisition criterion is not fair. Regarding necessary fluctuation in personnel, in experimental sciences, unlike probably to some other areas of sciences, certain maintenance of experimental know-how is essential to keep experimental know-how on a high level. This requires that fluctuation in personnel staff must not be 100 % but at a level between 75 – 85 % for instance, depending on the excellence and the uniqueness of research in the RU.
- When establishing programs and career regulations aimed at reaching high fluctuation after finishing the PhD, the EvUs or ROs should take care not to close all direct ways of transition from PhD to a staff position. The word “inbreeding”, though common in this context, is too negatively connoted, as in biology no inbreeding at all is favourable. And this is unlike to the situation in science. What is needed is fluctuation/mobility which can sufficiently be attained even when recruiting 20 % of the staff from the PhD sources of the RU. The word “inbreeding” should thus be substituted by “insufficient mobility/fluctuation” which targets the goal to be attained much more accurately.

5. Artificial definition of RUs, poor understanding by RUs of what is being expected

6. Sorry for the non-politically correct point of view:

- The use of European concepts for so-called positive points in a calibrated Europe, into a country with slightly different ways of life, in particular the PhD track which is not as secure as in western Europe, mainly economic reasons
- For some applications the lack of professionalism in filling the form and commenting on the various items proposed. Also the selection of papers to be evaluated, some being great, other being of a quite poorer quality.

7. The Registration Form asks for a lot of written information and detailed data that must be collected by the EvU and RU. I notice there is quite some variability over the RU's in quality and focus of their narratives as well as in the consistency of their data. It is possible that an excellent writer of a modestly performing RU with an up-beat presentation outshines another RU with actually better performance but lacklustre presentation. So, there should be some way of checking on the truth content of what has been written. In that respect local cross-checking by site visits is essential.

8. The information as provided is scattered and confusing. Speaking in simple university terms we are dealing with departments (EvU) and laboratories (RU). It would be wise to have two simple files (or books). The first file could contain an overall report on the level of departments with a chapter for each laboratory. In the Department part, aspects like overall mission, strategy, HR, etc. could be dealt with while in the chapters on the laboratories details on their scientific staff and laboratory achievements could be provided. It should be considered to include the H-factor of the researchers involved as well as the number of invited lectures given. The second file should contain all published papers including full author list and titles. Preferably a link to the one-line papers should be given. A division should be made between regular papers, proceeding papers and other contributions. Each laboratory should be given the opportunity to indicate their, say five, high ranked papers of which the laboratory thinks that they provide the best image of their laboratory. In this way a simple and clear overview will result.

9. Too many sub-aspects are addressed. The feeling of the RU is that writing of the self-assessment report requires a lot of effort, maybe too much effort. However, in my opinion some relevant important aspects are not addressed (see later). 2. Looking to the experience of the subject panel Environmental Engineering in the evaluation of the three RUs the impression exists that the relevance and importance of all questions is not always clearly understood to the RUs. 3. In the evaluation formats used by the panel, but also in the set-up of the self-assessment reports there is in fact no clear difference between the research executed at Universities and research executed at the Academia of Sciences. Of course, this is currently the typical situation in the Czech Republic. However, the Universities have another responsibility than the Academia of Sciences. At the Universities the research is an integrated part of the education programme. Research is primarily executed by MSc students and PhD students. At the Academia of Sciences this is according to my impression less. 4. If I look to the research fields and research output of the three RUs we have evaluated, I have the feeling that based on my own experience and my knowledge of the international research in the field of Environmental Engineering not all research is at high academic level. Part of the research is at a more technical level and strongly focused on practical application. This research can also be excellent and innovative research and can be very valuable for industry. However, results are often not or less suitable for publication in scientific journals with a specific format and structure. Results are often very useful for obtaining a patent. To assess this type of research correctly in the evaluation framework requires, in my opinion, more attention.

10. In my opinion the Metodika has a too strong administrative (bureaucratic) nature, too much numbers and from a distance¹
11. It seems to be rather problematic to exclude teaching from the evaluation for two reasons: (1) in universities it is very much interlinked with the research (not only by the few professors in smaller RUs but by postdocs teaching as “docent” and working on their second books); (2) in extra university RUs, where researchers - even not obliged to teach - often transfer first-hand knowledge into seminars and provide supervising work for doctoral students. Bibliometric reports do not help in the research fields humanities. A compromise found at the calibration meeting in May was found in a full list of books. During the panel meeting in July several matters have been discussed: (1) the need of a full bibliography of the RU for the given period; (2) to enlarge the list of outstanding outputs to more than 1-2 percent (10 books at max); (3) to provide international reviews of the chosen books; (4) to make clear to the RUs that not from all departments/sub-units products have to be on the list but the most outstanding ones; (5) to clarify the division of labour among potential panel members and their possible involvement in the peer reviewing process– which will cause greater problems in the future since the international “market” of reviewers and evaluators is limited due to the language barriers – thus please engage more colleagues from Slavic countries. (Take use of the Visegrad-factor). Scoring is needed, no doubt, but only A-E is too much like woodcarving. In order to express a certain tendency there should be possible scores like (1) B-plus or A-minus; (2) B to C or D to C. Needed is a better synchronization of order and meaning of the 5 criteria as asked for the self-assessment and in the evaluation reports. Going thru our notes during the panel meeting “jumps” had to be made for instance concerning performance - are there meant products or research strategies, belonging to research excellency as well as to research environment?
12. I think that at the beginning to none of the panel members was entirely clear what would be the meaning and the possible use of this pilot testing. The system is clearly based on the exact sciences, to be properly applied to the humanities still needs many improvements. It seems it should be used better recalibration with the exact sciences to ensure the recognition of excellence in the humanities. The emphasis on the aspects bibliometric is excessive and for the same reason, the system as a model (REF) has already been repeatedly criticized. Regards to the human sciences it does not seem entirely balanced the relationship between various types of research centers (universities, academy of sciences, archives, etc.). All research centers related to the Czech context (Czech literature, Czech history, etc.) will be inevitably penalized by definition of excellence used (Global leader). The question of the PhD is too different from the standards of other countries to be included. The panel members who are not familiar with the Czech system will necessarily consider a failure of the RU a common situation, leading to imbalances between panels. There is a real risk that the composition of the panels in the humanities consistently affects the final results. From this point of view, I was surprised by the absence of experts coming from central and eastern Europe. I myself had asked to be moved from the panel of literature because I believe that members of the panel should be not come from the same cultural context.
13. The system is in part (bibliometric report) too strongly leaning on the British system whose *negative* effects on research are quite obvious and have been object of research already. Teaching was not considered, but the quality of teaching should be part of such an

¹ This is my opinion. The point I want to stress is that an evaluation based on numbers and indicators alone has the risk to be not correct due to different interpretations of the definition of the numbers by the interviewed scientists and reviewers. I consider site visits essential.

assessment. Many of us had assessments “b-c” or “a-b”. It was difficult to express expectations and experience in the rating, even if we talked a lot about future (universities) and the past (academy, museum). As it is the system does, in the opinion of most of the panellists, not easily give Czech RUs a chance to get an A-level. This requires recalibration with natural sciences. The many “B”-ratings, on the other hand, are not very informative. A- or B+, B- should be made possible.

14. There were some questions about the real meaning of this evaluation. We were many times repeated, that the result of our evaluation has nothing to do with how the financial means will be devoted – something we can imagine at this stage of testing, but hard to believe. Sometimes we grade with “B”, knowing that the job done may be is not “outstanding”, but can’t be really much better considering the age of the Institution, the condition given, the necessity to “update” the discipline etc. The competition with some “hard science” disciplines may then appear very bitter... Teaching was not a part of the evaluation, some panellist regretted that. I partly agree, but if we do include other missions than research, we should also include even more. The results of NTM (National Technical Museum) are really biased when we can’t take in account the work done for exhibitions. Same type of remark for National archives. A big part of the outputs of the Historical Institute raise doubts, the biographical dictionary of this Institution, in its incomplete paper version, is presented as “outstanding work”, but we are not interested in catalogues provided by Museums or Archive... is it fair? I feel really uncomfortable with the ABC... grading, often chose to put “B-C”, or “C-B”. The double reviews were not often available. Some of them said more about the reviewer than about the reviewed book, I suppose that some reviewers didn’t really understand the meaning of the different grades. That will be a real problem to find enough reviewers in the future, and to keep on with the rule “the evaluation must be carried out by “independent researchers from foreign countries”.

15. The test evaluation has been based on three inputs: self-assessment report, bibliometric report, and list of publications, combined with the available excellent outputs of the RUs. In addition we visited some of the RUs evaluated. The three inputs can be prepared in more efficient way, to make the panels’ work more impartial and less time-consuming. First, the bibliometric reports are of limited value at present form, it could be made shorter and, perhaps, lists could also include the index categories. Panels will need complete lists of articles (with titles) and books (edited volumes and monographs) to evaluate the status and quality of RU’s research. As many as possible of the publications should be made available in physical form to panel members at an early stage of the evaluation process (pdf is fine, but...). To assess the quality of publications the anonymous peer-reviewing (available in pilot testing) should be completed with real individual book reviews. These should be supplied by the RUs themselves. When journal/ newspaper book reviews are not available (time-lag etc.) there should be at least two different anonymous peer reviews available for the panel members. In the testing project, clearly, the most important part of the evaluation were self-assessment reports. The template for self-assessment report is designed for hard sciences, which include many questions not relevant in humanities. Obviously, there were also misunderstandings regarding the selection of excellent outputs; some RUs chose “representative” outputs instead of “excellent”, to illustrate the variety (width) of the research, more than excellence. Yet this also illustrates RUs’ general difficulty in defining their excellent research. Evaluation process in this way can help both RUs and panels to reach agreement on what is excellent research in humanities. Also among panel members there were discussions on issues of excellent research and overall research performance; where is the borderline. I believe the

distinction between criteria 3 and 4 is useful, it enforces the RUs to state what is their best outputs and so to reach agreement on the standard of the RU's research quality in the process of writing self-assessment report. In panel discussions many of us expressed fear that the evaluation process on national level could be used/abused to reduce funding of humanities, if scores reached are not be as high as in hard sciences (the problematic scoring as global leader or international strong leader has been pointed above). It is extremely important for the development of evaluation tools to stress that evaluation results MUST NOT influence funding relations between hard sciences and humanities, and that the results of evaluation (scoring) are ONLY applicable among the research in humanities in EvUs and RUs, not between hard sciences and humanities.

16. In the actual templates we dealt with, it is not always easy to distinguish what concerns EvU vs. RU. So, you will find, sometimes, in our own report on JUCB, designations like the EvU/RU: we used them when we were now able to make the distinction. The difficulty, I think (following my former experience in Research Assessment), can be explained by the fact that the RU are not autonomous: particularly, on the financial level, they are dependent on the EvU (if not the central departments of the University). The second main difficulty, I met or felt, concerns the adequacy of some questions to the domains we deal (dealt) with: humanities, languages & literatures. It is obvious, at least in the case of Pardubice, that the Self-Assessment Report (below SAR) created difficulties; thus, sometimes, the writers "demonstrate their protest" in their answers (for instance when they have to assess their "competitiveness" namely at an international level, as they consider and say that their university is a "regional" one: see Pardubice's SAR Q 054 and 060). On a more "anecdotic" level, some questions in the SAR template may appear as redundant: see for instance Pardubice's SAR Q012c and d).
17. The Metodika system is too strong international. For most faculties in the humanities is not possible to reach the level "global player", because the Czech history, literature, language, and also the Middle European languages have just a scientific public in the Czech Republic and in the state in the neighbourhood. Nobody will read in South or North America in Czech language about the Czech history. In the natural sciences is possible to make a new formula, a new discovery etc. but in the national sciences like philology, history etc. that is not possible. So I apologize for a different grading in the humanities. Czech humanists can be compared with colleagues from Germany, Poland, French, and Hungary etc. but not with the whole world.
18. The risk of judging before adequately knowing, of evaluating before describing, especially if the only source of knowledge is the RU self-assessment and no on site visits have been accomplished. The risk of limiting the evaluation to the present situation, without knowing historical or contextual elements which could be helpful to explain what can or cannot be changed.
19. The Metodika shares the weaknesses of many other evaluation methodologies inspired by Anglo-Saxon models (and now increasingly modelled on the UK's Research Excellence Framework). It acts as though universities and research institutions are not rooted in the society and culture where they grew and that they can therefore all be evaluated in the same way. In reality, any evaluation methodology needs to take into account the specific cultural and historical context in which universities and research institutions have developed. In the case of the Czech Republic and other continental European countries, the role of the university as a bridge between generations needs to be central to the evaluation: How does research feed into teaching and the transmission of knowledge from generation to

