Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

The Why, What, How and When of Big Science evaluation. Perspectives on socio-economic assessment of research infrastructures

Silvia Vignetti

vignetti@csilmilano.com

CSIL

Manuela Cirilli

CERN, Knowledge Transfer Group

Julie de Brux

Citizing

Roger Eccleston

STFC – National Laboratories, Science and Technology Facilities Council

Alessandro Fazio

Joint Research Centre (JRC), European Commission

Research Article

Keywords: big science, research infrastructure, socio-economic evaluation, impact assessment, cost–benefit analysis, research policy

Posted Date: September 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10787602/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

The growing strategic and budgetary significance of large-scale research infrastructures (RIs) has sharpened the demand for rigorous socio-economic evaluation. Yet practice lags conspicuously behind expectation: existing evidence indicates that typically RIs do not conduct formal socio-economic evaluations of their facility and that where evaluation did occur it was typically driven by external funder requirements rather than by internal learning objectives. Drawing on a structured expert roundtable, this paper synthesises practitioner and policy thinking on four organising questions: why evaluation is conducted, what it should measure, how it should be executed, and when it should occur across the project life cycle. It does so by examining four interconnected dimensions: the rationales for evaluation; the core estimation challenges; the requirements for effective communication of results; and the governance of evaluation as a life-cycle activity. The central argument is that evaluation must move beyond episodic compliance exercises and be reconceived as an integrated analytical tool: embedded in the management of the RI from inception through decommissioning, supported by dedicated data infrastructure and governed by transparent methodological choices. A concluding section maps priority items for methodological development and for governance reform, including adequate resourcing and the development of inter-RI coordination mechanisms.

Introduction

Research infrastructures stand at a genuinely significant juncture in European policy. Three landmark documents published in the space of two years have collectively repositioned large-scale science at the heart of the European agenda. The Draghi report argued that closing Europe's innovation gap requires a structural increase in investment in research and large-scale infrastructures (Draghi 2024). The Heitor report went further, calling for a long-term strategy to secure investment in world-class research and technology infrastructures serving researchers, industry, and the public sector alike. The European Commission responded in September 2025 by adopting the first comprehensive European Strategy on Research and Technology Infrastructures, framing these facilities explicitly as the backbone of Europe's research and innovation system and as essential instruments for scientific excellence, competitiveness, and technological sovereignty (European Commission, 2025). The convergence across these documents is remarkable and the message consistent: Europe needs more investment in science, including big science.

The fiscal implications are substantial. The ESFRI roadmap, the intergovernmental framework that identifies priority large-scale facilities at the European level, implies investment needs of the order of €20–30 billion over the coming decade, with annual operating costs running to several additional billions (European Commission, 2025). Yet growing investment comes with growing conditionality. Policymakers are willing to support long-term, capital-intensive investments, but expect in return that these investments contribute demonstrably to economic growth, strengthen industrial ecosystems, support strategic autonomy, and generate visible societal benefits. The social contract underpinning public investment in large-scale science is increasingly conditional on formal socio-economic evaluation.

Against this backdrop, the evidence on evaluation practice is revealing. Evidence from the RI-PATHS project, a Horizon 2020 co-funded project that conducted in 2018 a comprehensive survey of evaluation practices among 191 research infrastructure managers, found that fewer than half had conducted a formal socio-economic evaluation of their facility, and of those that had, only half did so regularly rather than as a one-off exercise prompted by funder requirements. The gap between the policy expectations and the evaluative practice is evident.

This article is structured around four thematic sections derived from a structured expert roundtable: the rationales for conducting evaluation; the core estimation challenges; the requirements for communicating evaluation results effectively; and the governance of evaluation as a life-cycle activity. A cross-cutting section addresses the politics of evidence before the conclusion maps research and governance priorities.

1. Why we evaluate: a four-fold typology

There are at least three not mutually exclusive rationales for evaluating RI investment, each carrying different methodological implications.

