

A qualitative and multicriteria assessment of scientists: a perspective based on a case study of INRAE, France

Abstract: Psychosociology theories indicate that individual evaluation is ~~integral~~ to the recognition of professional activities. Building upon Christophe Dejours' ~~conclusions~~, this recognition is influenced by two complementary judgments: the “utility” judgment from ~~the~~ hierarchy and the “beauty” judgment from the peers. The aim of this paper is to elucidate how at INRAE individual assessment of scientists is conducted. This process follows a qualitative and multicriteria-based approach by peers, providing both ~~congrats~~ and advice to the evaluated scientists (the “beauty” judgment). Furthermore, we expound on how INRAE regularly adapts this process to the evolving landscape of research ~~practices~~, such as interdisciplinary collaboration or open science, assuring that assessments align with the current ~~approaches~~ of research activities.

Keywords: multicriteria assessment, open science, peers, qualitative evaluation, research assessment

Introduction: why an assessment of scientists?

Scientists undergo various types of evaluations throughout their careers, including during recruitment, promotions, calls for projects and at various point along their professional trajectory.

 France, the assessment of civil-servant scientists with permanent position is mandatory and governed by decree 83-1260 (30th December 1983¹) which sets out the statutory regulations applicable across French research organisations. This obligation provides an opportunity to devise a method of peer assessment that is beneficial to both the individual scientist and the research organization. This articles focusses on the routine assessments conducted throughout a scientist's career and does not address promotions.

In France, as well as in other countries, there are primarily two types of institutions for research and higher education: research institutes focused mainly on research and universities which have dual goals of education and research. French universities recruit scientists based on both their scientific expertise and their teaching specializations. However, in terms of assessment, there is typically no regular evaluation of scientists at French universities, except for promotion purposes. Conversely, in French research institutes such as INRAE (French national research institute for agriculture, food and environment) scientists are recruited based on their scientific expertise, research skills and projects. At INRAE, assessments are conducted every two years, independently of promotions, and most of other French research institutes follow a similar practice.

In this context, our initial focus will be on examining the methods through which assessments can be conducted, with a particular emphasis on the indispensable role of peers. Peers play a crucial role in facilitating a nuanced and tailored qualitative assessment of scientists. Consequently, we will delve into the comparison between quantitative and qualitative assessment approaches and explore the evolving criteria essential for qualitative evaluations. Lastly, we will underscore the significance of adopting a multi-criteria approach for ensuring equitable assessments of scientists within an applied research institute like INRAE.

How can we assess “work”? The word « work » encompassed multiple meanings such as « labour » and « artwork ». Evaluation must consider this dual aspects of work taking account both the execution of tasks and the creative process. Following Christophe Dejours' observations and conclusions (2003), the concept of “work” itself, encompassing both its labir and artistic dimensions, cannot be directly

1

<http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000316777&fastPos=1&fastReqId=1906599245&categorieLien=cid&oldAction=rechTexte>

assessed. Instead, only the *outcomes* of the work are observable and accessible to evaluation (Dejours and Deranty 2010, 2018). Assessing these results commonly involves quantitative approaches such as measurement and counting. However, ~~th~~ efforts, strategies, tricks and ingenuity invested in achieving predetermined objectives, ~~which constitutes the essence of work~~, are often invisible to evaluators. Dejours argues that the true nature of work must be evaluated through qualitative criteria, such as those involving “telling”, “relating”, “explaining”, and “clarifying” -verbs that convey a “story” or narrative. Consequently, peers, who share the same profession, understand the **intricacies**, face similar challenges and innovate to overcome obstacles, are best positioned to identify and assess the essence of “work” according to Dejours’s definition.

Peer-assessment of **scientists** work is crucial and indispensable for different reasons. Firstly, recognition is vital for fostering well-being in the work-place, and this applies to all workers, including scientists. As emphasized by Christophe Dejours, recognition entails two fundamental judgments: the “utility” and the “beauty” judgment, both of which are complementary and essential. The “utility” judgement, provided by the hierarchy, signifies that the activities and contributions of the worker are valuable to the organisation, ~~thereby~~ imbuing their work with significance. Conversely, the “beauty” judgment, offered by peers **know understand** the intricacies of the job because they performed similar tasks, is equally essential. Peers possess unique insights into the challenges faced by the assessed individual and can discern hidden aspects of their work, going ~~beyond~~ standard requirements to achieve objectives. Therefore, peer assessment ensure a holistic **understandingof** scientist’s contributions and provides valuable recognition taking account intrinsic values.

These conclusions, proposed by Christophe Dejours, are general and not specific to the assessment of scientist’s work. However, they provide a framework ~~whiching which INRAE~~ proposes an assessment procedure to strive for the most accurate qualitative assessment of scientists possible: an evaluation conducted by peers collegially, with the aim of providing advice rather than **penalties** or rewards, and **considering wide** range of missions and activities corresponding to personal and professional trajectories. To our knowledge, INRAE is unique among other French research institutes and universities in referencing these conclusions (Laaser and Karlsson 2021).

The aim of individual assessment is to evaluate the work, not only its outcomes. In practice, the peers at INRAE refers to the report, **stiving** to i) comprehend the individual’s position within the organization (team, lab, institute...), ii) **focus** on the activities and achievements, ~~emphasizing~~ the substance of what is document rather than the **quantitiy** of publications, and iii) analyse the person’s insight into their work. These factors are **deliberated** among committee members (see below), and subsequently, a message conveying their “beauty” judgement is communicated to the evaluated scientist.

Quantitative versus qualitative assessment

Most assessment methods implemented by organisations, including research institutions and funders, have traditionally focused (and are still) on quantitative criteria, which are effective for evaluating the outcomes of work rather than the work itself. Quantifying achievements, such as publications for scientists, has been (and still is) a common practice, but is now considered as imperfect and even unfair. It is challenging to account for the fact that certain achievements, such as publications, vary in difficulty depending on the discipline and the hypotheses being pursued. One approach to quantifying outputs has been the use of journal impact factor, which is based on the notion that publishing in a high ranked journal indicates superior compared to publishing in a lower-ranked journal, as it is generally more difficult to do so. In recent years, there has been a shift away from relying solely on quantitative parameters such as the number of publications, impact factor, H-index, and journal quartile ranking. In fact, several organisations, including DORA, (the San Francisco Declaration on Research Assessment²), have even prohibited their use. This move stems from the recognition that simple ranking offers a narrow perspective of the work being evaluated. Quantitative parameters only measure the outcomes of work, i. e. the quantitative objectives. They fail to capture how scientists conceptualize their ideas, address the goals of their study, and navigate challenges or novel situations. None of these aspects – such as a person’s ability to adapt to professional circumstances, search for solutions, or content with unpredictable results – are discerned by quantitative parameters. These capacities represent a broader view of “real work” as described by Christophe Dejours, involving the ability to confront the resistance of reality. Sometimes, this “real work” may remain hidden because it diverges from conventional approaches (and may even be breaking the rules), or because the situation compels the worker to take risks and defy odds). In the realm of research, this hidden work may encompass all the unsuccessful attempts or negative results - experiments that did not fail outright but did not confirm the initial hypothesis. If these “failures” are not acknowledged, it obscures the efforts and time invested in exploring various avenues to generate new knowledge. From an assessment perspective, it is crucial to provide scientists with the opportunity to communicate and elucidate these elements, which encapsulate the myriad approaches and strategies employed to achieve scientific objectives (which may eventually be translated into publication, for instance).