generation? How does it stimulate the intellectual life and technological development of the region where the institution is located? How does it build on local traditions and carve out new paths? The formulation of criteria in terms of 'global leadership' or being a 'strong international player' is often indicative of this type of approach that neglects the socio-cultural roots of institutions. What do we want Czech universities to do? To copy American and British universities that often have much larger funding and can attract a variety of foreign faculty? It is important to avoid giving the impression that the evaluation methodology compels the concerned institutions to break radically with the traditions of research in their own country. Such an impression may lead to resistance and unwillingness to participate. This would be tragic in view of the fact that the strength of universities in this part of the world in important ways resides in their organic link to society: their ability to attract faculty and students from the surrounding region; their capacity to educate and form the intellects of the population. Merely introducing peer reviewing will not remedy the problems caused by quantitative or bibliometric approaches. It may still discourage ground breaking, unorthodox and 'unfashionable' research, even when this research has substantial cognitive value. Panel members must therefore be carefully selected to ensure that they are open to a variety of approaches and innovative methodologies in the respective fields of research. Too many elements in the evaluation methodology are still coming from the exact sciences: counting the number of articles remains important, even though this has limited value for assessing research in the humanities. It is unclear what 'a driver for societal development' is, particularly in the case of the humanities whose very core is the transmission of knowledge and culture from generation to generation. The very notion of a 'research unit' needs specification in the case of HEIs, where much of the research in the humanities is carried out. A university department is a unit set up for the purpose of teaching a general subject; if it is evaluated as a RU, other criteria will have to be used than those that apply to RUs that have been organised in order to do research in a more narrowly targeted field. The evaluation exercise could be an incentive to start building research groups in the humanities, but then the departments should be allowed to decide which institutional unit they want to present as an RU for the evaluation. This might not correspond to an educationally oriented department in a university. Often it is difficult to find the right way to compare the different institutions. For instance, how can you compare an academy of sciences with a university? In the case of universities, it becomes extremely important that the research done by a research unit feeds into their teaching. The link between research and teaching is what allows universities to transmit knowledge from generation to generation. But this will also affect the publication record of the RUs in a university. The Academy of Sciences, in contrast, plays a different role. In so far as there is a need to focus on excellent research as one of the five criteria, the basis for evaluation should be rethought. A small number of selected publications will not reflect the overall quality of the research that is being done at an RU. The evaluation panel needs access to different kinds of publications of the RU: publications of different members; chapters from PhD dissertations; publications based on PhD dissertations.

3 Q-C. Please provide comments and/or suggestions you would like to add regarding the future Metodika?

Main panels

1. The difference between quality level A and quality level B seems to be very small for all 5 criteria. One might let the system work for some time and make possible changes later. However, already now an issue has appeared in a case where one RU has received some A grades, while another, comparable RU has not. If A and B grades were merged, this issue would disappear. One might possibly have to consider the need for one more level in the D and E region in cases of weaker RUs, for example outside the sciences. In future evaluations the purpose should be clearly specified for all involved. In order to ensure a relatively constant work load for the evaluation staff, one might consider a rolling evaluation plan, in which 1/5 of the units are evaluated every year. Another advantage in this system would be that improvements of the evaluation method, based on evaluation experiences, may be gradually introduced in the evaluation strategy.
2. I suppose it is too early to propose any significant changes or improvements. Let the system work for some time and make any changes later.
3. Definitely, having passed the discussion on this subject, I would recommend that part of real evaluation should be ½ day site-visit comparing self-evaluation, seeing main facilities of the RU/EvU and having chance to discuss with top-management and operational staff (PhDs etc.).
4. Method of this evaluation is completely in accordance with the recent programme declaration of the government of the CR. Comparison of Metodika to the current assessment of Czech Academy of Sciences outputs (is systematically carried out ; contemporary the 6th in order !) would be appreciated.
5. It will also be crucial for the future Methodology to identify how representative the set of key publications of the given faculty or institute is, and, especially, the relation of the set of these studies with the framework of continuous, conceptually based and systematically developing research of the evaluated workplace. (Otherwise, there is a threat of “buying” authors and publications that could “hide” the fact that the given workplace neither does relevant research, nor publishes at the required standard of quality. This phenomenon has already been identified in recent years.)

Furthermore, the **question of language** will be absolutely critical for the evaluation of the presented, representative publications: In essence, in Humanities – especially in the subjects that primarily or solely focus on examining Czech issues – it is **impossible or even absurd to limit evaluation to English**. In a Pan-European context, only a fraction of important excellent Humanities publications is published in English. On the other hand, it is obvious that despite undoubted relevance of the Czech published outputs in terms of Czech society (this is fully comparable with the German, English, Italian or, e.g. Polish situation), the principle of evaluation itself (its international comparison) **requires the need to submit publications primarily in key European languages**, i.e. in languages that are key for the given research topic area (e.g. a large portion of current international production on the history of papacy is published in Italian or is, at least, translated into Italian, similarly to e.g. a large part of work on Russian history, which is for example but not only published at the German history institute in Moscow in Russian).

Subject panels

1. It is important to define what should be the final goal of the evaluation. Depending on the answer to this question, RUs with completely different duties (Universities, Academy of Science, Private Companies, Governmental Institutions) should not be evaluated on the basis of an identical questionnaire.
 - A separation of EvU/RUs by groups of their main activity could be more useful and informative. (e.g. faculties of different universities; different institutes of the ASCR, different private institutes which are sponsored by research money from governmental/EU/public funds, etc.) or to compare single criteria (e.g. research excellence) between groups (e.g. research excellence of university based EvU vs. ASCR based EvUs. Many other questions (regarding PhD students etc. should be adapted to allow a fair comparison
 - Basis for a fair evaluation is that all ROs to be evaluated respond on a similar level of details. This was not the case due to various reasons.
 - The template could contain some headlines defining which aspects should be addressed in these reports even if this information is present in the general documents. This will facilitate attaining a uniform structure of the reports. These headlines in the reports should correspond to the headlines in the self-evaluation questionnaire.
 - The evaluation process (before the plenary discussion) involving a Chairman and one secondary reporter should be retained with a strong incentive to discussion beforehand between the two rapporteurs.
 - There is a need to have the total list of publications and the full list of researchers.
 - Site visits are essential, accompanied by discussions with the different categories of staff and should be developed. More time should be devoted to the visits taking into consideration the size of the RU.
 - For EvU containing several research units (RUs), it would require a comprehensive school plan and if possible, milestone goals.
 - Probably, it will be informative for private institutes to ask in self-assessment reports to describe shortly (if it possible) tasks of the contract projects and obtained results (as minimum, in some general words, in the case of confidentiality: the project was realized successfully and the final report was accepted by customer). They could contain a short narrative account of the value of their research to their customers and of the value of their outputs to the establishment of the unit reputation.
 - The amount of numbers and figures in the numerous tables handed out to the panel members is too large and could be reduced by ca. 70% which seems to be enough to obtain the essential information. A computer could make a digestion of all bibliometric data to confine these Tables and Figures to relevant numbers. Head counts and FTE numbers should be handled by computer and further used for “output/FTE” data, i.e. number of papers listed in the WoS divided by the FTE-researchers.
 - Not all outputs are of equal interest: Decisive are the numbers of WoS-listed publications, book monographs and book chapters, as well as patents. Conference proceedings are less important within a usual research environment. They should not be accounted in the same way as WoS-listed articles. In this respect the FTE assigned to pensioned/emeriti professors should be clarified and treated in a uniform way.
 - In order to clarify situation with PhD student training, it will be useful to present the specifications required to obtain a PhD (in general words) in the considered EvU/RU. Additional explanation will help to evaluate an effectiveness of PhD preparation in EvU.

The word “inbreeding”, though common in this context, is too negatively connoted, as in biology no inbreeding at all is favourable and should be substituted by “insufficient mobility/fluctuation” which targets the goal to be attained.

- When evaluating the success of a RU by the money gathered by competitive funding, it has to be considered that some funding institutions (e.g. EU) define local areas of preferred funding. Comparing RUs inside such preferred areas with those outside on the basis of a money-acquisition criterion is not fair.
 - Concerning the logistics, on the whole organization was satisfactory taking into account the character of pilot of Metodika. Documents were available on time on a specific website and all explanations/manual were provided on time with login/password information. The website was basic, but was sufficient and found efficient for the treatment of reports and exchanges of data. However, it is not sure that it could handle a large scale-up in a real evaluation procedure involving much more documents.
 - As a general suggestion, the administration of the evaluation procedure should be done in a more hierarchical manner to avoid a loss of information, duplicate or missing information or even contradictory advices and multiple emails. On one hand, panel members should have a single well-identified contact person for all questions related to the organization, documents, and accommodations.... The contact person may be the panel assistant. The panel assistant will assure the link with the other organizers and other staff in charge of the management and project coordination. On the other hand, for questions relative to sciences and evaluations, the contact person should only be the panel chair who in turn will be in contact with the main panel and project coordinator.
 - It would be better to find a procedure where experts do not have to pay in advance for their transportations or accommodations.
 - As improvement of the website, expert should be able to download all documents in a single archive instead of downloading the documents one by one. This is not only time consuming but also a possible source of mistakes.
2. Evaluate EvU as a whole with an umbrella document describing the strategy of the EvU and the position of the individual RUs within the organization. Enforce a SWOT analysis at the EvU level. Although this seemed to be required in the documentation, this was clearly not understood by the institutions. Describe output numbers at the sublevel of the RUs, with an section summarizing activity of individual members indicating what has been achieved and how the RU plans to operate in the future. A clear idea of future vision and strategy to achieve it must be provided.
 3. To start any self-assessment of an EvU and RU with their research mission and strategy being stated at the beginning of the period of review. I believe that research outputs of a RU should be assessed in relation to their own mission.
 - Reflection on external (environmental) factors affecting research developments over past period.
 - Response to outcomes of SWOT by RU and options for adjusting the research policy.
 - Anticipation of trends and developments over the next period, leading to corresponding adaptations in strategy.
 - When addressing research performance and career development of research faculty one could also include the degree of autonomy individual researchers have in managing their research and career advancements.
 - Research and education are supposed to be a holistic combination in a research driven university (“Humboldtian” model of higher education). This aspect, how research affects

academic educational programs and dissemination of knowledge to society, could be given a stronger accent as part of the societal relevance.

4. The list of questions was long. It would be better to use a short list of vital questions and for all the numerical data to provide a template for a table to filled in. In this way all units will provide the same information and the comparison becomes easier. Missing so far was a clear paragraph about research management. This should include i) how PhD students are evaluated, ii) how personnel is evaluated, iii) what is the mission, strategy and policy about continuity (or not) of research topics. Finally, funding strategies should be elaborated.
 - A bibliometric evaluation with the degree of detail as given is rather useful. Here we could recommend to make use extensively of WoS, SciVal, or Scopus. However, the evaluated laboratories should be able to check whether the input for the analysis is correct or, possibly, should be updated.
 - Information about the performance of PhDs could be included. For example, how many PhDs are finalizing in 4 years, 4.5 years, etc., and how many of the enrolled PhDs finally get their degree.
5. 1) Priority setting in the questions: what is very important, what is less important, what is not really important but might be useful to know. 2) Priority setting in the criteria. The new R&D Evaluation Methodology and Funding Principles is finally aimed to be used as a tool in decision making by governmental agencies about (governmental) funding. If information is available about governmental and national priority setting in research, it is advised to make this information available. 3) In the documents available to panel members for evaluation, Criterion 1 relates to RU and EvU. However, in the research unit evaluation report Criterion 1 only relates to the RU. In general it is important that the difference between questions regarding RU and EvU is always sufficiently clear. 4) Maybe less questions, however some important issues are not addressed (see later). 5) After reading the self- assessment report by an evaluation panel it might be useful to give the RU the possibility to provide additional information if the impression exists that questions are not sufficiently clear or if questions have not been answered in sufficiently detail. To that aim a site visit is very useful, however, is also very time-consuming. A good alternative is to invite 1 or 2 representatives of the RU (or EvU) to have a meeting with the evaluation panel. 6) Include some additional sub-questions that might be relevant for the future of the RU.
6. To get more feeling site visits are crucial. All the data don't have to go so much in detail as they do now. The classification in research fields in my opinion is also too detailed. Abstraction to fewer fields covering more the Czech situation and international connection could be considered.²
7. Once the feedback from the pilot testing will be taken by IPN into consideration and the decision is made to implement the new Metodika in CR the "news" about the system should be spread throughout the country soon in order to avoid a slow trickling thru by the "Czech experts", who's role as "observers" by the way during the testing phase proved to be not very officiant. They might be needed in the future to provide assistance to the panels in explaining the Czech peculiarities, but the constellation as met in July 2015 was at least questionable. All members of the panel were able to orient themselves in the Czech landscape – to be seen in the simple fact that the "talks" with the managements and the staff/researchers was held in many cases not in English but in Czech. This might not be the case in the future and more

² This remark was triggered by the fact that the output of the RU we reviewed was only for 1.5 % classified in the FoS the RU was classified in!

explanations will be necessary. The specialists attending the pilot testing should compile a sort of a manual for future panel members. What stays unclear can be clarified in one two-hour-session of the panels with the specialists at the beginning of the second day. Participation in the site visits (see below) is rather counterproductive since in a small historiography like the Czech one it creates the feeling of a “controlled discourse”.

