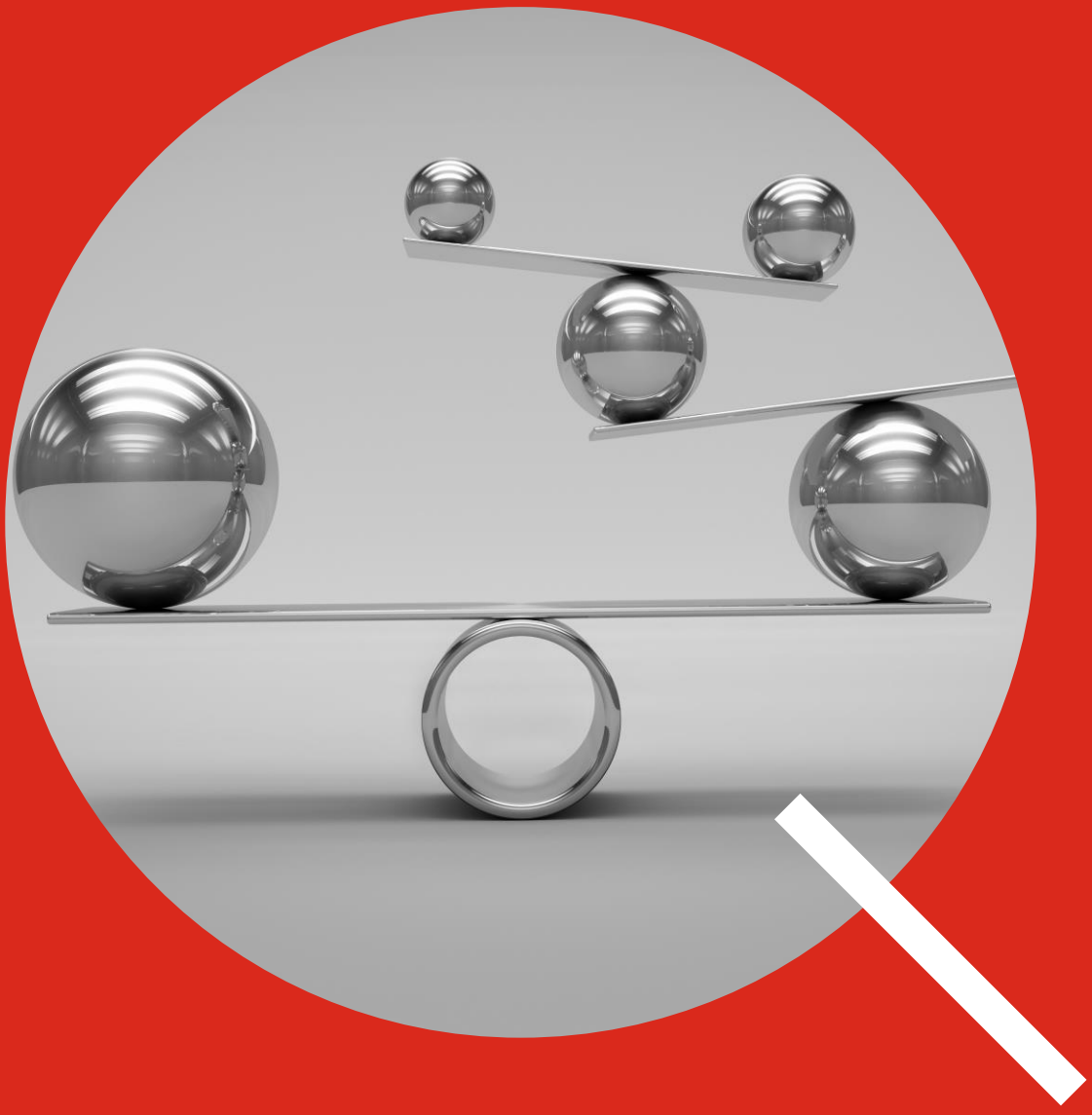
OFFICIAL-SENSITIVE

April 2026

Research into appraisal methods for R&D spending

International Case Studies

Final Report



SQW IFE

Contents

Executive Summary i
1. Introduction 1
2. UK R&D Context 3
3. Methodology and limitations 6
4. Appraisal approach in EU 9
5. Appraisal approach in Japan 17
6. Appraisal approach in Ireland 22
7. Appraisal approach in the US 27
8. Appraisal approach in Australia 35
9. Appraisal approach in Belgium 40
10. Appraisal approach in Sweden 44
11. Reflections 53

Annex A: References A-1
Annex B: Abbreviations and acronyms B-1
Annex C: Summary of insights from country cases C-1

Contact:
Tim Vorley
Tel: 01865 485839
email: tvorley@brookes.ac.uk

Approved by:
Rebecca Pates
Director
Date: 29/04/2026

Disclaimer
This report takes into account the particular instructions and requirements of our client. It is not intended for, and should not be relied upon by, any third party and no responsibility is undertaken to any third party.
Whilst SQW has used reasonable care and skill throughout, it is unable to warrant either the accuracy or completeness of information supplied by the client or third parties, and it does not accept responsibility for any legal, commercial or other consequences that arise from its use.

Executive Summary

1. In December 2025, SQW in partnership with the Innovation Research Caucus, Geoff White, Jonathan Blanchard Smith at SAMI Consulting and Jonathan Cook were commissioned by the Department for Science, Innovation and Technology (DSIT) to undertake research into appraisal methods for Research and Development (R&D) policy. This assignment included a desk review of evidence from seven international case studies to highlight the different approaches towards ex ante appraisal of R&D investments, led by the Innovation Research Caucus. The primary aim is to explore how other countries appraise R&D opportunities and identify insights that may support the UK government in developing business cases as part of R&D investment strategy.

2. The UK government has clear guidance on appraisal though the Green Book, differentiating between appraisal and evaluation. However, it is important to acknowledge that the nomenclature differs by country, and the term 'appraisal' is not used consistently across countries, while in others 'appraisal' and 'evaluation' are treated as synonymous.

3. The published literature on appraisal is limited compared to the more extant literature on evaluation. Following an initial screening of these literatures a long list of potential cases was developed from which seven cases were developed where evidence relating to appraisal was identified. This was supplemented with consultations with academics around the world who provided insights about how practice and terminology might differ in their respective contexts.

4. The paper describes how different countries approach appraising R&D to assess the expected impact of R&D investments, with a particular emphasis on public sector decision-making. Each international case study follows a standardised, five-part framework (agreed with DSIT) designed to ensure consistency and facilitate direct comparison across different countries. The international cases have been developed with the specific view to identifying insights relevant to the UK system, and any challenges noted may not be seen as challenges within their own systems.

5. Each case begins with a high-level overview of the national R&D landscape, providing a summary of the national context. This is followed by an overview of appraisal approaches, which examines the specific methodologies used by the country to appraise R&D investments. Within each case the strengths and limitations of these appraisal methods are considered, identifying good practice and challenges before reflecting on the data and capability requirements. Each case concludes with an overall assessment, synthesising the findings to highlight key characteristics and insights that could potentially be relevant in the UK context.

6. Table 1 (see annex C) summarises the findings the insights from across the cases. In terms of the use of logic models, while they are applied in the EU case there is only a partial use in the other cases. **There is only R&D specific guidance on appraisal in Japan,** and the prevailing delivery model is decentralised to the relevant funding and decision making bodies.

7. This paper demonstrates through the international case studies that the landscape of R&D appraisal is undergoing a fundamental shift from individual projects towards portfolios. A central finding of this research is that while traditional cost-benefit analyses remain a cornerstone of policy, they are increasingly insufficient for capturing the 'additionality' and transformative impacts of modern, mission-oriented investments. Consequently, **this study concludes that the UK is already a front runner in appraising R&D investment and is looking to develop more comprehensive appraisal tools to inform the development of business cases for R&D investment.**

1. Introduction

1.1 In December 2025, SQW in partnership with the Innovation Research Caucus, Geoff White, Jonathan Blanchard Smith at SAMI Consulting and Jonathan Cook were commissioned by the Department for Science, Innovation and Technology (DSIT) to undertake research into appraisal methods for Research and Development (R&D) policy. This assignment included a literature review of international approaches to appraisal, led by the Innovation Research Caucus.[1]

1.2 This report presents the findings from the international literature review. With a focus on R&D, which is defined as the systematic activities undertaken by companies or governments to innovate and create new products and services or improve existing ones, this report adds to our understanding of approaches to R&D appraisal outside of the UK.

1.3 Compared to ex post evaluation, the published literature on appraisal is limited with governments taking different approaches towards issuing formal appraisal guidance. The focus is national-level approaches across government. Where there is appraisal guidance specific to R&D this has been identified, although it was not generally found to be the case that methods were tailored. It is beyond the scope of this report to comprehensively consider guidance at a project or programme level, although where specific examples relating to projects or programmes were cited these are noted in the relevant international cases.

1.4 This report provides an international perspective with reflections on R&D appraisal practices in different countries, but **it is important to acknowledge that the nomenclature, understandably, differs by country**. The UK government has developed clear guidance on appraisal though the Green Book that differentiates between appraisal and evaluation (see Box 1-1).

Box 1-1: Simplified definitions of appraisal and evaluation

> Appraisal is concerned with assessing a proposal as it is developed, before it is approved and implemented. Evaluation involves assessing an intervention after it has been approved and implemented.

Source: Green Book

1.5 We offer the following notes on interpreting the case studies below.

- First, these definitions have been appropriated and applied consistently in the case studies regardless of whether the specific term "appraisal" is used in national guidance. Appraisal is understood in this report as formative stage in assessing whether an intervention should be progressed to implementation. Consequently, where mechanisms of ex ante assessment are

[1] The wider study involved a review of alternative appraisal tools, techniques and approaches identified by DSIT (an 'enhanced' theory of change, size of the prize, tipping point (break-even) analysis, risk quantification, futures thinking, KPIs and portfolio approaches) and a review of international appraisal frameworks and approaches.

described, we use the term appraisal and where they concern ex post assessment, we use the term evaluation.

- **Secondly, the concepts, and processes, of appraisal and evaluation are inherently linked.** As such, it is often important to discuss them in tandem as two sides to the same coin, so to speak. The Green Book primarily focuses on the first three stages of the ROAMEF cycle (rationale, objectives and appraisal) as key stages to identify a preferred policy options based on the estimated value of a given intervention. In most cases, guidance and practice links evaluation to appraisal as part of this cycle because the assumptions used in an appraisal informed decision-making process are based on lessons, feedback, and evidence generated from evaluating the outcomes of previous policies (and how well those aligned with their underlying expectations). A large part of the effectiveness of appraisal processes is how adept they are at accessing and incorporating evolving data and evidence to inform (or change) appraisal models. The UK's guidance on appraisal and evaluation avoids conceptual confusions, and through the ROAMEF cycle similarly seeks to ensure the closing (or continuation) of the loop with post-implementation evaluation intended to be fed into future programme design. As a result, it should be no surprise to see the term evaluation in the cases described below. When it does, we endeavour to make the connection to appraisal clear.

- **A third consideration is that these are all systems in the process of evolving.** Many are currently considering changes to appraisal processes or are in the middle of regular legislative refresh cycles. None of these systems are perfect, but all contain insights that have the potential to enhance the UK's own efforts to innovate its appraisal system.

- Finally, it is useful to approach the cases as nested systems, where multiple interlocking pieces of legislation, practice, and process interact to shape the environment within which R&D appraisal ultimately takes place. This often consists of a national level process that requires all new legislation, programming, or spending to complete checks designed to, for example, ensure cost effectiveness, proportionality, alignment with national priorities, and demonstrate awareness of and mitigation strategies for potential negative externalities. Depending on their size, R&D spending programmes are often subject to these processes. Regardless, these guidelines and policy practices shape the appraisal processes of science and innovation funding agencies and entities responsible to designing and delivering R&D funding. These entities often have their own appraisal processes, although the focus and level of detail available about these varies. Even in these cases, the same processes will apply across the science, research, and/or innovation portfolio, such that it is impossible to pinpoint processes that apply to R&D funding, as defined above, *exclusively*. Throughout the cases, we refer to and discuss the significance of this interlocking set of mechanisms on the logic that we are interested in all of the different mechanisms that are used to shape appraisal and appraisal environments.

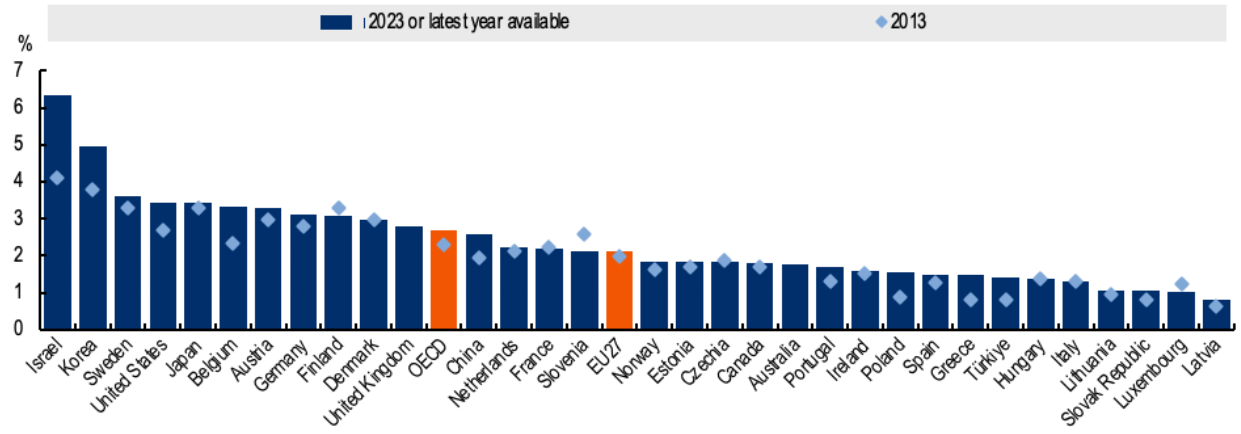
2. UK R&D Context

2.1 Before presenting the international case studies, this section frames the discussion by providing a portrait of the global R&D landscape. Understanding the effectiveness of policy appraisal requires a clear view of both the scale of investment and the tools used to benchmark success.

2.2 By setting out R&D expenditure expressed as a share of total GDP across developed nations and major economic blocs, provide necessary context to understand the UK's current position. A temporal perspective then shows the changes in UK R&D Expenditure as a percentage of GDP (1996 - 2022), identifying the key policy interventions - such as the introduction of R&D tax credits, the establishment of Innovate UK, and the Industrial Strategy.

2.3 It is important to understand where the UK sits among international peers in relation to its R&D expenditure expressed as a share of total GDP. Figure 2-1 reports this information for the UK and a set of developed countries as well as formalised groups of countries including the EU27 and the OECD.

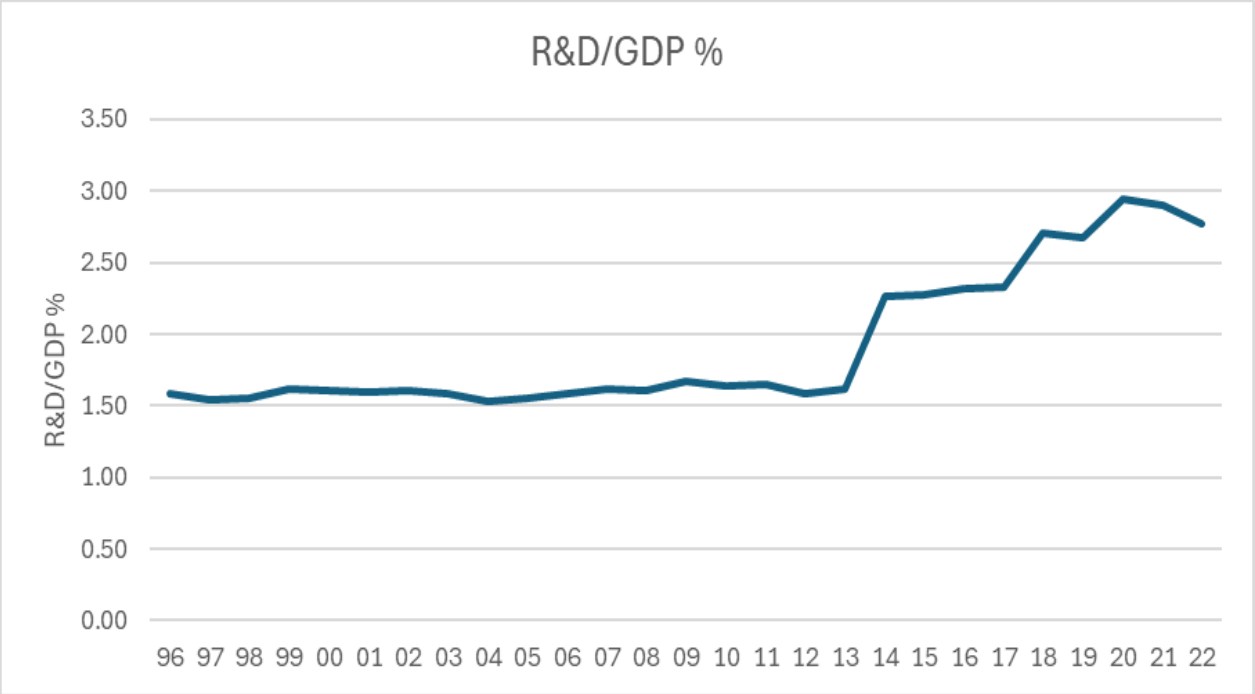
Figure 2-1: R&D as % GDP by Country



Source : https://stat.link/2orawq (Accessed April 22, 2026)

2.4 The UK share of GDP accounted for by R&D in 2023 was 2.80%. Of the countries considered in this paper, Belgium, Japan, and Sweden all have higher R&D / GDP shares with figures of 3.32%, 3.44% and 3.6% respectively. However, the UK demonstrated higher levels of investment in R&D than Australia (1.77%) and Ireland (1.58%). In a broader global sense, the UK is above the EU27 average of 2.13%, and the OECD average of 2.7%, but below the 3.44% reported for the US.

Figure 2-2: Time-Series UK R&D / GDP %, 1996 - 2021



Source: https://data.worldbank.org/indicator/GB.XPD.RSDV.GD.ZS

2.5 The UK also changed its definition of what counts as R&D in 2013 and how it is measured in national statistics. While the fundamental scientific definition of R&D - seeking an advance in science or technology - remained consistent, the 2013 changes fundamentally altered how R&D was claimed and incentivised, particularly for large companies. Reforms to national accounts (ESA 2010) saw R&D expenditure reclassified as an investment in assets, rather than just an expense, to better reflect its contribution to the economy.[23]

2.6 It is also appropriate to consider the evolution of UK R&D as prior to 2014 the UK had a low and constant R&D share of GDP that was between 1.5% and 1.6%. Due to definitional changes in the way UK R&D was defined and measured in 2013, from 2014 onwards, the UK share increased substantially and at present the new definition only covers the period 2018 onwards. The peak year between 1996 and 2022 was in fact the Covid-19 pandemic year when the UK R&D / GDP share was 2.9%.

2.7 In terms of the chronology of what occurred during this long time-series that might impact R&D spending, one of the earliest policy interventions was the SME R&D tax credit scheme which was introduced in April 2000. This was followed by the April 2002 large company R&D tax credit scheme introduction. February 2007 saw the formalisation of the Technology Strategy Board (TSB) and in August 2014 the TSB emerged as Innovate UK.

