

Institutional Responsible Research Assessment Efforts and the Response from SSH Disciplines – the Case of Masaryk University

Michal Petr^{1*}, Natálie Hílek¹, Petra Mořkovská¹, Monika Kuchlei Sieberová¹

*petr@rect.muni.cz

¹ Research Office, Masaryk University, Žerotínovo náměstí 9, 60177 Brno (Czech Republic)

1. Introduction

Since 2022, the Coalition for Advancing Research Assessment (CoARA) has framed the world's research assessment (RA) agenda with ten commitments for reform. It has become the most central initiative for responsible assessment. Nevertheless, how do we implement such general ideas in practice, and do they fit the disciplines' needs? In the same year, Masaryk University (MU) conducted the Internal Research and Doctoral Studies Evaluation (IRDE) developed with a strong focus on ideas of responsibility. In this simple research, we operate with the definition of RRA as an approach that "**incentivises, reflects and rewards the plural characteristics of high-quality research, in support of diverse and inclusive research cultures.**"¹ Implementation of general ideas (e.g., CoARA) differs in each country's legislative and historic-cultural context. On the example of the MU, we firstly want to describe the internal research evaluation system to comply with the principles of responsibility and secondly to document the response from SSH faculties to this system and to discuss the problem of assessing the quality of research in SSH and how it fits RRA ideas.

2. Research Assessment at the Masaryk University

Central evaluative activities at Masaryk University form a triad of components (Figure 1), each with its clearly defined role: formative **research evaluation** (IRDE), **funding** (performance indicators and contract), and **monitoring** (bibliometrics). Even though these are complementary, research evaluation is separate from funding and is not affected by bibliometrics, which is used for strategic analysis of the environment. We use a set of intentionally selective performance indicators for funding, but they have only a partial weight and specified role in the system. In doing so, we aim to create a culture of responsible assessment – an environment where we reduce undesirable research incentives.

¹ Curry, Stephen; de Rijcke, Sarah; Hatch, Anna; Pillay, Dorsamy (Gansen); van der Weijden, Inge; Wilsdon, James (2020). The changing role of funders in responsible research assessment: progress, obstacles and the way ahead. Research on Research Institute. Report. <https://doi.org/10.6084/m9.figshare.13227914.v1>