8. It should certainly be implemented the system of "double independence" (peer review + panel members) of judgment, necessary in the human sciences to ensure the independence of judgment. The peer reviewers cannot in any case coincide with the panel members. The results of the peer review should be sent to panel members before the meeting, otherwise do not make sense. The mechanism should be improved (modelled on that of the Academy of Sciences). It should not happen that a volume is evaluated A and E at the same time. The judgment of RU may not be so detached from the judgment of EvU. Very often the strategies on the basis the RUs are evaluated dare set at the level of EvU. Compared to the humanities the scale A-B-C does not warrant the real differences. The result will be too often B, thus preventing a real diversification. It would probably be useful to a number of examples (including foreign) to fully understand what is meant by A-B-C. You will probably also need to provide the use of A-, B+, B-, etc. If the five criteria seem homogeneous and will have the same weight, there is a large imbalance with the self-assessment, which devotes most of the questions on the first criteria.
9. Why shall an EvU register only one RU?
 - It is stated that the evaluation process will be “fully transparent” – the assessment of “excellent research” as it is not transparent at all, the coffee-grinding machine is...
 - Revision of the scale-system necessary (Gap between B and C is arbitrary as national/international combined with “strong” excludes medium international players.
 - Benchmark unclear: we discussed, without consensus: Shanghai-ranking, ivy-league, best European universities without Oxbridge...
 - In the assessment criteria (table 2, 6), collaboration should be defined more clearly (co-editing ...)
 - The self-assessment report is too small in size, and too poorly structured within longer answers
 - Too many abbreviations in the text, even if there is a table of abbreviations; as nothing is printed anyway, one can do without these many abbreviations.
 - The “hard copies of the above documents” (23) available during the panel session does not make much sense as anybody will have them printed or on a computer. It makes more sense to send them early or not to provide them at all.
10. The peer-review was the weak point of the evaluation. Lack of time, and lack of reviewers. I don't see why “reviewing” and “evaluating” should be more separated. We are in our scholar practice very often in the situation of having to review, and having to participate in the evaluation of a unit producing the output. But the indications how to write the review should be improved. May be including some more clear criteria than the empty stereotypes (rigour, originality etc.). We should also more clearly determine, what type of work deserves really “outstanding” (dictionary may be rigorously prepared, but is it an “original” work? What is its heuristic value?). In other words, that would be good to prepare a serious guideline for the reviewers. Some suggestions expressed by Alessandro should be examined; he was very enthusiastic about the ASCR model of reviewing (sometimes 3 reviewers for a single book). We should recall or specify in the guidelines for the Ru, that a reedition, or the translation of a book already published in another language many years ago is not a new output. The role of

the Czech experts was not very clear (maybe I missed something), especially, in case they don't really speak English, and communicate in German. I felt miserable; I know too well that I should understand... Information about cooperation should be improved. "International conference" with just the title and the numbers of participants, it doesn't bring clear picture of the nature of the exchange, the degree of "internationality" etc. Everybody in the panel knew rather well the Czech context of Research, so I can't really imagine, how it will be, if it is not the case. The sometimes perplexity of some of our colleagues of the other panel shows that it is not easy task...

11. So (considering what I have just written), I suggest (for B.1 and B.2) that, as well in the SAR as during the on-site visit, the distinction should be made very clearly: for instance, in the SAR, the criterion I ("Environment") concerns no doubt the EvU, what is not the case for the second one ("Membership of Research Community"), where the scientific domains have to be directly and specifically taken into account. During the on-site visit, the committee could meet separately members of the EvU and (then) members of the RU, as, for instance, we meet (met) separately doctorate students. On the other hand, and on a "lower" level, tips or advices could be given to the writers to avoid self-assessments like: "the number of publications is remarkable"; more generally, for the assessors as well the assessed, the advice could be given to prefer factual remarks, "observable" facts.
12. My suggestion is for the future of Metodika to improve a system, which will be stable in the next 10 years. The researchers needs clear criteria and a clear evaluating system. The Metodika have to decide about the form and content form the questionnaire for the EvU, but after that no more changes. (In Hungary and in Romania I see each year other formulars, templates etc. and this instability make the whole system frivolous. Another suggestion is, to have another template for the evaluation of books. The actually template ask about the quality of the book, but not for its position in the scientific world, not for the grade of recognition from other scientist. We evaluated Pardubice and Ceske Budejovice, so we read comments to some books, all the anonymous comment says, the books are good, but nobody says, if this book is the best in his topic, or this book has a good reputation in the Czech Republic. The evaluators need this information to score the outcomes form the RUs.
13. I would suggest that the publications chosen by the RU as especially representative of the research results should be listed, as well as the organised conferences and the Erasmus agreements. Each panel, especially those belonging to the Humanities area, should include a Bohemist, because of the relevance of Czech language both as object language and vehicular language in many researches.
14. Both for the self-assessment reports and the general evaluation, a larger framework is needed that makes the different criteria sensible and significant for a variety of research fields. It should be a framework that encourages and stimulates the Czech research community rather than discouraging it. The current methodology does not do that, as can be seen from the reluctance and even resistance to participate in it, already during the pilot-testing stage. The methodology also needs to take into account how ground breaking research can be stimulated through the criteria applied to evaluate it. Here is a suggestion: as a framework for the entire evaluation, both the EvU and the RU maybe asked to explain their vision for the research done at their institutions for the next 10 years. At the level of the RU, this would mean raising a different set of questions at the outset of the self-assessment report:
 - Where has your field of study come from? Here, the members of the RU would be asked to reflect on the growth process of their branch of research over the past ten years: Which different streams and strands of research have emerged? How would you locate your own

research in relation to these streams and strands? Where are you today in the development of this research?

- The next step would be to raise similar questions about the future of the field: Where do you project the field will be in ten years time from now? Which directions are the most promising for ground-breaking research? What would be the importance and impact of the results? Where in this picture would you situate the research that you would like to do during this decade?
- In contrast to tedious evaluation procedures, which will be resented and resisted by the Czech community of researchers (especially in the humanities), this type of exercise would be exciting to many researchers. They will be invited to reflect on where they wish to go in the next decade. Within such a framework, many of the criteria make sense or can be slightly reformulated. The questions of the self-assessment report could also be reformulated or restructured. Many researchers would be happy to try and provide 'benchmarks' that indicate success in reaching the projected research objectives for the next ten years. These benchmarks can then be assessed along the way, say, after five years during the next evaluation exercise.
- Research Environment: The research strategy of an EvU would then encompass the vision that different RUs have explained. It should tell the evaluators which research problems, programmes and agendas the EvU wants to invest in and why it thinks these are promising routes. The EvU should be asked to clearly state its goals. The strategy would then explain the different policies, measures and steps needed to attain these goals. For the HR management, the EvU can explain how its hiring, evaluation, funding, and promotion policies are related to the larger strategy. What will be its focus for building the Faculty or Institute in terms of the people employed?
- Membership of the National and International Research Community: Currently, when one asks RUs how they are embedded in a national and international network, one just expects lists of people, projects and institutions they have worked with and the output this has produced. In the framework of a ten-year vision for research, this becomes a different question altogether: Which scholars and research groups are you working with in order to solve the research questions you are busy with? How will these collaborations help you to achieve the research agenda and goals you have put for yourself? What role do your partners play in the advancement of knowledge in your field of study? How do you work together to address research problems? The answers to such questions would show the rationale for working with particular research groups. In many cases, it might be much more interesting for a Czech research group to work with an as yet unknown research group in another part of the world than to work with famous institutions or scholars. Otherwise, the evaluation methodology will only compel Czech RUs to imitate the Anglo-Saxon academic world as much as possible.
- Scientific Research Excellence: Providing excellent output would also need to be reconsidered within this framework. The publications provided should be located within the larger vision that an RU has on its future development over the next ten years. How do the publications reflect where the RU was and where it wishes to go in terms of its research realizations? Thus, the type of publications or 'output' would differ: say, a chapter from a recent PhD dissertation could be more telling than an article by one faculty member that was published in a Web of Science indexed journal some years ago.
- Overall Research Performance: Here, within the framework of a larger research vision, the question is not at all how much an RU has published. The question becomes how the publications are contributing to realizing the research vision and goals the RU has

formulated for the next ten years. Why do you publish in this journal rather than any other journal? How important is it for you to publish in international English-language journals or in other European and non-European languages? It could very well be the case in the humanities, for instance, that an RU orients itself towards the emerging scientific communities and intellectual movements in Asian countries. In that case, its choice of languages and publishers will be framed within this orientation. Such a vision could be very wise, given where the world is going.

- Societal Relevance: When the RU has explained its 10-year vision for research, the questions here can be better put: If your envisioned research succeeds, how would it contribute to reasonable reflection and debates on societal issues? What do you think will be the potential impact of your research results? How is your research anticipating societal and other challenges for the Czech Republic or humanity at large, which will come up during the next decades? How would its results help tackle these challenges?

4 Q-1. Your comments on evaluation on the RU level vs. EvU level.

Main Panels

1. No comment, it seems well balanced.
2. The EvU level comments were difficult to make because the EvU was itself not the subject of the evaluation; the RUs were really the unit of evaluation. An EvU is the sum of its parts, which may include RUs which perform well and those which don't, so an overall assessment of the EvU performance is meaningless unless it is possible to pull out some common factors across the RUs they submit.
3. I would only concentrate the evaluation on the RU level.
4. The size of the evaluated units must be clarified. Imagining evaluation of e.g. CU as a whole is completely absurd. However, evaluation of, for example, the Faculty of Arts, which is huge and has a very wide range of fields, would also be difficult. It is a gigantic organisation with many sections (history, linguistics, psychology...), in which it is already impossible to find a common denominator during current, internal university evaluations (a very similar situation is in e.g. the Faculty of Science at CU).

Subject panels

1. I did not in the end take part in any EvU write-ups as my assigned RU's were also EvU's. I only note that the EvU evaluation has a SWOT structure as compared to the 5 category scoring for the RU's and thus question whether the formats should be more in line with one another. It was not exactly clear to me what the ultimate consequences for the individual Institutions would be as a result of their evaluations. It is clearly useful for them to know their results and following the methodology of their self assessment indicates to them what is considered of "value" in assigning scores. This does give the Institution the ability to develop in the "favoured" direction. I would suggest reviewers are briefed on this point so that they fully understand the consequences of performing their work. For industrially oriented Institutions there is a problem of commercial secrecy of results. Indeed their best results may well be subject to confidentiality agreements. Additionally some of their work may be in the form of a valuable and much used service (such as a high quality catalyst testing service or a specialist analytical service) and will not necessarily generate an output in form of papers, patents or prototypes. A more comprehensive way of assessing the value of the work carried out by such Institutions needs to be considered. My suggestion would be to draw up a list of these possible additional outputs and request a narrative account of them in addition to the normal outputs.
2. Evaluation of RU separately from EvU is not fully effective in some cases. For example, well equipped analytical laboratory, as a rule, is also performing the function of service for other RU in the structure of whole EvU. In the case this RU "spends" working time for other RU on tasks which could be important, but it is reflected on outputs of this RU. Probably, evaluation of whole EvU can present more correct "picture".
3. Basis for a fair evaluation is that all ROs to be evaluated respond on a similar level of details. This was not the case due to various reasons. (Maybe there were already too many

independent evaluations during the last years/months.) It should be even more clearly distinguished – for the evaluators as well as the evaluated RU/EvU - which questions should be answered on the level of RU, which one on the level of the EvU. Evaluation report written by the panel-members: Please, provide already in the template some headlines defining which aspects should be addressed in these reports. This will facilitate attaining a uniform structure of the reports. These headlines in the reports should correspond to the headlines in the self-evaluation questionnaire. This has been widely done already, however, there are still place for some improvement as some confusion was observed e.g. regarding the placement of “research strategies”.