2.8 The launch of the 2017 Industrial Strategy set an ambitious target of reaching 2.4% of GDP by 2027, and while the UK has seen some improvement although this has not been reflected in growth and productivity. The UK's current Modern Industrial Strategy makes clear commitments to increasing

Office for National Statistics: Statistical Bulletin: Gross domestic expenditure on research and development, UK: 2013

[3] Office for National Statistics: Statistical Bulletin: Gross domestic expenditure on research and development, UK: 2013.

R&D investment as a central driver of economic growth. Positions R&D as one of the most important levers for productivity, competitiveness, and sectoral transformation.

2.9 The Green Book provides a robust foundation for appraisal in the UK. However the unique "high-risk, high-uncertainty" nature of R&D often sits uncomfortably within traditional cost-benefit analysis frameworks. This is consistent with Lord Willetts' Review which calls for supplementary guidance for 'science and technology' cases. This report considers how R&D appraisal is approached in the international case studies to inform this supplementary guidance.

3. Methodology and limitations

3.1 The review had three basic strands which were complementary to each other, including:

- An academic search engine review that encompassed Google Scholar and ResearchGate
- A general internet search
- A network call to established academics and policymaker colleagues.

Academic Literature Strand

3.2 The two most relevant and comprehensive academic search engines are Google Scholar and ResearchGate. However, in each case they are not limited to academic research papers but encompass a large amount of public policy reports and reports by think tanks and quasi-governmental agencies. Importantly, both also contain working papers which allows access to new and up-to-date research that is not formally published in academic journals.

3.3 We broadened our search to try and capture relevant research and documents that were not in the academic domain, for example, European Investment Bank (EIB), European Investment Fund (EIF), and OECD documentation and guidance relating to R&D policy and policy appraisal, and the Japanese Research Institute of Economy, Trade and Industry (RIETI), a policy think tank established in 2001. Its mission is to conduct theoretical and empirical research, maximise synergies with those engaged in policymaking, and to make policy proposals based on evidence derived from such research activities. The institute is highly regarded both in Japan and abroad for its activity.

On-Line Search Protocols

3.4 The first two strands used simple key word searches using the following search terms:

- R&D
- Public Policy
- Grants
- Tax Credits
- (ex ante) Public Policy Appraisal
- (ex post) Evaluation
- Innovation
- Vouchers
- Loans
- Loan Guarantees

3.5 In the first wave of search, the term R&D or Innovation preceded the use of the other individual terms, for example, R&D grants etc. In the second wave of search, we narrowed the focus to Public Policy Appraisal with and without R&D or Innovation to establish whether countries operated a general guidance framework for the appraisal of public policy schemes and interventions, or whether there were specific frameworks for R&D and innovation policy appraisal.

3.6 In total our search strands generated in the region of 100 documents relating to international aspects of R&D, public policy, the appraisal of public policy and the ex post evaluation of public policy across

a rich and diverse set of countries and supra national bodies such as the EU, European Investment Bank, OECD, The World Bank.

Network Call

3.7 These searches were supplemented with consultations with academics around the world who provided us with insights about how practice and terminology might differ in their contexts. The following individuals were contacted to both access and sense check information in the case study countries: in Japan (Professor Kenji Kutsuna, University of Tokyo), US (Professor Rebel Cole, Florida Atlantic University), Research Unit for Economics of Innovation and Industrial Dynamics, Sweden (Dr Sven-Olov Daunfeldt, Chief Economist at The Confederation of Swedish Enterprise), Belgium (Professor Marcus Dejardin, University of Namur), Ireland (Martina Lawless, Director (Chief Executive Officer) ESRI (Economic and Social Research Institute), and Australia (Professor Peter Carey, Deakin University).

3.8 These individuals were selected due to their close knowledge and experience of public policy and government in their own country (largely relating to R&D, innovation, financing, and public policy in these areas) and the alignment of their own research work or that of colleagues in their department, agency or institution. The response to the call for evidence was strong and international colleagues identified several relevant documents relating to public policy and how countries approached the appraisal process.

Methodological Limitations

3.9 The most significant limitation of this research relates to the fact that appraisal methodologies, to the extent that they are used, are not always public or published. Furthermore, where some appraisal processes (or equivalents) were available, there were still likely to be some government departments or agencies for which they were not. As a result, we cannot be sure that we have a comprehensive view of all of the ex ante R&D appraisal activities in each context. Also notable is that there is comparatively more published on ex post evaluations although these were clearly out of scope for this project. However, these can be quite useful as we have noted the dialogue between appraisal and ex post evaluation activities. Where relevant, we have explored ex post evaluation materials where they provide clues as to appraisal methodologies.

3.10 Note that variations between materials in each country means that the "Strengths and Limitations" sections for each country are not always directly comparable, as some are based on empirical evidence while others rely more heavily on qualitative stakeholder consultations.

3.11 Each of the cases is divided up into five sections designed to provide high level and more in-depth discussion of different aspects of the case.

- The **Overview** section describes the national context, providing insight into comparative R&D spending patterns, innovation delivery entities, and higher-level legislative appraisal mechanisms.

- The **Description of appraisal approaches** section delves into appraisal processes and mechanisms. Note that, per our discussion above, these reflections will often also touch on evaluation processes to the extent that these will shape the data available to inform appraisal decisions. This section will also comment, where relevant, about framework legislation governing spending and programme decisions that are applicable across government portfolios, including R&D spending.

- In the **Strengths and limitations of R&D appraisal** section, we more specifically draw out the advantages and disadvantages of the appraisal structures and processes that we were able to describe in the previous section. This section is typically more reflective. Where relevant, we draw out lessons for the UK but, for the most part, we identify more descriptively the elements that we consider strengths or weaknesses that can be used to inform the UK position, i.e. they may not be considered similarly in the context of their own systems.

- The **Data capability and requirements** section we explore data related questions, specifically focusing on the types of data collection and capabilities that are necessary to feed into appraisals and interpret downstream evaluation results.

- Finally, the **Overall assessment** section compiles our observations of the cases with an emphasis on drawing out lessons for the UK context.

4. Appraisal approach in EU

Overview of the EU R&D and Innovation System

4.1 The European Union operates through a multi-level governance system combining EU-level programmes with national and regional innovation policy, and is one of the largest and most complex public research and innovation systems globally, where its R&D policy aims to strengthen scientific excellence, increase economic competitiveness and address major societal challenges such as climate change, health and digital transformation. The EU's R&D and innovation appraisal system is built to manage a structural challenge: public R&D benefits are long-horizon, uncertain, and strongly shaped by spillovers and ecosystem heterogeneity, yet budget and policy decisions demand timely, comparable evidence.

4.2 For the EU's current R&D and innovation profile, the EU's 2025 brief (Benoit, et al., 2025) frames R&D intensity as the key headline indicator for EU performance and it reports long-run improvement (1.81% of GDP in 2000 to 2.22% in 2023), but still below comparator intensities (e.g., South Korea, the US, Japan and China). The brief also discusses the scale of EU R&D spending implied by the intensity figure (e.g., a 2023 R&D spend of approximately EUR 381bn and the associated investment gap to reach 3%), which is central to the EU's own appraisal narrative: "What would an additional euro of R&D investment buy?" and "How should it be raised and allocated?"

4.3 The 2025 European Innovation Scoreboard (EIS) is a comparative tool assessing the current R&D performance across countries but provides some important insights into the context within which RDI programme appraisal is interpreted. Methodologically, the EIS is built on multi-source measurement (venture capital, IP, publications/patents, Eurostat and national statistical offices), which implies that **EU appraisal depends on coordination across heterogeneous datasets and producers.**

4.4 EU R&D and innovation policy is primarily delivered through multiannual Framework programmes, the latest of which is Horizon Europe (2021-2027) with a budget of approximately EUR95.5bn (Benoit, et al. 2025). These programmes fund collaborative research projects, innovation partnerships and technology development across EU member states and associated countries, which aim to generate new knowledge, strengthen research capacity and facilitate the diffusion of innovation across the European economy. Horizon Europe, the current Framework Programme for 2021-2027, seeks to address several strategic objectives including strengthening science excellence, promoting industrial innovation and tackling global challenges such as climate change, digital transformation and public health (European Commission, 2023).

4.5 While European programmes are developed in compliance with the European Commission's Better Regulation Guidelines, it is the European Investment Bank (EIB) that sets and administers appraisals for R&D funding (among other areas of public investment). R&D funding falls broadly under its remit to support Private Sector Research, Development, Innovation and Digitalisation (RDI).

4.6 In recent years, most projects financed have been in the automotive, pharmaceutical, med-tech, clean-tech, industrial engineering, ICT and energy-intensive industries. This financing covers various eligible costs, such as salaries of researchers and technical/engineering staff, RDI consumables and materials, RDI equipment, outsourced RDI, costs for prototypes, and investments in RDI facilities. Financing is limited to activities up to the pre-commercial stage, However, it also supports pilot demonstration projects and first full-scale commercial production lines for breakthrough technologies, other Key Enabling Technologies and digital transformation projects. There is also funding for small and medium-sized enterprises and mid-caps, where the EIB makes investments to scale up innovative technologies and introduce them to market. The EIB generally focuses on identifying and prioritising projects leading to societal benefits where investment levels are adversely affected by material market failures, whether in the form of knowledge or environmental externalities, imperfect competition or incomplete markets.

Description of appraisal approaches

4.7 In general, the Better Regulation Guidelines (European Commission, 2021) define EU policy appraisal as an integrated set of instruments: regular evaluation/fitness checks assessing whether legislation and spending programmes are efficient, effective, relevant, coherent and add EU values; ex ante impact assessments analysing problems, objectives, options and trade-offs; stakeholder consultation as evidence input; and implementation/compliance support. The better regulation guidelines explicitly cover EU spending programmes both ex ante and ex post. It applies the "evaluation first" principle to spending, requiring the commission to check if existing funding programmes have delivered their intended results before proposing revisions or new ones. Any new spending proposal or significant revision must include an impact assessment to quantify expected costs and savings for businesses, citizens and public authorities. This is why ex post evaluation checks are interconnected with the ex ante impact assessments for any new policy/spending proposals. While the EU relies on complimentary approaches, listed above, for innovation policy appraisal, the EIB is responsible for the delivery of funding for R&D activity under Horizon Europe and other related schemes and is, as such, the architect and implementor of appraisal criteria.

4.8 The EIB guide (EIB 2023) frames economic appraisal as outlining the benefits and costs for society (beyond financial profitability) to determine viability. Its main approach centres on Cost-Benefit Analysis (CBA). The EIB guide (Ibid, 2) stresses that this kind of analysis can be incredibly useful in clarifying the aim of the project; estimating what will happen if the project goes ahead, and what will happen if it does not; evaluating whether the proposed project is the best option available; identifying whether the components of the project are the most efficient; estimating who will benefit or be disadvantaged from the project; quantifying the overall impact on government's fiscal position; determining whether the project is financially sustainable; assessing project risks; and, ultimately, giving decision-makers an informed view on whether the project is beneficial for society.

4.9 CBA measures the difference between the flow of costs and benefits with the project and those without (the "with-project" (WP) and "without-project" (WOP) scenarios, respectively), although the EIB recognises that there may also be several alternative "WP" approaches (such as a "do minimum" approach).

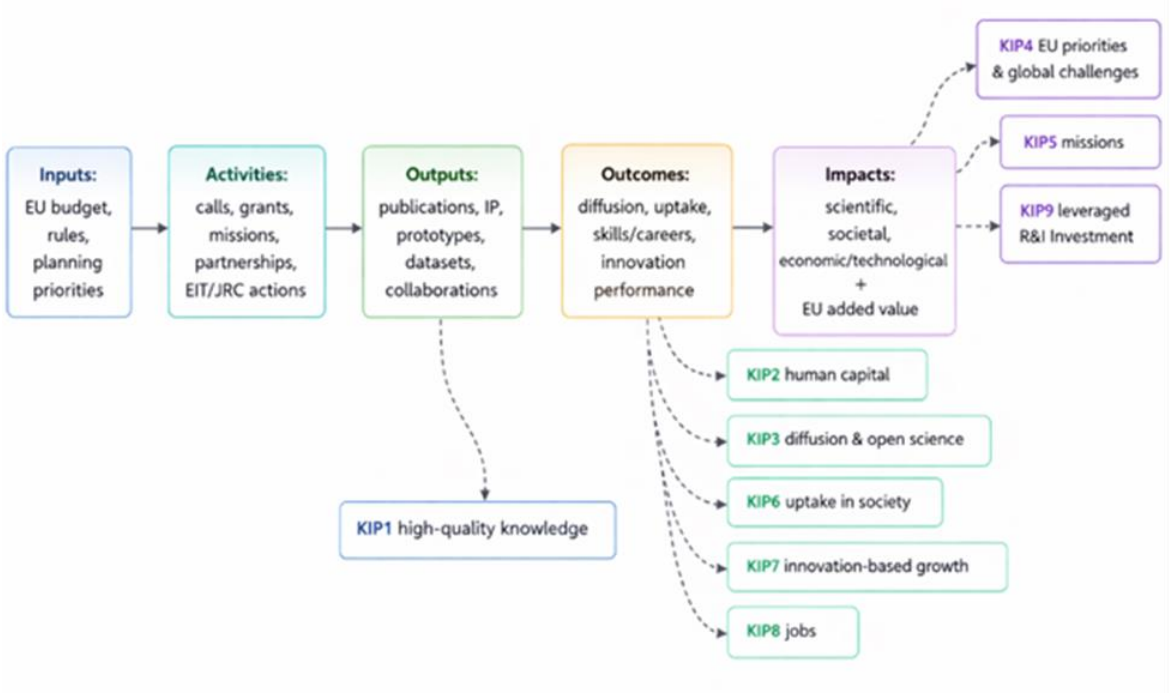
4.10 The cost and benefits of RDI projects are assessed by the EIB in relation to two different perspectives: the promoter and European society. **The EIB usually focuses on financial profitability for the promoter and economic profitability for society.** The two can be broken down as follows:

- Financial profitability evaluates the soundness of the project in advance and the extent to which it rationally allocates the promoter's resources. Financial profitability indicates the project's capability to generate future cash flows, which permits repayment of the investment undertaken by the promoter and compensation for the cost of capital invested. This assessment of financial viability gives an initial indication of positive economic impact. A project whose resources are properly allocated and expected to yield a positive return will likely support the promoter's long-term competitiveness and sustained profitability, thereby advancing wider economic growth, welfare and employment. Some projects may not be financially viable for the promoter in the medium term but are still judged to generate positive developments in the long term (e.g., "option" value), and have positive economic profitability because of their expected socioeconomic benefits (e.g., environmental protection and enhancement, climate adaptation and mitigation, and knowledge externalities).

- Consequently, an RDI project's costs and benefits for society are explicitly assessed by considering the project's externalities (positive or negative). These represent positive or negative effects on third parties (benefits or costs, respectively), which do not have monetary compensation. They are not reflected in financial accounts and not included in the project's financial profitability calculation.

4.11 It is important to note that this appraisal system exists in a broader context of impact evaluation that inevitably shapes the structure and content of business cases submitted to the EIB. At the programme level, EU research programmes are evaluated through structured monitoring and ex post evaluation framework. The Horizon Europe programme, for example, uses a comprehensive evidence framework (European Commission, 2023) designed to track programmes outputs, outcomes and impacts over time. It responds to the EU budget performance framework built from Horizon 2020 ex post evaluation lessons and aligned with Better Regulation Guidance. It frames evidence in three uses: "evidence to design" (strategic plans and work programmes), "evidence to inform" (ongoing monitoring and analysis) and "evidence to improve" (ex post evaluation). Specifically, the framework emphasises monitoring as an evidence base for ex post evaluation, not merely for reporting, but also to feed back into strategic planning and priority setting to inform decision making. It is built around two main sets of data and indicators in the Horizon Dashboard including Key Impact Pathways (KIPs) indicators and programme implementation/management data set out in the Horizon Europe Regulation. Figure 4-1 maps how the Horizon ex post evaluation framework works with the KIPs. KIPs are legally defined, should be time-sensitive and measure progress towards three primary impact objectives: science impact, societal impact and economic and technological impact. KIPs use proxy indicators spanning short, medium and longer terms including beyond programme duration. Indicators within these pathways track outcomes such as research publications, patent applications, industry collaborations, employment creation and societal benefits resulting from research activities.

Figure 4-1: Horizon ex post evaluation framework – Key Impact Pathways (KIP)



Strengths and limitations in R&D appraisal

Strengths – what works well to address the specific challenges of R&D appraisal

4.12 A core strength of the EU R&D appraisal system is that it functions in the context of and in conjunction with other elements of the ROAMEF cycle. While we have focused on the EIB R&D appraisal framework, it is important to note that this is informed by and interpreted in the light of other mechanisms, such as the Better Regulation Guidelines governing legislation and programme development and the EIS, which provides the framework for both ex ante assessment and ex post evaluation. This context ultimately defines a coherent policy cycle structure (ex ante impact assessment + ex post evaluation + consultation + implementation support) and specifically positions these as means to ensure EU intervention adds value and remains fit for purpose. These frameworks and guidelines translate to R&D spending by providing the appraisal standards and performance metrics that justify, guide and measure the impact of those funds.

4.13 In that vein, one of the EU's strongest R&D appraisal design features is the explicit coupling of ongoing monitoring, ex post evaluation criteria and legal obligation in Horizon Europe. The evidence framework (European Commission, 2023) specifies that KIPs plus implementation and management data form a comprehensive evidence base to support ongoing monitoring and progress towards impact and to support ex post evaluation across all five criteria including effectiveness, efficiency, relevance, coherence and EU added values. It also embeds transparency as a legal requirement through continuous monitoring and public availability of project data (Article 50). The Research and innovation Strategies for Smart Specialisation (RIS3) model also provides a structured template for regional evaluability that warns against strategies lacking clear intervention logic and indicator systems and set out monitoring and ex post evaluation roles (monitoring by implementers and

evaluation by independent experts), strengthening credibility for place-based innovation policy appraisal.

4.14 **The EU benefits from a sophisticated data infrastructure** including Eurostat, Community Innovation Survey, patent and publication datasets, etc. to support potentially more advanced methodological design and to underpin estimates.

Weaknesses – what are the limitations that cannot address the specific challenges of R&D appraisal

4.15 Despite these strengths, EU R&D/Innovation policy appraisal faces several methodological challenges as well. First, large-scale research programmes make it difficult to determine which outcomes are directly caused by EU funding, where such funding generates benefits through spillovers and system effects, yet many empirical approaches estimate beneficiary-level effects only. AI-Ajlani, et al., (2025) stresses that policymakers care about aggregate impacts on the system, which may differ from beneficiary-level causal effects due to spillovers, crowding-in/out and peer effects. It therefore treats spillovers as a first-order appraisal problem rather than a residual technicality.

4.16 Another limitation is long time horizons, where many research and innovation impacts targeted by EU R&D programmes may take decades to materialise, but ex post evaluations are often required earlier. The appraisal therefore needs to **include shorter-term surrogate outcomes and address the role of indicators such as KIPs in bridging the gap**. The evidence framework (European Commission, 2023) echoes this in operational terms: certain long-term KIP indicators have reporting frequencies beginning only later in the programme (e.g., yearly as from 2026 for specific world-class science tracking), showing real data lags even in an indicator-rich framework.

