Qualitative Evaluation of Academic Researchers in Computer Science: Practices and Reflections from CNRS

Pierre Senellart

• Pierre Senellart is with DI ENS, ENS, PSL University, CNRS, Inria; Paris, 75005, France. E-mail: pierre.senellart@ens.psl.eu.

Abstract: The evaluation of academic researchers often relies on quantitative metrics, despite their limitations. This paper examines the qualitative assessment approach used by the French National Center for Scientific Research (CNRS) in Computer Science (Section 6), as presented by a National Committee for Scientific Research (CoNRS) committee president. Rejecting bibliometrics like impact factors, CNRS emphasizes in-depth peer review of research quality, diverse outputs (e.g., software and datasets), and field-specific considerations. The process, aligned with the San Francisco Declaration on Research Assessment (DORA), faces challenges in scalability and subjectivity, addressed through committee diversity and structured deliberation. This work provides insights for institutions seeking fairer, more holistic research evaluation frameworks.

Key words: Research evaluation; Qualitative assessment; CNRS; DORA; Computer science

1 Introduction

The evaluation of academic researchers has long been a subject of intense debate within the scientific community. Traditional assessment methods that rely heavily on quantitative metrics—journal impact factors, citation counts, and h-index scores—have increasingly come under scrutiny for their limitations and potential distortions of research quality[1]. While these bibliometric indicators offer apparent objectivity and ease of comparison, their widespread adoption has led to unintended consequences, including the “publish or perish” mentality, gaming of citation systems, and the marginalization of important but less quantifiable research outputs[2]. This quantification paradox has sparked global movements advocating for more nuanced approaches to research assessment, most notably embodied in the San Francisco Declaration on Research Assessment (DORA)[3].

Within this context, the French National Center for Scientific Research (CNRS) presents a compelling alternative model. As Europe’s largest fundamental science agency, CNRS employs over 11,000 tenured researchers who hold unique full-time research positions without teaching obligations[4]. This distinctive structure, coupled with CNRS’s partnership with top French universities – including PSL University[5], which the author is affilated with – creates an environment where research evaluation can focus purely on scientific merit rather than conventional academic metrics. The CNRS system is particularly relevant for computer science, where research outputs extend far beyond traditional publications to include software, algorithms, datasets, and systems that often resist simple quantification[6].

This paper draws on the first-hand experience of CNRS’s evaluation practices, particularly in Section 6 (Theoretical and Core Computer Science) of its National *Committee* for Scientific Research (CoNRS). We aim to:

- Detail the principles and implementation of CNRS’s qualitative evaluation framework;

- Analyze its effectiveness in assessing computer science research, with special attention to non-traditional outputs;
- Extract transferable insights for improving research assessment in similar institutions globally.

By examining this established yet innovative approach, we contribute to the ongoing dialogue about creating fairer, more meaningful systems for recognizing research excellence—systems that value substance over symbolism and diversity over standardization.

2 Institutional Framework

2.1 The CNRS ecosystem

As Europe’s largest fundamental science agency, the French CNRS operates through a distinctive distributed model that blends institutional autonomy with interdisciplinary collaboration. Its organizational architecture comprises three interdependent tiers: the central governing body that sets strategic priorities, over 1,000 locally embedded research units conducting day-to-day investigations – often in partnership with universities and other research agencies such as Inria or INSERM – and the independent CoNRS that safeguards evaluation integrity. This structure enables CNRS to maintain France’s position at the forefront of theoretical computer science while accommodating diverse research cultures. As an illustration, the author is a Professor at ENS-PSL, a school of PSL University, and a member of the Computer Science departement (DI ENS) of ENS-PSL, which is a research unit affiliated with ENS-PSL, CNRS, as well as Inria. The CoNRS is in charge of the evaluation of the research conducted by CNRS researchers within DI ENS, independently from its central governing body.

2.2 CoNRS’s governance model

The CoNRS of the French CNRS is structured to balance academic autonomy with accountability through its governance model. Comprising 41 disciplinary sections, CoNRS ensures representation from both the national academic community, who elects the majority of its members, and the Ministry of Higher Education and Research, who appoints a minority to align with national priorities. Section 6, dedicated to Theoretical and Core Computer Science (renumbered Section 2 since September 2025), consists of 18 specialists (11 elected, 7 named by the Ministry for Higher Education) responsible for evaluating researchers in this domain. These members spend a significant portion of their time annually on assessments for recruitment, promotion, and regular evaluations. CoNRS emphasizes transparency in managing conflicts of interest, requiring members to abstain from discussions and decisions where conflicts exist. The committee also mitigates subjectivity through its diverse membership, ensuring a broad range of perspectives in the evaluation process. This model supports rigorous and fair evaluations while adapting to the evolving nature of computer science research.