4. RUs can be very small with little explanation what defines the borders of a RU within the EvU. In the full process, the EvUs should lead. It was not clear to this group whether any focus on the RU-level was useful. If the EvU was done properly it should have the RU in the larger context.
5. I have been involved in RU evaluations only. So, I cannot say much on effectiveness of evaluation at EvU level. I tend to believe that evaluations made at RU level will offer more selective & focused reviews for researchers in a particular field of science than aggregated reviews at EvU level. An evaluation at RU level will offer more specific advices for improvements in research output and performance.
6. Some nuisances for the evaluation panel. The first is the use of rather long acronyms with change only in the last part. This is confusing and could be replaced by a simple 3 (or 4) letter acronym or a number. More important, on the level of laboratories (RU) the evaluators are asked to write 100-200 words for each criterion plus another 100-200 words for conclusion and recommendation. That is 600-1200 words (Some journals have a maximum of 1000 words for a full paper). Even worse, for the department (EvU) level. A fair assessment can be given in in total 300-400 words.
7.
 - Positive aspects: Separate groups for each RU evaluators can concentrate a significant group of experts on the same domain for several RUs coming from various EvU.
 - Negative aspects: I think that the assessment of the RUs should not be separated from that of EVU for all members. Indeed, many elements of evaluation criteria are common to all RUs into each EvU (including overall strategy, interactions between RUs, some societal or international relations and especially governance). Apart the panel chair (and this is not certain), the experts have an extremely reduced vision of governance and global strategy of EvU.
 - Suggestion: A meeting of all experts in the same EvU could be organized to discuss specific points.
8. According to my opinion in the evaluation of the EvUs the same issues have to be addressed (more or less) as in the evaluation of the separate RUs. This also holds for the manner the evaluation is formulated. Only in that case it is possible to compare the various RUs in an EvU and to formulate an overall evaluation of the whole EvU. In the current methodology it might be useful to ask the panel chairs of all the RUs within a specific EvU to make a joint final evaluation of the EvU. To achieve a more balanced evaluation of an EvU it might be helpful to appoint subject panel chairs which are not only familiar within the research field of the specific RU that has to be evaluated but which have also some familiarity with the research fields of the others RUs within the EvU

9. In my case we only reviewed one RU, part of an EvU with 2 more RU's. From the self-assessment report of the EvU I got the strong impression that there is quite some overlap (collaboration) between the 3 RU's in a way that it is difficult to distinguish who do what. The site visit gave more insight in this.
10. I would go for RU-level-evaluation only since there the research work to be assessed is actually done. In any case on the first page of the self-reports a short outline of the RU's history (founded when, tradition lines, location within the local institutional scheme - here EvUs could be mentioned) is needed for foreign evaluators not being familiar with the Czech landscape.
11. There should be criteria or at least specific materials in order to understand the overall policy of the EvU. At present it is impossible to evaluate RU who have developed better strategies of EvU (it could happen and should be rewarded). Basically what we have done is an assessment of the RU and not of the EvU.
12. Lack of information about the history of the EvU. (JU is a young University, but not born from nothing...) Relation RU/EvU should be more clear.
13. The distinction between RU and EvU should be made clearer.
14. I think that the committees have to be aware that they have to distinguish (and I hope we have the means to distinguish always very clearly) the two levels.
15. The evaluation from the two levels is necessary. But I have a suggestion: Q014 – I would need a question Q014b – Description of research activities, objectives and achievements over the period 2010-14 of the RU. So we can separate the EvU from the RU and we will become a clearly answer about the RU
16. It is not very easy to distinguish the two levels. Sometimes one gets the impression that the EvU is too close to the RU to be independent in judgement and persuasive towards amelioration
17. In the case of both EvUs and RUs we evaluated, the research strategy was not clearly explained and elaborated. It should be clearer from the questions of the self-assessment report what is meant by research strategy. The above recommendations could help there. The self-assessment report should also ask the EvU how it provides support and incentives to its RUs and faculty to be active on an international level: for instance, how does the EvU give support for project writing for international calls? How does it provide support and incentives to publish in English? The questions about research environment and HR management could be slightly reformulated: How does the EvU try to create a vibrant and mutually supportive research community? How does it strive to create dedicated faculty members committed to the larger institution and its general interest, rather than people focussed completely on their own careers?

5 Q-2. Are there five assessment criteria relevant and properly formulated? Any suggestion for change?

Main Panels

1. The assessment criteria seem to be properly formulated. However, note the above discussion of the A and B grades.
2. The five assessment criteria are not discriminatory enough especially among A,B,C.
3. I think this is about right, and worked quite well in practice.
4. Selected criteria are adequate
5. Formulation of the Criteria A is probably adequate only to teams which are aspiring e.g. to win the Nobel prize. The other 4 criteria B, C, D, E, is the sufficient number.
6. Yes, they are. But for pure scientific evaluations the criterion “societal relevance” should not count.
7. The evaluation criteria and the five level classification grading system, as established in Methodology as suitable.

Subject Panels

1. I found the 5 criteria to be well formulated and not significantly overlapping. The further breakdown into sub elements is also helpful. In some categories a scale using descriptions using the words National and International is used with the methodology pointing out that these are used in a quality context and do not necessarily mean truly National or International. The intention is clear once it has been read a few times but could be a source of confusion until reviewers are familiar. Perhaps the terminology could be changed or some real examples given to illustrate the guidance. In general I feel that some example reviews might be useful for reviewers although there is a risk that phrases used in them might be plagiarised and thus overly constrain the free thoughts of the reviewers.
2. On my opinion, the five criteria are properly formulated in general. They are good starting point to evaluate quality of research and to formulate the common opinion of panel members. I have one doubt on effectiveness of “Quality level - A - The RU is a Global Leader”. I think that this description can be applied to rare organisations in the world. On my vision, a the “Quality level - B-The RU is a Strong International Player” may be the highest rating.
3. This worked ok. RU funding profile needs to be clear and to be clear by individuals.
4. Criterion 1 although well-formulated covers a number of items and many questions have to be addressed to indicate on the overall criterion. Criteria 3 and 4 are not totally independent from each another. Criterion 3 is largely based on the evaluation of the “production examples” selected by the RU while criterion 4 covers all with performance, both scientific and financial “powers” of the unit: criterion 3 appears somehow as a part of criterion 4, which is actually quite large
5. Yes. No reason to suggest changes
6. The criteria used to grade are in principle ok but could be refined somewhat. For example, criterion I should contain also the research management aspects. For criterion II the

description B (strong international player, as such ok) and description C (strong national player, as such also ok) have a large gap. This could be for description C something like: strong national player with some international recognition. Further a clear distinction between the purely scientific and more societal/environmental aspects could be made. For example, criterion III: Scientific Research Quality (so as to avoid the word excellent as a judgmental adjective) and criterion IV: Interaction with society and environment. Finally, criterion V could be an overall rating, not being necessarily an exact average of the other grades.

7. Positive aspects: Yes in my opinion. Suggestion: Perhaps it would be interesting to not limit the very good research units at the B level, because it is very hard to get an "A" position given the excellence of criteria for this note. The "B" could be replaced by an "A" and "A" by A * "or" A + ". It is true that it is the same, but it may be better!
8. The five criteria are relevant, however the weight of the relevance of these criteria are in my opinion different. The focus of a research organisation is to achieve output. The output can be formulated as societal relevance of the RU activities in terms of their reach and significance to the benefit of industry, the public sector and other societal actors, but also as the contribution of the RU activities to the scientific knowledge field. There is also a relationship between the societal relevance and the scientific relevance. Taking into account this integrated output Criteria 3 and 4 are the most important ones in my opinion. Criterion 5 deals with the overall performance, however, as formulated it does not include all aspects as mentioned in Criteria 3 and 4. It, however, includes the additional aspect of research productivity versus the size of the research unit in terms of FTE. Criteria 1 and 2 are primarily dealing with aspects that show the quality of the organisational structure and the quality of the researchers, but are not directly related to the output.
9. In my opinion the 5 assessment criteria cover the important aspects of the evaluation quite well.
10. Every single criterion should to be clear defined in terms what is belonging to it (see Prof. 6.1_Hengerer's remarks) and this both (1) to the RUs for the self-reports and (2) to the panel members for scoring. Important for the environment criterion is the question of sanctions for the case the internal quality control (criterion I) shows non-fulfilments.
11. The five criteria could work, but should be refined. Also in this case the use of examples on the correct meaning of environment, excellence etc. it might be useful. I. It required the list of people employed, otherwise it is impossible to understand some particular changes that sometimes occur suddenly in the RU. II. It should be made clear that what is meant by collaboration, differentiating partnerships that lead to the publication of a book from those based on institutional projects. III. It is the criterion which needs more work. There have been some misunderstandings about the meaning of the research excellence. If, as I believe, it means only the excellence outputs should be explained to the RU that they should not necessarily cover all areas studied. It must be increased the number, 4 is a number that can fit for patents, not for scientific monographs. They are to be determined with great care the rules under which the panel members may change the judgment of peer reviewers (it can of course happen only in exceptional cases). It should be clear to the peer reviewers that a book of great value in the national context cannot be penalized because it does not address the issue in the international context. IV. Little significant are the bibliometric data, much more useful it turned out the "Overview tables...", but must be taken well in advance to the panel members. You have to specify much better what is meant by research. V. This is the more complex criterion for the humanities. Personally I find strange that may have the same weight

as other criteria, the gap with the exact sciences in this case is really great. A book dedicated to Czech literature will never have the same social relevance of a search of equal quality in the medical field, for example.

- Criterion I: Research environment -Clearer definition of “environment” needed, a casuistic should probably evolve. Between C and D: What about international players who are no strong national players? The scale should be revised. It should be clearly stated if ivy-league-like RUs are global leaders only, or if a European level should be taken as reference
- Criterion II: Membership -It was unclear if awards should be integrated in this criterion. It should clearly be stated that publications are not meant here – aren’t they? But what about cooperation in edited volumes? Invitation to conferences? We learnt that students abroad could only be integrated in the self-assessment report if they stayed three months, but not if they stayed twice for two months; should be more flexible. Cooperation in ERC projects: to be rated here or under “overall output”?
- Criterion III: research excellence -In our group, more than one person took this for “excellence of the overall output” even if it was stated that the excellent output was meant. During the meeting with the RU, researchers stated that they preferred the “coffee-grinding” system as this system is transparent, whereas an assessment by experts is not (see our example: an “A” and an “E” for one single book) . It should clearly be stated for the RUs if they are expected to submit excellent single outputs or if they are expected to submit a sample representing an excellent spectrum of interests, languages etc. The whole range of books and the articles should be available for the assessment-team. A panel can evaluate the quality of the output if given one day time. The External evaluation, reviews, and the panel should be combined. Some (not all) required that in any book on an A-level, the Anglophone literature should be discussed.
- Criterion IV: overall research performance - More flexibility needed in the self-assessment report, internal grant systems, for example. Many of us put the awards here. Esp. in the case of archives and museums, the problem of (limited) research (publication) within an scientific institution should be addressed; some in our group were of an opinion that inventories should not be rated as research, but non-scientific work; some (a minority) were of an opinion that editions should not be rated as scientific work in this framework, too. I found the formalities for the reports on excellent outputs too tedious (quota words / pages).
- Criterion V: unclear, if publications are “interaction”

12. Since the question of PhD is so present at the very beginning (which also influence a lot the following), a short presentation of the situation in Czech Republic would be helpful for the panellist (with stress on the differences between different types of institutions). I am not sure that the structure of Criterion I (coming from a “hard science” model) is well adapted to social sciences and humanities. There should be more specified criteria to differentiate the collaborations. Criterion III is a little bit redundant with criterion IV. The explanations following the selection of outstanding books are not enough informative. There should be more effort to “situate” the selected book in the national and international production in the field (tell that the book is good, recalling the contents is not enough). Even in Social sciences and Humanities articles may be sometimes more important than books. Thus they shouldn’t be a priori expelled from Criterion III, or at least be more registered somewhere (Criterion IV?), and not only as a number. It appeared during the panel, that mention of FP7 is not enough. It should be specified whether the RU is participant or initiator. Criterion V. The

indications should be improved, with some keywords may be. I noticed that historians didn't really mention the (good) work they often do with Film and TV.

13. The criteria are of key significance for evaluation of RUs' scientific and social relevance. But they should be defined more in accordance with humanities' peculiarity and it should be made clear for the teams writing the self-assessment reports what the criteria means for their evaluation. Apparently, there was misunderstanding regarding criteria 3 and 4. Also scoring could be adjusted to specific situation in humanities. Scoring, as formulated in the pilot testing, does not always reflect the specific conditions in humanities (also see above)
14. I think that sometimes the criteria can be difficult to be distinguished, since the issues addressed are very near, for instance criteria II, III and IV: I had some difficulties to distinguish very clearly these 3 criteria. On the other hand, I think that other criteria could be proposed, both to the assessors (in their final report) and the assessed (in their SAR), on: Ties between research and training: for instance, it may be asked if and how members of the RU are involved in defining the contents/conceiving the curricula of the trainings; if/how the specific domains of research clearly are identifiable in trainings. Doctoral training and more generally doctoral policy: for instance which is the student participation in the "life" of the RU, that is to say, do they elect/send representatives? And more generally "democratic life" of the RU (elections, governance, direction, executive/scientific boards, assemblies...).
15. The five assessment criteria are very good, I have no proposal to change anything. My suggestion: On the second page we need the date. Page 3: Please explain what is the FTE. (Full time equivalent?) Page 5, Q013c – I think this question is not necessary, because overall has the doctoral students and the supervisors the same rights and obligations
16. In my opinion, the proposed criteria are in themselves relevant and properly formulated, let us say criteria at 360°, i.e. on an absolute scale. As a matter of fact, such a scale has to be conceived and used in a prudent perspective, i.e. according to the local agibilia. I mean, it is not expectable that a young university can become a world leader in ten or twenty years, in a discipline which has been professed in all the continents since centuries. On the other hand, it is hardly conceivable that a really pioneering RU in such a short period and in an almost unknown place can be acknowledged as the first one, the global leader, because of the relatively young fame it has gained. Furthermore, there are some subjects which are so specific, not world-wide known (e.g. Czech literature, Czech language, Czech history), that their global leadership needs to be better defined: it is rather a matter of national leadership, in connection with some international competitive units.
17. Some of the criteria are inadequately formulated: for instance, it hardly makes sense to speak of 'global leadership' when one is evaluating the HR management and research strategy of an EvU in relation to its RUs. There is also a problem with the sub-criteria in case of research excellence and research performance. 'Originality', 'significance' and 'rigour' can often conflict. For instance, some rigorous work may be insignificant and unoriginal. Some potentially significant and original work may not at this stage of development be very rigorous. The self-assessment reports should include more information about the attempts of the RU to acquire international and national funding. How many applications did they submit? How many were successful? How does this compare to the general success rate of applications to the funding programmes in question? Why do the members of the RU think some applications have failed? Also see the above recommendations.