The Coalition for Advancing Research Assessment (CoARA)³ campaigns for an “assessment of research, researchers and research organisations that recognises the diverse outputs, practices and activities that maximise the quality and impact of research. This requires basing assessment primarily on qualitative judgement, for which peer review is central, supported by responsible use of quantitative indicators”. This is a highly significant position statement signed by hundreds of European universities and institute,

² <https://sfedora.org/read/>

³ <https://coara.eu/> COARA: Agreement on Reforming Research Assessment

including INRAE, which prevailing the paradigm of quantitative assessment of research. In our view, the push for a paradigm shift originated from by bottom-up initiatives such as DORA and the Leiden Manifesto⁴, followed by top-down decisions at international European and national levels, and further reinforced by new bottom-up initiatives like Peer Community In (PCI)⁵. The overarching goal is to transition towards qualitative assessment methods.

The current paradigm of quantitative research assessment emerged in the 1980's, influenced by the theory of new public management, which emphasizes the use indicators to define research excellence. Consequently, employing bibliometric methods made it convenient for non-scientist stakeholders, such as policy makers and administrative managers, to rank, to compare individuals and organizations using a standardized set of indicators. Gingras (1996) clearly discusses bibliometry's application in research, highlighting both negative aspects, particularly when applied to individual assessment, and positive aspects, such as tracking the temporal dynamics of science. Hence, while indicators can prove valuable in specific contexts, they continue to play a central in assessment procedures.

Indicators are very valuable tools for ranking. A recent open discussion held by DORA⁶ focused on the potential negative repercussions of rankings in evaluation practices, particularly in connection with a capitalist approach to research activities. Krystian Szadkowski has linked the capitalist economy to the functioning of science, where liberalism provides a framework for competing for research funding and influences the ranking and prestige of institutions and researchers, illustrating the interconnectedness of these aspects. Indeed, there are numerous instances worldwide where governments prioritize "profitable" research, emphasizing "value for money" in research assessment, as evidenced by examples such as Martin (2011) in England, and the work of Gingras and Khelifaoui (2021) concerning French medical research. However, evaluation could serve as a catalyst for unraveling this complexity. Ranking, being ubiquitous and intertwined with capitalist economies, is a cornerstone of quantitative assessment methods. This poses challenges to public research, prompting several initiatives to advocate for a more thoughtful methods to ranking's use. While we will not delve into the topic of international rankings of universities and research institutions here (as our focus is on individual scientist assessment), we can highlight initiatives like the "More Than Our Ranks" (MTOR) initiative⁷ (<https://inorms.net/more-than-our-rank/>) within the the International Network of Research Management Societies (INFORMS). MTOR aims to equip institutions with tools to evaluate all their activities independently of rankings.

⁴ <http://www.leidenmanifesto.org/>

⁵ <https://peercommunityin.org/>

⁶ <https://sfdora.org/2023/06/07/to-rank-not-to-rank-or-to-rank-responsibly/>

⁷ <https://inorms.net/more-than-our-rank/>

Peers for a qualitative assessment scientists work

The challenge facing research organisations, their managers and scientists today is to adopt an assessment method that is more **suited** than strictly quantitative methods. The goal is to capture the **« hidden »** activities of researchers, which includes not only their successes, but also their difficulties, failures and strategies for overcoming obstacles – elements that truly represent their work beyond just the outcomes. Peers play a central role in this process, as they bear the responsibility of recognizing and evaluating these aspects.

Following the decree of 1983, INRAE established a process of regular ~~scientist assessment~~. Initially, the evaluation consisted mainly of a report and list of productions analysed by committees (see below) which provided feedback to the scientists. While this framework is still in use today, there have been changes in the content and the structure of the report and productions, with several additional elements being ~~developed~~ (see below). One significant reorientation occurred when France established a National division for research organization ~~assessments~~ in 2006, known as AERES, later becoming HCERES⁸). INRAE then began developing adapted criteria for research institutes based on finalized objectives. This marked the first instance of employing a multi-criteria approach to assess scientists which continues to be utilized today (see below). This shift was crucial for tracking the evolution of scientist's various ~~activities of a~~ at INRAE. Nonetheless, challenges persist, including the integration of assessment for open science ~~practices~~, of interdisciplinary research, scientific integrity, and qualitative assessment - goals central to this perspective.

Nowadays at INRAE, individual assessment of scientists is performed by groups of peers called “Specialized Scientific Commissions” (SSC), organized by disciplines or groups of disciplines. The 13 SSC cover all types of disciplines present at INRAE (**Table 1**), and each INRAE scientist selects the SSC that corresponds best to her or his activities. Each of the 13 INRAE SSC is a group of 20-24 scientists nominated or elected for four years - half of whom do not belong to INRAE, and is headed by a president who is external to INRAE. Within a committee, there are representatives from different categories of groups: gender balance is respected, as well as age categories, aiming to balance “junior” and “senior” assessors in order to capitalise on their different views. Each SSC follows precise guidelines by INRAE for several years (Direction de l’Evaluation, 2023). When peers meet to discuss the different dossiers, a representative of the Evaluation Department is present to guaranty the process, ensuring proper operation according to SSC rules, particularly to prevent any misuse of quantitative criteria and to respect the 25 criteria of discrimination prohibited by French law (article 225-1 of the penal code, 2022).

⁸ <https://www.hceres.fr/en>

Table 1: List of the 13 INRAE Specialized Scientific Commissions (SSC). Alphabetical order.