4.17 Wirkierman, et al., (2018) looking at the National Innovation Systems (NIS) framework and the diversity of European innovation systems, argues that there is “no one-size-fits-all” and countries’ idiosyncrasies and heterogeneities mean that success depends on the complex combination of conditions rather than a universal model. The system complexity also leads to pervasive heterogeneity, and appraisal strategies should respect policy heterogeneity rather than assuming uniform treatment effects. Moreover, multi-level governance also increases the complexity of policy appraisal. The innovation outcomes should be linked to regional ecosystem maturity, and investment mixes should differ across region categories, therefore, the innovation divide requires diffusion and skills as well as support for applied research and uptake. This implies that EU appraisal cannot rely on a single success measure and must confront distributional and place-based trade-offs.

Data and capability requirements

4.18 The Horizon Europe evidence framework (European Commission, 2023) spells out a modern ex post evaluation track: administrative data collected via proposal and reporting templates, enriched across the programme lifecycle, and integrated across multiple IT systems (Participant Data Management; submission/evaluation; complaints; synchronised grant management; integrated workflow for grants/experts/audit; and a central eGrants data warehouse).

4.19 The evidence framework indicates that programme appraisal depends on linking administrative project data to wider statistical infrastructures and mentions use of R&D indicators from Eurostat, the World Bank, and external bibliometric/open science infrastructures for KIP measurement. For network-style appraisal, it references monitoring of collaborative links and network analytics as part of implementation and management data. Although this administrative data collection and linkage is often associated with looking backward (ex post), the framework uses this data to directly improve the design and appraisal of future R&D spending. The framework is designed specifically to bridge the gap between ex post evaluation and ex ante appraisal by creating a continuous evidence loop. Specifically, the framework ensures that the lesson(s) learnt from previous funding cycles are integrated into the Impact Assessments required for future strategic plans. By tracking KIP indicators with the with established administrative data framework, the Commission can better predict which types of R&D investments have a higher probability of success. Additionally, data from the "Participant Data Management" and "eGrants" systems helps policymakers challenge the clarity and plausibility of new programmes objectives before they are launched,

4.20 At the system level, the 2025 EIS reveals the additional muti-source measurement stack needed for comparative innovation appraisal: it credits data inputs including venture capital expenditure from Invest Europe, trademarks and designs from the European Union Intellectual Property Office, publications/patents providers, Eurostat units and national statistical offices, and Joint Research Centre (JRC) auditing for composite indicators. These data dependencies imply that EU appraisal relies not only on programme IT data systems but also on long-run measurement governance and standardisation. Wirkierman, et al., (2018) state that looking at NIS frameworks also provides an additional micro-data requirement for heterogeneity-aware appraisal, which builds its empirical mapping on Eurostat Community Innovation Survey 2014 (CIS2014) micro-aggregated firm data and relates firm innovation strategies/performance to institutional characteristics of national innovation systems, suggesting that robust assessments of policy effects across contexts require firm-level innovation survey evidence, not only grant reporting outputs.

4.21 The RIS3 framework (European Commission, 2012) requires regionally disaggregated context, result, and output indicators and demands that these be anchored in a clear intervention logic to avoid "common evaluation failures" in strategies without clear intended results and causal mechanisms. Finally, the EIB guide (2023)'s appraisal framework implies specific counterfactual and externality data needs: without-project scenarios; valuation parameters for externalities (including environmental and knowledge spillovers); risk and uncertainty parameters; and for research infrastructures, quantified benefit channels such as human capital, publications, spin-offs, and patents.

4.22 Apart from the data access and linkages, there is a high need for analytical and institutional capabilities in the EU R&D policy appraisal. In this sense, a rich level of accessible and relevant data supports the appraisal process. In addition, AI-Ajlani, et al., (2025) implies a high bar of ex post evaluation skills handling spillovers and aggregate effects, long-run outcomes including surrogate variables, mediation pathways, programme heterogeneity and econometric pitfalls in common methods such as difference in difference (DID). The Horizon Europe evidence framework (European Commission, 2023) also implies data engineering and applied analytics capabilities to maintain and exploit the integrated IT/data stack and to operate public-facing dashboards and internal analytics.

Institutionally, Better Regulation Guidelines (European Commission, 2021) emphasises the need for dedicated resources and training and for centres of expertise within DGs and central services, which is an explicit capacity prerequisite rather than an assumed backdrop. The RIS3 framework (European Commission, 2012) adds the institutional requirement of independent evaluators for credible strategy evaluation. The Horizon Europe regulation also requires interim evaluation assistance from independent experts selected transparently.

Overall assessment

4.23 Overall review of the EU R&D policy appraisal system shows that it is strong, combining clear overall policy guidance and criteria that shape legislation and programming with explicit processes for ex ante appraisal across a wide portfolio of public spending, including R&D. This also takes into account a fuller appreciation of the timing of benefits attributable to policy interventions.

4.24 It also includes mechanisms to systematically translate ex post evaluation findings to continuously inform the development of ex ante appraisal frameworks: the Horizon Europe evidence framework treats monitoring, identifiers and data linkage as a permanent infrastructure that supports mid-course programme adjustment, public transparency and legally mandated ex post evaluation. Ex post evaluations are used as evidence to justify estimates in subsequent ex ante cases. However, the EU's own ex post evaluation scholarship then provides a candid warning: even sophisticated indicator frameworks cannot by themselves resolve fundamental causal challenges such as spillovers, displacement, long lags and programme heterogeneity; appraisal must be designed to address these properly. This is a recognition that ex post evaluation alone may be insufficient to fully appraise R&D policy options.

4.25 As for UK policymakers, there are a few implications to be drawn from the review of the EU R&D policy appraisal:

- First, the UK could adopt the time-phased "impact pathways" for major UK R&D portfolios with staged proxy indicators and explicit lag recognition to allow appraisal cases to consider the full time-horizon of potential policy effects and impact. Horizon Europe's KIPs are designed to reflect a non-linear R&D impact chain and separate short/medium/long indicators with some long-term evidence only appearing years into the programme. UK programmes can learn from the design principle: commit to proxy indicators when impacts lag, rather than letting "impact" be an ex post narrative.

- **Second, it may be appropriate to build a monitoring system as ongoing evaluation infrastructure, not as a reporting overhead.** The Horizon framework shows that credible ex post evaluation depends on integrated grant IT data systems, structured templates and a unified data warehouse with external validation and dashboarding. The UK policy appraisal framework could also budget and govern these systems upfront rather than rely on retrospective data assembly. It will be useful to mandate identifier-first data design (researcher IDs, DOIs, organisational identifiers) to enable low-burden longitudinal tracking.

- Third, the UK can also learn from the EU by making portfolio ex post evaluation and heterogeneity analysis routine for large R&D funding projects. The RIS3-style evaluability principles could also be applied to UK regional-based or targeted innovation policy. The cohesion policy emphasises that investment mixes should differ by ecosystem maturity and that diffusion and skills are needed alongside applied research and uptake.

- Fourth, the EU ex post evaluation paper also warns beneficiary effects can diverge from aggregate impacts due to spillovers and crowding effects and the EIB appraisal framework shows how to define without-project scenarios and value externalities while accounting for risk/ uncertainty. **The UK programme can integrate these elements explicitly rather than treating spillovers as an untested assumption.**

5. Appraisal approach in Japan

Overview of the country

5.1 The overview review of Japan's current R&D policy shows that it has evolved from a historically strong, large-firm incremental innovation model toward a more mission-oriented and system-directed innovation policy framework. This shift has been driven by structural economic stagnation since the 1990s, technological transitions and increasing recognition that traditional industrial R&D models were less suited to emerging science and technology driven sectors (Larrue, 2021). The Japanese policy response has therefore emphasised long-term strategic planning, coordination through central government "control tower" structures notably within the Cabinet Office and Council for Science, Technology and Innovation (CSTI), and the deployment of large-scale, goal-oriented innovation programmes.

5.2 Specifically, Japan's policy mix is characterised by a combination of mission-oriented programme (e.g., the Strategic Innovation Promotion Programme (SIPP), ImPAC, Moonshot), business R&D tax incentives, SME innovation support, including Small Business Innovation Research (SBIR)-style schemes,[4] and public innovation financing instruments delivered through multiple agencies. Larrue (2021) highlighted that the mission-oriented programmes are explicitly designed to address societal challenges and require coordination across ministries, regulatory domains and research communities. This implies that appraisal cannot be limited to single-project outputs; rather, it must progress toward complex, long-term societal and technological goals. At the same time, Japan's long-standing reliance on supply-side innovation support and strong corporate R&D capacity means that policy appraisal must grapple with questions of additionality and system transformation rather than simple increases in R&D expenditure. Intarakumnerd (2013) reviewed Japanese financing policy for innovation since 1990s and explicitly notes that Japan historically focuses more on supply-side support mechanisms and less on demand-side instruments such as innovation-oriented procurement, which affects both policy effectiveness and the ability to observe downstream impacts.

5.3 Overall, Japan's R&D policy context suggests a system with high strategic ambition and institutional sophistication, but also significant complexity in governance and policy layering which directly shapes how appraisal is conducted and used. Japan is a suitable comparator for the UK when examining appraisal methods for R&D investment. It combines a highly advanced innovation system, a strong culture of evidence-based programme design, and has policy challenges that mirror the UK's. While there are some distinctive aspects to the R&D system in Japan, this serves to highlight meaningful similarities and differences, yet as a large, mature and international R&D system is comparable with the UK. The strong culture of structured evidence-based appraisal is also relevant in this case.

[4] SBIR-style schemes are competitive, government-backed, non-dilutive funding programs designed to foster innovation by supporting SMEs in high-tech R&D with commercial potential. Example of such scheme in the UK is the Innovate UK Contracts for Innovation (formerly known as the Small Business Research Initiative or SBRI).

Description of appraisal approaches

5.4 **Japan stands out among advanced innovation systems for having a formalised, codified national framework specifically for assessing government-funded R&D.** This is both ex ante and ex post in nature, covering the full lifecycle of R&D policy appraisal and assessments. **This suggests an R&D specific approach rather than relying solely on general public policy appraisal guidance.** The 2012 National Guidelines for Evaluating Government-Funded R&D published by Cabinet Office in Japan (2012) establish a system-wide evaluation (both ex ante and ex post) structure that operates across multiple layers including projects, programmes, organisations and broader policy measures. Importantly, these guidelines embed evaluation within a Plan-Do-Check-Action (PDCA) cycle, signalling that **appraisal is intended not merely as accountability but as an iterative management tool for improving R&D policy effectiveness and prioritisation.** The guidelines set out the need to **clarify relationships between different policy layers and for outcome-oriented goals, which indicates a sophisticated understanding of R&D as a multi-level system rather than a collection of isolated interventions.** Methodologically, Japan's official approach is notably pluralistic. While quantitative indicators such as outputs, outcomes and efficiency metrics are encouraged to improve objectivity, the guidelines stress the caution against over-reliance on metrics alone, particularly for fundamental research where impacts are uncertain and long-term. Instead, qualitative expert judgement, international benchmarking and forward-looking assessments of scientific potential are emphasised (Cabinet Office Japan, 2012). This reflects a recognition of intrinsic uncertainty of R&D investments and the limitations of purely economic appraisal models in practice.

5.5 Beyond formal guidelines, **the appraisal of major mission-oriented programmes follows a portfolio and stage-gate logic.** For instance, Larrue (2021) describes how programmes such as Moonshot are assessed through three mechanisms: diversified project portfolios, milestone reviews and structured discontinuation where progress is insufficient. **The core insight from such Moonshot-style programme is that its three mechanisms all carry implicit ex ante appraisal logic that conventional CBA frameworks miss.** Rather than asking "which single project has a positive NPV?", the ex ante question becomes "what combination of high-risk bets, with what spread of approaches, gives the programme the best chance of reaching the mission goal?" This is **a real-options framing rather than a standard cost-benefit one.** Crucially, the milestone review and discontinuation architecture must be designed before funding begins. The 'stopping rules', the authority to rebalance the portfolio, and the conditions under which the causal theory would be revised are all highly relevant ex ante appraisal questions. This means **a Moonshot-informed business case should specify not just the expected causal pathway to impact, but the observable signals that would trigger continuation, revision or termination, which effectively embedding the evaluation framework into the appraisal itself from the outset.** Such approaches align more closely with innovation portfolio management than traditional cost-benefit appraisal. The financing policy review paper (Intarakumnerd, 2013) suggests that, in practice, assessment across innovation programmes can be uneven and fragmented. Monitoring and assessment practices differ significantly across agencies and there are concerns about reliance on self-evaluation and limited independence of evaluators in some cases, which creates a gap between the formal sophistication of national framework and the operational reality of programme-level appraisal.

Strengths and limitations in R&D appraisal

Strengths – what works well to address the specific challenges of R&D appraisal

5.6 A key strength of Japan's appraisal approach lies in its **formal institutionalisation of R&D appraisal and evaluation as a core government function.** The national guidelines clearly define evaluator responsibilities, transparency expectations and the role of ex ante and ex post evaluation in accountability to taxpayers, which indicates **that appraisal is embedded within governance processes rather than treated as an external audit exercise.** Another major strength is the system's methodological flexibility, which **combines quantitative indicators with qualitative expert judgment.** Such a methodological approach avoids the common pitfall of metric-driven assessment that can distort research incentives or penalise high-risk innovation, which is particularly appropriate for mission-oriented and frontier research programmes where success pathways are uncertain. In this respect, the focus on appraisal that is informed by clear assessment guidance allows a mission and portfolio level focus that transcends individual policies.

5.7 The national guidelines also indicate that Japan's framework directly addresses a common failure in R&D systems – multiple layers of administration evaluating the same activity – by recommending elimination of overlaps, sharing results and using electronic systems to streamline processes (Cabinet Office Japan, 2012). Japan also demonstrates an advanced understanding of portfolio-level assessment, where **mission programmes are designed with stage gates, diversified project portfolios, and acceptance of failure as part of high-risk innovation management**. **This allows appraisal to focus on overall mission progress rather than binary project success** (Larrue, 2021). Moreover, there is growing attention to SME/start-up instruments with accumulating descriptive evidence. For instance, Japan's SBIR has tracked spending and award counts over time, and such programme footprints can be a foundation for more structured evaluation (e.g., counterfactuals, cohorts, longitudinal tracking) (Kato and Coad, 2024).

Weaknesses– what are the limitations that cannot address the specific challenges of R&D appraisal

5.8 However, the review of Japan's R&D policy also highlights structural limitations. First, the governance complexity arising from multiple ministries, councils and coordination bodies involved R&D funding, appraisal and evaluation. This can increase the transaction costs and blur boundaries between strategy/coordination vs. implementation/management, and create accountability issues for ex post evaluation outcomes (Larrue, 2021). **Insufficient coordination between agencies** such as METI and MEXT[5] further fragments innovation financing and **make system-wide appraisal difficult**, particularly for policies spanning the full innovation lifecycle (Intarakumnerd, 2013). The second limitation relates to evaluation (ex ante and ex post) independence and credibility risks. Intarakumnerd (2013) notes that monitoring and evaluation are often not systematic, where self-evaluation is common and even third-party evaluation may still be closely tied to implementing

[5] Ministry of Education, Culture, Sports, Science and Technology (MEXT) and Ministry of Economy, Trade and Industry (METI) are key Japanese government ministries that collaborate to advance research, technology, and industrial policy. MEXT focuses on basic research and education, while METI focuses on commercialisation and industrial application.

organisations, potentially reducing objectivity and the willingness to terminate underperforming programmes.

5.9 Larrue (2021) also highlight the constraints of transparency and international learning, where assessments of major programmes such as SIPP and ImPAC are often conducted by Japanese experts and not widely disseminated internationally, limiting peer scrutiny and cross-country policy learning.

5.10 Finally, there are also programme design issues where the SME support is not always aligned with dynamism goals. The SBIR monograph argues that Japan's SBIR scheme differs from the US in that it targets relatively older SMEs rather than dynamic start-ups, illustrating how appraisal outcomes may reflect structural targeting choices rather than implementation performance, complicating interpretation of effectiveness (Kato and Coad, 2024).

Data and capability requirements

5.11 The 2012 National Guidelines for Evaluating Government-Funded R&D published by Cabinet Office, Japan (2012) calls for systematic storage of R&D results and ex post evaluation outputs in standardised formats, inter-organisational sharing of evaluation information, and the use of electronic systems to streamline application, screening and disclosure processes. This implies a highly developed conception of the infrastructure needed for effective R&D assessment. This also shows that **Japan views evaluation (ex ante and ex post) as a data-intensive, system-wide function rather than just a project-level assessment**. Specifically, the guidelines emphasise building dedicated evaluation divisions, securing budgets for evaluation functions, and expanding the evaluator pool including both domestic and international experts with research/management experience, while managing evaluator burden. Such provision reflects recognition that **appraisal capacity depends on institutional resourcing and human capital not just methodological guidance**.

5.12 Robust appraisal (including counterfactual analysis) requires firm-level datasets on R&D, productivity, patents, employment and financial performance, matched with policy exposure variables and crucially datasets capturing supply-chain and ownership networks for investigating potential spillover effect. The ability to integrate these datasets also enables ex post evaluation of indirect spillovers and social returns moving beyond narrow beneficiary metrics. The **empirical studies drawing on Japan's corporate survey and tax administrative data have demonstrably informed iterative policy design** in the sense that Japan has repeatedly revised credit rates, caps, SME thresholds, and the open innovation credit structure in ways that reflect heterogeneous firm-level findings. However, such justification rests primarily on the existence of positive externalities rather than a formally quantified net social benefit calculation. The challenge of incorporating the ex post evaluation into ex ante appraisal approach more effectively might be tackled by the better linked microdata availability. Moreover, the currently available network or supply chain datasets are limited by a lack of transaction values or cross-border linkages, which constrains the precision in spillover estimation. These broader challenges in the evaluation of innovation impacts is commonly observed in the literature despite the use of advanced statistical methods.

5.13 Finally, for the specific capabilities that are needed for the Japanese system work in practice, it requires choosing appropriate indicators, balancing qualitative and quantitative judgement and avoiding perverse incentives for the evaluation design. For mission-oriented programmes, it requires portfolio/programme management capability (Larrue, 2021). The RIETI report (2024) reviewing the UK government's policy evaluation compared with Japan's case suggests that **appraisal capability is increasingly being professionalised through Japan's broader Evidence-Based Policymaking (EBPM) agenda,** including institutional reforms, administrative review processes and dedicated analytical centres. This indicates a strategic effort to embed analytical expertise within government decision-making.

Overall assessment

5.14 **Overall, Japan's core strength is that it has both (a) a formal, system-wide R&D evaluation structure with the national R&D evaluation guidelines, and (b) advanced programme forms (mission-oriented portfolio) that acknowledge uncertainty and embed structured stopping/learning mechanisms.** However, the effectiveness of the Japanese system is constrained by governance and coordination complexity, uneven assessment independence across agencies, and limited dissemination of major programme ex post evaluations findings, reducing the impact of appraisal on strategic decision-making.