The first rationale is grounded in economic theory. Growth theory establishes that sustained economic growth is not a natural or automatic process: it requires that specific conditions be met, chief among them the capacity of an economy to generate innovation (Aghion and Howitt, 2023; Aghion et al. 2025). If innovation is the engine of long-term development, then the infrastructures that support frontier scientific research are, in principle, a necessary condition for it. Yet, investments in scientific research do not automatically produce socially valuable outcomes, and the type, direction, and magnitude of impacts depend on the choices made in designing, governing, and evaluating the investment itself. Investments in research infrastructures are justified by market failure considerations: they produce scientific knowledge which is subject to public-goods logic (its social returns systematically exceed what private investors can appropriate) and generate technological spillovers and knowledge transfers that private actors have no incentive to internalise. These market failure characteristics justify public funding in principle, but do not guarantee its social value in practice. Public investment in research infrastructures is warranted only to the extent that the value of the positive effects and externalities it produces outweighs its costs and negative externalities, a condition that cannot be assumed and must be tested. Evaluation is the instrument through which that test is conducted.

A second rationale concerns social legitimacy. The need to demonstrate, to citizens as much as to funders, that the societal value of large-scale science investment is commensurate with its cost makes evaluation the primary instrument through which that justification is rendered visible and open to public scrutiny. RI cannot afford to remain in the ivory tower: as public scrutiny of large-scale investment intensifies, the ability to communicate the value of big science in terms that are meaningful to non-specialist audiences becomes a condition of continued public support, not merely a communications exercise. The ZAD (Zones à Défendre) movement in France, a series of sustained protest occupations

against large infrastructure projects, illustrates the political consequences of perceived procedural exclusion and the failure to engage citizens in the justification of major public investments. In this context, evaluation serves an accountability function directed at citizens as well as funders. The normative ideal is therefore an iterative relationship between public deliberation and socio-economic analysis, in which the assessment is developed in dialogue with civil society rather than presented as a foregone conclusion.

A third rationale is practical. Where genuine alternatives exist in project design (across variants, phasing, or scope, as is typically the case for large and complex facilities at the frontier of technological development), evaluation can guide choices toward the option of highest net social value. This decision-aid function is the most demanding in analytical terms but arguably the most consequential, since it requires evaluation to be conducted before irreversible design and procurement commitments are made. It also implies that evaluation must be sufficiently granular and credible to discriminate meaningfully between alternatives rather than merely to justify a predetermined choice. The same logic extends beyond the investment decision itself: tracking impacts throughout the operational life of a facility, at mid-term and ex-post, allows managers and funders to identify what has worked, what has not, and where course corrections are warranted.

Good institutional practice requires attending to all three rationales simultaneously. The most important point is that evaluation must not be merely a legitimisation exercise: systematically feeding findings back into project management and organisational practice across the portfolio of facilities operated by a given institution is equally essential. In this perspective, France stands as a good example. A 2012 law and a 2013 implementing decree (French Law 2012; Decree 2013) mandate socio-economic evaluation for any state investment exceeding €20 million, with mandatory counter-expertise commissioned by the Secrétariat Général Pour l'Investissement (SGPI) (an office attached to the Prime Minister with the specific mandate of reviewing the analytical quality of evaluations) for projects above the €100 million threshold (French Law 2012; Decree 2013). This approach institutionalises the assessment of investments in a way that encourages the project owner to design for socio-economic benefit generation through a common, national framework. The SGPI model is instructive beyond the French context: by institutionalising counter-expertise as a standard feature of the process, it treats economic evaluation not as a bureaucratic hurdle but as an iterative scientific activity, explicitly acknowledging the incompleteness of available tools and creating institutional incentives for methodological innovation rather than formal compliance.

2. Estimation challenges: attribution, intangibles and the evolving methodological frontier

The objects of impact estimation in RI evaluation are substantially more contested than the rationale for evaluation itself. Three interconnected issues are worth mentioning: the distinction between attribution and contribution, the challenge of valuing intangible impacts and the methodological frontier, which is itself actively developing.