Agronomy, Livestock and Forest
Animal Biology
Ecology, Population biology and ecosystems dynamics
Economics, Sociology and Management science
Environmental sciences: earth, water and atmosphere
Interactions between Pests, Symbionts or Commensals with their Hosts
Mathematics, Informatics, Digital sciences, Artificial intelligence and Robotics
Microbiology, Microbial ecosystems, Agri-food systems, Biotechnology
Nutrition and Toxicology
Plant and Animal Genetics
Plant Integrative Biology
Research support and management
Science and food, materials and bio-based products engineering, Materials, Residual Resources

Peers (either internal or external to INRAE) are not remunerated for their work. INRAE considers that this is part of the “expertise” criteria (see below) participating in peer evaluation within the SSC. The average cost of one SSC per year is approximately 8000 euros for reimbursement, based on 3 to 4 days of meetings per SSC, with no interview (see below, conclusions) and for assessing approximately 700 scientists per year.

The SSCs produce independent assessments, regardless of the INRAE hierarchy. They provide advice which is collectively discussed. Specific referees, chosen from among the members of the SSC, are appointed for each evaluated scientist but remains unknown to her or him. This confidentiality is important as the judgment of “beauty” is thus given by a community of peers rather **then** by only one peer, in order to enhance the value and the significance of the assessment. The outcome of this process is a personal and dedicated assessment provided through a written message every two or three years by peers of the discipline to the evaluated INRAE scientists 

An advice-based assessment

As mentioned earlier, scientist assessment is based on the “beauty” judgment made by peers. INRAE’s aim is not to punish or reward, but to provide advice in a humanely manner and with good intentions. The advice typically strikes a balance between acknowledging positive aspects evaluated by the peers and offering opinions on the scientist’s choices (such as methods), the dynamics and relevance of

research, or potential future **directions** (for example in terms of collaboration). This general advice pertains to the trajectory of the evaluated scientists and may vary between junior **and** senior scientists. For example, the message might include the following elements and sentences: a contextual statement like “You work on the effect of...”, or “You are involved in projects aiming at...”; followed by a series of sentences congratulating **activities** (specific results, management of important projects, involvement in education if applicable, or other specific activities like open science); and finally a series of sentences offering advice or discussing elements of discussion concerning the near future, project orientation, or trajectory of the future career. When the situation indicates elements of degradation (see below), this can be delicately mentioned in the message. For example: “The committee has identified a critical issue concerning your publications since you have not **publish** since...”, or “**Regarding the relationship within your group which seems to limit your possibility to...**” In this context, the committee will inform your hierarchy to assist you in resolving this issue”. A considerable amount of time is taken to craft these messages in order to ensure clarity and limit potential misinterpretation by the assessed scientist. One particular consideration is the coherence of their work with the strategy of INRAE, even though this specific point, theoretically, is more a judgement of “utility”. Since the commissions work independently and autonomously in their assessment of “beauty”, there is no direct correlation with the judgment of “utility”. The latter is determined by laboratory directors or senior managers, and the evaluated researcher must include the opinion of their laboratory director in their file following a personal interview. This opinion and interview constitute a judgment of “utility”. In other words, throughout the assessment process, the researcher receives the two necessary opinions for recognition of their work (from peers and from the hierarchy), but these opinions are provided independently. However, there are indirect connections, as i) the hierarchy receives the messages delivered by the commissions to the researchers, and ii) the commissions may contact the hierarchy in case of difficulties encountered by the researchers (see below).

Upon reviewing the entire file written by the evaluated scientist, and considering the recommendation to present it in a narrative mode, peers can identify **various** types of difficulties the evaluated individual might have encountered. These difficulties may be evident either because the evaluated person describes a challenging situation in the document, or because SSC members perceive a lack of **dynamisms** or motivation. This action enables the hierarchy to take appropriate steps (e. g. contacting the scientist, the head of the lab) with the assistance of professional human resources personnel at INRAE to understand and support the individual and the laboratory in resolving the situation as much as possible. In **Table 2**, we provide an example of types of such situations detected in 2020 (involving 60 concerned scientists) and 2021 (involving 74 concerned scientists) out of a total of 1454 assessed scientists. The majority of these situations (78%) are easily resolved within a few months: a simple discussion between the scientist and their hierarchy is often sufficient to clarify the issue and find a resolution. In other cases, both parties agree that a change is necessary, and various solutions are ~~explored~~, such as a change of team or

laboratory, or even reassignment. Resolving these cases can take a year or even more. And very rare instances, the situation is severely degraded (for any reason), prompting a comprehensive analysis involving multiple human resources professionals.

Table 2: Types of difficulties identified by the peers, for 2020 et 2021. This corresponds to 105 different assessments (scientists). For each scientist, the difficulty might fit with different types.

Type	Junior scientists	Senior scientist	Total
Weak project	50	5	55
Lack of autonomy, motivation, dynamism...	14	1	15
Level of activity (lack of scientific output, dispersed activities...)	37	2	39
Working environment, conflicts, medical issue, overwork...	49	17	66
Non-delivered document or incomplete records	7	4	11

The results in Table 2 clearly highlight that "junior scientists" are more frequently identified with difficulties (157 occurrences versus 29). Several hypotheses can be proposed to explain this phenomenon. Firstly, it is possible that both INRAE and peers pay closer attention to junior scientists who are still establishing themselves as permanent researchers and thus may require additional support and guidance. Secondly, junior scientists are no longer in the post-doctoral phase and are expected to manage various aspects of research, such as proposing projects, seeking funding, and supervising students. This transition to a more comprehensive role in research may pose challenges for some junior scientists. Thirdly, the process of adapting to a new environment, new research topics, and the demands of a permanent scientific position can disrupt the ability of junior scientists to fully engage in their research activities during the initial years. Additionally, settling into a permanent scientific role may coincide with personal life changes, further affecting their capacity to fully immerse themselves in their work. Lastly, junior scientists at INRAE undergo assessment three times in five years, which is more frequent than senior scientists who are assessed only twice. This increased frequency of assessment for junior scientists may lead to a higher likelihood of identifying difficulties or challenges they may encounter. Overall, these factors contribute to the higher frequency of identified difficulties among junior scientists compared to their senior counterparts.

During the discussion among committees, different views on specific cases may arise. While there is no specific procedures in place to manage these discussions, the role of the committee president becomes crucial. Their personality and ability to foster an environment of attentive listening, and benevolence are instrumental in managing any conflicts that may arise and in reaching consensus decisions. If an unresolved conflict does occur between members of a committee (which to our knowledge has happened

only once in the last five years) then the Evaluation Department steps in to act as mediator between the involved assessors.