5.15 As for the implications on UK's policy appraisal, **Japan's "Portfolio + stage-gate" logic could be used to complement VfM model.** Moonshot's package-level risk management and stopping rules offer a useful model for UK R&D missions where project failure is expected, where appraisal criteria and assessment design should match the innovation risk profile. Japan's emphasis on avoiding duplicative evaluation and building shared evaluation infrastructure (including to inform appraisals) could also transfer to UK's case considering the UK often faces challenges with fragmented evidence across funders/programmes. Japan's guidelines explicitly recommend standardised storage and cross-organisation sharing of ex post evaluation information to reduce burden and improve reuse - this is an operational lesson rather than a constitutional one. Japan's own review report of the UK (RIETI, 2024) shows that the UK historically leaned heavily on ex ante cost–benefit logic and is now strengthening ex post evaluation integration and causal evidence use (e.g., through the evaluation taskforce). That framing is useful for a UK audience because it positions **Japan an inspiration for "how to build a stronger learning loop".**

6. Appraisal approach in Ireland

Overview of the country

6.1 Based on the European Innovation Scoreboard (2025), Ireland is classified as a "Strong Innovator", which scored 123.1% of the EU average and ranked 5th. Specifically, the EIS table shows Ireland's top-tier performance in population with tertiary education, sales and employment impacts, sales of innovation and labour productivity. It also shows strong linkages driven by the innovative SMEs collaborating and public-private co-publications due to targeted instruments (e.g., Disruptive Technologies Innovation Fund; innovation partnerships; emerging cluster frameworks). Such strong "system connectivity" implies that useful intermediate outcomes for evaluation may include network indicators, collaboration outputs and diffusion pathways – not only patents and productivity.

6.2 However, despite the high innovation outcomes and linkages, Ireland's R&D country profile shows a distinctive policy-mix imbalance. There is very low public-sector R&D expenditure, 16.7% of EU average (rank 26) and heavy reliance on tax incentives. Although direct and indirect government support of business R&D is high, the indirect measures represent around 80% of the total business R&D support. The Finance Bill (2025) reinforces how central tax-based support remains in Ireland's policy approach, which proposes a rate increase in the R&D tax credit and a higher first-year payment threshold. A tax credit lowers the user cost of R&D. The alternatives are grants or soft loans which also lower the user cost of R&D, but this system leans heavily towards the former in the R&D space with grants the preferred alternative for research and talent development.

6.3 Ireland is a strategically relevant comparator for the UK when examining appraisal methods for R&D investment. The two countries share deep structural similarities in their innovation systems, but Ireland also offers distinctive features that make comparison analytically useful. The industrial mix (i.e. advanced manufacturing, life sciences, digital, financial services) and a shared research culture (i.e. university centred research systems with strong international collaboration) are comparable in both the UK and Ireland. Ireland has rapidly increased its R&D intensity over the past decade, which aligns well with the UK.

Description of appraisal approaches

6.4 To better understand Ireland's R&D appraisal approaches, this case reviewed two types of government documents: general government policy development and evaluation approaches, and instrument-level appraisal built into the administration and review cycle of the R&D tax credit.

6.5 First, Ireland's Policy Handbook (2024) is explicit that policy advice must be logical and evidence-based, and that policy proposals brought to the government should set out impacts across a range of dimensions. It outlines **a four-step appraisal and design process that requires a documentation of policy demand, data and evidence review, evidence-based design, and then policy selection.** Most importantly for appraisal practice, the handbook provides a structured resource set across three "pillars" (Data and Evidence, Feasibility and Implementation, and Legitimacy), including a list of tasks, tools and processes that are intended to make policy evaluable by design. Under Data and

Evidence, it highlights institutional resources such as the Central Statistical Office (CSO) administrative "data-room" service based on linked data, enhanced data-sharing through a data governance board, and CSO "pathfinder projects" based on linked data. It also explicitly references capability training in Cost Benefit Analysis and Counterfactual Impact Evaluation alongside statistical tooling. Under Feasibility and Implementation, it highlights formal appraisal/review mechanisms like Regulatory Impact Assessment (RIA), Public Spending Code and Spending Review Process that structure government decision-making. Finally, the handbook also includes a "Design with Best Advice" checklist which requires a logic model, identification of risks and sensitivity/scenario analysis. **Overall, the general policy handbook set up an appraisal culture that addresses that risks/assumptions must be tested and data access must be treated as an enabling condition for ex post evaluation not just a "nice to have".**

6.6 Given that tax credit is Ireland's dominant public R&D support instrument, its government framework is central to understanding the R&D appraisal in practice. Ireland's R&D Credit Guidelines (2015) set out concrete eligibility appraisal and compliance assurance, and state that to claim the credit a company must satisfy two essential tests: the Science Test to qualify R&D activity; and the Accounting Test (proper tracking/accounting of costs) to ensure that contemporaneous records are available for Revenue inspection. They also lay out what documentation must exist to evidence uncertainty, advancement and systematic investigation. The guideline creates a micro-level appraisal regime which addresses that firms must articulate and evidence the logic and conduct of their R&D activities.

6.7 At the policy level, in the 2025 public consultation paper on the R&D Tax Credit and on options to support innovation, the Department of Finance set the policy objective clearly: increase business R&D to support innovation and productivity and keep Ireland attractive for investment. It explicitly states that the credit is reviewed regularly under the Tax Expenditure Guidelines to ensure effectiveness and alignment with policy objectives. The consultation report also provides a narrative of how the regime has evolved through iterative reforms and periodic evaluation, including the most recent evaluation in 2022 and the Finance Act 2022 restructuring linked to OECD Pillar Two treatment of refundable credits. Finance Bill (2025) further shows appraisal and policy adjustment in action. It proposed increasing the credit rate, raising first-year payment threshold and introducing a 95% deeming rule for staff time allocation. These changes reflect ongoing policy tuning informed by administrative and behavioural considerations.

Strengths and limitations in R&D appraisal

Strengths – what works well to address the specific challenges of R&D appraisal

6.8 First, Ireland's policy handbook (2024) suggests that policy appraisal is seen as a capability to be built not just a compliance step. **The emphasis on logic models and explicit theories of change, risk and sensitivity analysis and early identification of data and evidence needs** suggests that Ireland policy processes aim to make programmes ex post evaluation-ready from the outset to support initial appraisal.

6.9 The second strength lies in the **assurance of the strong evidentiary standards at firm level**. Revenue's detailed documentation requirements for the tax credit mean that qualifying R&D activities must be systematically described, evidenced and costed. The science and accounting tests force claimants to articulate uncertainty, advancement and systematic investigation and to maintain traceable costing and time allocation. This creates a structured evidence base that supports both compliance and potential evaluation. For these purposes, uncertainty is defined as:

- uncertainty as to whether a particular goal can be achieved, or
- uncertainty (from a scientific or technological perspective) in relation to alternative methods that will meet desired specifications such as cost, reliability or reproducibility.

6.10 This high evidence standard for supported firms allows feedback to inform the appraisal of existing policy interventions and the preparation of new appraisal cases.

6.11 In addition, the Department of Finance Consultation and Finance Bill reforms demonstrate an iterative policy cycle. The credit is routinely reviewed, evaluated and adjusted through legislation and it explicitly uses consultation to gather evidence on effectiveness and design. **This indicates that appraisal is not purely retrospective but feeds forward into instrument design, suggesting a learning-oriented policy system.** This is particularly important for R&D appraisal as R&D policy operates in a rapidly changing technological and fiscal environment. An iterative approach allows adaptation while preserving continuity of core instruments. In one sense, Consultations can inform future policy appraisal and increases the potential for a case to be built for a reconfiguration of an existing policy instrument.

Weaknesses– what are the limitations that cannot address the specific challenges of R&D appraisal

6.12 **In practice, additionality is inherently difficult to assess.** As addressed in the EIS, Ireland heavily relies on tax incentives while direct public-sector R&D expenditure is low. This means that ex post evaluation must answer counterfactual behavioural questions rather than simply track spending. Counterfactuals are often part of the appraisal process, but not in this case. That they only come in later and are difficult to assess is a significant weakness of this system. The Consultation's explicit questions about what would have happened in the absence of the credit highlighted this challenge. The requirements of estimations of counterfactual firm behaviour induce a complex econometric challenge, which requires sophisticated econometric methods and integrated longitudinal datasets. Specifically for R&D appraisals, indirect instruments blur the boundary between policy impact and underlying economic structure. High-performance sectors may claim credits because they are already R&D intensive not because of the policy impact.

6.13 The administrative complexity also constrains policy and appraisal. Specifically, in the consultation report (2025), stakeholder feedback emphasises complexity in the tax credit regime. Policymakers explicitly avoid broadening the definition of R&D to include innovation because it would increase uncertainty and reduce administrability. Therefore, such complexity could limit both participation and obscure measurement. Smaller firms may be deterred from claiming, skewing the observed distribution of beneficiaries. The complex rules may also generate inconsistent interpretations and

documentation. This suggests that a narrow and focused approach to defining eligibility for specific programmes and schemes should be built into scheme appraisal.

6.14 An EU policy paper (Al-Ajlani et al., 2025) noted that R&D policy evaluation often measures beneficiary impacts, but policymakers care about system/aggregate impacts. Spillovers, peer effects and crowding effects mean that beneficiary-level impacts can drive divergence from aggregate impacts and long-term impacts may differ from short-term indicators. Ireland's high linkage and collaboration scores suggest such networked innovation system effects are likely important but difficult to measure. This suggests that appraisal needs to widen its remit and consideration the wider system that firms are embedded within.

Data and capability requirements

6.15 The review of Ireland's documents collectively imply that effective and robust R&D appraisal depends on linked data infrastructure, counterfactual evaluation capacity and high-quality operational evidence from the tax system. Specifically, administrative microdata, CSO data-room, linked-data services, data sharing governance and linked-data pathfinders are explicitly highlighted as evidence enablers, allowing firm behaviour to be linked with outcomes. This is directly relevant for R&D appraisal as many key assessment questions cannot be answered without the system-level data infrastructure. For instance, to assess the impact of the R&D tax credit on productivity or employment requires linking tax claim data with firm performance, labour and sector datasets over time. The structured and linked data will also support the counterfactual evaluation design (e.g., matching, Difference-in-Difference (DID), requiring firm-level panel data), longitudinal tracking of outcomes.

6.16 To answer the counterfactual questions in the policy assessment, advanced analytical skills are needed. The policy handbook recognises that data alone is insufficient and addresses the need for analytical capability, including training in Cost Benefit Analysis (CBA), counterfactual impact assessment and advanced statistical tools such as R and Stata. These methods are precisely those required to evaluate indirect instruments like tax credits, where impacts must be inferred from behavioural responses rather than direct outputs. Sufficient policy assessment needs econometric and quasi-experimental tools, and the ability to interpret intermediate outcomes that predict long-term effects.

Overall assessment

6.17 One important feature of R&D expenditure in Ireland is that private sector firms accounted for 84% of the spending in 2023, and this was primarily led by foreign-owned enterprises. In this respect, within an economy where private investment is high, public R&D investment has historically lagged below the EU average.[6] Overall, Ireland combines strong innovation performance but has a distinctive R&D policy mix with low public-sector R&D expenditure and heavy reliance on indirect business support, particularly the R&D tax credit. **Ireland's policy appraisal approach is structurally coherent but methodologically demanding. This structure shifts appraisal away from**

[6] Central Statistical Office: https://www.cso.ie/en/releasesandpublications/ep/p-berd/businessexpenditureonresearchanddevelopment2023final-2024preliminary/keyfindings/#:~:text=April%202025%20%2C%2011am-,Key%20Findings,compared%20with%2037%25%20in%202021.

programme delivery metrics toward behavioural additionality, spillover and system-level impact. In the UK, the system tends to focus on individual schemes and programmes. The specific R&D policy appraisal combines a system-level evaluability framework with a dominant tax-based instrument governed through clear evidentiary standards and periodic review. Thus, individual policy instruments are then assessed at a higher level with all potential policy options. The key strengths lie in data readiness, compliance discipline and iterative reform, while the central challenge exists in assessing additionality and spillovers in a tax-credit-led ecosystem.

6.18 The UK Green Book frames appraisal as options appraisal and evaluation across the policy cycle requiring clear objectives, credible option development, assessment of cost/benefits/risks, and monitoring and ex post evaluation. Ireland's policy development guidance aligns closely with this logic. Ireland's experience reinforces key Green Book principles: build evaluation into policy design; ensure tax incentives are evidence-ready and administratively workable; prioritise counterfactual analysis for VfM assessments; and complement programme assessment with system-level analysis of spillovers and longer-term impacts. **The latter point of considering system level analysis extends the Green Book approach.**

7. Appraisal approach in the US

Overview of the country

7.1 The United States invested approximately $940bn in R&D in 2023, representing roughly 29% of global R&D expenditure – the narrowest margin of leadership since the mid-1990s (Zimmermann, 2025). The OECD Main Science and Technology Indicators also show R&D intensity at 3.6% of GDP, placing the US fourth globally behind Israel (6.35%), South Korea (4.96%) and Sweden (3.60%) (OECD, 2025a). A key feature of R&D in the US is the dominance of business sector funding, which counts for approximately 75% of all US R&D, while the federal government's share has declined from 67% in the early 1960s to just roughly 18% in 2023 (Zimmermann, 2025). Such a shift reflects both the extraordinary growth of corporate R&D concentrated in five industries led by information technology and chemicals manufacturing, and the secular decline of federal investment. The composition of US R&D by type reveals a further structural concentration with experimental development absorbing 67% of total spending, applied research 18% and basic research only 15%. Between 2012 and 2023, the business sector's share of basic research funding grew from 21% to 35%, while the federal share fell from 52% to 41%, raising concerns about whether market-driven basic research adequately serves the public good.

7.2 Overall, the US innovation system is characterised by what Atkinson (2020) terms the "innovation success triangle", which identifies three interdependent determinants of national innovation performance: the business environment; the trade, tax and regulatory environment; and the innovation policy environment. It argues that despite the strong business environment in the US, the innovation policy environment has deteriorated to become the weakest leg of the triangle. **The governance of federal R&D is deliberately fragmented across more than 23 agencies with independent R&D budgets, each pursuing mission-specific research objectives.** Specifically, the department of Defence accounts for approximately 62% of the FY2026 R&D budget request followed by the National Institutes of Health (15%), Department of Energy (8%), NASA (6%) and the National Science Foundation (2%) (CRS, 2025).

7.3 In the US, coordination is provided by two executive institutions: The Office of Management and Budget (OMB) and The Office of Science and Technology Policy (OSTP), with both together issuing the annual R&D budget priorities memorandum that constitutes the primary strategic steering mechanism (Vought and Kratsios, 2025). In practice, Federal agencies send R&D priority recommendations to OSTP and OMB, which form the basis for the joint memorandum. OSTP then provides guidance as agencies develop their budgets, focusing on agencies with larger R&D budgets and cross-agency priorities. OSTP advises on the quality and alignment of agency budget proposals with presidential priorities, while OMB decides what goes into the president's budget proposal, though in many cases OMB defers to OSTP. OMB evaluates agency budget submissions and sets cross-government analytical requirements through circulars including A-94 (benefit-cost analysis) and A-11 (budget preparation). An OSTP provides scientific and technological advice to the president. Congress retains ultimate control through authorisation and appropriations, and the tension between executive direction and congressional allocation is a defining feature of the system. Wood (2025) presents a complementary analytical framework, characterising the US as a "Polycentric Innovation

State" - a hybrid system combining market mechanisms with federal mission-oriented policies and state-level initiatives. This three-level structure (federal, state, and private sector) provides resilience: even when federal funding contracts, state governments pursue independent innovation strategies and private sector investment continues.

7.4 **The OMB Circular A-94 (Guidelines and Discount Rates for Benefit-Cost Analysis of Federal Programs) provides the formal methodological framework for the prospective appraisal of federal spending programmes**, which was substantially revised in November 2023. However, the circular provides generic guidance applicable to all federal programmes while **there are no R&D specific provisions**. It does not set out methods for estimating knowledge spillovers, managing diversified research portfolios, valuing staged investments under radical uncertainty, or recognising basic research as a public good with characteristics fundamentally different from standard government expenditure.

Description of appraisal approaches

7.5 In terms of the R&D policy appraisal approach in practice in the US, **the primary R&D priority-setting mechanism is the annual joint OMB-OSTP memorandum** issued to all federal agencies with significant R&D budgets. The most recent iteration, M-25-34/NSTM-2 established five R&D budgetary priorities for FY2027: (1) ensuring American leadership in critical and emerging technologies; (2) unleashing American energy dominance and exploring new frontiers; (3) strengthening American security; (4) strengthening American health and biotechnology; and (5) assuring continued space dominance (Vought and Kratsios, 2025). The memorandum also specifies five cross-cutting actions: restoring "Gold Standard Science"; building the STEM workforce; expanding research infrastructure; revitalising the science and technology ecosystem through public-private partnerships; and focusing on high-value research efforts that translate into strategic advantage.

7.6 Critically, this **process constitutes strategic direction rather than formal appraisal, which only operates as a political-strategic instrument,** shaping agency budget submissions through directional guidance rather than analytical discipline. The memorandum does not require agencies to demonstrate positive Net Present Value (NPV) or benefit-cost ratios for their R&D investment. **The contrast with A-94's methodological sophistication is stark: the two instruments exist in largely separate institutional worlds.** Atkinson (2020) provides the analytical context for this gap, arguing that the United States "has no national, coordinated innovation policy system" and that the innovation policy environment is fundamentally weakened by the absence of systematic priority-setting informed by comparative return analysis. The annual memorandum partially addresses coordination but substitutes political preference for analytical rigour.

7.7 In practice, **R&D funding decisions are made through agency-specific mechanisms that vary substantially by mission and research type.** For basic research, scientific peer review constitutes the primary selection process. The National Science Foundation (NSF) evaluates proposals against two criteria established by the National Science Board: intellectual merit (the potential to advance knowledge) and broader impacts (the potential to benefit society) (NSF, 2025). Each proposal is assessed by external reviewers and programme officers who also consider portfolio balance. The

National Institutes of Health employ a similar but more structured process: since January 2025, proposals are assessed against three factors - importance of the research, rigor and feasibility, and expertise and resources - with the third evaluated on a binary "sufficient/not sufficient" basis to reduce reputational bias (NIH, 2025). **DARPA represents a fundamentally different appraisal model.** Rather than peer review, DARPA empowers approximately 120 programme managers to identify breakthrough opportunities and fund performers at their discretion. **Programme managers are expected to reject projects for insufficient ambition, and the agency explicitly tolerates failure:** approximately 85% of projects fail to achieve their planned goals, on the rationale that if none failed, the portfolio would not be sufficiently ambitious (Bonvillian, 2006). There is no project-level benefit-cost ratio requirement. Instead, DARPA employs the qualitative "Heilmeier Catechism" - a series of questions about what is being attempted, what is new, who cares, and what difference success would make. **This represents appraisal by expert judgement rather than formal analytical method.**

7.8 **Mission-oriented agencies frame R&D in terms of mission requirements rather than social return.** The Department of Energy's national laboratories pursue research programmes aligned with energy, environmental, and national security missions. The Department of Defence funds R&D through a structured system from basic research (Budget Activity 1) through experimental development (Budget Activity 6-7), with **investment decisions driven by capability gaps and threat assessments rather than civilian benefit-cost analysis.** The Small Business Innovation Research (SBIR) and Small Business Technology Transfer (STTR) programmes represent a distinctive technology transfer mechanism. Federal agencies with extramural R&D budgets exceeding $100m must allocate 3.2% to SBIR grants, supporting early-stage commercialisation in a three-phase model. These programmes, alongside the Bayh-Dole Act (1980) and subsequent Technology Transfer Acts referenced in Atkinson (2020), constitute the institutional architecture connecting federally funded research to commercial application.