2.3 Disciplinary partitioning in computer science

CNRS’s bifurcation of computer science into Section 6 (theoretical foundations) and Section 7 (applied systems) – since September 2025, respectively Sections 2 and 3 – reflects a deliberate epistemological strategy to accommodate the field’s methodological diversity, but also a practical concern of splitting a field that has become too large for a single evaluation committee. Section 6 encompasses areas like cryptography and formal methods where contributions are assessed through mathematical proof rigor and conceptual novelty, but also networking and software engineering, giving a large space to implementations and systems. Section 7 covers areas such as machine learning, robotics, or signal processing, with connections to both applied mathematics and to engineering and technological transfer. This division proves particularly consequential in promotion cases, where a candidate's work at the interface of the areas of both sections may undergo parallel reviews. A representative cross-sectional evaluation of a researcher developing verified neural networks requires weighting between Section 6's emphasis on verification completeness and Section 7's focus on model accuracy – an example of interdisciplinary assessment.

3 Principles of Qualitative Evaluation

3.1 The case against quantitative metrics

The CNRS evaluation system's rejection of quantitative metrics stems from well-documented systemic flaws that compromise research assessment validity. Bibliometric indicators, particularly prevalent in computer science evaluation, suffer from three fundamental limitations. First, their susceptibility to manipulation undermines reliability – a concern highlighted by recent cases of fabricated publications on ResearchGate with counterfeit citations to inflate metrics. Second, cross-disciplinary incomparability renders them ineffective for a multidisciplinary organization like CNRS; the expected annual output for a theoretical computer scientist (e.g., 1–2 substantial proofs) may radically differ from that of a machine learning researcher (often 5–10 experimental papers, sometimes within large teams). Third, emerging risks from AI-assisted evaluation tools have proven unacceptable, both due to privacy violations (processing confidential career dossiers without consent) and the impossibility of auditing algorithmic decision-making processes. These limitations collectively justify CoNRS's categorical exclusion of impact factors, h-indices, and citation counts from evaluation criteria.

3.2 Implementing the DORA principles

As a signatory to the DORA, CNRS has operationalized its principles through concrete institutional policies. Most radically, the complete decoupling of publication venue prestige from research quality assessment represents a structural departure from global academic norms. In practice, this means a seminal paper presented at a small regional workshop would theoreticlaly receive equal consideration to one published in a top-tier journal, provided their intellectual contributions prove equivalent upon full-text review. The declaration's emphasis on evaluating "research on its own merits" manifests in Section 6's requirement for candidates to submit narrative research statements contextualizing their contributions beyond publication lists. However, full adherence to DORA's ideal of exhaustive output review remains aspirational – a tension explored in Chapter 4's discussion of implementation challenges.

3.3 A multi-dimensional evaluation framework

As illustrated in Fig. 1, CNRS's qualitative framework evaluates researchers across six interdependent dimensions, each calibrated for computer science's unique characteristics. Research quality assessment prioritizes long-term impact over immediate productivity, valuing foundational theoretical advances (like new complexity class separations) equally with applied breakthroughs (such as verified cryptographic implementations). The coherence of a researcher's agenda receives particular scrutiny, with evaluators tracing thematic progression across publications and software artifacts. Community engagement metrics extend beyond conventional service roles to include substantive contributions like OpenReview peer critiques and standardization body participation – activities that reflect computer science's conference-centric culture, particularly in theoretical fields where program committee roles carry substantial weight. For senior researchers, leadership manifests through both academic supervision outcomes (including PhD completions) that are systematically tracked, and institutional stewardship, such as directing shared research infrastructure.

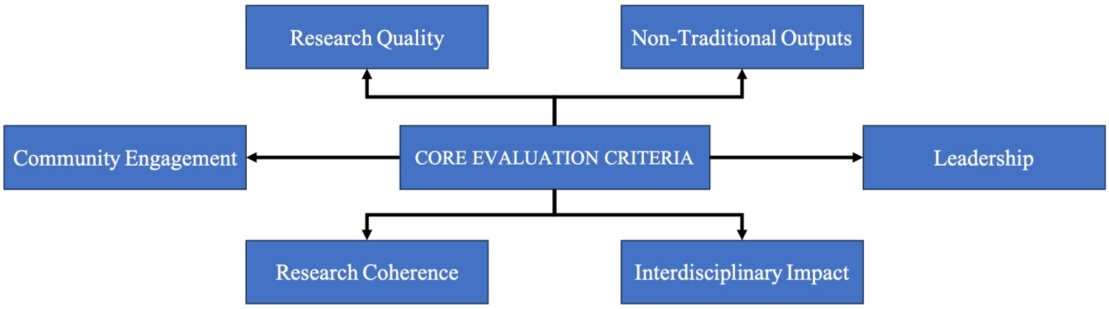

Fig.1 Multi-dimensional qualitative evaluation framework for computer science researchers at CNRS. adapted from CoNRS Section 6 evaluation guidelines (2023).