6 Q-3. Formulation of questions and structuring of the self-assessment report according to five criteria. Were all the questions (asked of the Research Units) useful? Would it be more advisable to limit the questions to major around ten and have an additional data-gathering section?

Main panels

1. The self-assessment report is the principal material for the evaluation, but the quality of the reports presented by RUs vary a great deal. Clearly, the scientific productivity can to a large extent be evaluated on the basis of bibliometric data, while the assessment of all other criteria rely on self-evaluation. Note that since the present reports were prepared for this pilot exercise only, without any consequences for future financing, some units may not have paid sufficient attention to the report preparation.
2. As above (Q1 and Q2) – yes, I very much agree with this
3. Questions seems to be adequate
4. Formulation of questions is O.K.
5. I am pleased with the formulation of questions according the five criteria.
6. We recommend limiting the “self-assessment report” to the minimum necessary for basic information and orientation of the referees, and to reflect more deeply on the “objectivity” of the information on key results, e.g. based on reviews in expert printed media. An issue is the question of increased standardisation of self-reflection (the details of the questions asked) and, subsequently, determining the criteria of evaluation of “non-scientific” work of the individual evaluated unit, i.e. the work that is meant for the general public. Achieving reasonable unification, and, as a result, comparability of these sources, is crucial to both the evaluation procedure and its results.

Subject panels

1. The questions are well formulated. For industrial oriented institutions some additional questions to gain insight into the value of the research outputs would be valuable. Some broad assessment of the order of magnitude of the financial value which discoveries are thought to have yielded on implementation would help in assessing how valuable the results of research have been. The results of the SWOT analysis can be applicable to any of the 5 categories.
2. I think that general structure of self-assessment report is optimal in order to evaluate main parameters of EvU activity, to understand strong and weak sites etc. Unfortunately, self-assessment reports that have been presented on evaluation differed by quality in preparation. Some questions proposed for answering had no presentation in some reports or some figures (numbers) on activity had not desirable additional explanation.
3. Representation of academic communities in outputs is very poor, and should be improved by including a table with outputs of individuals. At no stage were the RUs asked to give us an overview of what they thought their research activity was. It was impossible from the

provided paperwork to get a feeling for who was in the RU, what they did, and how they related to each other. Gender data are only useful if presented for more aspects including career stage, completion statistics, promotion etc. PhD completion data need to be included, real PhD duration, drop-out rate etc. It would have been helpful for the paperwork to indicate who was active internationally such as invitations to conferences.

4. For criteria 4, evaluation of the budget and capacities of an RU is sometimes difficult because of the definitions – some of them are little understandable to the reviewer (no professional European commissioner), i.e. between the various sources of funding (“competitive state..”) All questions useful? More or less, yes. Limit the number of questions? Probably not. Would have preferred that the questions were expressed a little bit further the politically correct, uniformed way of considering things. Also sometimes being better formulated to avoid misunderstanding (I had some too), allowing the RU to fill the various items with more accurate data
5. The questions are OK. Though I would like to see the research mission and strategy of an RU articulated up-front in a self-assessment report. Why? The mission & strategy (most likely) affect the way the research environment has been structured and the resulting research performance and quality.
6. Q001 to Q003 : Missing information on governance (detailed flowchart, organization of life of the laboratory, distribution of budgets, organization of internal seminars, health and safety, quality procedure, ...)
Q009 or Q010 : The tables are not always clear. For example, why there are so many differences between the FTE and headcounts (We had explanation at the beginning of the test and ongoing). Note: It would be necessary to show the teaching and administrative loads for any researchers and/or teachers
Q012 : OK, all right
Q013 : OK, but it lacks information on the fate of the doctors (work, postdoc ...), for example 1 or 2 years after supporting their PhD.
Q014 and Q015 : This very important part deserves to be more structured, by team and/or by topics ... A diagram would be necessary and useful for evaluators.
Q016 :OK, all right
Q017 to Q019 : Clarifications are generally little or not met. It should be clarified by questions
Q020 to Q027 : OK, all right
Q028 and Q029 : You should ask more details on methods of distribution of budgets between sub-teams. Is there a "pot" common?
Q030 to Q033 : The questions are perhaps not sufficiently raised, since only one RU (on three) replied. Yet it is important to know the facilities laboratories have, especially if there is no site visit.
Q034 : The question is poorly understood. It is indeed a scientific strategy and it must be said
Q035 to Q044 : OK, all right
Criteria III : It's fine to ask a selection, but experts need a complete list of research outputs by classifying them as in table Q045
Q045 to Q054 : OK, all right
Q055 and Q057 : OK, all right
Q056 and Q058 : To put more in evidence when appropriate. It is really important now that this type of societal criteria have a very significant weight. I think as much as the excellent

outputs.

Q059 and Q060: OK, all right

SWOT analysis: OK, all right

7. It is advisable to reduce the number of questions, for example questions dealing with the career development of PhD students. In general it is more or less standard at foreign a university that after finishing a PhD research trajectory a new job is found in the industry, at another research institute or at another university. It is especially important that during the PhD trajectory the student can learn the tools how to work with colleagues in joint projects, how to find a new job, how to set up business, how to apply for a new position. A continuation to keep employed in the own institute or university should be an exception. Additional questions that might be useful to include in the self-assessment report are:-Vision of the RU on its current and future research: Mission? Motivation? Who are the clients? Wishes of the clients? Future research plans? How to achieve a top position? Who are the competitors? Research strategy?-Visibility of the RU: Visibility to students? Visibility to the industry (potential clients)? How to convince companies to cooperate? Visibility to Society, including governmental agencies? Participation in the Public Debate? How to improve the visibility?-Clearness about the size, significance and output of the research programme. This is especially important for RUs with strongly different sub-research programmes. What are the core sub-research programmes? Additional research projects? Relationship between these programmes? List of all output?-Relationship with alumni (Dr, MSc) and work field (research leaders, CEOs, government): Structural and formalized contacts? Advising regarding research programme, priority setting, visibility of the RU, contacts with potential clients, networking?-Inbreeding aspects: How to avoid a high level of inbreeding?-Fate of PhD students: Reasons for ending a PhD research trajectory?
8. In my opinion the EvU should take more responsibility and control of the assessment report in a way that the report is more systematic, concise, homogeneous and avoiding overlap. Also a common vision with strategy for the future on EvU level and differentiated to the RU level is needed.
9. A set of questions is clearly helpful, should be of course mirrored in the evaluation reports templates. A section for additional data sounds reasonable. The section "prizes awarded" should be divided in international and national and accompanied with a sort of a regist who "decorated" the researcher (what for instance is the relevance of a prize awarded by the Austrian Friends of Mallorca"?). One question should concern the "unique selling point" of the RU, another (mostly for universities) the placement-rate of their graduates in research centres home and abroad. As to the RU's own journals: please ask for the number of subscribers (home and abroad). Concerning the RU's publication strategies in cases they function as publishing house ("Indoor publishers" like HU, USD, NA) the question is how open and attractive they are for researchers from outside; in addition here and in all other cases please ask for contracts for book series with "main" and "leading" publishers home and abroad (including the language question).
10. No, not all the information was equally important. For the humanities little value has for example questions about the research infrastructure. What is lacking is a brief presentation of the history of the institutions, in some cases necessary to understand the linked between RU and EvU. I already mentioned the problems with the issue of PhD. In some cases the RU have detected the little space available. Absolutely necessary for the human sciences is the list of scientific articles published (especially when abroad in prestigious journals).

11. The questions of the self-assessment do not always fit to the five criteria; some are unclear, many do not leave enough place. Missing: information about the history of the faculty; the notion of strategy should be clarified; “plans” (where would you like to see the RU in 10 years?) The text is very much too small and hardly readable without zooming in, making it hard to read the printed version). Full list of articles needed, too, not only books
12. Some questions are not very adapted to social sciences and humanities. The question about the strategy (at least in the answers) is linked to the targets and outputs, not so much to the use of the existing tools in scholar practise. Thus the role (and the nature) of “international” conferences in the dispositive is unclear. Many remarks above belong also to this 3rd question.
13. Yes, I think that some questions/issues could be gathered.
14. Is very good.
15. About this and the previous point, an intermediate level between B and C could be inserted: before becoming strong international players, national players should be encouraged to begin with international terms of comparison and international calls, challenges etc., giving proof of their commitment on this level even before being credited with an evident capability at this stage. Furthermore, I think it can be difficult for inside players to identify what is expected abroad from those who are acknowledged as belonging to world famous traditions, like Prague school linguistics. To take these premises for granted, keeping silence about them, seems awkward: either for consent and progress or for dissent and diverging motivation, the local paradigm seems not to be forgettable and reducible to silence

7 Q-4. Do you consider on-site visits important or even indispensable? Would an interview of RU staff members by the panel in a neutral location be an option instead?

Main panels

1. The on-site visits are generally regarded as useful but their organization is difficult and the visits are time consuming, Shorter visits than half a day will hardly add sufficient new information, and even short visits would combined take much time for the panel members. It would usually be difficult to find panel members agreeing with such a short durations. Another possibility would be to split the panel in several smaller teams, who could visit different RUs. As a low-cost compromise in cases where no site visits are made, the panels should be given a chance to clarify outstanding questions in a meeting with an RU (EvU) manager.
2. On site visits enhance the influence of human factor - the visitors could be negatively or positively influenced by random effects (e.g. English proficiency of persons they meet, personal sympathy or antipathy etc...). I am not convinced they are really necessary
3. I have mixed feelings about this, as discussed when we met. If a full review of all of Czech academic activity is planned, for a single date, I doubt very much whether site visits would be possible, and doing some but not all is unfair. I think it comes down to a decision on the purpose of the review: assessment for the purpose of distribution of funding (as in the UK) or assessment for the purpose of supporting development (as I understand is the case in France). In the UK case, the emphasis is on fairness, and assessment is based as much as possible on peer review of outputs. There is little need for interviews or visits. This is easier to manage but is not very supportive of development.
 - I found the visit I went on very useful in helping me understand the context of the response the RU had made, but I'm not sure it would have helped me rank that RU against others.
 - The point made that the visit allowed the panel to understand the RU's written submission better is not persuasive – in the end, the RUs have to learn to write their submissions better!
4. I found very important to visit the RU/EvU facilities and meet management and staff members on their site. It definitely objectivises the self-assessment.
5. Panellists' personal on-site visit of RUs/EvUs as well as face to face discussion with management and staff members are crucial point of the evaluation methodology!
6. Site visits are indispensable for any kind of evaluation!

Subject panels

1. There is no doubt that the one site visit I made was valuable and other reviewers seem to have had the same opinion. However the visit we made was not structured by us, we simply followed an agenda proposed by the Institution visited. I had formulated a few general questions before hand. The logistics of having a site visit to every Institution are considerable and a decision on whether it should be at RU or EvU level would also be needed. The logistics of the latter are even more daunting. It would be valuable to incorporate some site visits into

the evaluation process but this would have to be done with care. Some form of schedule would need to be drawn up whereby perhaps visits were carried out say every 3rd evaluation. However that would make a very long time scale presuming that evaluations are done every 5 years. Alternatively the choice of a small representative percentage could be made and as part of the process of choosing the Institutions themselves could be asked whether they would like a visit. This question could also include a description by the Institution of any specific things they would like to cover during such a visit. I would also suggest that if site visits are made a part of the system that a set of possible questions for reviewers to choose from are drafted and also a suggested agenda and list of persons to interview. One of the things we picked up was that the Institutions probably delegate a lot of the questionnaire to administrative staff whereas a lot of the subject matter concerns the actual research about which questions are better to be answered by the academic staff. Also Institutions typically have a lot of competing administrative work on grants and funding so that the evaluation may be seen as an additional burden especially if there is no tangible outcome. Adding a visit to the process may not be welcome.