In conclusion, the role of peers is crucial in providing scientists with recognition through a “beauty” judgement and in proactively identifying and addressing potential to prevent situation from escalating. We place particular emphasis on supporting junior scientists during their initial years at INRAE. This involved conducting three assessments within the first five years following their recruitment, aiming to assist them in stabilizing their projects and initial achievements.

Evolution of qualitative assessment criteria

Qualitative assessment by SSC members is grounded in a framework and guidelines provided to scientists, aimed at guiding their writing towards a “storytelling mode” centred on i) facts and achievements, and ii) a reflexive analysis of the activity, including successes, failures, and difficulties (Direction de l’Evaluation, 2023).

Assessment solely based only on (inappropriate) quantitative metrics appears to be the simplest way for an automatic or administrative evaluation by non-experts, individuals who lack the ability to analyse the true quality of work and its outcomes. There are several ~~examples~~ of misapplication of quantitative indicators in the research assessment of individuals. One example pertains to impact factors and the number of citations, which often poorly correlate with the fundamental criteria expected for high-quality research: quality research: statistical robustness, the value of declared data, and replicability (Dougherty and Horne 2022). The transition from quantitative metrics to a qualitative approach on a broader scale may necessitate the definition and implementation of “qualitative indicators” In 2019, Wouters et al. (2019) proposed a framework comprising approximately 150 indicators and assessed their relevance for various types of evaluations, including infrastructures, research and funding organisations, individual researcher activities, career advancement, and recruitment. While this approach aids in defining qualitative indicators, there is a risk of succumbing to the temptation to assess ~~quantitatively~~ with an excessively lengthy list of indicators, ~~potentially leading~~ to a ~~regression into~~ the paradox of using a quantitative approach for a qualitative assessment.

Subjectivity refers to how individuals' perceptions and interests may influence assessment outcomes. Quantitative assessment likely aims to minimize subjectivity. However, it is important to note that much of the literature on subjectivity in work evaluation pertains to performances, indicating that we are still far from achieving a comprehensive assessment that transcends mere "performances." (Tran and Järvinen 2022). **Subjectivity in qualitative approaches is well recognized particularly in fields such as sociology and humanities (Bumbuc 2016).** It is crucial for everyone involved in qualitative assessment to be aware of when **subjectivity** comes into play and to address these potential biases when evaluating

research quality. Employing methods such as cross-checking analyses and seeking multiple perspective on a situation, commonly referred to as triangulation (Fichten & Dreier 2003), can help mitigate the adverse effect of subjectivity. Therefore, i) the use of indicators are essential to minimize the risk of subjectivity during qualitative assessment, and ii) employing a multicriteria approach, as done at INRAE, can further reduce this risk by balancing each criterion, ultimately allowing defining the profile of each scientist based on the distribution of their various types of activities. However, subjectivity cannot be entirely eliminated and is inherently part of the individual qualitative assessment process, prompting questions such as, "Is my judgment correct, especially when considered collectively within my SSC?" **By its very nature, this question lacks a definitive answer; instead, consensus within the group determines the validity of the judgment made.**

For an **organizarion** like INRAE, qualitative assessment must consider potential new directions in research practices. Recently, INRAE has chosen to enhance its activities and visibility in various area, which are elaborated below : expertise, partnerships, interdisciplinarity, and open science.

Expertise and support for public policies: the goal of expertise is to provide scientific  technical knowledge, tools and methods to stakeholders responsible for public policies including ministries, agencies, local authorities, European and international institutions, and universities. These resources aid in informing, designing, implementing, and evaluating public policies. At INRAE, expertise activities manifest in various forms such as collective scientific expertise, foresight studies, research for and on public policies, training, participating in working groups and public bodies, and the establishment and management of observatories or databases. These aspects are outlined in guidelines provided to scientists who are asked to describe these activities in their assessment files. Peers evaluate how these expertise activities ~~salignuit~~ with the three other types of activities, assess their coherence with the scientist's personal trajectory, examine their outputs and determine if they have been effectively disseminated to the appropriate audience.

Research in partnership with a view to contribute to all forms of innovation: innovation stems from ~~diversed~~ partnerships involving research or training institutions, technical centers, agricultural or agro-industrial institutes, competitiveness clusters, public and private economic entities and civil society organizations. The objective is to facilitate the co-construction of the value creation process among all project stakeholders. INRAE advocates for the concept of diverse of innovations, meaning that research may innovate to address economic, political, environmental, societal or health-related issues. Collaborations entails producing outputs with others that are enriched and different from what could have been achieved independently. Therefore, it is crucial in terms of assessment that researchers ~~explicitely~~ articulate their partnership approach in terms of co-design, co-construction and co-realization, with long-term programs punctuated by more focused projects. These partnership should address and respond to questions concerning original and beneficial research.

Practices of interdisciplinarity: by definition, a partnership aims to foster innovation, and achieve more collectively by leveraging differences, in ideas, skills, expertise, and resources. Collaboration involves working with individuals who may come from diverse scientific backgrounds. The success of an interdisciplinary partnership hinges on the ability to facilitate dialogue among individuals from ~~various~~ disciplines. Whether the partnership is academic, private, public or/and involves citizen participation, at either national or international levels, interdisciplinarity must be actively fostered in the processes of co-construction and co-realization. Considering interdisciplinarity in the assessment process is crucial in order to recognize the inherent costs associated with this interdisciplinary efforts and to focus, within this context, on the quality of research inquiry and the relevance of this approach. However, ~~practising~~ interdisciplinarity in terms of assessment can sometimes perceived as a disadvantage as it may blur professional identities and introduce complexities associated with belonging to ~~multiple~~ social groups (Negro and Leung 2013). A recent study delved deeper into this issue and revealed interdisciplinarity is often penalized due to reinforcement of social boundaries (Fini et al. 2023). At INRAE, since each SSC is ~~centred~~ on a specific discipline, there might be difficulties in assessing scientists who work at the interface of multiple disciplines. Peers within a particular SSC may not be fully equipped to deliver comprehensive and tailored “beauty” judgment. As a solution, INRAE allows scientists to be assessed by two different SSC, covering the disciplines relevant to their research (for example mathematics and ecology). In ~~thi~~ way, the evaluated scientist receives two complementary “beauty judgment” from each discipline. While this approach mitigates the issue of social boundaries mentioned by Fini et al. (2023), it remains a proxy for assessing interdisciplinarity work since each SSC evaluates only one discipline, ~~potentially~~ overlooking the true capacity to work at disciplinary interface.