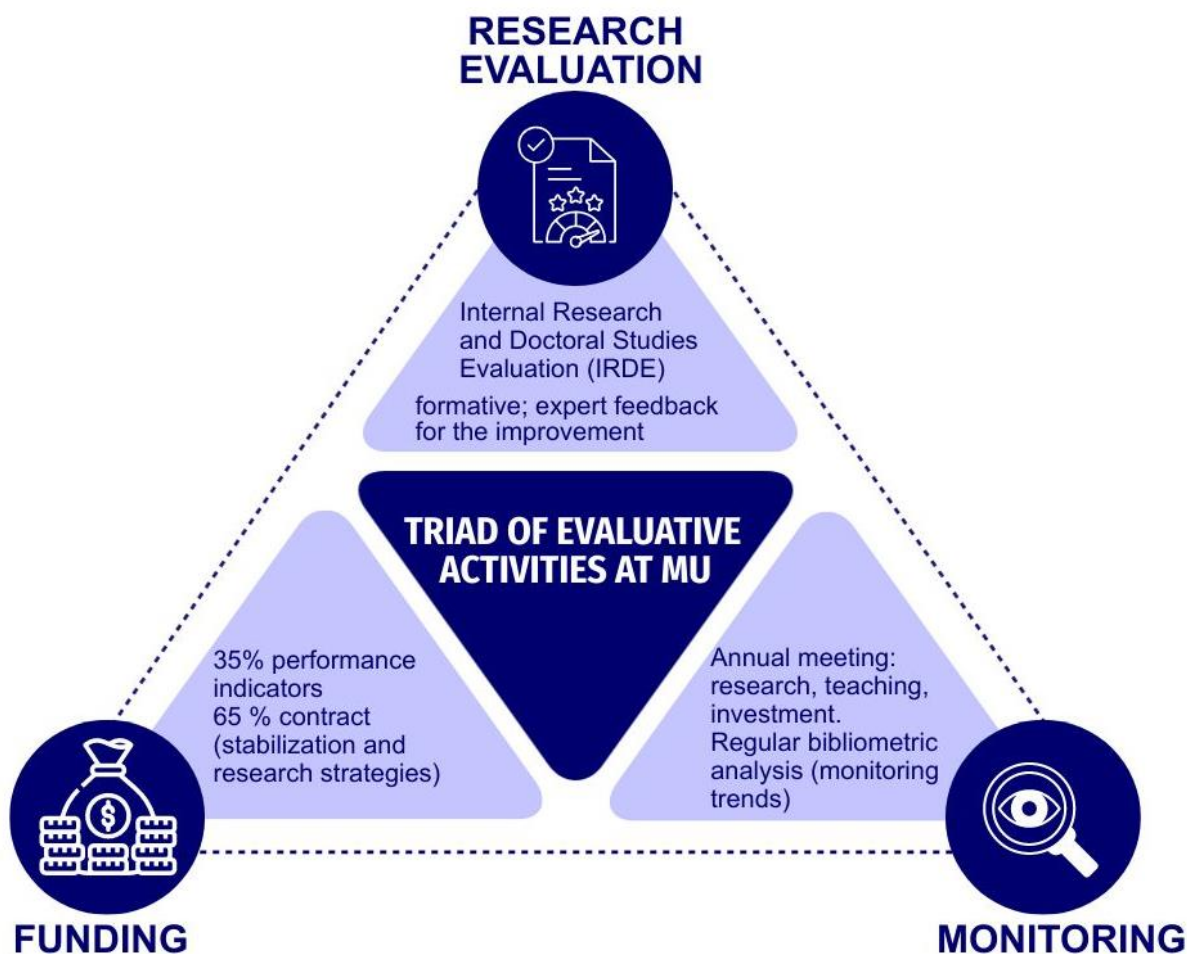


Figure 1 System of evaluation activities at the Masaryk University

Internal Research and Doctoral Studies Evaluation (IRDE) 2022

IRDE has an exclusively **formative** mission: it provides departments and doctoral programmes with expert feedback. Thus, IRDE is designed for their use and benefit, without direct financial incentives and unencumbered by bibliometric parameters. IRDE operates in a 5-year period, and the evaluated unit is twofold: **department** and **doctoral programme**. Since 2019, we have been developing the overall design and the criteria with uncompromising attention to the principles of responsible evaluation; we were inspired by recent best practices such as SEP in the Netherlands and SCOPE framework and made it CoARA compatible. This regard is unique in the historic-political context of the higher education system in the Czech Republic. The key features of the concept of responsibility are:

- Faculties/institutes participated in the planning and development.
- High level of flexibility: the setup and composition of panels were entirely up to the choice of the faculty/institute (disciplinary / broad faculty panel); self-evaluation reports of units are predominantly narrative and customizable, allowing a full range of information relevant in diverse disciplines.
- Bibliometrics is intentionally reduced and panoramic and only aims to deliver basic underlying publication statistics.

- Evaluating the evaluation phase and excessive communication of results: consultation with the creators of the SCOPE protocol, a conference dedicated to responsible evaluation, survey, and summary report critically evaluating the IRDE.

Such a system has proved to have several positive effects so far. A high level of flexibility, inclusivity, and missing link to money made the evaluation broadly accepted and led to the transformation of the evaluation culture. Some faculties kept their panel for the future as persistent ISAB.

IRDE and funding operate as separate processes. However, faculties, as an impact of evaluation, newly prepare a **strategic plan** that IRDE results may inform and that is linked to the significant part of the core budget for the research – **contract** (Figure 2).