A fundamental conceptual difficulty concerns the attribution of outcomes to a specific infrastructure investment. Research infrastructures operate within innovation ecosystems characterised by long time horizons, multiple interacting actors and highly nonlinear processes of knowledge diffusion. Insisting on clean causal attribution, that is establishing that a specific investment produced a particular technology or economic outcome, systematically misrepresents how scientific knowledge propagates through society. The history of satellite navigation technology is instructive in this respect. In 1957, two researchers at the Applied Physics Laboratory used a rudimentary radio receiver to pick up the signal transmitted by Sputnik. Applying the Doppler effect, they calculated the satellite's orbit from a known fixed point on Earth. The US Department of Defense recognised that the inverse problem, computing a position on Earth from a known satellite orbit, was equally tractable and funded the development of a navigation system for submarine positioning. The eventual outcome was the Global Positioning System (GPS), launched as NAVSTAR in 1978 and fully operational with a 24-satellite constellation by 1993. Patents, commercial agreements and companies subsequently emerged around GPS technologies and their applications. The relevant point, however, is that these downstream commercialisation pathways were neither visible nor determinative at the time of the original scientific observation or during the earliest stages of public investment. They arose later, as consequences of a long and cumulative process of publicly funded research, technological development and diffusion, rather than as clearly identifiable objectives at its outset. The policy implication is therefore a conceptual reorientation from attribution, which demands a traceable causal line from an initial investment to a specific final outcome, towards contribution analysis, which asks how an investment enriched the scientific and technological ecosystem from which later outcomes emerged. The challenge is that funding decisions remain the prerogative of individual institutions, each with a legitimate interest in demonstrating the distinct contribution of its own investment. This creates a tension at the heart of any serious attempt to move from institution-specific attribution towards ecosystem-level evaluation.There is then the challenge of intangibles. Several categories of impact that are manifestly real from a social welfare perspective resist monetary quantification with existing techniques. Science diplomacy provides a prominent example: CERN's role during the Cold War as a space of scientific exchange between researchers from NATO and Warsaw Pact countries contributed to channels of political communication that had value independent of any technological output, yet no established evaluation methodology can assign a credible monetary equivalent to it. Cultural transmission, the influence of large-scale science on public scientific literacy, on career choices across successive cohorts of students, and on the prestige of European science globally, is similarly resistant to monetisation. While major facilities are beginning systematically recording cultural impacts embedded in school visits, teacher-training programmes and public engagement activities, the diffuse, transgenerational effects of big science on collective scientific culture remain beyond the reach of current evaluative instruments.

Some of those challenges reflect the broader state of the discipline. Socio-economic evaluation of research infrastructures is itself a relatively recent and actively developing field. Some methodological building blocks are now well established: cost-benefit analysis frameworks, input-output modelling, and standard approaches to quantifying human capital formation and technological spillovers are routinely

applied and reasonably well understood. Others remain genuinely open research questions. The valuation of scientific publications is a prominent example: current practice relies on production-cost proxies that measure what it costs to produce a paper rather, a pragmatic starting point which however may not fully grasp the actual social value of scientific knowledge. The application of option value concepts may offer a theoretically grounded way to account for the high-impact, low-probability nature of frontier scientific discoveries. Advanced data analytics and econometric identification strategies, linking research outputs to downstream outcomes such as productivity growth and innovation adoption, open further avenues for moving from proxy measurement toward genuine causal estimation of social value. In this sense, large and complex research infrastructure projects offer a distinctive opportunity: not only to advance the frontier of physics, or biology, or astronomy, but to advance the frontier of economics as well.

3. Communicating evaluation results

Methodological choices in RI evaluation involve an irreducible tension among three desiderata: analytical rigour, cross-facility comparability and communicability to non-specialist audiences including ministers, parliamentarians, and the public. Movement toward any one vertex of this triangle typically requires concessions at the others.

Accountability requires accessibility: an evaluation that is methodologically rigorous but intelligible only to specialists fails this function. Apparent methodological sophistication can function as a form of opacity: the more elaborate the measurement model, the less visible the normative assumptions embedded within it become. A highly complex evaluation may inspire unwarranted confidence in its conclusions while obscuring the value judgements (about counterfactual scenario, the scope and definition of relevant impacts and so on) on which its results critically depend. Conversely, reducing the evaluation to a single number achieves communicability at the cost of suppressing information that is essential to good decision-making. Neither extreme is satisfactory; the art lies in calibrating complexity to the decision context and audience.