Practices in open science: open science is a comprehensive approach aimed at enhancing the ~~reproducibility~~, transparency and robustness of research (Susi et al., 2022). Over the ~~pas~~ few years, there have been a significant international and European effort to promote open science practices, particularly concerning publications (open access), data, code and computer programs, and citizen science. Many international scientific institutions and universities have endorsed various manifestos, such as DORA (<https://sfedora.org/read/>) and Leiden (Hicks et al; 2015). ~~Additionally~~, several countries, including the European Community, have developed roadmaps to encourage scientists to ~~adapot~~ these new practices. The overarching principle behind this engagement is that the scientific content of an article holds greater importance than publication metrics, and that all the data should adhere to the FAIR principle⁹ and be well-described through metadata.

In **february** 2022 during an Open Science European Conference (OSEC) held in Paris, France, a significant number of European universities, research organisations (including INRAE) and funders

⁹ FAIR data are data which meet principles of findability, accessibility, interoperability, and reusability.

signed the “Paris Call on Research Assessment”¹⁰. The objective of this initiative is to reinforce the shared European vision regarding the recognition of quality and the diverse impacts of research that adhere to the highest standards of ethics and integrity. It aims to value the diversity of research activities, and recognize not only research outputs but also the proper conduct of research. Research organizations, including INRAE, now have a clearly defined framework for evaluating open science practices. In that context, we conducted a benchmark analysis to examine how different countries and organizations incorporate open science practices. This benchmark utilized a corpus of twenty documents produced by various international organizations, states or universities (such as Bristol, UCL, Utrecht) spanning from 2015 to 2022 (**Table 3**).

Universities	
Utrecht University	Recognition and Rewards Vision - 2021
Delft University of Technology	Open Science Programme 2020-2024 Research and Education in the Open Era Evaluation 2021 & Work plan 2022 - 2021
University of Bristol	Academic Promotions Framework 2021-2022 - 2021
Universities Norway	NOR-CAM - A toolbox for recognition and rewards in academic careers - 2021
Maastricht University	Room for everyone’s talent at Maastricht University - 2020
University College London	UCL Academic Careers Framework - 2018
University Medical Center Utrecht	Guide for reviewers/evaluators that use the UMC Utrecht indicators for impact - 2016
Ghent University	Vision Statement For Evaluating Research At Ghent University - 2016
University of Glasgow	Academic Promotion Criteria
Others entities	
YUFE Alliance	Open Science Assessment and Incentives at the YUFE Alliance - 2022
LERU	A Pathway towards Multidimensional Academic Careers - 2022
VNSU, KNAW, NWO	Strategy Evaluation Protocol – 2021. Standard Evaluation Protocol - 2015
DORA, EUA, SPARC Europe	Reimagining Academic Career Assessment: Stories of innovation and change - 2021
TJNK, TSV	Good practice in researcher evaluation. Recommendation for the responsible evaluation of a researcher in finland - 2020
FOLEC	Towards A Transformation Of Scientific Research Assessment In Latin America And The Caribbean - 2020
VNSU, KNAW, NWO and ZonMw	Room for everyone's talent - 2019
European Commission	Evaluation of Research Careers fully acknowledging Open Science Practices - 2017

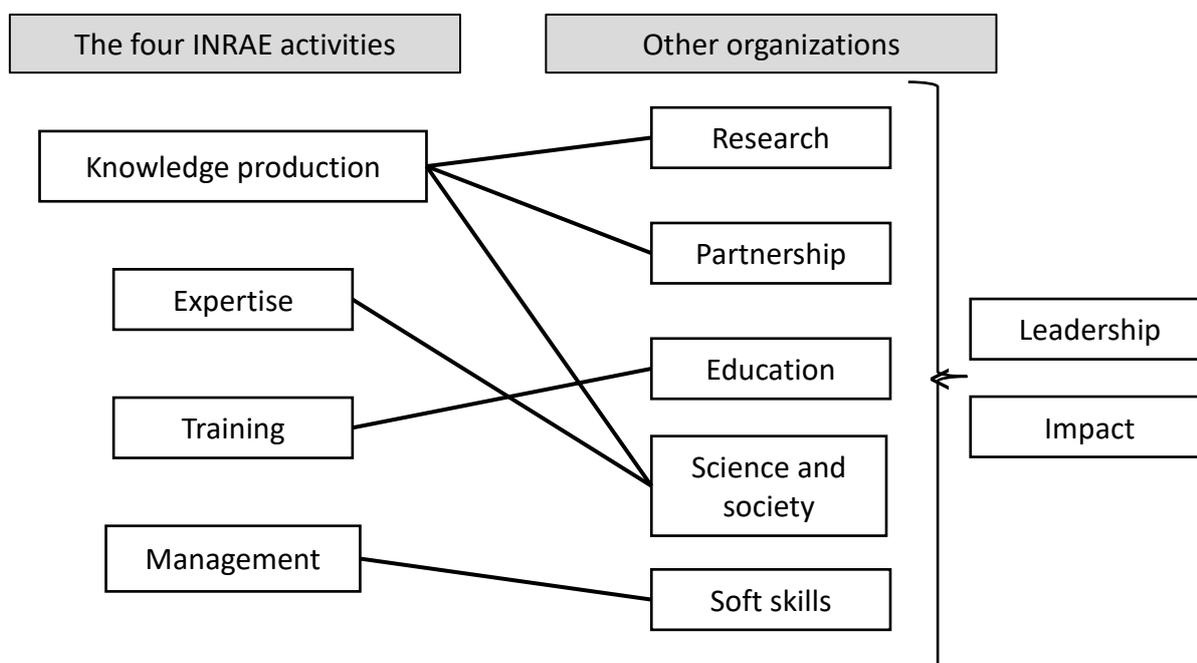
Table 3: Corpus used to compare and contrast scientist assessment at INRAE and other organizations. YUFE: Young Universities of the Future Europe. LERU: League of European Research Universities. KNAW, NWO and VSNU: association of universities in the Netherlands, which had already signed the DORA declaration. DORA: Declaration of San Francisco. EUA: European University Association. TJNK: Finish Committee for Public Information. TSV: Federation of Finnish Learned Societies. FOLEC: Latin American Forum on Research Assessment. ZonMw: The Netherlands Organisation for Health Research and Development.

¹⁰ <https://osec2022.eu/paris-call/>

Our benchmarking **primarily** focuses on universities, national roadmaps, and clusters of organisations such as the League of European Research Universities, and the European Commission, which have broader objectives than INRAE. We encountered challenges in augmenting this corpus with data from other scientific organizations that share similar objectives to INRAE.