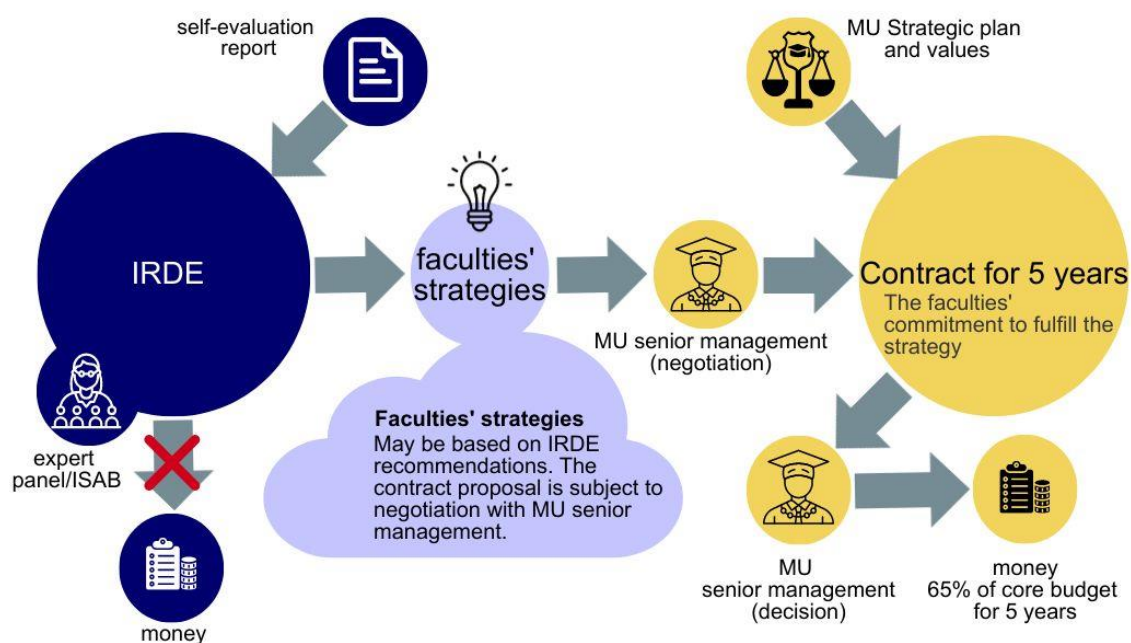


Figure 2 Impact of IRDE

3. Responsible assessment in SSH

In this simple research, we want to disentangle the drawbacks and benefits of this kind of institutional assessment for SSH disciplines. Publication patterns, channels, and types are much more heterogeneous in SSH disciplines than in science, technology, and medicine (STM). Research assessment in SSH is a persistent issue that is discussed in various communities (e.g., ENRESSH). Here, we want to **disentangle if implementing of the above-defined principles of RRA can bridge the gap to SSH and to what extent it fits SSH disciplines.**

We conducted six semi-structured interviews with MU SSH faculties' vice-deans for research: Faculty of Social Sciences (**FSS**), Faculty of Arts (**ARTS**), Faculty of Education (**PED**), Faculty of Business and Administration (**ECON**), Faculty of Law (**LAW**), Faculty of Sports Studies (**SPORT**).

Each vice-dean represents the full range of disciplines in their faculty. We asked four default questions – see Findings.

1. What is the notion of quality in your disciplines, and how can it be represented in the assessment?
2. To what extent does the concept of RRA correspond to your faculty's notion of how disciplines should be evaluated? What does responsibility mean in the assessment in SSH?
3. To what extent was the design of IRDE sufficient for the notion of quality in SSH and for the responsibility in assessment in SSH?
4. As for the systemic impacts on academia, how did your faculty respond to the IRDE?

4. Findings

1) What is the notion of quality in your disciplines, and how can it be represented in the assessment?

The question on the quality notion in SSH should have made a baseline for the attitude to the assessment model in the aspect of diversity and concerning one of the main functions of research assessment – identifying research quality. In the responses to the quality notions, we see two trajectories. The first and major trajectory accepts journal publishing and journal-level indicators as a general and convenient signal of quality and the effort to publish in community-recognized prestigious journals as a wanted publication pattern (FSS, ECON). However, they concurrently admit differences in this attitude between sub-disciplines (ARTS) or teams/individuals (LAW, SPORT). Secondly, societal relevance, i.e., the ability to address research topics with some influence on society, or "meaningfulness" is perceived as central (PED). In contrast, publications are considered one of more forms of communication in the field of Education, reserved for empirical research methods: *"...to narrow it down to them would be much like reducing the other ways in which the discipline works and communicates."* (PED)

Two respondents (ARTS, LAW) distinguish between different sub-disciplines and their attitude to the concept of quality: *"In law, there is certainly no consensus on this as a field, i.e., what is quality there are very different perspectives. Let us say a national group thinks that the goal of a researcher is to contribute to the development of the Czech legal system. Then, a group that recognizes the specifics of the law says that the goal is to contribute to knowledge on an international scale, which is done through the best journals and publishers."* (LAW)