2. I feel that on-site visit is very important to clarify some questions and problems what are raised from reading of self-assessment report. Maybe, visits are not so important for evaluation of small EvU/RU and RU which occupy (secondary) service position in EvU / scientific organization. But in the case of "serious" RU as a part of large EvU the site visits are very desirable.
3. The site visits should be unnecessary as long as the panel is give clear information about the structures in Czech academia, and adequate paperwork is provided. Unfortunately, this was not the case which may also in the future necessitate an efficient mechanism of site visits. These should then be at the level of the EvU. At the level of a RU, the site visit is not effective. The problem at the moment is that the RUs appear to be defined ad hoc, and sometimes is only one research professor heading a larger group of semi-independent researchers. Visits were not an efficient use of panel time. EvU teams could come to the panel.
4. Absolutely!! This was not possible for the three RU in chemical engineering, so we just could assess things from forms that were often rapidly filled, sometimes with contradictions, also to better assess the way they're working, the available room, building, infrastructures and facilities they actually have. Also talk with people, better understand their point of view as PhD students, technicians or first rank professors.
5. Yes indispensable. It is a much needed independent check by the reviewers of an RU. The findings at an on-site visit complement the supplied written text and data in the Self-Assessment report of a RU.
6. Direct contact with the evaluated laboratories is almost mandatory. In the case of many laboratories, this could be achieved by having, say, two researchers of that laboratory visiting the committee instead of the other way around. In case of a site visit it is wise to include a discussion with some PhD students in absence of the scientific staff.
7. It's completely essential. It is not necessary to visit everything, but we must at the very least a discussion with the leaders of the team, and then with a group of permanent staff without managers, then with graduate without permanents. One or two local visits (necessarily) selected by the research team can then be carried out and a visit of some heavy and semi-heavy equipment (function of available time)

8. An on-site visit is useful, however, requires an intensive preparation and planning from both sides and is time-consuming. An interview of RU staff members by the panel is a very good alternative. I believe that it is always useful, even necessary, that there will be a contact moment between the evaluators and the RU to discuss the self-evaluation report. The evaluation panel has then the opportunity to ask questions and to inform the RU staff members about relevant issues that have not been addressed in the self-assessment reports. The RU staff members have the opportunity to give additional information or to elucidate aspects that have been described only very briefly or very vaguely.
9. Our on-site visits to the RU was crucial to get a good insight and feeling of the structure and governance of sub-units of the RU, of their accommodation and research infrastructure, of the drive and passion of scientist and students and of the satisfaction of the students. We consider on-site visits indispensable for a good evaluation
10. As stated above, talks with the academic leadership/management and the staff/researchers of the RUs is one of the three main sources for the panel's assessment and thus indispensable. The practical issue of course is what I called the "Reisezirkus" with regard to the high number of RUs. I suggest to visit the "big players" and to invite "delegations" from the smaller ones composed after a given "key": Directors/chairs 2-3, researchers 5-6 of all "sorts" (senior and junior), 2-3 doctoral students, 1-2 technical staff – in sum: 10-15 at max.
11. At first I did not think that would be important, the reality has however revealed that a lot of information was not so clear. It was very important to understand the relationship between RU and EvU and their strategies. The representation of the RU was too narrow; it would serve a more qualified representation. The only real chance to realize profitably this meeting is to organize a meeting of qualified representatives of the RU with the panel members in Prague.
12. Useful, as many questions raise during the assessment, and important points do not make it into the self-assessment report [internal financing of excellent research, for example] indispensable, if the evaluation shall have (as it should) a formative function of advising the RUs . If site-visits are considered too complicated, expensive or time consuming: they are not necessary in that the evaluation team needs to travel; it would be sufficient if the faculties send their representatives to Prague for a meeting of about three hours.
13. Visit on the spot brings a lot. We can hear things that don't appear in the written report, at least not so explicitly. In USD and NA we had the opportunity to see the outputs (not only their PDF). Possibility of exchanging, clarifying, with actors of the RU. A space of formation too? Useful to hear also question from RU's researcher about the meaning of the Evaluation.
14. On-site visits are very important, perhaps even indispensable, part of the evaluation process. Presentation of publications and interviews with RU's staff provided an important complement to self-assessment. Interviews modified, partly, the overall view of RUs. Less important, perhaps, was the presentation of research facilities (library, offices). An efficient alternative to on-site visit could be to invite some representatives of the RUs' staff (leaders, PhD students, postdocs) to panel meeting for interviews.
15. Yes. I think that during the on-site visits (indispensable indeed), we could have an independent meeting with the RU (as we had with the students).
16. The on-site visits are important, in all the countries, that I know good (Romania, Germany, Hungary) they will do.
17. Important for sure, if not even indispensable, because self-awareness and others' observations can mutually support a more complex and complete evaluation; a symbolic

representation such as a written self-assessment does not make useless or superfluous a direct observation of place, people, structures.

18. The site visit turned out to be extremely important in the case of our subject panel. It gave an impression of the RU very different from the self-assessment report. Speaking to the PhD students, in particular, turned out to be very helpful in getting a clearer picture of the RU, its research and its future potential. The fact that they could receive the evaluation panel at the premises of their own departments played a positive role. However, if site visits are too cost and labour-intensive, meeting a group of faculty and PhD students in a neutral location would also be helpful for the evaluation panel.

8 Q-5. Do you consider grading of selected outputs by independent reviewers useful?


Main panels

1. Yes, this feature is very useful.
2. Very much so. Essential in my view. However, the reviewers COULD be members of the panel. It is important to have at least two views of each output.
3. Yes
4. Yes, without any doubt.
5. Yes, I do.
6. **The set of key publications** should be, in some form, confronted with the overall production of the evaluated institution, among other things, considering its size. The idea that, for example, the huge Faculty of Arts at CU will be relevantly represented (and its scientific output will be relevantly shown) with 10 publications within three to five years is absurd. For a small institute or museum, this can be completely relevant. Maybe a rule that is used in the CU to evaluate the best scientific monographs annually could be used; every faculty can propose a clearly defined proportion of its output (one third annually), published in every given year, to be evaluated by the committee. This ensures that the representation of the publications of the individual faculties is substantially different.

Subject panels

1. The grading by expert and independent reviewers is essential. It is difficult for the reviewers of the Institution to be specialist in the subject matter of the nominated excellent outputs. The comparison of results from reviews by one or two reviewers shows good agreement so that the cost of review could be cut considerably by electing to use just one reviewer. Mobilising sufficient specialised reviewers is likely to be a problem even if they are paid. The peer reviewed journals will have extensive lists of prospective reviewers for specialist papers and I suspect that they will already have made assessments on paper quality during their internal review process. Whether any of this information is available is questionable, it is probably kept confidential within the Journals' editorial systems. As I understand it the reviews simply assign a score on a 5 point scale but do not record any detail as to why the score was assigned. It may be worth thinking about including a set of key criteria such as originality, readability/presentation, and relevance to the field, excellence of the scientific method etc. but on a two or three point scale in support of the overall score.
2. On my vision, additional review of selected published outputs is not so productive for evaluation. The bibliometric analysis proposed by Metodika is enough to formulate general opinion.
3. 1% of the publication output results in a limited selection of the unit's activity, with a significant percentage of the submitted papers relating to a single individual and/or embedded and highly dependent on external collaborations. This should be changed and an overview should be provided for the output numbers of each professor and associate professor of the RU. The independent reviewers were of some use though the panel read all the outputs.

4. Yes. 70% of the evaluation made by independent reviewers seems correct to me, i.e. done with care and onsciousness. More important (sometimes troublesome) the selection of the output made by the RU!
5. In principle it is a good way. But in practice it was not as useful as it could have been. The generic key words in this grading exercise are: originality, significance and rigour. The narratives by the independent reviewers did not refer to these generic qualities at all in their justification of scoring. May be, the reviewers did account for these aspects in their minds or on an intuitive level, but it was not made explicit in the narratives.
6. The grading of a selection of papers by experts is considered as useless. Papers are reviewed and, when deemed necessary, will be assessed by the panel members. Time and money used for this can be spent better otherwise.
7. Yes to my opinion, it is useful.
8. In my opinion grading by independent reviewers is not really necessary and probably also not very useful. Grading by different independent reviewers can result in a judgement that is not fully in line with the criteria used by the members of the subject panel. There is also no possibility to have an interaction with the subject panel members. The panel members have also the expertise to grade selected outputs. If the grading of the specific selected outputs is also done by the panel members, then a more balanced evaluation of the total output can be expected.
9. We consider grading of selected outputs by independent reviewers useful because it contributes to making the grading more objective.
10. In principle yes (not only for practical reasons and the work load), but I am not sure if future panel members should not be involved. The fact that some of our panel members did so helped to assess the quality and to rectify very diverse gradings by external reviewers. Perhaps panel members should give a “third voice” in this cases – precondition: the review process starts longer time in advance.
11. The question is misplaced, they are absolutely essential to leave no absolute power to the panel members, who have not in the humanities the support of bibliometric data. In the final system of evaluation should be absolutely implemented functions already used by the Academy of Sciences (reviews of books published, peer review evaluation process that comes- if necessary even after four judgments-to an objective assessment, etc.).
12. It is useful if, in case assessments are to disparate, a third opinion is asked for (see proposition Alessandro 6.1_Catalano who insisted on the merits of the evaluation of the AVCR. The evaluation-team needs access to all books and articles during the evaluation, they should be made available; let the experts pass half a day with them, this will give a rather reliable assessment. The evaluation-team should get the national and international reviews of books of the RU; as reviewing takes up to five years, the period for reviews should be longer than the report-period.
13. Grading or evaluating? The comment is the more important than grade. Why reduce the review to 200 hundred words? What means “independent”? Non-Czech researcher able to read 500 hundred pages in Czech is very often somewhat close to the Czech. I wrote already on this issue (see above). The small per cent of research outputs is sometimes misleading. On the other hand it says something about the Unit (we can see what is supposed to be the very best, and try to find out why).

- 
14. Grading was important even if in some cases the grades varied very much, which, in fact, can be a sign of originality of the output. But the grading of anonymous peer reviewers should be completed by book reviews, if possible, supplied by the RUs.
 15. Yes, above all when there are no specialists of precise domains in the committee.
 16. Yes, this is useful, but the reviewers have to tell not just the quality of the outputs, but also the recognition of the books in the Czech Republic.
 17. Yes, I do
 18. In the case of our panel, it did not turn out to be useful. The reports of the independent reviewer came very late. Most of the grading was not credible and not well supported by arguments, but just gave very high scores and glowing remarks. The fact that the panel members could not read the selected outputs themselves (because they did not know the Czech language) posed major problems for evaluating these outputs. In general, we think at least some of the panel members should be able to read the publications and evaluate them, in order to come to a credible scoring for this criterion.

9 Q-6. Do you agree with assessment of only a small percent of research outputs or would you prefer to receive a full publication list? Would you be happy with a quality judgement based purely on bibliometrics, rather than independent review of selected publications?

Main panels

1. I didn't like the small self-selection of outputs. I would go for ALL outputs (the full publication list) OR a set maximum number for each academic (in the UK it's 4 within the 5 year period). This avoids the situation that the response is dominated by one or two prolific authors.
2. Bibliometric should be the basis. An independent review of selected publications (being ranked above certain level) is recommended.
3. Yes, I do. Full list of publications is available using databases, like WoS, WoK, Scopus, etc. Only selected publications (not in impacted journals, like books, monographs, conference proceedings,...) could be reviewed independently.
4. I would like to receive a full publication list and to choose the publications, which deserve an evaluation.
5. This point is connected with the issue of the ratio of monographs vs. journal studies. Both are, of course, meant as publications in "West European" publishing houses and journals. The committee in the pilot panel stated in the discussions that it is "impossible" for the evaluated institutions to have so far predominantly presented monographs as their best work (other publication outputs have been practically a priori discarded) and that no articles in foreign scientific magazines have been represented in any of the evaluated material (let us remind that the RIV a priori eliminates studies from conferences and other proceedings – including very prestigious foreign publications – as "unscientific", unless this concerns proceedings of the large, international congresses. In this respect, the evaluation of output of e.g. Czech historiography varies substantially from the evaluation of German or French historiography).

Even if we omit the fact that practically none of the current foreign referees, who required this from Czech historiography, have any of their own studies printed in magazines in West European countries that are not their country of origin, it still remains important to discuss this topic seriously. However, we would like to emphasise that for operational purposes, it is a question of the time limit in which, after determining the actual "hard" request for this type of publication output, the evaluated workplaces will realistically be able to submit it. This concerns mainly concentrating funds and re-focusing their current network. There will be zero added scientific value of these publication outputs in comparison to the current publications. Nowadays, it is more productive to publish foreign language studies of Czech scientists online, if possible on international renowned platforms, or in connection with them.

Purely "bohemian-centric" humanities disciplines (e.g. fields such as older Czech history, history of Czech literature, grammar and linguistics of Czech language) are, however, and will be limited in this respect by a narrow interest in their outputs of foreign (international) experts. Therefore, the pilot evaluation potentially carries the danger of undervaluing or underestimating truly "excellent" and scientifically beneficial outputs, if they are to have such objectively limited impact and are published "only" in Czech and locally. As a result, the

parameters determined by the Methodology should consider this and subsequently, the foreign expert referees with enough insight should be able to equally take into account also “excellent” and in foreign reviews positively rated publication outputs published in Czech (German, Polish) language in Czech expert journals and magazines.