7.9 A-94 is formally triggered when agencies submit budget proposals to OMB under Circular A-11, which requires all agencies with annual R&D funding exceeding $10m to categorise spending as basic research, applied research, or experimental development. The categorisation serves primarily as a statistical and classification function, not as a trigger for different depths of analytical justification. The three categories carry different definitions but the same reporting obligations. However, in practice, basic research faces the lightest justification burden (peer review typically suffices), while experimental development attracts more scrutiny through overlapping procurement and investment review frameworks — but this differentiation emerges from the broader system rather than from A-11's categorisation requirements themselves. In principle, agencies should demonstrate that proposed investments generate positive discounted net benefits. In practice, applying benefit-cost analysis to R&D encounters three fundamental methodological obstacles. First, the knowledge spillovers that constitute the primary economic justification for public R&D are extraordinarily difficult to monetise ex ante. Second, A-94's framework accommodates risk through probability-weighted expected values and sensitivity analysis, but the outcomes of fundamental research are often not merely uncertain in magnitude. Third, R&D investments create the option - but not the obligation - to pursue commercialisation if early results prove promising. The discount rate also remains a critical parameter. Even at the revised 3.1% real rate, discounting systematically penalises

basic research whose benefits materialise over decades relative to applied investments with nearer-term payoffs.

7.10 The R&D tax credit, another key R&D policy instrument in the US, codified at 26 U.S.C. §41, was first enacted in the Economic Recovery Tax Act of 1981 and made permanent by the PATH Act of 2015. It provides a credit of 14–20% of qualified research expenditures above a base amount. However, as Atkinson (2020) observes, US R&D tax incentives are minimal compared with most OECD and BRIC nations. A critical recent development was the Tax Cuts and Jobs Act of 2017 provision - effective from January 2022 - requiring companies to capitalise and amortise domestic R&D expenses over five years rather than deducting them immediately. This effectively increased the tax cost of R&D. The One Big Beautiful Bill Act (P.L. 119-21, signed 4 July 2025) permanently restored immediate expensing for domestic R&D under a new Section 174A at an estimated ten-year cost of $139bn (JCT, 2025). This episode illustrates how the regulatory and tax environment - the second leg of Atkinson's triangle - can either support or undermine innovation incentives through mechanisms entirely separate from the formal appraisal framework.

Strengths and limitations in R&D appraisal

Strengths – what works well to address the specific challenges of R&D appraisal

7.11 **The most significant strength of the US approach is its practical flexibility,** that it allows strategic priority-setting through the annual memorandum to operate alongside scientific peer review without imposing rigid benefit-cost requirements on basic research. The differentiation between political direction for strategic themes, and expert judgement for project selection, formal analysis for large programmes where feasible has proved broadly functional despite its lack of formal design. The 2023 revision of A-94 represents genuine methodological advancement. The **reduction in the social discount rate** from 7% to 3.1% substantially improves the treatment of long-term investments, including R&D. The introduction of distributional weighting, explicit treatment of the marginal cost of public funds, and the framework for certainty-equivalent calculations under risk aversion all represent conceptual improvements over the 1992 predecessor. The **requirement for sensitivity analysis, publication of assumptions, and prospective planning of retrospective studies** promotes transparency.

7.12 Mission-oriented institutions, particularly DARPA, represent a **proven model of appraisal by expert judgement that has produced transformative technologies** including the internet, GPS, and mRNA vaccine technology. The DARPA model demonstrates that formal benefit-cost analysis is not the only or necessarily the best approach to selecting high-risk, high-reward research investments. Its institutional features (empowered programme managers, term limits creating urgency, explicit tolerance for failure, flat hierarchy) constitute a coherent alternative appraisal philosophy that has been replicated internationally, including through the UK's ARIA. The strength of US private sector R&D (Wood, 2025; Zimmermann, 2025) means that public appraisal failures are partially compensated by market-driven investment. The polycentric structure identified by Wood (2025) provides additional resilience through state-level innovation initiatives that continue regardless of federal policy shifts. Finally, **the proportionality principle embedded in A-94, which encouraging agencies to tailor analytical depth to the scale and significance of the investment at issue,**

represents a pragmatic concession to the reality that full benefit-cost analysis is neither feasible nor cost-effective for every R&D decision. A-94 states that social benefits and costs, and not only the benefits and costs to the Federal Government, should be the basis for assessing government programs or policies that have effects on private citizens or other levels of government. This **takes into account incremental, intermediate, and international effects, as well as intangible costs, multiplier effects, and environmental costs.** These are principles the UK might usefully adopt more explicitly.

Weaknesses – what are the limitations that cannot address the specific challenges of R&D appraisal

7.13 The most fundamental weakness is that **A-94 contains no R&D-specific provisions.** Despite R&D constituting the primary mechanism through which governments invest in long-term productivity growth, the central appraisal framework treats it identically to building a road or procuring equipment. There is no methodology for estimating knowledge spillovers, no portfolio optimisation framework for managing diversified research investments, no option value methodology for staged research programmes, and no guidance on the distinctive public goods characteristics of basic research.

7.14 Moreover, the priority-setting within a narrative of national security might not appear to be analytically grounded. M-25-34 frames all five priorities within a narrative of national security and American economic power, with explicit references to "making America great" (Vought and Kratsios, 2025). While political direction for publicly funded research is legitimate and necessary, the memorandum provides no evidence base, comparative return analysis, or systematic assessment of whether the identified priorities represent the highest-value use of constrained public R&D resources.

7.15 The appraisal system has not prevented underinvestment in federal R&D. Federal R&D spending has declined from approximately 1.75% of GDP in the 1960s to roughly 0.6% recently (Atkinson, 2020; Zimmermann, 2025). An appraisal framework that permits - or fails to flag - this degree of underinvestment relative to estimated optimal levels has not fulfilled its purpose. Using causal identification from postwar changes in US federal R&D appropriations, Fieldhouse and Mertens (2024) estimate social returns to federal nondefense R&D of 140-210% and conclude that the estimates imply substantial underfunding of nondefense R&D. Their 2025 companion paper reinforces this finding demonstrating that these high social return estimates align broadly with the wider literature on returns to private and total R&D spending, and warning that future fiscal consolidations targeting nondefense R&D could impose larger adverse effects on productivity and economic growth (Fieldhouse and Mertens, 2025).

7.16 **Fragmentation across 23-plus agencies with independent R&D budgets also inhibits cross-portfolio optimisation. No institutional mechanism exists to assess whether the aggregate federal R&D portfolio is optimally allocated across research types, time horizons, or risk profiles.** The annual memorandum provides thematic direction but does not constitute portfolio management. Atkinson (2020) argues that the US has no national, coordinated innovation policy system.

Data and capability requirements

7.17 The US possesses **one of the world's most comprehensive R&D statistical infrastructures,** centred on the National Center for Science and Engineering Statistics (NCSES) within the National Science Foundation. NCSES is one of 13 principal federal statistical agencies, serving as the nation's leading provider of statistical data on the U.S. science and engineering enterprise (NCSES, 2024). NCSES conducts the annual Survey of Federal Funds for R&D, collecting agency-level R&D obligations data from 25 agencies covering approximately 99% of all federally sponsored R&D. Circular A-11 requires agencies to categorise R&D spending by character of work - basic research, applied research, and experimental development - using definitions aligned with the OECD Frascati Manual (OECD, 2015). This classification framework, harmonised with international standards since the OMB's 2016 update, enables international comparisons. The AAAS R&D Budget and Policy Program provides independent analytical capability, synthesising federal budget data with international comparisons.

7.18 The STAR METRICS programme (Science and Technology for America's Reinvestment: Measuring the Effect of Research on Innovation, Competitiveness and Science), launched in 2010 as a joint effort of NIH, NSF, DOE, and OSTP, represents an ambitious attempt to document outcomes and public benefits of federal R&D investment. However, the programme's own description acknowledged that there is little thoroughly conclusive evidence that investment in research produces positive economic effects, illustrating the fundamental difficulty of R&D evaluation (as distinct from appraisal) that constrains the evidence base available for prospective decision-making.

7.19 The analytical capabilities required for rigorous R&D appraisal include spillover estimation through econometric methods, real options pricing models for staged investments, patent citation analysis and bibliometric assessment of knowledge flows, and input-output modelling of R&D's macroeconomic impacts**. A-94 itself recognises the capability gap, stating that "OMB urges agencies to invest in the analytical capabilities to calculate such measures" for certainty-equivalent valuations** (OMB, 2023). This is an acknowledgement that many agencies lack the technical capacity to implement even the general provisions of the circular, let alone any R&D-specific extensions.

7.20 One potential barrier to appraisal in the US is the perception that the market makes better resource allocation decisions than the public sector and that this might help to explain the decline in public R&D investment over recent years. In this respect, the existence of market failures is one that will naturally be sorted by the market mechanism over time without the need for direct government intervention other than in creating a level playing field where firms are able to make their choices about how much R&D to conduct.

7.21 The Foundations for Evidence-Based Policymaking Act of 2018 (P.L. 115-435) established a **statutory framework for building analytical capacity,** requiring each federal agency to designate an Evaluation Officer, a Statistical Official, and a Chief Data Officer, and to publish systematic evidence-building plans identifying policy questions the agency intends to address. This infrastructure supports ex post evaluation rather than ex ante appraisal, but it creates institutional foundations upon which R&D appraisal capabilities could be built. The fragmented agency structure, however, means that **analytical capabilities are developed independently within each agency rather than coordinated across the federal research portfolio.**

Overall assessment

7.22 **Overall, the United States exhibits a disconnect between appraisal capacity and appraisal practice in R&D.** The country possesses a recently modernised general appraisal framework (A-94), the world's most comprehensive R&D statistical infrastructure (NCSES), a deep academic literature quantifying R&D returns, and a Congressional Budget Office capable of sophisticated dynamic scoring of R&D impacts. Yet **the actual process by which federal R&D investments are decided operates through a three-tier system in which none of these analytical resources are systematically deployed.** At the first tier, strategic and political direction through the annual OMB-OSTP memorandum establishes thematic priorities without analytical grounding. At the second tier, scientific peer review at agencies such as NSF and NIH selects individual projects based on scientific merit and feasibility, without reference to economic returns. At the third tier, **formal benefit-cost requirements under A-94 exist in principle but are rarely applied to R&D in practice, because the framework lacks R&D-specific provisions** and the methodological obstacles to monetising basic research benefits are formidable.

7.23 The UK's own appraisal framework (the HM Treasury Green Book) shares certain features with the US system, including a general benefit-cost framework that was developed for conventional public spending but not for promoting inherently uncertain science and innovation (as Lord Willetts observed in his 2024 independent review of the DSIT business case process). The Green Book, like A-94, contains no R&D-specific supplementary guidance.

7.24 As for the implications for the UK's policy appraisal, there are some positive lessons emerging from the US experience. First, **portfolio-level appraisal provides an essential complement to project-level analysis.** The DSIT (2025) assessment that £1 of civil public R&D generates £8 in net economic benefits for the UK and the CBO's finding that federal R&D generates $11.50 per dollar over 30 years in the US[7] establishes a systemic presumption in favour of R&D that project-level benefit-cost ratios cannot capture.

7.25 The US experience demonstrates **the value of differentiated appraisal by research mode.** The institutional diversity of the US system - NSF's peer review, DARPA's empowered programme managers, DOE's mission-oriented laboratory system - implicitly provides different appraisal cultures for different research types. The UK could develop explicit differentiated appraisal requirements in supplementary Green Book guidance, recognising that the appropriate analytical standard for different types of research can differ fundamentally (e.g., applied research and experimental development).

7.26 **The annual strategic priority refresh provided by the OMB-OSTP memorandum has practical value in its directness and regular cadence.** The UK could consider a regular cross-government

[7] Note that this figure is used primarily as a macroeconomic scenario tool, not as an ex ante appraisal instrument. It is used to defend or attack aggregate R&D budget levels in Congressional debate, not to choose between specific programmes. CBO itself frames it as illustrative macroeconomic modelling rather than a decision-support tool for allocating between funding options, and there is no mechanism in the US system that translates this aggregate multiplier into programme-level appraisal.

R&D priority statement, coordinated between HM Treasury and DSIT, that provides clearer strategic direction to UKRI, ARIA, and departmental R&D programmes.

7.27 In addition, **the US move from a 7% to a 3.1% discount rate demonstrates the material significance of this parameter for R&D appraisal.** The independent Green Book discount rate review commissioned by HM Treasury could address whether R&D investments warrant specific discount rate treatment.

7.28 **A-94's uncertainty treatment framework also offers a pragmatic model for handling the radical uncertainty inherent in basic research without either demanding impossible precision or abandoning analytical discipline entirely.** A-94's transparency requirements provide a model for UK R&D business cases that could improve accountability without imposing spurious quantification.

8. Appraisal approach in Australia

Overview of the country

8.1 Australia's R&D profile is characterised by low intensity, with total R&D spend (GERD) at $45,082m and as 1.69% of GDP (Australia Bureau of Statistics 2023-2024). Although this is an increase of 0.03% since 2021-22, the figure is below the OECD average and Australia's position has slipped over the long run relative to other countries. However, during 2023-2024, business expenditure on R&D (BERD) grew strongly to $24bn, up 18% from 2021-22, where 42% was in the research field of information and computing sciences. There are calls for the government to aim for a higher total country investment in R&D closer to 3% as a proportion of GDP (Science and Technology Australia 2023). In 2024-25, government budget allocations for research and development (GRARD) – a funder-based R&D metric reported in the annual Science, Research and Innovation (SRI) budget tables - is estimated at $10.1bn, which comprise budgeted R&D expenditures from departments funds, R&D support from administered funds, and R&D tax incentives (Prior and Brennan, 2025).

8.2 Specifically, Australia's R&D tax incentive (RDTI) provides a refundable offset for smaller businesses and a tiered intensity-linked non-refundable offset for larger companies. This structure is generous compared to support for SMEs globally and aims to enhance innovation even for unprofitable companies. In addition, the R&D support in Australia is shaped by the major, long-horizon industrial policy agenda, focusing on priority industries. Public R&D support is also often offered through mission-style and sectoral funding channels. In terms of the R&D policy direction and reform narratives at the system-level, the strategic examination consultation panel (Commonwealth of Australia, 2025) emphasise the need to shift focus from "R&D as an input" toward an R&D and innovation pipeline that support movement "from discovery to commercialisation at scale" with fewer fragmented programs, clearer national focus, and stronger translation capacity.

8.3 Australia has a mature, internationally connected R&D system with appraisal and evaluation practices that closely align with the UK and Green Book principles. The national innovation agencies in Australia include the Australian Research Council (ARC), the National Health and Medical Research Council (NHMRC) and the Commonwealth Scientific and Industrial Research Organisation (CSIRO), which is the national science agency and innovation catalyst all of which use structured ex ante appraisal, clear logic model-based programme design, as well as impact focused ex post evaluation frameworks. Australia also faces similar strategic challenges around private R&D, improving productivity through innovation, and delivering mission-oriented investments in areas like the UK (e.g., clean energy, health and digital technologies). As in the UK, there are also strong university–industry collaboration cultures.

Description of appraisal approaches

8.4 Overall, the policy appraisal approach in Australia is multi-layered with Cabinet process discipline, whole-of-government impact analysis for major proposals and regulatory impact analysis for substantive rule changes. **Specifically, the Cabinet Handbook (2022) does not prescribe methods like cost-benefit analysis, but it institutionalises how proposals are prepared and**

challenged. In practice, Ministers are expected to ensure timely and thoughtful consultation, so that Cabinet is well informed and well advised, and differences are resolved or clearly surfaced to enable collective decision-making. This is effectively the procedural infrastructure that forces cross-portfolio scrutiny before decisions. This primarily governs policies and major funding programme decisions. Proposals that require new government spending (like a new R&D grant scheme) must go through the Budget process, which involves intense scrutiny by the Expenditure Review Committee (ERC) to ensure the proposal aligns with government priorities and is fiscally responsible.

8.5 The parliamentary library brief (2025) on Australia's Impact Analysis Framework, used to assess the consequences of policy changes, describes a structured impact assessment (IA) process overseen by Office of Impact Analysis: (1) the oversight body determines whether IA is needed; (2) the department prepares IA; (3) the oversight body accesses quality of analysis and process; (4) IA goes to decision-makers; and IAs/assessments are later published. Crucially, IA is framed around seven questions, including explicit requirements to define the problem and data availability, assess options, estimate net benefits, consult and specify ex post evaluation against success metrics. However, the brief also pointed out that IA is selectively triggered, where only less than 3% of the preliminary assessments between 2022-2024 required an IA, and of those IA needed cases, only about 70% were actually completed. This matters for R&D policy assessment, as many innovation interventions can be administratively "small" yet systematically important – so what falls below threshold may escape deeper scrutiny. Where an IA is not done or is inadequate, the framework uses Post-Implementation Reviews (PIRs) as a corrective, which is often required within two to five years for decisions judged to have substantial or widespread economic impact. The detailed Australian Government Guide to policy impact analysis (Commonwealth of Australia, 2023) is even more explicit that an insufficient IA should not be used as a basis for decision-making and triggers a PIR requirement within two years.

8.6 Regulatory Impact Analysis (RIS) is mandatory for substantive regulatory policy change, and Australian Government Guide to RIS (2020) stresses it should be developed early as the starting point of policy development. The guide also addresses that regulation should not be a default and policy makers should examine genuine options (including the non-regulatory ones) and recommend the option with the greatest net benefit. It strongly pushes for early RIS development, positioning the seven RIS questions as the starting point of the policy journey. It also describes a central oversight/support function (OBPR, now renamed the OIA following the 2023 reform) that coaches agencies, supports estimation including CBA and burden measurement, and reports on compliance and post-implementation reviews via a public website.

8.7 Overall, in practice, the framework creates a multi-layered policy appraisal in Australia, where (1) Cabinet process ensuring consultation and collective challenge; (2) IA deep appraisal for major proposals with publication and quality tiers; (3) RIS deep appraisal for substantive regulatory changes, with explicit burden measurement expectations. Note that while this process seems generic it is a comprehensive system that applies to funding programmes, regulations, and broad policies, though the specific "track" used depends on the nature of the proposal. The first layer described here (Cabinet process) is for policies and funding proposals, the second layer (IA) is for major proposal and policies, while the last layer (RIS) is for regulation substantive rule changes. They could all relate to the R&D funding, policy or regulations.

Strengths and limitations in R&D appraisal

Strengths – what works well to address the specific challenges of R&D appraisal

8.8 First of all, **policy appraisal in Australia has strong formal standards and quality assurance.** Across IA and RIS, policy appraisal is not "freeform", and is structured around seven questions, explicitly requiring option comparison, net benefit reasoning, consultation, and a plan for implementation and ex post evaluation. **The system also tries to guard quality through graded tiers from insufficient, to adequate, to good practice, then to exemplary, and uses PIRs when analysis is insufficient**.

8.9 Moreover, **the policy appraisal process has strong transparency** and is useful in parliamentary scrutiny. The IA brief argues IA can support not only executive decision-making but also parliamentary scrutiny as the assessments is published after decisions. For regulations tabled in Parliament, the IA must be included in explanatory materials. In the case of R&D policy appraisal, that design feature matters for R&D policy legitimacy, given that innovation policy often involves uncertain returns and distributional trade-offs, where explanatory materials can be politically stabilising.

Weaknesses – what are the limitations that cannot address the specific challenges of R&D appraisal

8.10 **The IA framework is criticised for underutilisation,** and only a small share of proposals triggers full IA (ANAO, 2024). For R&D policy, this raises a classic risk: many interventions are delivered as multiple small program tweaks (grant rules, eligibility, reporting, procurement settings, tax incentive parameters). Individually they may not hit the IA threshold, but collectively they can reshape incentives and administrative burden.