Most respondents immediately began describing external signs demonstrating the desired performance rather than being confident in formulating "what is quality". We see this as an influence of external evaluation criteria (e.g., national systems): *"One group of researchers prefers bibliometrics, but I have to say that it may not be a natural tendency, it is rather an influence of external evaluation. In other words, even these people go to a Q1 journal because they know that the journal is Q1 and because it will bring them some points."* (LAW) These criteria, such as national assessment, may subtly stimulate the formation of a view of what constitutes actual disciplinary quality demands: *"Primarily, we follow disciplinary practices, but of course, we also*

have to play a game of national methodology. However, we do not perceive it strictly because it roughly corresponds to the disciplinary practices regarding the interpretation of bibliographic information." (FSS) Except PED, also ECON was quite confident in formulating the quality concept: "... peer recognition, that's probably the primary thing... The discipline has a particularly clear idea of what is the quality, they just direct it to a certain segment of journals, and they're interested in what's going on there. ...but being in WoS is in general no sign of quality at all." (ECON)

2) To what extent does the concept of RRA correspond to your faculty's notion of how disciplines should be evaluated? What does responsibility mean in the assessment in SSH?

According to ECON, the concept of RRA responds well to how the discipline wants to be evaluated. However, there can be some contradiction in the perception of evaluating in a national group, where researchers make their assumptions about how the quality in economics should be evaluated. ECON said there is no need for new evaluation approaches; quality can be found in the publication record; it does not matter how many citations a researcher has but what are his or her best ones. When it comes to diversity and inclusivity, ECON sees it as an essential part when it comes to the community of the faculty (e.g., hiring) but not necessarily as a part of the research: *"When researchers conduct research on Czech data for a Czech audience, they will not get it published in an excellent international journal, and that is ok, because they are focused more on social relevance of their research". (ECON)*

ARTS claimed that the concept of RRA fits the idea of people who are being evaluated, but something else is needed to fit the idea of ARTS management for evaluation. On one hand, ARTS appreciates the inclusivity and diversity given in evaluations, but on the other hand, they ask what the purpose of this kind of evaluation is: *"Many institutes have grown on the ruins of what was here before 1989 and have not gone through any wave of internationalization, have not connected with the world, and are very much encapsulated. For example, the Department of Theatre Studies publishes its journal. The evaluators were very positive about that. However, when you look at the performance evaluation, about 70 % of outputs are published in that journal, which seems to be a structural problem. It seems to be terribly asymmetrical". (ARTS)* This type of inclusive evaluation according to ARTS cannot help their national field to internationalize and break the encapsulation. ARTS also reflected the possibility of choosing evaluators, as some departments chose more critical evaluators. However, overall, they could not use the evaluation results at the faculty level as the evaluation was too diverse.

Similarly, LAW sees inclusive and diverse evaluation as something that can hold them back and appreciates one-sided researchers oriented only on national or international research. At the same time, researchers in Western countries can do both. Even though most researchers would agree on the RRA concept, the problem is that everybody seems to adopt that definition based on what they have already been doing. LAW also questioned the qualitative evaluation methods and concluded that there is no effective way to conduct it even though they admit they need it. LAW reflects on the question of choice of the evaluators, who are not critical enough and can hold back the development of the faculty.

Valuable experience with rather strict evaluators has SPORT, which got feedback from the evaluators that the excessive diversity and inclusivity in their research is stopping them from chasing excellence. *"I think that diversity and inclusivity need to be circumscribed because if it is boundless, it has no direction and no goal."* (SPORT) SPORT is also looking for methods to detect research quality, currently mostly for bibliometrics indicators and grants. However, they are determined to assess the role of researchers in the faculty and society, international involvement, cooperation with organizations, community engagement, etc. These factors, however, need to be evaluated qualitatively because not every appearance on local TV has the same value as cooperation with an army, which leads to significant research projects.

Rather a critical view on the RRA concept is held by PED: *"The idea of some great all-encompassing diversity is then an overwhelming issue that should rather be channelled by promoting a so-called problem-oriented approach in research."* PED sees the problem-oriented approach as concentrating on the focus of the research, the orientation towards problems, and its social relevance with impact on practice. Responsible evaluation should evaluate these research factors; therefore, the evaluation needs to be qualitative, and the evaluators must be experts from the given discipline.