As a starting point, we examined how the four main activities used to structure the assessment of INRAE scientists (i.e., production of knowledge, expertise, training and management) cross align with criteria considered by other international organizations. Initially, we observed that the four activities encompass all types of activities identified in assessment procedures within an open science context. In **Figure 3**, we illustrate the correspondence between INRAE categories and those of other **organization**. Despite **differences** in vocabulary (for instance, the term “Education” is more common in universities than “Training”) a strong **alignment** is evident. However, there are terms not explicitly **mentioned** in INRAE activities such as “Soft Skills” or “Leadership”. **We interpret these terms as referring to skills that permeate all types of activities, and are challenging to assess due to their focus on behaviour and abilities.**

additionally, while “Impact” assessment is proposed by other organizations, it is not directly addressed by INRAE. INRAE has developed and implemented tools (referred to as ASIRPA¹¹) to define, describe and measure the impact of research, but not at an individual level. This focuses on economic, societal, political, environmental and health impacts of research on the mid to long term based on projects or long-term research initiatives (Joly and Matt 2017; Joly et al. 2019). INRAE maintains that impact assessment cannot be attributed solely to one individual (the scientist), as impact is systemic and involves multiple actions and stakeholder.



¹¹ https://www6.inrae.fr/asirpa_eng/

Figure 1: comparison of categories used for assessment of research activities between INRAE (left) and other organizations (right).

Based on the different axes outlined in **Figure 1**, there is not a significant discrepancy in the mode of research assessment between INRAE and other international organizations. In fact, INRAE is one of the organisation that has established **specific** procedures and criteria for assessing open science practices. Practically, we recommend the individual being evaluated should mention their potential involvement in open science by i) listing new products specific to bibliodiversity (e.g. preprints, data repositories...), ii) describing actions toward achieving FAIR data principle, iii) outlining their strategy in open science (e.g. preference for diamond or golden journals...), and if applicable, iv) explaining their personal involvement or actions in open science initiatives such as participation in Peer Community In, engaging in open peer reviews, processes for sharing data, or involvement in citizen science projects.

Open science practices raises questions regarding scientific integrity due to the expansion of target audiences, widespread dissemination of research results and increased production. Consequently, there is a risk of misuse of these information, such as its propagation through social networks or preprints being misconstrued as validated science. This necessitates heightened vigilance regarding ethical and deontological considerations (Shaw 2003), as well as a focus on transparency and traceability of research processes. Moreover, there must be greater attention paid to research data, their management, and, when appropriate, their sharing. Regarding the assessment of scientific integrity, INRAE provides scientists with the opportunity to articulate in their report how they uphold their scientific integrity in general, and more specifically, in relation with open science practices. This allows individuals to reflect on their adherence to ethical principles and demonstrate their commitment to maintaining the integrity of their research endeavors.

A multicriteria assessment

There are several reasons for considering different criteria during the assessment of scientists. Firstly, INRAE is a research institute that integrates basic and applied approaches to achieve final objectives for society. This encompasses various disciplines and expertise, each requiring specific criteria, ranging from agronomists working with farmers to molecular biologist in the laboratory. Secondly, scientists' missions extend beyond knowledge production to encompass expertise, education, and management. Thirdly, scientists's missions evolve throughout their careers, with senior scientists often becoming increasingly involved in management roles.

In the mid-2000s, INRA and Irstea (the two founding organizations of INRAE) established and participated in the inter-institutional working group on the evaluation of the finalized research, known as EREFIN ("Evaluation de la REcherche FINalisée" which stands in French, translated to "Assessment of Finalized Research"). The objective was to promote a comprehensive evaluation that goes beyond

simply assessing the production and dissemination of new knowledge to include other missions such as expertise, training, and contribution to scientific culture. This initiative resulted in the development of tools that provide a framework for various activities, a catalog of possible potential outputs and descriptors, as well as assessment criteria (EREFIN 2011). Today, these tools remain largely utilized in various organizations and evaluation throughout France.

At INRAE, building upon the framework provided by EREFIN, we encourage scientists to report their engagement in various anticipated activities through four primary categories (**Figure 2**):

- Production of knowledge.
- Expertise and knowledge mobilization.
- Training through research, initial and continuing training.
- Animation or direction of institutional groups, major instruments, resources, programs or networks.

This list represents various potential areas of actions, subjected to the “beauty” judgment of the peers within the SSC; there is no expectation for any individual scientist to engage in all of these activities. Consequently, peers will not criticize a scientist for not participating in all four activities. Instead, peers are tasked with evaluating how scientists manage their diverse activities and whether this aligns with their objectives and career stage. At INRAE, assessment also involves examining scientists’ trajectory, recognizing that junior and senior scientists may allocate their time differently among the four different main types of activity. Depending on factors such as age, experience, career path, and scientific fields, researchers may engage with these four main types of activities to varying degrees (**Figure 3**).

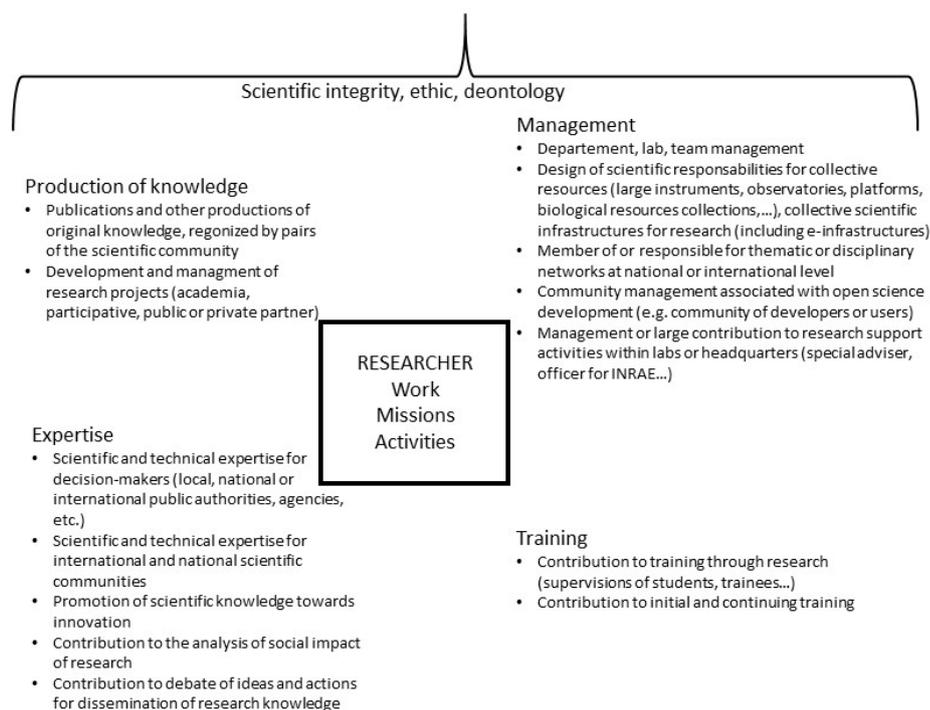


Figure 2: description of the four main types of activities at INRAE that scientists under assessment can mention and develop in their report.