Subject panels

6. Both would be appreciated (in spite of the larger work required...), Bibliometrics done in a generic way can sometimes lead to non-sense since the scientific “production” largely differs from one area to another. The domain “chemical engineering” covers quite a large number of topics, nearly to nanoscience or advances materials, thermodynamics, to technology and energy (also from oil chemistry to biomass etc...). The possible production of a paper does not correspond to the same amount of work and progress of knowledge in any of the above topics, so.. Citation data may be quite “worse”...
7. In my opinion the assessment of only a small percentage of the research outputs is insufficient to make a well-balanced judgement of the quality of the RU. Also a quality judgement based purely on bibliometrics or on the results of an independent review of selected publications is in my opinion not a reliable method for the formulation of a well –balanced judgement of the RU. There are several reasons for this opinion. The output of an RU not only exists of scientific publications in scientific journals but often comprises several other issues such as patents, reports for industry, and, especially for universities, well-educated PhD and MSc students. It is in general recognized that the quality of a university is often measured also on basis of the quality of the MSc and PhD graduates. In the education and also in the judgement of MSc and PhD students the Master theses or Doctors theses play a crucial role. These theses are characteristic for the quality of the research issues and the research competence of the graduated MSc and PhD researchers. In the evaluation of the RU we have observed that often the research programme is very much scattered and that the coherence between the different research issues is not always visible. And that the scientific publications in scientific journals are often mainly dealing with only one key research issue or one sub-research programme. I think it will be important that the evaluation panel is informed about all types of outcomes ,also about reports for companies. As already mentioned in the previous it is not always possible and useful to publish results of research into scientific journals. Finally we have also the important criterion of Societal relevance of the research. This is also a type of output that not always can be formulated and assessed as a scientific publication in a scientific journal.
8. In any case, it is a method not even useful in the humanities. All additions in red are useless, if not even harmful as regards the humanities
9. Both. We missed the full publications lists in our assessments. The way to assess in both cases is not the same, but they are complementary. Would you be happy with a quality judgement based purely on bibliometrics, rather than independent review of selected publications? No (see above). Reviews of selected outputs are indispensable for a quality (precise) judgment too. But the RU have to express/explain the reasons of their choice of selected publications, clearly. And I think that, in case of selection of publications, the common works (as editions of books) should be preferred to individual works.
10. We would prefer to receive a full publication list. This information is necessary. A quality judgement in our fields of study cannot be based on bibliometrics.

10Q-7. Do you consider bibliometric analysis important for your disciplinary area? What are your comments and/or suggestions to the content of the bibliometric report?

Main panels

1. Bibliometric reports are an indispensable part of scientific evaluations. One might even add a similar analysis of the Czech grant database (CEP). If each RU lists its members, an analysis of financed projects may be very easy.
2. This was helpful, but threw up oddities which were difficult to explain (and also included one or two obvious mistakes). I think it is useful for the panels to have but as additional information, rather than the primary source. It will probably be more effective when more units are included. I personally found the citations and comparison with the Czech average in the field useful. The RUs need to see all data used in their evaluation, I suggest.
3. Both the bibliometric analysis and patent application search are typically applied in the field of chemistry/chemical engineering.
4. They are not relevant for the humanities.
5. **Bibliometric analysis** of live production (i.e. not retrospective research performed in hindsight and using relevant methodological resources) **are irrelevant for humanities**, because the form of the standard (prestigious) publication output here – and all over Europe- is not an article in foreign expert printed media, but a monograph published with a “good” publisher, participation in international research and conference projects and their – mostly in the form of proceedings – outputs (here, a full discussion could be made about the phenomenon of the “collective monograph” as cover for a bibliographically properly adapted journal).

Subject panels

1. The bibliometric analysis is important. Both individual researchers and Institutions publishing record of scientific papers is commonly regarded as a key indicator of their success. As commented earlier it would be helpful if some of the key indicator ratios for example describing output productivity could be pre-calculated and shown in a table for the full spectrum of Institutions being assessed so that the reviewer has an easy way of zeroing in on the key performance ratios as well as putting these into perspective with those of all the Institutions under evaluation.
 2. Bibliometric analysis is important for every disciplinary area. It is serious instrument to evaluate scientific activity of RU and to understand a position of RU on national and international “field” of specific area of science.
 3. The bibliometric report was of very good quality and very useful
 4. Important? To a slight extent, but the basis for evaluation. The bibliometric report looks really complete, exhaustive, but (sorry) nearly technocratic.. In some groups, it is the strategy to over publish, in other not.. In spite of the famous word “publish or perish”, one may prefer to have assessment tools not only given in terms of numbers of papers or citations. Reputations, contra realization can be effective criteria too
 5. Yes, the bibliometric analysis played a significant role in getting a good understanding of the amount, nature and variety of the research outputs. · No comments / suggestions on its content.
-

6. A bibliometric evaluation with the degree of detail as given is rather useful. Here we could recommend to make use extensively of WoS, SciVal, or Scopus. However, the evaluated laboratories should be able to check whether the input for the analysis is correct or, possibly, should be updated. Information about the performance of PhDs could be included. For example, how many PhDs are finalizing in 4 years, 4.5 years, etc., and how many of the enrolled PhDs finally get their degree.
7. Yes, it is perfect as it is presented
8. In my opinion it is very important, however, not the only important aspect. In case of Universities important aspects are also the quality of the MSc and PhD graduates. What is the opinion of Alumni about their education programme (including also the research issues)? What is the opinion of the industrial leaders about the quality of the MSc and PhD graduates? In case of non-university research institutes it is very important how the results of the research are validated and used by the industry. What are your comments and/or suggestions to the content of the bibliometric report? Keep it up to date and complete, but keep it also simple. Only really relevant information, not too many details.
9. Since the results of the R&D evaluation are used for a basic allocation formula for funding and for establishing the scientific excellence of RU's a bibliometric analysis which allows for an accurate comparison is important. However the bibliometric analysis as used in the present evaluation goes into far too much detail (e.g. 10%, 25%, lower 50% upper 50%). In addition Access to a complete list of publication with a scientific bibliometric analysis and van Raan Index³ would be a better basis than the selected papers used in the present evaluation.
10. As stated: rather useless for humanities
11. Bibliometric analysis are not important at all, much more useful were the "Overview tables...", but must be taken well in advance to the panel members
12. Not important, not even useful in the field of humanities, just the descriptive statistics. What would be helpful, is a comprehensive list of publications, articles and chapters in edited volumes included, by author and by journal/editor
13. No. The bibliometrics is not significant. We didn't use it during the panel. The complete list was very useful. More information on outputs needed: national and international reviews (could be very helpful in some cases...)
14. Obviously, in humanities the bibliometric analysis does not say much. Outputs in history are 1/ articles published in scientific journals, 2/monographs, edited volumes, chapters in edited volumes. Regarding the group 1 (journal articles) bibliometric analysis can be of use if journals are included in indexes for humanities (e g THOMPSON, ERIH), but also preferably including index grading of journals (e g ABC). This indicates, at minimum, the editorial quality and scientific impact of the journal. Regarding group 2 (books) bibliometric report does not say anything. Here the evaluation must be completed by individual assessments by panel members, anonymous peer-reviews, and book reviews. In Sweden and Norway bibliometric

³ His work covers the development of scientific and technological fields; analysis of scientific progress and assessment of scientific performance; statistical properties of indicators, mapping of science and technology and in particular research related to important socio-economic issues; science as a 'self-organizing' cognitive ecosystem. In this work he uses elements of physics and mathematics, particularly complex systems. This is done very professional in an objective way and allows comparative evaluation on the basis of RU's, EvU's, national and international. His system does however not include the Czech literature data base. It compares within disciplines. The system takes also into account different culture of publishing for different disciplines. <http://www.cwts.nl/tvr/>

analysis of books is based on grading of publishing houses (from Cambridge UP to individual publishing). Nevertheless, even this has met much criticism from the side of scientific community. Do you consider a calibration exercise a necessary and useful part of the Metodika? Yes, calibration meetings helped panel members to identify and discuss differences in evaluation, and to reach an agreement on the quality of research.

15. Yes. In this way, the committee would be able to see and assess the forms/media of publications. What are your comments and/or suggestions to the content of the bibliometric report? We could ask for details on: (articles published in) national vs. international journals
16. The bibliometric analysis is not relevant in my area, in the humanities. If we make a book, for example a dictionary or a book about the history, the citations will come later not direct after the apparition from the book. Its necessary to know, how many articles and monographs was written in the RU, but the detailed analysis – like in the bibliometrical report from Ceske Budejovice – don't help the evaluators, to make a clearly image about the RU.
17. At least I do not yet consider bibliometric reports sufficiently reliable.
18. The bibliometric reports were largely useless in the case of the subject panel Philosophy, Ethics and Religion. They did not give us any picture of the publication records of the RUs. Without the necessary information about the titles and themes of the publications and the journals or publishers where they have been published, numbers of publications are uninformative. Dividing the number of publications by the number of researchers does not give a genuine indication of how vibrant the RU may be.

11 Q-8. Do you consider a calibration exercise a necessary and useful part of the Metodika?

Main panels

1. The calibration exercise is very useful for setting common standards for the evaluation, although it may not have worked perfectly in this pilot exercise. One might consider placing it later, right after the start of the evaluation itself, when issues and problems are better understood by all panel members.
2. In my opinion the calibration exercise is essential for success of Metodika.
3. Yes, but in practice I think this is very difficult to achieve. In my view, it can only work well if a reasonably large number of units are assessed together, so that they can be ranked, and the ranking compared with a similar ranking for another discipline. With the small numbers of units considered in this pilot exercise it was difficult to do any sensible calibration.
4. Absolutely! Calibration exercise should not only to remind basic approaches aligning interpretation of levels of assessment in each category but should also achieve minimisation of subjectivity during the assessment.
5. Calibration exercise is very useful for evaluation methodology development.
6. Yes, I do.
7. Calibration meetings were very important and helpful when verifying the positives and negatives of the Methodology and in order to unify the evaluation procedures of the panels.

Subject panels

1. It is essential for each field to conduct some sort of cross calibration rather than just relying on the results generated by each review panel. Possible techniques might include consideration of the rankings to see if there appear to be anomalies, examination of outlying scores (i.e. very high or very low scores). Also the total average score could be calculated or the distribution plotted in a chart to see if it appears “normal”.
2. Yes!
3. Yes, as a start but there is also a need to calibrate afterwards.
4. Yes, absolutely, for better knowledge of things (country, topics, methods, ethics, significant criteria..), also to get the panel members started in the area.
5. Yes, both necessary and useful. Different cultural, professional and linguistic backgrounds of reviewers tends to play a role in scoring and how outcomes are formulated. It is important to be able to reconcile ‘harsh’ and ‘lenient’ ways of scoring. Also it is important to harmonize the sometimes rather different formulations of reviewers especially when they have essentially very similar perceptions of research quality and performance of an RU.
6. The calibration exercise is not really discussed but its impact seems to be limited. A clear description of the quality characteristics for each of the criteria employed is of larger importance (see the remarks on this before).
7. Yes

8. I think a calibration exercise is useful. Especially for preparing the EvU report a more or less uniform reporting/format of the evaluation of the relevant RUs belonging to the EvU is necessary.
9. The calibration exercise has proven to be a necessary and useful part of the Metodika. With the experience of this calibration the Metodika can be improved. It also shows that a as broad as possible basis of consensus of all involved units is an absolute requirement for success of the evaluation. The process could also be made less bureaucratic⁴
10. Yes, at least in the first 2-3 future rounds, and please in panels consisting of 5 members and more for two of them in advance in order to find a sort of a routine.
11. Absolutely necessary. I believe in deed that some issues (the sense of being a Global leader for example) deserve a further calibration exercise.
12. Absolutely necessary.
13. Definitely. Many concepts must be discussed and recalibrated for the field of Soc Sciences and Humanities. "Global leader" and so on.
14. Yes
15. Yes, it was useful and necessary.
16. Yes, I do for reasons of the same kind of those mentioned in (2).
17. As currently conceived, the calibration exercise did not play a significant role in the evaluation process. The weighting of the criteria seemed arbitrary and was inadequate to distinguish between different kinds of institutions such as universities and the academy of sciences.

⁴ As a reviewer you have to go through a lot of background reading, getting familiar with many concepts and definitions characteristic for the Metodika. This goes till small details. The solidness of the Metodika depends from the interpretation of these definitions and concepts by scientists and reviewers. In my opinion this leads to overanalyses particularly in the case of scholarly output.

12Q-9. What are your comments to the pilot testing (overall organization, materials submitted, panel meetings logistics)?