FSS considers the RRA concept rather vague; they use bibliometrics to evaluate most of their production in basic research, and for the rest, they monitor monographs, applied research, and the third role of the researchers. For FSS is crucial to consider the specifics of the different research disciplines at their faculty.

3) To what extent was the design of IRDE sufficient for the notion of quality in SSH and for the responsibility in assessment in SSH?

Respondents agreed that IRDE was well set up and sufficiently flexible. It allowed faculty to adjust self-evaluation report sections to capture disciplinary diversity and quality. *"The setting was inclusive because faculty could significantly tailor the assessment to what law (field) thought was important. Evaluators' questions were directed toward the specificity of law."* (LAW). However, faculties did not always use this potential because they did not have experience with similar assessments. *"IRDE gave much leeway, but it could not be fully exploited because it was the first time. There was no previous experience with it."* (PED) For example in retrospect and given the evaluators' reactions, PED would modify the form of the evaluation reports, which were too structured.

The vice deans of 3 faculties (LAW, SPORT, and ARTS) mentioned that diversity and inclusivity may have been considered to such an extent that the evaluation of the quality of research in the field has been blurred. *"I honestly do not think that it (IRDE) was a completely true reflection of where the faculty stands. It certainly was inclusive, but maybe I am speaking purely for myself here now, that it was maybe too inclusive..."* (LAW), *"The inclusivity was indeed there, but it is much worse to work with it at the institution level!"* (ARTS) The experience of ARTS and LAW shows that diversity and inclusivity can be well applicable to the evaluation of academics. However, it is difficult to work with such information at the faculty level.

The committee's composition (evaluators) played a crucial role in the IRDE. The question is whether some dissatisfaction with the evaluation results was related to the choice of evaluators. Two respondents (ARTS, LAW) did not select strict or top researchers for disciplines: *"In many cases, the evaluators were such a good granddads."* (ARTS) *"The group that evaluated the Faculty of Law was more in the national direction, there were certainly not some superstars of the international scene, which also exist in law."* (LAW) On the other hand, SPORT, who also mentions too much inclusivity, was satisfied with the results of the evaluation, perhaps thanks to the appropriate evaluators: *"We invited European stars in the discipline, and they gave us tough feedback, saying that we are just weak because that is the way it is. We were too diverse and inclusive, and that did not lead to having science geared towards excellence..."* (SPORT) PED also mentions the selection of the evaluators as both a key and complex task. The importance of the selection of evaluators is confirmed by the ECON, who mentions that the benchmark for capturing quality in the field of economics was the evaluators, or in other words, recognized experts determine what is of quality in the discipline and what is not. This is consistent with ECON's leaning towards bibliometric indicators at the journal level as a general and appropriate signal of quality, where quality is assured by peer review.

4) As for the systemic impacts on academia, how did your faculty respond to the IRDE?

Most faculties have, to some extent, implemented the recommendation from IRDE in internal documents and research strategies or have seen a direct impact on the internal culture and the way of thinking of the faculty representatives. SPORT proceeded to the most fundamental restructuring based on the IRDE results, creating fewer new departments around defined research areas. Despite the recommendation for a narrower research focus, ECON decided not to change the faculty structure but created research themes that will be supported gradually. The reason was also not to jeopardize existing staff and to keep the structure according to the educational area. PED did not proceed with any of the structural reforms: *"... that type of assessment (IRDE) is not the calibre to trigger things like that... it can incentivise!"*(PED)

Some faculties set up support mechanisms that will lead to the achievement of the strategic objectives. FSS modified the internal faculty evaluation to correspond more closely with the IRDE focus and adjusted the weights for excellent publications. SPORT established the "Scientific Leaders" position within new departments primarily responsible for the department's scientific strategy. Despite not recognizing a real impact at the departmental level, ARTS decided to change the internal funding system based on publications. The funding system was also changed at ECON towards the possibility of creating financial reserves to support newly defined research topics.

LAW stated that the recommendations were partially translated into strategic plans, but no major transformative shifts existed. This could be due to a good rating as an evaluation result, which, in the personal opinion of vice-dean, does not correspond with reality and was presumably an effect of the choice of particular committee members or insufficient boundaries set by the Rector's Office

The impact on the internal culture was primarily mentioned by SPORT *"But I think that in general people have accepted it and that IRDE has helped us to really kick off an increase in quality ..."*

(SPORT) and PED, where the evaluation process itself has shifted the mindset of individuals in the faculty.