3.4 Recognizing non-traditional research outputs

Computer science's distinct scholarly products necessitate evaluation methods beyond conventional publication analysis. Software artifacts undergo rigorous assessment, with open-source projects evaluated based on technical quality (measured through code review), documentation completeness, and community adoption metrics like GitHub stars or dependency graphs. Section 6's evaluation committee recommended explicit benchmarks for dataset contributions, weighting factors such as reusability (documentation standards) and scientific impact (on further research works). These practices align with the CRediT taxonomy's granular role attribution, which helps disambiguate individual contributions in collaborative projects – a critical need given computer science's median authorship of 4.2 per paper. This expanded output recognition reflects the field's evolving scholarly practices while maintaining rigorous quality standards through peer-informed evaluation.

4 Implementation Challenges

4.1 The ideal-practice gap in qualitative evaluation

The CNRS evaluation system embodies a fundamental tension between the theoretical ideals of comprehensive qualitative assessment and the practical constraints of large-scale research evaluation. DORA, and the way it is perceived by CNRS governance, advocates for complete disregard of publication venues, requiring instead a full examination of research content. In practice, however, Section 6 committee members face severe limitations in implementing this ideal. With approximately 300 active researchers and hundreds of annual recruitment cases, exhaustive reading of all outputs proves operationally impossible. This forces pragmatic compromises: candidates are asked to self-select their top 3–5 publications for in-depth review, while other works receive only cursory examination. The committee also strategically employs proxy indicators – such as conference prestige or community awards—as indirect measures of quality, despite recognizing these violate pure DORA principles. These adaptations reveal the inherent scalability challenges of qualitative methods in national-scale evaluation systems.

4.2 Managing subjectivity through structural design

The subjective nature of qualitative evaluation necessitates careful institutional design to maintain legitimacy. Section 6's 18-member committee deliberately combines researchers with divergent expertise—from theoretical cryptographers to systems engineers—to balance disciplinary biases. This diversity manifests in concrete evaluation practices: when assessing a researcher in formal methods, the committee strives to make sure that at least one member specializes in verification, while others provide cross-disciplinary perspective. Decision-making procedures further mitigate subjectivity through consensus-seeking deliberations where all members must articulate their reasoning before any vote occurs.

4.3 Institutionalizing conflict resolution

CoNRS codifies conflict of interest management through explicit, pre-defined rules that address the French academic community's dense collaboration networks. The most frequent conflicts arise from co-authorship (covering the past five years) and institutional affiliations, which constitute the primary grounds for recusal. The standardized resolution protocol requires conflicted members to both abstain from discussion and (if allowed by the specific rules covering the evaluation) physically leave the deliberation room during affected evaluations. For particularly sensitive cases – such as evaluating a committee member's doctoral student for a recruitment – the committee member in question temporarily withdraws from the entire evaluation process. These mechanisms aim to maintain trust while acknowledging the unavoidable entanglement of France's small but elite computer science community. This leads to fewer formal complaints over evaluation decisions and, hopefully, to a better acceptance by the community at large of the evaluation process.

5 Case Study: Decision-Making in Section 6

5.1 The four-phase evaluation process

The CoNRS Section 6 evaluation cycle implements a rigorous multi-stage assessment protocol that balances depth with operational feasibility. During the 2025 hiring round for eight permanent junior positions, the committee processed 152 complete dossiers through sequential refinement phases. In the initial documentary review, candidates submitted structured dossier, including research statements and research plans, accompanied by self-selected publication portfolios. Each dossier was evaluated independently by two committee members, chosen to best fit their perspective. For the 32 shortlisted candidates, an additional third reviewer was assigned, and technical interviews focused on probing fundamental contributions rather than presentational polish. Deliberations involved multiple days of discussion on each candidate, to finally obtain a list of 17 candidates, including a main list of 8 selected candidates and a supplementary list. The final ranking, in the CNRS recruitment process, is elaborated by a final jury involving both the governance of CNRS and CoNRS members – this resulted, as is common in practice, in confirming the top 8 candidates to whom an offer was made to join a CNRS research unit.