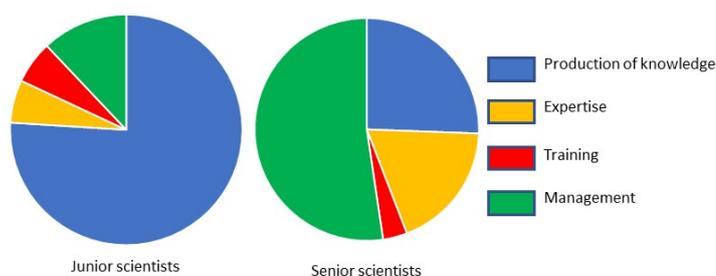


Figure 3: example of repartition of the activities of junior and senior INRAE scientists. CRCN: Chargé de Recherche Classe Normale (first degree of scientist recruitment at INRAE). DREX: Directeur de recherche classe Exceptionnelle (last degree of promotion at INRAE). Based on 4989 responses of scientists on a period of 2015-2021 years.

Conclusions and tracks for the near future

On July 20, 2022, the “Agreement on reforming research assessment” was published by the European Commission (via Science Europe, the Coalition for Advancing Research Assessment CoARA)¹², after ~~consultating~~ more than 350 organisations from 40 countries. This declaration outlines 10 commitments that signatories, including INRAE, pledge to uphold and implement. Many of these commitments are already integrated into INRAE’s individual researcher assessment processes, such as recognizing **divers** contributions, conducting qualitative evaluation, discontinuing inappropriate metrics, allocating resources for assessment reform, multi- and tailored criteria, and ensuring transparency in assessment processes. However, there is still room for improvement, including enhancing the exchange of practices with partners, communicating progress made (as done in this paper), and evaluating our ~~practices~~. The following points highlight represent the evolution that we believe is important for INRAE to continually consider, and even anticipate, changes in context and practices of being a scientist.

Training for assessment. Emphasizing several missions, such as involvement in expertise and support for public policy, as well as the promotion of open science practices, may be perceived by researchers as an unwelcome “top-down” imposition that adds to their workload without apparent benefit. Additionally, in certain disciplines, such as biology/medicine compared to mathematics, the adoption of such practices may be slow due to structural or historical factors. Advocacy and training are ~~therefor~~ essential for the researchers to understand the benefits of these practices. Similarly, providing training to peer members of the SSC is crucial to assist them in applying an appropriate “beauty” judgment regarding these new practices.

A shared assessment between organizations. INRAE does not claim to be the first in France or Europe to undertake this endeavor. However, we are actively engaged in the ~~qualitative~~ assessment process based on the “beauty” judgment. In response to top-down initiatives from Europe and national ministries, INRAE collaborate with other universities and research institutes in various national ~~committees~~. For example, INRAE, CNRS, and Inserm share common objectives aim at improving qualitative ~~assessment~~ based on international (DORA), European (CoARA), and national directives. This collaborative effort is essential for real progress in ~~assessment~~ criteria because the globalization of science requires paradigm shifts be embraced by a vast majority of scientific organizations.

It is essential that all stakeholders involved in assessment share common main objectives and use similar processes with appropriate indicators. Assessment of researchers occurs not only regularly by specific commissions such as INRAE SSC, but also at ~~keypoints~~ of their careers during recruitment or promotion, and the different juries have to apply the same general guidelines towards qualitative and

¹² <https://www.scienceurope.org/our-resources/agreement-reforming-research-assessment/>

multicriteria evaluation, including open science practices. Assessment of collective structures or projects must also share a common base. In France, HCERES (the national agency for research and education assessment of organisations and laboratories) and ANR (national funding agency for research) use a multicriteria approach for the assessment of organizations and projects, respectively, and recognize open science contributions, reducing the use of metrics such as H index and impact factor (the European ERC committees proceed the similarly)¹³. These are signs that assessment processes are changing and that a general common framework is shared at all assessment levels of researchers activities, funding and career development. Thus, the assessment procedure at INRAE i) follows what is usually performed in other research organizations but ii) is original in its way ~~centering~~ on a multicriteria approach and a qualitative assessment adapted to a “beauty” judgment given by peers and iii) could be easily transferable to other institutes and universities (DORA Case study, 2023).

Impact of evaluation for scientists and INRAE. Individual assessment of scientists has two main outputs: one for INRAE in terms of identification of potential new scientific breakthroughs, and for the scientist in terms of professional trajectory and development. In short, these outputs concern the “evaluation of evaluation”, or the impact of assessment. For the Institute, the evaluation reports compiled by the researchers represent an unparalleled and highly valuable database from which both quantitative and qualitative analyses can be conducted. These analyses serve to assess the current state (strengths and weaknesses) of research expression at INRAE and to identify opportunities or risks for the future. INRAE is currently focused on i) determining the questions to which evaluation data could provide answers, and ii) developing the tools to extract and analyze the information. For the scientists, ASIRPA's ex-post tools (Joly and Matt 2017) could be adapted to analyze the impact of specific projects on researchers' careers, aiding them in gaining perspective on their career trajectory and impact, thus clarifying their professional future (scientific directions, mission orientation, etc.).