8.11 Another criticism when applied to the R&D policy is that **the R&D "policy mix" is hard to appraise when the framework is fragmented and compliance heavy.** The Strategic Examination consultation findings (Commonwealth of Australia, 2025) are blunt: stakeholders describe "proliferation of fragmented initiatives" plus "excessive and cumbersome rules and regulations," and argue that complexity is slowing the R&D system and creating a secondary market of consultants and grant writers. This is not just a delivery problem; it's an appraisal problem - because fragmented instruments make it difficult to attribute outcomes and to run clean assessments, and heavy compliance costs can reduce true additionality. The argument is that a dense mix of R&D funding and policy instruments can make it difficult to attribute impact, which influences the validity of assumptions used in appraisal. The stakeholders at the Strategic Examination consultation also argue that the current funding models, incentives and research metrics can favour "safe, incremental research," discouraging bold experimentation and entrepreneurship. **That critique points to a deeper appraisal challenge: conventional "net benefit" calculations and short-term performance metrics can undervalue high-risk, long-horizon R&D portfolios unless appraisal frameworks explicitly adjust for uncertainty, option value, and spillovers.**

8.12 On the appraisal machinery, the combination of IA/RIS questions, oversight, quality tiers, publication norms, and PIR backstops is a robust design aimed at improving investment and regulatory policy decisions and accountability. However, the Strategic Examination consultation responses suggest the binding constraint for R&D policy effectiveness may increasingly be fragmentation and instability in the policy mix; regulatory and administrative burden; incentives that reward outputs/inputs over value and adoption; weak scaling and commercialisation pathways; risk aversion in public systems (including procurement), reducing adoption of new-to-market innovation.

Data and capability requirements

8.13 **In general, the macro or system data foundation for R&D seems to be well established.** The Parliamentary Library update emphasises routine system metrics (GERD; government R&D support) and references government efforts to compile innovation statistics and dashboards. The ABS business R&D publication provides a granular industry distribution of BERD and other structural indicators. Overall, the macro-level R&D statistics released by the Australian Bureau of Statistics provide a solid monitoring base.

8.14 However, there is active debate on how to measure what actually matters for RD&I outcomes. Specifically, the reform agenda pushes toward outcome-focus measurement. The Strategic Examination findings (Commonwealth of Australia, 2025) explicitly call out that the system "needs to measure what matters - even when it's hard," warning against measuring only what is easy and comfortable. It argues for a greater focus on outcomes (including economic growth) rather than inputs. Such outcome-focused measurement requires additional data and capabilities beyond the macro statistics like GERD and BERD. For instance, pipeline tracking from discovery through adoption not just publication/patents are needed. Additional commercialisation and scale metrics of R&D, which could link to productivity and value capture, could be provided. R&D tax incentives also require data on administrative burden and compliance-cost measurement. While these metrics may only seem relevant to ex post analysis, availability of more sensitive to a variety of costs is also important to feed into ex ante decision making.

8.15 The Strategic Examination emphasises that STEM skills need to be complemented by "commercial acumen, investment analysis, business model innovation," and go-to-market capability, implying that R&D policy success depends on skills embedded across the ecosystem (government, universities, business, investors). **In appraisal terms, this suggests assessment teams need capacity to judge not only scientific excellence but also the strength of translation pathways, market creation, and the role of procurement and regulation in adoption.** This reinforces a theme that has emerged in other cases that emphasises that the effectiveness of an appraisal system is not only tied to process and methodology but, crucially, also to data quality and availability as well as assessors with the skills to effectively seek out, analyse, and interpret appropriate data to inform appraisals.

Overall assessment

8.16 From the review of Australia's policy guidance documents, Australia looks like a country where the formal machinery of appraisal is relatively strong, but the R&D system itself is signalling that appraisal and measurement need to evolve to support a more mission- and impact-driven R&D

pipeline. Overall, Australia has the scaffolding for rigorous appraisal but needs to apply it in a more system-wide, outcomes-oriented way that can reduce complexity, sharpen priorities, and measure impact across the full R&D and innovation pipeline.

8.17 Specifically, the UK context shares similarities to Australia's IA framework with less coverage and weak cumulative effects. The UK also has strong appraisal guidance, but many incremental policy changes (eligibility tweaks, reporting requirements, program reshaping) can accumulate into large burdens and distorted incentives without ever being treated as a single major reform. **A transferable lesson is to complement "major proposal appraisal" with portfolio-level appraisal - periodic synthesis that looks across the entire R&D support mix.** The Strategic Examination consultation findings (Commonwealth of Australia, 2025) specifically suggest the policy to move beyond individual project appraisals and look at the entire R&D support mix to address fragmentation and declining R&D intensity. A pertinent comment from the Strategic Examination of the R&D system was that "there are too many programs, spread too thinly" and that "strategies formulated by federal, state and territories governments should be coherent, related, synergistic".

8.18 Australia's consultation findings repeatedly link innovation underperformance to cumbersome rules and a compliance-heavy environment, including around the R&D Tax Incentive and fragmented funding navigation. The UK has comparable debates (e.g., administrative costs of grant systems, innovation finance access, complexity of support landscape). **Australia's Strategic Examination calls for measuring what matters (outcomes, economic growth) rather than what's easy (inputs)**. This is highly transferable to the UK context, where there is also pressure to show productivity and growth impact. The caution is that the Strategic Examination simultaneously stresses that high-quality basic research is a cornerstone that must be sustained. The insight for the UK would be to build appraisal frameworks that separate (a) frontier research quality and capability building from (b) translation/adoption impact - so impact demands do not unintentionally bias funding toward short-term, low-risk projects.

8.19 Australia's stakeholders argue that current models and metrics favour safe, incremental work and reflect a low appetite for risk and fear of failure. The UK's innovation agencies also face the same conceptual issue: if appraisal is overly "audit-like," it discourages high-variance bets. Therefore, the UK policy appraisal needs to accommodate risk and long-time horizons in innovation policy. The Australia consultation review calls for a "healthy appetite for risk" and long-term investment models, arguing that incrementalism has failed to address structural economic shifts. It advocates for shifting away from short-term grant cycles toward predictable, patient capital, and reforming the R&D Tax Incentive to better support high-risk projects. Australia's consultation notes suggest government purchasing power could be used more strategically, and that risk aversion in procurement disadvantages new-to-market innovation. This is directly transferable to the UK, where public procurement is often discussed as a lever for innovation diffusion. The lesson is to appraise R&D policy not just as "funding research," but as a full set of levers including procurement, regulation, and market formation.

9. Appraisal approach in Belgium

Overview of the country

9.1 Belgium boasts one of the EU's highest R&D intensities, with investments reaching 3.43% of GDP by 2022. Belgium is classified as a "Strong Innovator" in the European Innovation Scoreboard (2025), with performance well above the EU average on many R&D and innovation indicators. It ranks first in business sector and innovation expenditures per person employed, with very high scores for direct/indirect government support of business R&D. Collaboration within the innovation system is also identified as a relative strength, supported by policy instruments such as competitiveness clusters and strategic applied research centres (often at regional level). The EIS profile highlights strong performance in public-private co-publications and innovative SMEs collaborating with others. However, Belgian firms, as those in many other countries, are not always able to translate innovation through to market impact, although the building blocks are in place. It may simply be a timing issue as market penetration of new innovative products and services can take a number of years to manifest themselves at the firm level.

9.2 Belgium's governance of R&D and innovation policy is shaped by its federal structure, where competences are divided between federal and federated (regional/community) levels. This creates constraints for policy evaluation practices which are harder to implement consistently across different levels of government (OECD, 2023). R&D policy is highly decentralised, split between federal and regional governments (Flanders, Wallonia, Brussels). Federal policy primarily focuses on tax-based incentives including 'innovative income deduction', partial withholding tax exemption for R&D staff, and R&D tax credits, while regions drive direct funding for research projects.

9.3 Belgium works as a comparator for the UK because it shares structural similarities in its R&D system. Belgium uses comparable appraisal and evaluation approaches and operates within a European policy environment that emphasises evidence-based investment. Belgium's R&D is significantly above the EU average and reflects the UK's ambition for higher-intensity R&D investment. In addition, Belgium uses a) ex ante appraisal; b) counterfactual evaluation; and c) logic-model-based programme design.

Description of appraisal approaches

9.4 **Belgium does not have a single, dedicated R&D appraisal framework**. Instead, R&D policy appraisal is embedded within the general public policy appraisal and evaluation systems. For instance, the Regulatory Impact Assessment (RIA/AIR) which is mandatory since 2013 for drafting legislation, ministerial decrees submitted to the Council of Ministers (OECD, 2023), structured around 21 themes with emphasis including SMEs, gender, administrative burdens, etc. **Instruments such as RIAs are required for all legislation, including spending allocations, and so these guidelines are de facto part of the appraisal process.** The implementation, management and reporting of RIA is carried out by the Impact Assessment Committee (IAC). The current RIA framework was introduced to assess specific economic, social and environmental impacts of public policy and this also applies to R&D related incentives (OECD, 2025). **In practice, RIAs are often**

described as largely qualitative, conducted late in the process and rarely using more advanced techniques such as cost-benefit analysis or cost-effectiveness analysis.

9.5 There are also ex post reviews of regulations, which are intended to assess whether such regulations remain effective and justified. The Policy Evaluation Framework (ex ante and ex post) **focuses on assessing the relevance, effectiveness, efficiency and impact**, coordinated partially through 'FPS BOSA' (Directorate-General for Budget and Policy Evaluations). Spending Reviews, also increasingly used since 2020, which links ex post evaluation evidence to budgetary decision-making. The logic is simple in that if an ex post evaluation shows that a scheme generated good returns for the taxpayer, then the spending review would take note of this in its allocations for future funding. Finally, the Evidence-Informed Policymaking (EIPM) ecosystem relies on administrative data, research organisations, advisory bodies, and academia to assess public policy in general (OECD, 2025).

9.6 In practice, existing appraisal tools apply to public policies and spending in general, including R&D measures, but they are not systematically embedded upstream in R&D policy design and not tailored to R&D specific programme evaluation. First, coalition agreement sets policy direction and R&D priorities are often fixed early for consensus reasons. Then, with the evidence requested from internal analysis, advising councils, federal planning bureau or academia, policies are developed within ministries. The ex ante regulatory impact assessment is often conducted late in the process and perceived as a box-ticking exercise with rather limited influence on design. Due to the limited coordination across ministries, the ex post evaluation is often conducted selectively, which offers limited feedback into new policy.

Strengths and limitations in R&D appraisal

Strengths – what works well to address the specific challenges of R&D appraisal

9.7 Overall, the appraisal tools exist and are embedded in both the regulatory process and budget process (spending reviews), both of which influence spending decisions, which signal an **evidence-informed policy making orientation**. Policy appraisal in general has strong legal and institutional basis with mandatory RIA and formal recognition of policy evaluation (both ex ante and ex post). The strong academic system, competent analytical units in several FPS and rich administrative datasets provide a good supply of evidence.

9.8 **Policy appraisal in practice also links evaluation to government spending (including R&D) which incentivises VfM thinking.** Belgium has high business R&D intensity relative to its EU peers and strong collaboration indicators, which provide useful foundations for results-based R&D policy appraisal.

Weaknesses – what are the limitations that cannot address the specific challenges of R&D appraisal

9.9 Belgium's public policy appraisal and evaluation system has been criticised for the limited use of hard data, comparative analysis of alternatives or advanced methods like cost-benefit analysis or cost-effectiveness analysis, and lack of mechanism to ensure proportionality of analysis (OECD, 2025).

While Belgium has rich administrative and statistical data, access to linked datasets for research and evaluation is difficult which has limited advanced impact analysis.

9.10 Across ministries, few administrations can conduct robust policy assessments. OECD notes several **resourcing constraints inside the coordination bodies**, which undermines the ability to coordinate and uplift evaluation across government. The ministries are not directed to clearly define the problem, develop SMART objectives or compare options. Overall, policy appraisal in Belgium is reportedly limited by the **weak coordination across ministries and policy domains**. There is a lack of strong central structure ensuring consistency or learning across sectors.

Data and capability requirements

9.11 There are several systemic needs relevant to the R&D appraisal approach in Belgium relating to both data and the skills and capabilities. As noted above, although Belgium has rich administrative and statistical data, there are challenges of accessing the data and linking datasets for research and assessment. This has limited the use of advanced impact analysis. As pointed out in OECD (2025b), OECD OURData results indicate **Belgium is below the OECD average for data accessibility and government support for reuse**. There are issues including limited access to raw survey microdata and challenges linking datasets. Under the current structure with multiple data integrators (federal and regional), it creates access frictions. Therefore, to inform evidence-based appraisal and improve the current framework, clearer, more transparent data access and linkage would be needed. In addition, the EIPM diagnostic (OECD, 2025b) describes interoperability and resource issues around staffing also limited Belgium's ability to take full advantage of its own data for analysis and policymaking.

Overall assessment

9.12 Overall, Belgium shows very strong innovation system performance and investment indicators (especially business R&D and innovation spending), yet **the appraisal system is relatively fragmented and constrained** by late/formalistic RIA, limited ex post learning inside the executive, and capacity and data constraints.

9.13 The UK has a relatively more centralised government structure and a longer tradition of formal appraisal (e.g., Green Book), which structures policy design, appraisal, monitoring, and evaluation through common methodologies such as cost–benefit analysis and formal assessment planning. This provides consistency and clarity, particularly for large and complex R&D investments, but can be resource-intensive and procedurally demanding. Belgium, by contrast, does not operate a single, dedicated R&D appraisal framework and instead assesses policies through a combination of Regulatory Impact Assessments, policy assessments, spending reviews, and expert advice. This approach offers more flexibility and allows appraisal to adapt to political and institutional contexts but can result in uneven application and weaker integration across the policy cycle. While the UK emphasises standardisation and analytical depth, Belgium's system reflects a more decentralised and consensus-oriented model, with appraisal practices shaped by coordination across multiple institutions. Thus, one of the differences between Belgium and the UK is that the former seeks to

establish wider political consensus across agencies and institutions, and the latter is more informed by experts and structured evidence.

9.14 The review on Belgium highlighted that policy appraisal tools need constant reinforcement of analytical culture, early-stage appraisal and incentives for departments to use evidence to shape not just justify R&D policy decisions. It also highlighted that coordination across government is critical. It generates a consensus early in the process, rather than policy being developed within individual departments and excluding others that also have an interest in this policy area. However, the innovation system itself is performing well and investment is strong, which might suggest that with a more consistent and comprehensive appraisal approach and more funding allocated to evaluation, policymakers would be better informed about what is driving the strong performance of the Belgian innovation system.

10. Appraisal approach in Sweden

Overview of the country

10.1 According to the European Innovation Scoreboard (EIS) 2025 country profile, Sweden is a highly R&D-intensive, networked innovation economy, an "Innovation Leader" ranked first in the EU with a summary innovation index of 138.1 and a very strong business R&D intensity, public-private collaboration, and patenting. In the Swedish government's latest research and innovation bill in 2024/25, Sweden reports total R&D expenditure of SEK 223,626m in 2023, corresponding to 3.6% of GDP and exceeding the EU-wide 3% target, where 74% of R&D was performed by the business sector, 22% by higher education sector and 4% by other sectors including government and private non-profit (Swedish Government, 2024a). However, there are also notable weaknesses in the Swedish R&D profile. Specifically, job-to-job mobility of human resources in science and technology has declined sharply, and it shows poor resource productivity and lower share of highly cited publications, relatively low participation rate in Horizon Europe (European Commission, 2025). These trends are central to the current government R&D strategy plans. At the policy level, the Swedish government's latest research and innovation bill sets the direction for 2025-2028, anchored in "excellence, internationalisation and innovation" and the ambition that Sweden be among the world's leading R&D nations (Swedish Government, 2024a). Specifically, it proposes a cumulative increase of SEK 6.5bn in 2025-2028 for research and innovation.

10.2 **As for the appraisal framework, it comprises several interlocking institutional and regulatory layers.** First, the government sets strategic directions and budgetary framework through Research and Innovation Bills issued approximately every four years. Research and Innovation Bills are strategically tight but operationally flexible. While the government defines the "destination" by earmarking large sums for specific strategic areas and infrastructure, it delegates the final project-level allocation to independent funding agencies and universities. After the bills have been passed, research and innovation funding is distributed through five principle agencies, each with significant operational autonomy with Vinnova (the Swedish Innovation Agency) as the primary vehicle for strategic and challenge-driven programmes; Vetenskapsrådet (Swedish Research Council) focusing on investigator-initiated basic research evaluated through peer review; and Formas, Forte, and the Swedish Energy Agency addressing sector-specific research domains. Each agency incorporates ex ante impact assessment to comply with formal regulatory requirements and designs its own ex post evaluation structure with significant methodological autonomy. The agencies are legally autonomous administrative authorities. Their individual project funding decisions are internal administrative decisions, made independently based on their own mandates, criteria, and programme frameworks. They do not submit project-level appraisals to central government for approval, and the government does not direct or approve individual funding decisions. Central government accountability runs through agency annual reports and parliamentary oversight of the agencies' aggregate performance, not through project-by-project appraisal submitted upward.

10.3 In general, the formal regulatory scaffolding for ex ante appraisal was updated in 2024 the Ordinance (2024:183) on Impact Assessment (Swedish Government, 2024b), which **requires all government agencies to prepare and document impact assessment** before deciding on regulations or

submitting proposals for new legislation and funding programmes. **It requires that major programme decisions include a problem analysis, consequences of inaction, alternative options, expected costs and revenue, compliance costs, and critically a specification of how and when consequences will be evaluated**. As for the methodological guidance, Ekonomistyrningsverket (ESV), the State Treasury, has been tasked with developing a comprehensive framework by April 2026. The accompanying governance regulation handbook (Statsrådsberedningen, 2025) provides operational guidance for agencies involved in regulatory drafting and explicitly references compliance with the 2024 impact assessment ordinance. Together, the Ordinance and the governance handbook constitute Sweden's formal ex ante appraisal framework, which is then supplemented at programme level by specific agencies.

10.4 For the UK, Sweden offers a good comparative case. First, the Magenta Book and HM Treasury's Green Book frame UK appraisal also face challenges between theory-based approaches suited to complex system interventions and quantitative VfM assessment demanded by the treasury. Sweden's R&D excellence makes it a suitable high-quality benchmark with insights likely to be relevant and transferable to the UK. Sweden's experience with the Strategic Innovation Programmes (SIPs), the Meta-evaluation (ex post) of those programmes, and the subsequent[8] shift toward excellence cluster each offer transferable lessons for UK programmes seeking to balance transformative ambitions, accountability and adaptive governance.

Description of appraisal approaches

10.5 Overall, Sweden's R&D/innovation policy appraisal framework appears to be a stack of interlocking institutional and regulatory layers with (a) the Research and Innovation Bill cycle; (b) Vinnova's Impact Logic Model; (c) the Strategic Innovation Programmes (SIPs) evaluation cycle; (d) Vetenskapsrådet's Peer Review.