5. Conclusion

SSH faculties' response to this evaluation system model has diverse manifestations. As for organizational changes, two SSH faculties transformed their traditional structures (department/chair) towards STM-type research groups and topics (SPORT, ECON), which may boost strategic planning in research. Elsewhere, the evaluation has led to differentiated requirements for departments and research support thereof (ARTS).

For all SSH faculties, IRDE is a setup and a concept allowing a high level of flexibility and inclusivity to cover the full complexity of the disciplines' methods, patterns, and practices. **It is very much welcomed and appreciated.** However, bibliometric indicators still play the role of a stimulant towards strengthening the perceived quality of research with external (national) evaluations, especially when compared with the European West. This causes a critical controversy, especially in SSH in central and eastern European countries, where the internationalization, visibility, and impact lag somewhat behind their Western and Nordic counterparts. Here, SSH researchers perceive as necessary, first of all, to achieve the desired level of performance, however, communicated through journal-level indicators with their limitations and reductiveness in mind, before approaching the values of responsible evaluation, multilingualism and protecting local excellence: *"According to the academics from abroad we have spoken to, it is an automatic requirement there, and it is a level we just have to get to. People here make excuses that it is a terrible pressure and that it cannot be done, but they just do not want to do it. We just have to get to the level that is in our research area, and that's the standard."* (SPORT)

Social Sciences, economic disciplines, and part of the humanities see journal publishing and related metrics as central for identifying quality, as they are considered usual in STM fields. A question arises if a part of the reforming process should be to revise and rethink the evaluation criteria. Some patterns may be outdated even if we consider them valid (e.g., the prestige of WoS-indexed journals).

At the same time, SSH faculties require high flexibility and inclusiveness in listing all relevant research and teaching activities. Discussions and evaluations in SSH often consider outcomes such as necessary diversity, multilingualism, etc. Based on our experience and feedback, we assume that it is not only the criteria but the whole system, including the incentives and consequences set up, that are crucial for an effective and responsible SSH evaluation. For example, direct financial consequences can devalue even a very inclusive evaluation. Further, the management of an institution, which usually deals with various academic and organizational issues, usually wants research evaluation to provide straightforward answers to what are, in reality, very complex questions. This legitimate requirement must be balanced with the formative function of the evaluation. At the same time, the inclusivity and openness of the evaluation may result in gaming with the panel members and a somewhat blurred interpretation of the evaluation results, which contrasts managerial needs and prevents management from using them correctly.

The SSH evaluation undergoes a paradox. The need for an inclusive and diversified approach to evaluation in SSH is strong, and some disciplines (education and law) require these values to be even stronger. Nevertheless, the requirements from and the feeling of responsibility to the external evaluation, whether it is direct (state budget for research) or merely psychological (national-wide performance analyses), remain very strong. The bibliometric indicator is here, in fact, the most substantial criterion.

From the survey and experiences from the IRDE, we can formulate the following conclusions and recommendations for implementing RRA principles in real life, sometimes not only in SSH but also in other disciplines:

- To be meaningful, evaluation reform must occur not only at the central level of the institution but also at all levels and across all contexts (hiring, promotion, research evaluation, etc.).
- The values of inclusivity and diversity are welcomed but must not relativize different evaluation missions for different interest groups. These values must be conceptualized in the evaluation system. RRA commitments are not to be used very straightforwardly but according to the desired mission and purpose of the evaluation.
- Despite recent trends and publication patterns in SSH, the journal-level metrics remain essential in some SSH fields as a signal of quality and benchmarking tool for achieving desired performance in compared to EU-15.
- The focus on formulating values for evaluation and notions of quality in different disciplines² is needed to allow respective interest groups to benefit from evaluation results and, concurrently, as a protection against the possible inadequate influence of metrics.
- The new evaluation at MU was broadly accepted and (mostly) honest and led to the transformation of the evaluation culture.

² Ochsner, M. (2022). Identifying research quality in the social sciences (pp. 48–66). In Engels, T. C. E., & Kulczycki, E. (Eds.). (2021). Handbook on Research Assessment in the Social Sciences. ISBN 9781800372542.