Main panels

1. The pilot testing was very well organised and the organisers should be congratulated for their effort. However, the pilot evaluation included only about 30 RUs. Much will be different when hundreds are evaluated. The web system developed for the evaluation is generally good.
2. The pilot testing was very well organised. The organisers made excellent job and should be congratulated especially for the selection of panel chairs.
3. I think it was well organised. Clearly some reviews of papers hadn't been completed, which would be necessary in a full exercise. I also had the impression that the bibliometrics came rather late. The Main Panel can't do its job until it has the results of the Subject Panels, so timing needs thinking about. In the UK system, the Subject Panels and the Main Panels meet separately, with perhaps a week in between, giving time for results to be brought together for the Main Panel to see.
4. I believe that pilot testing is necessary and unavoidable part of introduction of the new methodology for any of business activity. Overall organisation was excellent as well as way and timeliness of materials submitted and panel meetings logistics. The only remark is to necessity of better description and definition of role of panellists, esp. members of the main panel.
5. I appreciated very much good organization, materials distribution, panel meetings, etc.
6. Despite the time pressure it was very well organized by the Metodika-team.
7. **The pilot test was perfectly prepared**, both in terms of the outstanding organisational work, as well as managing the events and the assistants' work. Minor problems occurred when logging in to the system, but these were not the fault of the organizers and were, with their help, overcome.

Subject panels

1. The pilot testing was well organised with excellent facilities. There was an issue of whether delegates should each be equipped with a computer and logged in to the system. I had some problems with accidentally generating multiple copies of reports when making amendments. The evaluations of the EO's were a bit late on the scene but apparently due to difficulties finding reviewers. The secure computer system performed well and is fairly easy to use. I suggest to check the logout function as I have the feeling that it sometimes leaves you logged in. This is because if you reopen a page from browser history you sometimes are straight back in the system even though you logged out. We had some problems with the Czech keyboard,

which slowed down revision work. Usual solution is to have a switch routine (such as CTRL + another specially nominated key) to switch between German, English or other system keyboard arrangements. Finally the catering arrangements were good and the celebratory champagne at the end was highly appreciated.

2. Overall process and logistics were organized very well thanks to organizers.
3. Pilot testing was well organized with the exception of the administrative support provided to the panel only being at a mechanics level rather than also providing advice on the Czech University and Research funding structures. Of course there were unexpected issues to deal with but this was to be expected.

The panels needed to be supported by effective administrative staff who understand the process, the university system, the national and international research scene including funding, being able to provide immediate feedback/support for the panel.

Site visits should not be necessary if the paper work is properly completed. Visits of the leadership of EvU to a central place in Czech to meet the panel could be considered as an alternative for the lengthy traveling and site visits.

The organization should implement an effective organization on travels and accommodation so that panel members do not need to spend their own time and money on this, and get into complex schemes of refund.

In a social context, it would have been advisable to organize a dinner with all panels.

On the organization, consider organizing the hotels, travel by a separate entity so that panel members are not bothered by going into a complex reimbursement scheme.

4. High commitment and availability of the staff working for the job (professors, engineers, administratives) in a sincere, friendly atmosphere. The material submitted by the RU was of “moderate” quality, I mean 50% of the items have been filled in a non-professional way, with sometime no control or supervision of the RU leaders. Probably also because some RU’s were not totally aware that the “game” could be sort of valuable training for them, for next time when the evaluation is done for real. Lack of time, not for us but for the staff working on the project, with sometimes a succession of contradictory messages or request, or improvisations (in particular in Prague, last July. Sorry)
5. My main problem was to find out what I was supposed to do as a reviewer. Lots of documentation was offered, explaining and justifying the METODIKA approach. It was written from the perspectives and needs of the METODIKA developers. After wrestling through this documentation I managed to summarize my review activities on a single page (“Work Table” on next page as an example). Starting with a straight “What to DO” document for reviewers could be more effective, while METODIKA development stuff serves as back ground reading.
6. With respect to the organization of the pilot, logistics and arranging facilities were done excellently. However, there is some concern about two aspects of the information provided. First, the amount of information provided beforehand was extensive and will probably not be read completely by future evaluators, if done in the same way in future. There was also a

certain degree of repletion/updating involved. Probably it is best that, after acceptance of the invitation by future evaluators, a complete package of information is provided which states concisely the purpose and details with a reminder relatively shortly before the event itself. Second and related to the first (and with no offense to all, but in particular the ladies, involved), the relatively large stream of emails before the event does not make the procedure necessarily clearer. We do understand that the event of 30-06/02-07 was a pilot, but for the future, a more limited amount of emails would be beneficial.

7. Perfect, but it is a little unfortunate to ask the experts to take care of their tickets and have to settle hotel bills
8. I look back to my participation in the pilot testing with great satisfaction and pleasure. I have experienced an open and constructive atmosphere to do our work. The meetings have been very well prepared. Also the secretarial support we got was very useful and efficient!
9. The pilot testing was well organized, there was quite some overlap in the background material submitted and this should be more concise and more coordinated to make it more homogeneous. The logistics were fine except that for me I stepped in late in the process.
10. IPN has a highly professional Metodika team – was a real pleasure to take part in the pilot testing. A great deal you have ahead- Viel Erfolg!!!
11. The overall organization in Prague was perfect. The material submitted arrived too late and incomplete. It should be provided the opportunity to print them so that they remain legible and should not be read to the computer. The logistics was very good.
12. overall organization very good ,materials submitted -very good in general, however: I would have very much appreciated to get printed material (at least the self-assessment report, the list of books, the bibliometric report; we are probably all at 70 hours per week, and doing this kind of work does not necessarily need the computer all the time). The writing in the self-assessment is very much too small; a readable layout would make it easier. Panel meetings logistics-Perfect
13. Space to discuss in the panel and outside. Perfect organisation, bravo!
14. Many suggestions above relate to the issues of materials submitted, on-site visits, logistics, etc. In general, the organization worked very well and I would like to thank you for the organization. Materials submitted were easily accessible, the website was very good. Personally, I would appreciate to get schedule of meetings more in advance, we are all working within tight time-schemes. But I understand the stressful situation in which your team was in May and June. Logistics that suppose that panel members stay for all costs (travel and accommodation) in advance is not good, and it may negatively influence the willingness of people to participate; of course, this is not the fault of Metodika project but the issue for the EU funding agency. The location in Prague 6 and the library was perfect, near to airport and at reasonable costs. Perhaps the July evaluation meeting could be shortened to three days and, as suggested above, on-site visits replaced by inviting RU's staffs for interviews. I

have been using Skype for meetings of small international groups for a long time and have never complained, even if I understand the advantage of meeting panel members face to face. Skype meeting cannot replace the final evaluation meeting but can be a good complement in future.

15. Everything was very well organized and we were very well received. It was a great pleasure to work in this environment. Děkuji vám!
16. It was perfect, everything was perfectly organized.
17. My comments are quite positive as far as the whole purposes of the project and its tools, quality of cooperation (both general and within the panel), efficiency. My personal difficulty at the beginning was especially that one of understanding the order or succession to be given to the submitted materials and the required level of information to be assured before the general and panel meetings in Prague and on site. The explanations received in Prague made the task possible and more than reasonable. The feeling of full responsibility and total freedom we got when received highly qualified our cooperation.
18. The overall organization was excellent and the team at the Ministry did a fantastic job. The same goes for the logistics of the panel meetings. There were some problems with the materials submitted, such as the self-assessment reports, the excellent research output, and the peer review reports for those. We think this resulted from misunderstanding and resistance towards the evaluation methodology. We feel such resistance should not be dismissed as refusal to change or an indication of the fossilization of Czech universities and research institutions. Especially in the humanities, the resistance may be propelled by a sense that there is something very wrong with this type of evaluation methodology: that it harms universities and research institutions rather than stimulating them. The Ministry could organize debates and consultations with the people concerned about such potential effects. If it takes the concerns of humanities researchers seriously, it will be much more successful in getting them on board for the entire evaluation methodology.

13 Other comments

Main panels

- More work needs doing on how the RUs submit their PhD numbers. We spent a lot of time trying to unravel this. This is clearly complicated by part-timers, and even more so by the fact that some PhD students are enrolled in universities but can be trained somewhere else. I think the organisers need to work out what they really want to know here, and simplify the information-gathering to achieve it.
- There needs to be agreement on how to count collaborations – if the intention is to promote collaboration, which I think is the case, then it is best to double-count collaborative publications, for example, i.e. both sides of the collaboration can claim it.
- I noted that some panel members made assumptions about the order of authors, e.g. that the first or last author was more important. I think this is risky. I would suggest that an instruction is given that panels do not make any assumption about author order.
- I would omit review papers from consideration on the grounds that they are not original research.
- I would provide for the panels a calculation of research income per investigator (or researcher) and number of PhD completions per investigator, because these are very useful comparators.
- Several panel members suggested that “research management” details were missing: e.g. appraisal of staff.
- In general the texts from RUs were not very good. In particular, they need reminding “don’t just describe the process for arriving at a strategy but describe the strategy itself.”
- In a full-scale exercise, the overseas panellists will need an explanation of the basic research landscape in the Czech Republic – funding sources, staff salaries, equipment funding, etc.
- I think there should be a greater emphasis placed on confidentiality of information. It shouldn’t be possible for RUs to identify the performance of any individual from the assessments they are given.
- **In conclusion**, we summarise that for the purposes of a future routine evaluation in humanities, it will be necessary to clarify how representative and structurally comparable the presented set of key publications of each of the workplace is, to solve the issue of language (it may be common that the referees invited will be confronted with languages that they are unfamiliar with) and field profile, i.e. a wider competency of the referees and their “insight” into Czech conditions, as well as their willingness and objective ability to assess a wide range of topics and how these topics were dealt with in the scientific studies presented, will largely influence the results of the evaluation, “freed” from the supposedly “hard”, i.e. easily quantifiable parameters of the “coffee grinder” system.

It will be necessary to define further criteria determining both national and international benefits of the evaluated institutions and their “societal relevance” more precisely.

When fulfilling these conditions (including accepting the above mentioned specific comments and after certain adjustments, we consider it potentially possible to also apply the Methodology tested to evaluating humanities in the future.

Evaluation aspects per Criterion

Assessment criteria	Sub-criteria (key words)	Evaluation aspects (Aggregated)	Input documents
I. Research environment	Quality of research management (including HR management for faculty and PhD programs)	<ul style="list-style-type: none"> • Yield & rate of completion of PhD projects • Development / training programs of PhD students • Postdocs & visiting researchers • Career evaluations and advancement of faculty • Faculty diversity & international exchanges of researchers • Sources and stability of funding • Quality of research facilities 	SAR ⁵ : Q9-34 SAR: SWOT
	Effectiveness of research strategy	<ul style="list-style-type: none"> • Articulation / mission / coherence • Focus / grouping in themes • Funding abilities • Competitive positioning • Beneficiaries of research • Limiting factors 	
II. Membership of the global and national research community	National research: participation, collaborations and recognition	<ul style="list-style-type: none"> • Partnerships with joint outputs • Collaborations with industry • Scientific positions, honours & prizes • Membership of academic advisory boards 	SAR: Q35-38 SAR: Q39-44
	International research: participation, collaborations and recognition	<ul style="list-style-type: none"> • Partnerships with institutions abroad • Associated outputs and visitors • Study trips abroad • Editorial boards & international conferences 	
III. Scientific research excellence	<i>Originality, significance, rigour</i>	<ul style="list-style-type: none"> • <i>Scholarly (selected best papers)</i> • <i>Non-scholarly (by expert opinions)</i> 	Integrated Report on Excellent Outputs
IV. Overall research performance	Research output	<ul style="list-style-type: none"> • Scholarly output rates (J, B, C, D) (Jimp, Jour-25%, CitImpE28, cited10%, cited-25%) • Non-scholarly output rates (patents, designs, pilot plants) • Value for advancement of research 	SAR: Q45-46 Bibliometric rep. SAR: Q47-54
	Competitiveness in research	<ul style="list-style-type: none"> • National competitive funding • Funding by competitive EU programs • Income from contract research • National positioning in research field • International positioning in field • Perceived limiting factors 	
V. Societal relevance	<i>Reach and significance</i>	<ul style="list-style-type: none"> • Knowledge and technology transfer activities; e.g. by research for non-academic societal target groups • Incubators, clusters, spin-offs • Membership Boards / advisory bodies • Marks of recognition by society • Societal value of activities 	SAR: Q55-60

⁵ SAR Self-assessment report

Pilotní ověření návrhu nové metodiky hodnocení výzkumných organizací
Samostatný doplňující dokument 8
Komentáře členů hlavních a oborových panelů
k metodice hodnocení a pilotnímu ověření

Vydává Ministerstvo školství, mládeže a tělovýchovy, Karmelitská 7, Praha 1
Individuální projekt národní pro oblast terciárního vzdělávání, výzkumu a vývoje:
Efektivní systém hodnocení a financování výzkumu, vývoje a inovací (IPN Metodika)
www.metodika.reformy-msmt.cz

Praha 2015