There is persistent institutional pressure, from funders, governing bodies, and communications departments alike, to distil an evaluation into a headline economic multiplier: a single ratio of aggregate expenditures generated by the investment. This practice is methodologically problematic on multiple grounds. It implies a precision that the underlying estimates cannot support; it creates incentives for reverse-engineering the analysis to achieve a credible-sounding number; it forecloses the probabilistic reasoning that large-scale investments under genuine uncertainty actually require; and it invites unfounded comparisons across facilities that differ in type, disciplinary domain, maturity, and context. A more defensible approach, increasingly adopted in frontier evaluation practice, is to replace point-estimate multipliers with probability distributions over outcomes, as recently done for the feasibility study for the Future Circular Collider (FCC) (Benedikt et al., 2025).

The evaluator's transparency obligation follows as a corollary: assumptions must be stated explicitly and publicly; sensitivity analyses must be conducted over the parameters to which conclusions are most sensitive; and uncertainty must be communicated as substantive information about the investment decision rather than suppressed in the interests of rhetorical simplicity. The SGPI's explicit acknowledgement that current methods are incomplete, and its treatment of methodological innovation as a positive feature of counter-expertise rather than a liability, models a form of institutional epistemic humility that has broader applicability.

4 Evaluation as a life-cycle activity

A central observation is that ex-ante, mid-term, and ex-post evaluations are typically conducted as independent, disconnected exercises: the analytical assumptions of the ex-ante appraisal are rarely revisited at mid-term; the findings of ex-post evaluations rarely inform the framing of successor facility assessments. The result is an evaluative activity that fails to accumulate institutional knowledge over time.

Best practice requires evaluation to be conceptualised and resourced from project inception, not merely to inform the initial investment decision but to establish the baseline data, the counterfactual framework, and the analytical model that mid-term and ex-post evaluations will subsequently require. There is consensus that evaluation should be understood as an evolving analytical model that is maintained as a living companion to the project, updated iteratively as design assumptions are refined, empirical evidence accumulates, and the policy context shifts. A facility that begins thinking about evaluation only after it has commenced operations faces an irreversible data problem: baseline conditions are unobserved, counterfactual comparison groups are difficult to construct, and the causal chain linking investment to outcome has already partially played out without being recorded.

A structural obstacle to life-cycle evaluation identified in the roundtable is that the data systems of research infrastructures are designed for operational and administrative purposes rather than for socio-economic assessment. Procurement classification schemes do not map onto the national accounts categories used by economists; employment and personnel records are maintained according to HR rather than research-system taxonomies; data retention periods are calibrated to audit requirements rather than to the decades-long time horizons over which RI impacts materialise; and GDPR constraints complicate longitudinal tracking of individuals whose careers were shaped by facility employment or training. A productive analogy is the data curation model established by the ATLAS and CMS experiments at CERN, two of the four major detectors on the Large Hadron Collider, which maintain structured, richly documented, publicly accessible datasets specifically engineered for downstream scientific use by the broader research community. An equivalent curatorial commitment to evaluation-relevant data (for procurement records, personnel histories, knowledge transfer agreements, publication outputs) would substantially reduce the cost and increase the reliability of socio-economic assessment.

A promising institutional mechanism was discussed in the EIROforum Technology Transfer Group which has been exploring the introduction of contractual clauses in industry partnership and licensing agreements requiring partners to share downstream data on publication outputs, further innovation, and commercial outcomes with the originating RI. Such clauses would begin to address one of the most persistent empirical gaps in impact assessment: the near-total invisibility of what happens to scientific and technological knowledge after it leaves the facility.

Evaluation practice may therefore be understood as an evolving analytical companion to the project rather than a one-off bureaucratic exercise. This reconceptualisation of the evaluation function as has significant practical implications. Monitoring and evaluation activities are chronically under-resourced, particularly in facilities established before explicit M&E requirements became standard funder conditions. The appropriate normative principle is that evaluation resources should be embedded in project budgets from the point of initial capitalisation, treated as a recurrent capital cost analogous to calibration, maintenance, and scientific operations rather than as a discretionary overhead subject to periodic cuts.

5 The politics of evidence

An overarching theme running through the roundtable discussion is that evaluation is irreducibly a political activity as well as an analytical one. What counts as a good evaluation, which impact categories to include and which to exclude, how to define the perimeter of the project under assessment is not determined by analytical technique alone but involves value judgements that are legitimately contestable.