5.2 Differentiated output mechanisms

The evaluation process yields tailored outcomes calibrated to decision contexts. In recruitment setting, the output is a (partially ordered) list of candidates, which is provided to the final admission jury, which usually confirms and refines this list into a total order. The list takes into account strategic scientific topics put forward by CNRS for the competition (in 2025, quantum computing, software science, ecoresponsible computer science, and cybersecurity) and maintaining an adequate diversity of the profiles of the candidates selected. By contrast, biennial research evaluations prioritize developmental feedback over binary outcomes. This may involve pointing out strong aspects of the work conducted by the researcher in question, as well as suggesting ways to further their develop their career (e.g., taking on more PhD students, accepting national duties, exploring the connection of their research with a neighboring area, changing publication venues). The committee also strives adapting the evaluation to specific cases (different practices in different areas, specificities of interdisciplinary research, or specifics of a researchers' life, such as returning from parental leave), offering adapted recommendations when possible.

6 Discussion and Concluding Perspectives

6.1 Theoretical and practical implications

The CNRS evaluation model offers three transformative insights for computer science research assessment globally. First, its demonstrated viability disproves the common objection that qualitative methods cannot scale – Section 6's successful management of 300+ active researchers evaluated every 2.5 years (averaging 120 evaluations annually, in addition to the hundreds of dossiers submitted for recruitment) establishes that rigorous peer review remains operationally feasible even for large communities. Second, the framework's adaptability to computer science's unique scholarly outputs provides a template for other fields grappling with non-traditional research products; integration of software artifacts as a formally recognized evaluation category is a case in point: demonstrated impact cases (e.g., cryptographic libraries adopted in industry) have significantly influenced evaluation outcomes and exemplify this extensibility. Third, the system's resilience to gaming contrasts sharply with metric-dependent models, as witnessed by its overwhelming acceptance by the research community.

6.2 Limitations and institutional preconditions

The model's effectiveness hinges on specific cultural and structural conditions that may limit direct transferability. The French civil service framework provides unique stability, with CNRS researchers' tenured status reducing evaluation stakes after recruitment compared to the "up-or-out" systems prevalent in Anglo-American universities. Section 6's evaluation process requires substantial effort per case and also demands exceptional service commitment from members – a conservative estimate of the effort required is around two months of active work yearly for regular members, and more for the committee president and members involved in additional evaluation committees. Furthermore, the system's reliance on committee diversity assumes sufficient disciplinary critical mass, posing challenges for smaller national systems or emerging fields lacking established expert pools. These constraints suggest hybrid models may be necessary for adaptation, perhaps combining periodic qualitative evaluations with lighter interim metrics.

6.3 Concluding recommendations

Four principles emerge for institutions considering qualitative evaluation reform: (1) Field-

Specific Calibration – the CNRS bifurcation between Sections 6 and 7 demonstrates how evaluation must align with disciplinary epistemology rather than impose universal standards; (2) Transparent Subjectivity – by openly acknowledging and structuring committee diversity (e.g., between theoretical and applied research), systems can transform bias management from weakness to strength; (3) Developmental Orientation – the documented efficacy of CNRS's detailed feedback suggests evaluation should primarily serve researcher growth rather than mere selection; (4) Infrastructure Investment – the evaluation process necessitates substantial organizational infrastructure, including appropriate software infrastructure and support staff for dossier management and deliberation coordination to sustain the intensive process. As artificial intelligence threatens to exacerbate metric gaming, these human-centered approaches offer a robust alternative for sustaining research integrity – one that measures what matters rather than merely what can be counted.

References

[1] Hicks D, Wouters P, Waltman L, et al. Bibliometrics: the Leiden manifesto for research metrics[J]. Nature, 2015, 520(7548): 429-431.

[2] Brembs B, Button K, Munafò M. Deep impact: unintended consequences of journal rank[J]. Frontiers in Human Neuroscience, 2013, 7: 291.

[3] DORA. San Francisco declaration on research assessment[EB/OL]. [2012-01-01]. https://sfdora.org/.

[4] Centre National de la Recherche Scientifique. CNRS annual institutional report[EB/OL]. [2023-01-01]. https://www.cnrs.fr/en/annual-report.

[5] PSL. Paris Sciences et Lettres university academic profiles[EB/OL]. [2023-01-01]. https://www.psl.eu/en.

[6] Howison J, Bullard J. Software in the scientific literature: problems with seeing, finding, and using software mentioned in the biology literature[J]. Journal of the Association for Information Science and Technology, 2016, 67(9): 2137-2155.

Pierre Senellart is a Professor of Computer Science at École Normale Supérieure, PSL University, Paris, France, the the director of the Graduate Program in Computer Science of PSL University, and Vice-President of PSL University in charge of digital infrastructure and IT convergence. His research interests focus on practical and theoretical aspects of data management, including Web mining, information extraction, and uncertainty and provenance management. He holds a chair at the PR[AI]RIE Paris School of AI institute and served between 2021 and 2025 as the President of Section 6 (Core Computer Science) of France's National Committee for Scientific Research, in charge of the hiring and evaluation of CNRS researchers. Additionally, he advises the scientific and ethical committee monitoring university admission platforms in France.