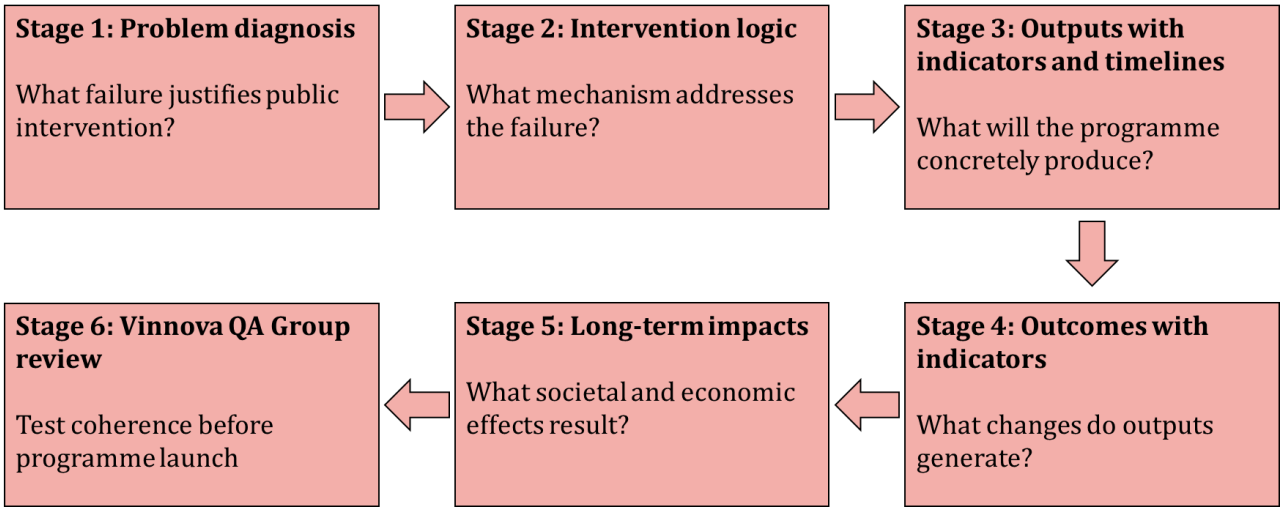
10.6 In practice, the Research and Innovation Bill cycle (issued approximately every four years) represents Sweden's primary strategic appraisal mechanism, which draws on commissioned analyses from funding agencies, university sector inputs and international benchmarking to set priorities and the budget framework. For instance, the latest 2024/25 Bill (Swedish Government, 2024a) references the report on European competitiveness (Draghi, 2024), OECD data on declining citation impact and Horizon Europe performance, and European Innovation Scoreboard ranking to justify its investment programme. It also establishes parameters within which agencies design their own ex ante assessment and ex post evaluation activities but does not prescribe specific methodologies. In the current 2025-2028 Innovation Bill cycle, spending is planned to receive an SEK 6.5bn (£530m) boost (OECD STIP Compass, 2025).

10.7 The Swedish Innovation Agency, Vinnova, published its foundational methodology document VINNOVA Analysis VA 2008:01, which focuses on impact and rests on a four-part integrated quality assurance model (Vinnova, 2008). **Specifically, the impact logic assessment is conducted ex ante, designing and testing the expected causal chain from input through activities, outputs and**

[8] The Excellenskluster (excellence clusters) are specified in the 2024 Research and Innovation Bill as a new joint programme instrument co-owned by Vetenskapsrådet and Vinnova with two tracks. The Vetenskapsrådet track is defined as an open competition for future technology areas and the Vinnova track focuses on strategic technology areas.

outcomes to long-term impacts before the launch of the programme. In practice, the impact logic diagram (see below, Figure 10-1) involves: (1) specifying the problem diagnosis – what market failure, system failure or coordination failure justifies public intervention? (2) mapping the intervention logic – what specific activities will address that failure, through what mechanism? (3) specifying outputs with indicators and timelines – what will the programme produce, by when and how will this be measured? (4) specifying intermediate outcomes with indicators – what behavioural and structural changes do those outputs need to generate? and (5) specifying long-term impacts.

Figure 10-1: Vinnova's impact logic assessment



10.8 One of the notable elements of this model is its focus on additionality. Vinnova's impact logic model emphasises three types of additionalities:

- Input additionality means that companies decide to invest more in R&D because of public support than they otherwise would do.
- Output additionality is related to outcomes from R&D spending. The output may be an increased number of scientific publications, more patents, more prototypes, new business plans and new goods or services. Impacts can include increased turnover, an increase in value added, more jobs, increased productivity and increased exports.
- Behavioural additionality focuses on shifting firm behaviour through public support, such as increased R&D investment, new partnerships, or enhanced capability for innovation. Behavioural additionality is considered the most important but the hardest to measure.

10.9 By exploring multiple vectors of additionality, Vinnova builds in a different perspective on the expected economic benefits of public investment in R&D than other appraisal methods summarised so far.

10.10 In addition, there has been recent update and development of the above Vinnova's impact logic model since Vinnova joined in the Transformative Innovation Policy Consortium9 around 2016-2017. The

[9] The Transformative Innovation Policy Consortium is international collaborative research and practice network coordinated by the Science Policy Research Unit (SPRU) at the University of Sussex, involving innovation agencies and research institutes from Sweden, Colombia, Norway, Finland, South Africa and other countries. https://tipconsortium.net/

engagement has involved joint research projects, peer learning exchanges with other member agencies and co-designed policy experiments in which Vinnova ran real programme interventions as a structured experiment with built-in evaluation from the outset. **The most concrete influence on ex ante appraisal is the progressive incorporation of Theory of Change (TOC) into Vinnova's programme design requirements since 2019** (Rohracher, Coenen and Kordas, 2023). Before this, Vinnova's impact logic framework was essentially a structured logic model: a forward-mapping tool specifying the intended causal chain from inputs through to impacts but not the backward-mapping from desired impact that a full TOC demands. In practice this has meant that for certain programmes (e.g., particularly the Challenge-Driven Innovation programme and the mission experiments), Vinnova requires applicants and programme designers to articulate not just what a programme would produce, but what theory of system change it was testing, what assumptions about actor behaviour, institutional dynamics, and market formation underpinned that theory, and how the programme design would generate evidence to confirm or challenge those assumptions.

10.11 While this appraisal method firmly centres impact goals in relation to nationally developed innovation strategies (such as those elaborated in the Research and Innovation Bill cycle), it also aims to assess cases against their contribution to other EU goals. For example, for its Challenge-Driven Innovation programme, Vinnova requires applicants to develop an impact logic mapping project goals, method, and metrics against the UN's 2023 Agenda SDGs, using Nesta's 'Standards of Evidence' as the methodological reference (Vinnova, 2025). This implies a formative, learning-oriented programme design approach beyond just monitoring outputs.

10.12 The Strategic Innovation Programme (SIP) evaluation cycle (with mandatory three-, six- and nine-year reviews) is Sweden's most developed and extensively documented appraisal approach for complex, collaborative innovation programmes. The seventeen SIPs (jointly funded by Vinnova, the Swedish Energy Agency and Formas) are subject to mandatory on-going evaluations every three years, with broader six-year evaluation serving as the primary gateway for continued funding (Vinnova, 2025). Here, **the ongoing evaluation will inform subsequent appraisal which comes at the funding continuation gate decisions**. The mandatory three-year evaluations do function as structured ex ante appraisal for the next phase, and they produce action plans and recommendations that explicitly condition what a programme must do in its next stage to receive continued funding. In this narrow sense, the evaluation cycle does perform an ex ante function which resets the programme's objectives and indicators before next funding period begins. A meta-evaluation on all seventeen six-year SIP evaluations employed a mixed methods design including document studies, self-evaluations, registry analyses, funding analyses, social network analyses, bibliometric analyses, in-depth interviews, web surveys and expert assessment (Vinnova, 2024). The general conclusion was that the instrument had proven effective for building national capacity and international competitiveness with strong results in R&D collaboration, network formation, and knowledge creation, but mixed evidence in contributing to socio-technical system transitions. The meta-evaluation (Vinnova, 2024) also informed the 2024 Research and Innovation Bill's decision not to open new SIP rounds and to replace the instrument with excellence clusters. This shows that evaluation framework in Sweden is essential and did shape the ex ante design rationale of the successor instrument.

10.13 **As for Vetenskapsrådet (Swedish Research Council), it employs international peer review as its appraisal approach.** For national-level research quality evaluation, the Council employs panels of international experts assessing a combination of bibliometric data, selected publications, national statistics, and case studies, rated on a five-point quality and impact scale (Vetenskapsrådet,2025). These national-level research quality evaluations are field-level assessments of the state of a whole discipline in Sweden, and their primary purpose is strategic intelligence for the research system. This provides Vetenskapsrådet, the government and the broader research community with a rigorous international benchmarked picture of where Swedish research is strong, where it is losing ground and where capacity needs to be built or sustained. The findings from these evaluations feed into the ex ante policy design at the national level. For resource allocation, a quality-based model was proposed in 2022 and subsequently evolved in the 2024/25 Bill into a competitive model rewarding institutions on the basis of top-10% cited publications, success in European competitive funding and doctoral research outputs (Swedish Government, 2024a). This quality-based allocation model is specifically about directing government grants to universities and higher education institutions as a block, giving universities discretion over how to deploy the funding internally.

10.14 For the regulatory impact assessment, The Ordinance for Impact Assessment (2024:183) mandates that an impact assessment includes the problem and desired change, consequences under "no action", alternatives and pro/cons, and justification of the preferred option. Specifically, the impact assessment must include: (1) description and calculations of costs and revenues for the state, municipalities, regions, companies and individuals; (2) description and if possible calculations of other relevant consequences; (3) measures taken to ensure proportionality avoid unnecessary cost/restriction; (4) assessment of entry-into-force timing and information needs; and (5) a description of how and when consequences can be evaluated. The requirement on quantifying costs, revenues implies that cost-benefit and cost-effectiveness type of appraisal methods are emphasised in Swedish appraisal practice.

Strengths and limitations in R&D appraisal

Strengths – what works well to address the specific challenges of R&D appraisal

10.15 **The core strength of the Swedish appraisal framework rests on the methodological sophistication,** and it has been at the international forefront for complex innovation programmes. Vinnova's active engagement with the Transformative Innovation Policy (TIP) Consortium and associated working groups has placed it at the frontier of experimental appraisal approaches (Schwaag Serger et al., 2023). Within the Swedish context and applied to Vinnova's BioInnovation SIP, Haddad and Bergek (2023) demonstrate how the integrated TIP evaluation (equivalent to appraisal) framework can be operationalised. The framework involves three components: an ex ante Technology Innovation System (TIS) assessment to define the initial system state; a TOC specifying the causal pathways through which the programme should contribute to system change; and a Realist Evaluation (RE) mechanism for understanding what works for whom under which conditions. Developed by Pawson and Tilley, [10] it focuses on Context-Mechanism-Outcome (CMO) configurations to explain how, and why, interventions produce intended or unintended results within complex

[10] See: https://assets.publishing.service.gov.uk/media/60f7fdf7d3bf7f56824cc634/Brief_introduction_to_realist_evaluation.pdf

systems. Haddad, et al. (2024) deploying this approach across three Vinnova programmes, produced the most empirically grounded cross-case comparison of transformative innovation policy evaluation in the international literature.

10.16 Another strength relates to the formal regulatory appraisal requirements. The Ordinance for Impact Assessment (Swedish Government, 2024b) creates **a statutory requirement to plan ex post evaluation at the point of programme design for all regulatory decisions**. The governance handbook (Statsrådsberedningen, 2025) further reinforces this by embedding consultation and Regulatory Council scrutiny requirements into the legislative process. The institutional evaluation cycles (the mandatory three/six/nine-year SIP evaluation cycle) also create a systematic accountability in which ex post evaluation findings have genuine consequences for programme continuation.

10.17 **Sweden's capacity is also underpinned by high-quality data and statistic infrastructure,** with Statistics Sweden providing comprehensive R&D expenditure data by sector and performer, the SwePub bibliometric database enabling national publication and citation tracking, Vetenskapsrådet's biennial Swedish Research Barometer benchmarking national R&D performance against international comparators, and the Community Innovation Survey proving firm-level innovation data. Such strong data assets support the multi-method SIP evaluation approach and the quality-allocation indicator used in the university funding model. As above, **this is relevant because this data ultimately feeds into the assumptions that underpin appraisal cost benefit estimates.**

Weaknesses – what are the limitations that cannot address the specific challenges of R&D appraisal

10.18 The Swedish National Audit Office's 2020 examination, RIR 2020:6 (Riksrevisionen, 2020), identified serious weaknesses in the SIP appraisal approach. First, many SIPs lacked clearly expressed expectations and were deficient in terms of impact logic despite Vinnova's formal framework. Second, conditions for review were weak - programmes were evaluated on their own merits but not compared with each other, undermining portfolio-level accountability. Third, agencies followed evaluators' recommendations to continue funding without documenting their own considerations, creating an accountability gap between the ex post evaluation evidence, appraisal and the funding decision. This overreliance on external assessments means that administrative considerations rarely figure into decisions to continue R&D funding programmes. Fourth, no systematic assessments had been made of the programme portfolio as a whole - the appropriate number of programmes, their strategic complementarities, or portfolio design relative to government objectives. Fifth, external experts contracted by agencies exerted major influence over programme content without agencies systematically documenting independent assessments before decisions. Sixth, the government's reporting of ex post evaluation conclusions to the Riksdag was found to be unclear and incomplete, with overall conclusions dispersed across agency annual reports rather than consolidated in parliamentary reporting (Riksrevisionen, 2020). These findings reveal a persistent gap between the formal apparatus of assessment and genuine accountability.

10.19 Despite the strength and strong R&D profile in Sweden, a weakness appears in the framework is the gap between mission/transformative policy ambitions and traditional ex post evaluation practices.

Despite changes to appraisal and evaluation in Sweden, it remains mostly summative and embedded in established accountability structures, which can create tensions while not necessarily delivering the intended change. Similarly, another study looking at Sweden's innovation programme evaluations found that even where transformative innovation policy principles are adopted, **appraisal strategies tend to rely on logic models rather than fully articulated transformative Theory of Change (TOCs) that specify causal mechanisms and contextual influences,** limiting the ability to assess system change and unintended effects (Haddad, et al., 2024). Moreover, in the funding policy appraisal, the risk of narrow metrics is acknowledged by the Swedish Government: it notes **the challenge of measuring quality and that indicator-based allocations capture some quality dimensions but not others.**

10.20 A significant gap in Swedish innovation policy evaluation is the limited capacity for counterfactual assessment and relies on within cohort change analysis to inform appraisal. Unlike the UK, where Innovate UK links programme data with ONS Business Structure Database, Companies House, and HMRC administrative data to construct matched comparison groups, Sweden has conducted very few counterfactual impact ex post evaluations of innovation programmes. This is a structural limitation noted by OECD (2023b) in its recommendations on optimising public research funding, which explicitly called for investment in register-linked evaluation datasets.

10.21 The OECD evidence indicates that strategic or targeted funding can also struggle to shift actual institutional behaviour without aligned incentives and governance. It shows that SFO design produced limited industry involvement in practice and that the scheme's "more-of-the-same" character was influenced by consultation and context, reinforcing conservation tendencies (OECD, 2016). Evidence also suggests that the quality of ex ante appraisals is uneven in practice. Several Swedish Regulatory Council opinions conclude that impact assessments do not meet the requirements of The Ordinance (2024:183) framework, citing missing or insufficient analysis for affected firms (Swedish Government, 2024).

Data and capability requirements

10.22 **Effective appraisal of Sweden R&D innovation policy requires strong quantitative data across multiple dimensions.** At the input level, Statistics Sweden's R&D surveys - covering gross expenditure on R&D by performer and source of funds, aligned with the OECD Frascati Manual - provide internationally comparable expenditure data updated annually (Statistics Sweden, 2025). These are supplemented by Vetenskapsrådet's Swedish Research Barometer, which tracks bibliometric indicators including citation impact, publication counts, field-normalised indicators, ERC applications, and researcher career pipeline metrics (Vetenskapsrådet, 2023). Patent data from EPO and PCT registers feed the European Innovation Scoreboard indicators, and the Community Innovation Survey provides firm-level data on innovation activities and collaboration patterns enabling some econometric analysis.

10.23 The specific requirements listed in the impact assessment Ordinance (2024:183) demand baseline data and counterfactual logic, distributional information and abilities to model compliance/administrative burdens and broader effects. The Swedish Government Regulation handbook (Statsrådsberedningen, 2025) also emphasises that early consideration of what

competence is needed (legal, economic calculations, statistical procession), and notes that agencies may need support from other agencies or external contractors for specialised analyses. The Government assignment to Swedish Financial Management Authority for guidance development also formalises this capability logic, specifically stating that assessing consequences would require specialised expertise and methodological support and require ESV to draw on a wide range of domain agencies to cover different consequence categories.

10.24 Assessing the transformative dimensions of Swedish innovation policy requires qualitatively different data from standard R&D statistics. Technology Innovation Systems (TIS) analysis — as applied in Haddad and Bergek's (2023) integrated framework — requires actor landscape mapping, institutional characterisation, network structure analysis, and assessment of system functions including knowledge development, market formation, mobilisation of resources, and creation of legitimacy. Collecting and maintaining this data requires dedicated research effort rather than routine administrative data collection.

10.25 The SIP meta-evaluation (Vinnova, 2024) addressed this through social network analysis mapping collaboration linkages across 2,700 unique actors, combined with expert assessments and stakeholder interviews. However, this data was collected retrospectively for ex post evaluation purposes rather than being maintained as a continuous monitoring asset. A critical capability gap identified by Rohracher, Coenen and Kordas (2023) is that system-level monitoring infrastructure — tracking actor network evolution, technology readiness trajectories, and regime change indicators over the twelve-year SIP lifecycle — has not been institutionalised within agencies' regular data architectures. It is therefore dependent on ex post evaluation-specific data collection that is expensive, inconsistent, and difficult to compare across programmes and use to inform appraisals.

Overall assessment

10.26 Sweden's R&D and innovation policy appraisal framework operates **a decentralised multi-agency architecture with significant methodological autonomy**, underpinned by a strengthened statutory ex ante impact assessment regime (Swedish Government, 2024b) and procedural requirements codified in the governance handbook (Statsrådsberedningen, 2025). It is also in a state of constant evolution as **each policy cycle aims to improve on the linkages between appraisal and accountability**. At programme level, the SIP mandatory independent ex post evaluation cycle with genuine funding consequences provides a structured accountability rhythm. The 2024 SIP meta-evaluation (Vinnova, 2024) is exemplary in its scope, methodological transparency, and portfolio-level ambition. Against these strengths, the framework faces structural challenges. The Riksrevisionen (2020) audit documented **persistent accountability deficits**: weak documentation of agency decision-making, absence of portfolio-level assessment, and inadequate parliamentary reporting. The theory–practice gap identified by Rohracher, Coenen and Kordas (2023) and Haddad, Wise and Arnold (2024) means that Sweden's transformative innovation policy ambitions consistently outrun the evaluation frameworks deployed to assess them. In this respect, improved frameworks would support better appraisal, although, due to wider uncertainty, may not fully resolve the appraisal challenges. The Ordinance (2024:183) (Swedish Government, 2024b), while a significant step, applies primarily to regulatory and legislative decisions and does not yet reach the programme design choices through which most strategic decisions are made. And the 2024/25:60

Bill's shift to excellence clusters (Swedish Government, 2024a) creates a successor instrument without a specified appraisal/evaluation architecture, risking the replication of early SIP evaluation gaps (OECD, 2016).

10.27 As for the transferable implications for the UK, Sweden's experience offers the UK two structural lessons that go beyond methodology:

- First, a sophisticated formal appraisal architecture - whether Sweden's statutory impact assessment ordinance or the UK's Green Book - only generates genuine accountability if agencies are required to document their own independent assessments and act visibly on ex post evaluation findings. The Riksrevisionen (2020) found that Vinnova routinely followed evaluators' recommendations without recording its own reasoning, and that no portfolio-level assessment existed across the SIP programmes. The UK faces an analogous risk: UKRI's evaluation strategy is well-designed on paper, but the NAO (2025) has similarly noted gaps in thematic and portfolio-level evaluation. More specifically, the UK generally focuses on single policy appraisal rather than looking across a portfolio of schemes and programmes simultaneously. Sweden's mandatory three/six/nine-year SIP evaluation cycle, with funding continuation explicitly contingent on evaluation outcomes, offers a replicable model for UKRI's longer-duration challenge and strategic technology programmes - provided it is paired with stronger requirements for documented agency decision rationales and consolidated parliamentary reporting.