Towards assessment by interview. INRAE currently has a roadmap that includes several items, such as integrating elements of scientific integrity into scientists's reports. However, it remains challenging to define criteria or indicators for this integration (Moher et al. 2020). One option could be to incorporate an interview between the assessed scientist and the SSC; this approach might allow for a deeper exploration the scientist embodies scientific integrity across four main aspects: reliability, honesty, respect and ~~responsability~~. Additionally, interviews could encompass other aspects related to **soft-skills**, which are ~~more easy~~ to appreciate through ~~oral exchange~~. Interviews would also provide an opportunity to engage scientific discussions, further aiding ~~scientists~~ in their professional trajectories. Although scientists are encouraged to mention difficulties or failures they may have encountered in their report,

¹³ In French: <https://openscience.pasteur.fr/2022/01/17/le-hceres-signe-dora-et-fait-le-choix-dune-evaluation-multi-criteres-et-plus-qualitative/#626207478e82ce3af6eb/archives>, <https://anr.fr/fr/actualites-de-lanr/details/news/lanr-en-soutien-dune-science-ouverte/>, page 6 of <https://anr.fr/fileadmin/documents/2022/ANR-COP-2021-2025.pdf>

an ~~interview~~ may be better suited for explaining these delicate situations. There numerous advantages to considering interviews in the individual **assessment** process. However, two obstacles currently exist: firstly, the number of scientists at INRAE (approximately 2,500) makes organizing such interviews challenging, and secondly, interviews inherently breach the confidentiality rule of assessment. Nevertheless, one potential compromise for conducting these interviews could involve selecting a subset of assessed scientists based on specific career periods, thereby accepting that confidentiality may be breached in these cases.

Environmental impact. The evolution of evaluation procedures and criteria is necessary as it aligns with the **evolving** missions and work methods of scientists. The **rofessionnal** environment influences how scientists carry out their job. A simple example is the impact of the Sars-Cov-2 epidemic on in-house work and **telecommuting**. However, on a deeper level, broader environmental and societal contexts (not directly related to scientific activities) may prompt new ways of considering scientists's role. At INRAE, some scientists are mindful of the impact of their activities on climate change and carbon footprint. Consequently, they may adjust their experimental plans to be more energy-efficient, reduce international air travels. These changes in ~~practices~~ could affect the scale of experiments and/or international collaborations. It is too early to determine whether those potential changes will become long-term practices, but it illustrates of how assessment procedures might evolve to account for the environmental impact of research activities.

Conflict of interest disclosure

The authors declare they have no conflict of interest relating to the content of this article. **XXX** was recommender to PCI Zoology and member of the PCI Geomics board till December 2023.

Funding

No specific funding was allocated for this perspective.

References

- Bumbuc S, 2016. About subjectivity in qualitative data interpretation. International Conference Knowledge-based organization, Vol. XXII No 2
- Dejours C, 2003. L'évaluation du travail à l'épreuve du réel : critique des fondements de l'évaluation. Editions QUAE, ISBN 978-2-7592-2460-9, pp 62

Dejours C & Deranty JP, 2010, The Centrality of Work, *Critical Horizons* 11: 167–80.

Dejours C & Deranty JP, 2018, *The Return of Work in Critical Theory: Self, Society, Politics* (New Directions in Critical Theory Book 56)

Direction de l'Evaluation, 2023. *CSS Guide Book 2020-2024 INRAE 2020-2024*. Hal INRAE (hal-04202855)

DORA case study, 2023, National Research Institute for Agriculture, Food and the Environment (INRAE), France, in preparation

Dougherty MR & Horne Z, 2022. Citation counts and journal impact factors do not capture some indicators of research quality in the behavioural and brain sciences. *Royal Society Open Science*, 9(8), 220334. <https://doi.org/10.1098/rsos.220334>

EREFIN, 2011. *Evaluation des collectifs de recherche : un cadre qui intègre l'ensemble de leurs activités*, internal report. DOI 1.5746699115890933e12

Fichten W & Dreier B, 2003. Triangulation of subjectivity. *Forum Qualitative Sozialforschung Forum: Qualitative Social Research*, 4(2). <https://doi.org/10.17169/fqs-4.2.702>

Fini R, Jourdan J, Perkmann M & Toschi L, 2023. A new take on the categorical imperative : gatekeeping, boundary maintenance, and evaluation penalties in science. *Organization Science*, 34(3), 1090-1110. <https://doi.org/10.1287/orsc.2022.1610>

Gingras Y, 2016. *Bibliometrics and research evaluation : Uses and abuses*. The MIT Press.

Gingras Y & Khelfaoui M, 2021 *L'effet SIGAPS: la recherche médicale française sous l'emprise de l'évaluation comptable*. *Zilsel: Science, technique, société* (8), 145-174.

Hicks S D, Wouters P, Waltman L, de Rijcke S, Rafols I, 2015. The Leiden manifesto for research metrics, *Nature*, 520, p. 429-431

Joly PB, Matt M, 2017. Towards a new generation of research impact assessment approaches. *J Technol Transf*, <https://doi.org/10.1007/s10961-017-9601-0>.

Joly PB, Matt M, Robinson DKR, 2019. Research Impact Assessment: from ex post to real-time assessment, *Fetval Journal*, 47, 35-40, <https://doi.org/10.22163/fteval.2019.326>

Laaser K & Karlsson JC, 2021. Towards a Sociology of Meaningful Work, *Work, Employment and Society* Volume 36, Issue 5 <https://doi.org/10.1177/09500170211055998>

Martin BR, 2011. The Research excellence framework and the ‘impact agenda’: are we creating a Frankenstein monster? *Research Evaluation*, 20(3), 247-254. <https://doi.org/10.3152/095820211X13118583635693>

Moher D, Bouter L, Kleinert S, Glasziou P, Sham MH, Barbour V, et al., 2020. The Hong Kong principles for assessing researchers: fostering research integrity. *PLoS Biol* 18(7): e3000737. <https://doi.org/10.1371/journal.pbio.3000737>

Negro G & Leung MD, 2013. “Actual” and perceptual effects of category spanning. *Organization Science* (on line) <https://doi.org/10.1287/orsc.1120.0764>

Shaw IF, 2003. Ethics in qualitative research and evaluation. *Journal of Social Work* 3(1) 9-29
DOI:10.1177/1468017303003001002

Susi T, Heintz M, Hnatkova E, Koch W, Leptin M, Andler M, Masia M, Garfinkel M, 2022. Centrality of researchers in reforming research assessment: Routes to improve research by aligning rewards with Open Science practices. ISE, online report <https://initiative-se.eu/paper-research-assessment/>

Tran, T-V & Järvinen J, 2022. Understanding the concept of subjectivity in performance evaluation and its effects on perceived procedural justice across contexts. *Accounting & Finance* 62 (3) <https://doi.org/10.1111/acfi.12916>

Wouters P, Ràfols I, Oancea A, Caroline S, Kamerlin L, Holbrook JB, Jacob M, 2019. Indicator Frameworks for Fostering Open Knowledge Practices in Science and Scholarship. Report. European Commission Directorate-General for Research & Innovation, Directorate G — Research doi: 10.2777/445286