This political dimension carries the risk that methodological choices are made, consciously or unconsciously, to produce results that serve institutional interests rather than to honestly represent the social value of the investment. The pressure on facility managers to report favourable multipliers, noted in section 2, is one manifestation of this risk. Another is the systematic bias toward proximate, quantifiable impacts that may undervalue the long-horizon, low-probability discoveries (of which GPS remains the most celebrated example) that represent the highest expected social value of fundamental science investment.

The French SGPI model represents one possible governance response: by mandating counter-expertise conducted by an analytically independent office attached to the head of government rather than to the sponsoring ministry, it creates an institutional structure in which the political incentives of sponsoring agencies do not go unchecked. This institutional architecture, combining mandatory evaluation with independent scrutiny and explicit acknowledgement of methodological incompleteness, offers a template that merits examination at the European level.

Conclusion and research agenda

Three conclusions stand out from the discussion. First, methodological humility is not a weakness but a professional obligation. The tools available for socio-economic evaluation of research infrastructures are improving but remain imperfect: they systematically struggle with long-horizon impacts, intangible benefits, and the nonlinear pathways through which scientific knowledge generates social value. Acknowledging these limits openly, rather than concealing them behind easy-to-communicate numbers, is both more honest and, ultimately, more credible.

Second, the governance architecture surrounding evaluation matters as much as the methodology. The institutional incentives should reward transparency, support iterative refinement of analytical assumptions, and position evaluation as a tool for improving project design and sustaining an open dialogue with funders and the public, not as a compliance exercise to be cleared once and forgotten.

Several significant questions remain open. How can contribution analysis be operationalised rigorously without ending into unfalsifiable narrative? What governance architectures can sustain longitudinal data infrastructure across the multi-decade life cycles of major facilities? Can stochastic methods for representing uncertain, transformative discoveries be made accessible to non-specialist decision-makers without sacrificing analytical integrity? And how can the results of socio-economic assessments be translated into simplified but accurate narratives that effectively communicate the value of research infrastructures to the wider public? These are not merely academic questions: the quality of public investment decisions in large-scale science depends on how well the evaluation community answers them.

Declarations

Competing interests

The authors have no relevant financial or non-financial interests to disclose.

Author Contribution

S. Vignetti: conceptualisation, project administration, writing — original draft, writing — review and editing. M. Cirilli: conceptualisation, writing — review and editing. J. de Brux: conceptualisation, writing — review and editing. R. Eccleston: conceptualisation, writing — review and editing. A. Fazio: conceptualisation, writing — review and editing. All authors read and approved the final manuscript.

Data availability

This article is based on a recorded expert roundtable. No primary quantitative datasets were generated or analysed.

References

1. Aghion, P. and Howitt, P. (2023), The Creative Destruction Approach to Growth Economics, European Review, 31(4), pp. 312–325. https://doi.org/10.1017/S1062798723000212.
2. Aghion, P., Bergeaud, A., Boppart, T., Klenow, P.J. and Li, H. (2025), *Good Rents versus Bad Rents: R&D Misallocation and Growth*, NBER Working Paper No. 34190, National Bureau of Economic Research, Cambridge, MA.
3. Benedikt, M., Zimmermann, F., Auchmann, B. et al. (2025) Future Circular Collider Feasibility Study Report. Eur. Phys. J. Spec. Top. 234, 5113–5383. https://doi.org/10.1140/epjs/s11734-025-01958-5
4. Décret n° 2013 – 1211 du 23 décembre 2013 relatif à la procédure d'évaluation des investissements publics en application de l'article 17 de la loi n° 2012 – 1558 du 31 décembre 2012 de programmation des finances publiques pour les années 2012 à 2017
5. Draghi M (2024) *The future of European competitiveness*. European Commission, Brussels.
6. European Commission: Directorate-General for Research and Innovation, (2024) *Align, act, accelerate – Research, technology and innovation to boost European competitiveness*, Publications Office of the European Union, https://data.europa.eu/doi/10.2777/9106236
7. European Commission (2025) *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: A European Strategy on Research and Technology Infrastructures.* COM(2025)497 final/2. Brussels, European Commission.
8. French Law 2012 – 1558 (2012) Loi no. 2012 – 1558 du 31 décembre 2012 de programmation et de gouvernance des finances publiques. République française.

Footnotes

1. https://ri-paths-tool.eu/en