- Second, Sweden's decade of work on transformative innovation policy evaluation points directly to a gap in UK practice. For example, neither the Green Book nor UKRI's evaluation strategy currently addresses whether innovation programmes are contributing to the right kind of systemic change — what Haddad and Bergek (2023) call directionality — as distinct from efficiently executing their initial objectives. As the UK scales up mission-oriented programmes through UKRI and Innovate UK, it will face the same **theory-practice gap** that Rohracher, Coenen and Kordas (2023) document in Sweden: transformative policy ambitions outrunning evaluation frameworks designed for an earlier paradigm. The practical implication is that it is worth considering the development of a complementary track — drawing on Sweden's integrated TIP evaluation framework — that deploys systems-mapping, Theory of Change, and realist evaluation methods for complex, long-duration programmes where traditional counterfactual attribution is methodologically untenable. Sweden's experience is that **this requires sustained investment in interdisciplinary appraisal and evaluation capability within the funding agency, not merely better framework documents.**

11. Reflections

11.1 The review of R&D innovation policy appraisal systems across the European Union, Japan, Ireland, the United States, Australia, Belgium, and Sweden establishes one overarching finding: **no comparator country has fully resolved the fundamental tension between the methodological demands of rigorous prospective appraisal and the inherent characteristics of research and innovation investment** — long and variable time horizons, deep uncertainty, non-linear impact pathways, pervasive knowledge spillovers, and the public goods nature of basic science. The UK is not, therefore, starting from an inferior position. First, the UK faces a precise and formally identified (by official reviews) set of appraisal gaps: no R&D-specific appraisal guidance at the national level, and no mandated Theory of Change or logic model for ex ante appraisal decisions (even though this is advised). While the UK's situation is comparatively well developed, there are opportunities to strengthen policy appraisal.

11.2 The UK's current appraisal framework, which is centred on the HM Treasury Green Book's Five Case Model and social cost-benefit analysis, is more centralised, more consistently applied, and more structurally coherent than most comparator systems. Belgium's framework is fragmented across Regulatory Impact Assessments areas, spending reviews, and expert advisory processes without common methodology. Australia's formal machinery is structurally sound but does not focus explicitly on outcomes and system-level impacts. Sweden's statutory impact assessment regime, despite its sophistication, generates persistent accountability deficits in practice, with the Riksrevisionen (2020) finding weak documentation of agency decision-making and no portfolio-level assessment across major strategic innovation programmes. Ireland's system is structurally coherent around its dominant tax credit instrument but faces deep methodological challenges in assessing additionality. The United States operates a three-tier system in which formal cost-benefit requirements, strategic political direction, and scientific peer review exist in largely separate institutional worlds.

11.3 The UK is a leader in terms of the maturity and application of appraisal, although the international cases highlight several potential areas for further consideration in relation to the current UK appraisal framework. The focus of these insights to enhance appraisal in the UK are both within and beyond the immediate scope of R&D appraisal guidance, including:

- Adopting ***differentiated appraisal requirements*** could provide a more nuanced understanding of R&D investments. While there is proportionality guidance, the UK Green Book does not differentiate between types of R&D investments despite their different risk profiles, time horizons, and benefit structures. Japan's national R&D evaluation guidelines (ex ante and ex post) explicitly distinguish between basic, applied, and mission-oriented development, with different analytical standards and accountability mechanisms for each. Moonshot programme's stage-gate logic and portfolio-level risk management provide a working model for high-uncertainty investment that has no current UK equivalent outside ARIA itself. Sweden's SIP evaluation cycle with genuine funding consequences at three, six, and nine-year intervals, shows how long-duration programmes can be held accountable without demanding premature monetised returns. The supplementary Green Book guidance could specify differentiated appraisal requirements for

at minimum four modes: curiosity-driven basic research, strategic applied research, technology translation, and high-risk frontier investment. ARIA's model, already institutionally embedded, provides a natural template for the last of these that could be formally codified and its lessons disseminated.

- By ***treating monitoring as a reporting overhead*** rather than analytical infrastructure could benefit understanding and insights. The EU's Horizon Europe framework is instructive: it mandates integrated grant data systems, identifier-first data design (researcher IDs, DOIs, organisational identifiers), and a unified data warehouse from programme inception — not retrospective data assembly when ex post evaluation is due. Key Impact Pathways are designed around a non-linear impact chain with explicit lag recognition, acknowledging that definitive evidence may only emerge years after delivery. The practical implication for the UK is to mandate, at programme design stage, the use of researcher identifiers (ORCID), Digital Object Identifiers for outputs, and organisational identifiers (ROR) that enable low-burden longitudinal tracking — building the data infrastructure that evaluation will subsequently require rather than assembling retrospective evidence when spending reviews demand accountability. Budgeting explicitly for monitoring systems at programme inception, following the EU model, would address this structural deficit.

- The ***treatment of uncertainty merits consideration*** when appraising R&D investments. Japan's approach is interesting as rather than demanding precision where it is difficult to achieve, it recommends that evaluation design should explicitly match the innovation risk profile of the programme, with stage-gate logic, portfolio-level risk management, and structured stopping rules replacing project-level benefit-cost ratios for high-uncertainty investments. Japan's appraisal process allows for uncertainty by deferring to process where precision in terms of hard metrics is not expected or required. Australia's consultation findings, which link innovation underperformance directly to risk aversion and fear of failure fostered by audit-style appraisal, reinforce the same point from a different angle. The UK could introduce real options analysis as a standard complement to social cost-benefit analysis for staged R&D investments, recognising the value of managerial flexibility — the option to continue, scale, pivot, or stop — as a quantifiable element of programme design that conventional NPV calculations cannot capture. In this respect, real options analysis (ROA) captures the value of managerial flexibility and strategic adaptability in the face of uncertainty, which traditional Net Present Value (NPV) analysis typically ignores. Real options analysis is, by definition, a quantifiable method, but one that allows for strategic options such as deferring, expanding or abandoning. It is used for high-uncertainty, high-volatility projects where decisions are made sequentially rather than all at once. While NPV assumes a fixed, linear path for a project, real options recognise that management can make decisions over time, reacting to new information by expanding, delaying, or abandoning projects.

- Greater ***cross-portfolio coordination and system-level appraisal*** across government department, devolved administrations, and non-departmental public bodies (e.g., UKRI). This is the polycentric governance challenge identified by Wood (2025) in the US context and by Riksrevisionen (2020) in Sweden — where a decentralised multi-agency architecture generates programme-level accountability without cross-portfolio optimisation. Japan explicitly recommends standardised storage and cross-organisation sharing of evaluation information to

reduce burden and improve reuse in appraisal, noting that fragmented evaluation across funders and programmes generates duplication of analytical effort without enabling the comparative assessment of relative returns across the portfolio. Australia's Strategic Examination similarly calls for portfolio-level appraisal to complement major proposal appraisal for individual programmes.

11.4 The insights from across the seven cases points to a single essential condition for appraisal reform to deliver genuine improvement: the analytical framework must connect to the decisions that determine the scale and direction of R&D investment, not merely to the documentation that surrounds them. Sweden demonstrates the risk of disconnection through a statutory framework Riksrevisionen (2020) found isolated from actual agency reasoning. Belgium demonstrates it through a formally complete system that lacks the analytical culture needed to shape rather than justify decisions. Australia demonstrates it through machinery that is structurally sound but insufficiently outcomes-oriented to prevent declining risk appetite.

11.5 For the UK, the institutional moment is genuinely favourable, the Green Book discount rate is under independent review, ARIA provides a live institutional experiment in alternative appraisal models, and the DSIT portfolio-level evidence base exists and is being strengthened. **The challenge is to ensure that the methodology, once designed, is embedded in the institutional processes through which the UK's science budget is allocated, justified, and held accountable**. That requires not only better guidance documents but stronger requirements for documented decision rationales, consolidated parliamentary reporting, and periodic independent review of whether aggregate public R&D investment is — in both scale and composition — commensurate with the evidence on social returns and the demands of long-run national competitiveness.

Annex A: References

Al-Ajlani, H., Mitra, A., Kasy, M. and Niakaros, K. (2025) Measuring Success: Common Challenges in Assessing the Impact of Large-scale R&I Funding Programmes. Brussels: European Commission.

ANAO (2024) *Administration of the Impact Analysis Framework*, Australian National Audit Office Available at: https://www.anao.gov.au/sites/default/files/2025-05/Auditor-General_Report_2024-25_33.pdf

Atkinson, R.D. (2020) *Understanding the U.S. national innovation system, 2020*. Washington, DC: Information Technology and Innovation Foundation. Available at: https://itif.org/publications/2020/11/02/understanding-us-national-innovation-system-2020/

Australian Bureau of Statistics. (2023-24). *Research and Experimental Development, Businesses, Australia*. ABS. Available at: https://www.abs.gov.au/statistics/industry/technology-and-innovation/research-and-experimental-development-businesses-australia/latest-release.

Bajgar, M., Ito, K. & Timmis, J. (2025). "R&D Spillovers through Buyer-supplier Networks", *RIETI Discussion papers 25127*, Research Institute of Economy, Trade and Industry (RIETI). Available at: https://ideas.repec.org/p/eti/dpaper/25127.html

Benoit, F. & Karvounaraki, A. & Stevenson, A. & Ravet, J., (2025). "Shaping the Future: EU R&D Investments Explained", *EU research and innovation paper series KI-01-25-091-EN-N*, Directorate General for Research and Innovation (DG RTD) of the European Commission. Available at: https://ideas.repec.org/p/eug/wpaper/ki-01-25-091-en-n.html

Bonvillian, W.B. (2006) 'The connected science model for innovation: the DARPA role', in National Research Council, *21st century innovation systems for Japan and the United States*. Washington, DC: National Academies Press.

Cabinet Office, Government of Japan. (2012). National Guidelines for Evaluating Government Funded R&D.

Commonwealth of Australia. (2020) *Australian Government Guide to Regulatory Impact Analysis 2nd Edition,* Commonwealth of Australia, Department of the Prime Minister and Cabinet.

Commonwealth of Australia. (2022) Cabinet Handbook 15th Edition, Commonwealth of Australia, Department of the Prime Minister and Cabinet.

Commonwealth of Australia. (2023) *The Australian Government Guide to Policy Impact Analysis,* Commonwealth of Australia, Department of the Prime Minister and Cabinet.

Commonwealth of Australia (2025) Strategic Examination of R&D: discussion paper consultation findings and analysis, Strategic Examination of R&D independent expert panel.

CRS (2025) *Federal research and development (R&D) funding: FY2026*. R48694. Washington, DC: Congressional Research Service. Available at: https://www.congress.gov/crs-product/R48694

DSIT (2025) *The value of public R&D*. London: Department for Science, Innovation and Technology. Available at: https://www.gov.uk/government/publications/the-value-of-public-rd/the-value-of-public-rd

European Commission (2012) Guide to Research and Innovation Strategies for Smart Specialisation (RIS3). Brussels: European Commission.

European Commission (2021) *Better Regulation Guidelines*. Brussels: European Commission.

European Commission (2023) Evidence Framework on Monitoring and Evaluation of Horizon Europe. Brussels: European Commission.

European Commission (2025a) *Shaping the Future: EU R&D Investments Explained*. Brussels: European Commission.

European Commission (2025b) *European Innovation Scoreboard 2025*. Brussels: European Commission.

European Investment Bank (2023) *The Economic Appraisal of Investment Projects at the EIB*. Luxembourg: European Investment Bank.

Fieldhouse, A.J. and Mertens, K. (2024) The return to government R&D: evidence from U.S. appropriations shocks, *Federal Reserve Bank of Dallas Working Paper,* No. 2305

Fieldhouse, A.J. and Mertens, K. (2025) The social returns to public R&D, *NBER Working Paper* No. 33780.

Government of Ireland (2024) Policy Handbook: A 'real world' guide to policy development,

Government of Sweden (2024) Forskning och innovation för framtid, nyfikenhet och nytta (Research and innovation for the future). Stockholm: Government Offices of Sweden.

Haddad, C.R. and Bergek, A. (2023) Towards an integrated framework for evaluating transformative innovation policy, *Research Policy*, 52(2), 104676

Haddad, C.R., Wise, E. and Arnold, E. (2024) 'When theory meets practice in transformative innovation policy evaluation: experiences from Sweden', *Research Evaluation*, 33

HM Treasury (2020) *The Magenta Book: Central Government Guidance on Evaluation*. London: HM Treasury. Available at: https://www.gov.uk/government/publications/the-magenta-book

HM Treasury (2025) *Green Book Review 2025: Findings and Actions*. London: HM Treasury. Available at: https://www.gov.uk/government/publications/green-book-review-2025-findings-and-actions

Intarakumnerd, P. (2013). Japanese Financing Policies for Innovation Since the 1990s. STI Policy Review, 4 (2), 55-73.

Kato, M. and Coad, A. (2024), "Japan's SBIR Scheme", *Annals of Science and Technology Policy*: Vol. 8, No. 1–2, pp 77–93. DOI:10.1561/110.00000028-6.

Larrue, P. (2021), "Mission-oriented innovation policy in Japan: Challenges, opportunities and future options", *OECD Science, Technology and Industry Policy Papers*, No. 106, OECD Publishing, Paris, https://doi.org/10.1787/a93ac4d4-en.

National Audit Office (NAO) (2025) UK Research and Innovation: Providing Support Through Grants. HC 875. London: NAO.

NCSES (2025) *Federal funds for research and development: FYs 2024–26*. NSF 25-329. Alexandria, VA: National Center for Science and Engineering Statistics. Available at: https://ncses.nsf.gov/pubs/nsf25329

NIH (2025) *Simplified peer review framework*. Bethesda, MD: National Institutes of Health. Available at: https://grants.nih.gov/policy-and-compliance/policy-topics/peer-review/simplifying-review/framework

NSF (2025) *Merit review*. Alexandria, VA: National Science Foundation. Available at: https://www.nsf.gov/funding/merit-review

OECD (2015) Frascati Manual 2015: guidelines for collecting and reporting data on research and experimental development. Paris: OECD Publishing.

OECD (2016). *OECD Reviews of Innovation Policy: Sweden 2016*, OECD Publishing, Paris. Available at: http://dx.doi.org/10.1787/9789264249998-en

OECD (2020), Regulatory Impact Assessment, OECD Best Practice Principles for Regulatory Policy, OECD Publishing, Paris, https://doi.org/10.1787/7a9638cb-en.

OECD (2023a) Public research funding in Sweden: Optimising the system in response to multiple demands, *OECD Science, Technology and Industry Policy Papers,* No 148, OECD Publishing, Paris

OECD (2023b) OECD Science, Technology and Innovation Outlook 2023. Paris: OECD Publishing

OECD (2023c) Improving decision making through policy evaluation in Belgium, OECD Public Governance Policy Papers No.31

OECD (2025a) *Main science and technology indicators, March 2025*. Paris: OECD Publishing. Available at: https://www.oecd.org/en/data/datasets/main-science-and-technology-indicators.html

OECD (2025b) Building capacity for evidence-informed policymaking in governance and public administration in Belgium,

OECD STIP Compass (2025) *Research and Innovation Policy Bill 2025-2028 - Research and Innovation for Future, Curiosity and Benefit (Sweden).* Available at: https://stip.oecd.org/stip/interactive-dashboards/policy-initiatives/2025%2Fdata%2FpolicyInitiatives%2F99997408

OMB (2016) *Circular A-11, Section 84: schedule C*. Washington, DC: Office of Management and Budget.

OMB (2023) *Circular A-94: guidelines and discount rates for benefit-cost analysis of federal programs*. Revised November 2023. Washington, DC: Office of Management and Budget. Available at: https://www.whitehouse.gov/wp-content/uploads/2023/11/CircularA-94.pdf

Parliament of Australia (2025) Australia's Impact Analysis Framework, Parliament of Australia, Policy Brief, 2025-26

Public Health England (2021) *A brief introduction to realist evaluation.* Available at: https://assets.publishing.service.gov.uk/media/60f7fdf7d3bf7f56824cc634/Brief_introduction_to_realist_evaluation.pdf

RIETI (2024) Deepening evidence-based policymaking in the UK: Focusing on the Policy Evaluation Task Force, *RIETI Policy Discussion Paper Series 24-P-008*

Riksrevisionen (2020) *Innovation genom samverkan — statens satsningar på strategiska samverkans- och innovationsprogram* [Innovation through collaboration — state investments in strategic collaboration and innovation programmes], RiR 2020:6. Stockholm: Riksrevisionen.

Prior, R. and Brennan, T. (2025) R&D and innovation in Australia: 2024 update, Parliament of Australia, Research Paper Series, 2024-25

Rohracher, H., Coenen, L. and Kordas, O. (2023) 'Mission incomplete: Layered practices of monitoring and evaluation in Swedish transformative innovation policy', *Science and Public Policy, 50(2), pp. 336–349.*

Science and Technology Australia (2023) Set a 3% R&D target and double the dollars for discoveries Available at: https://scienceandtechnologyaustralia.org.au/uni-accord-panel-set-rd-target-and-double-research-funds/

Schwaag Serger, S., Dachs, B., Kivimaa, P., Lazarevic, D., Lukkarinen, J., Stenberg, L. and Weber, M. (2023) 'Transformative innovation policy in practice in Austria, Finland and Sweden', *OECD Science, Technology and Industry Policy Papers, No. 156. Paris: OECD Publishing*

Statsrådsberedningen (2025) *Myndigheternas föreskrifter – SB PM 2025:4*. Stockholm: Statsrådsberedningen (Prime Minister's Office).

Swedish Government (2024a) *Forskning och innovation för framtid, nyfikenhet och nytta* [Research and innovation for the future, curiosity and utility], Prop. 2024/25:60. Stockholm: Regeringskansliet.

Swedish Government (2024b) *Förordning (2024:183) om konsekvensutredningar* [Ordinance (2024:183) on Impact Assessments], SFS 2024:183. Stockholm: Ministry of Finance.

Statistics Sweden (2025) Research and Development in Sweden 2024 — Preliminary Statistics. Stockholm: Statistics Sweden (SCB).

Vetenskapsrådet (2023) *The Swedish Research Barometer 2023*. Stockholm: Swedish Research Council.

Vetenskapsrådet (2025) *National Evaluations of the Results of Research*. Stockholm: Swedish Research Council.

Vinnova (2008) *VINNOVA's Focus on Impact* [VINNOVA Analysis VA 2008:01]. Stockholm: Vinnova.

Vinnova (2024) Meta-evaluation of the Swedish Strategic Innovation Programmes. Stockholm: Vinnova/Technopolis Group.

Vinnova (2025) *Strategic Innovation Programmes*. Stockholm: Vinnova.

Vought, R.T. and Kratsios, M.J. (2025) Memorandum M-25-34/NSTM-2: fiscal year (FY) 2027 administration research and development budget priorities and cross-cutting actions. Washington, DC: Office of Management and Budget and Office of Science and Technology Policy, 23 September. Available at: https://www.whitehouse.gov/wp-content/uploads/2025/09/M-25-34-NSTM-2-Fiscal-Year-FY-2027-Administration-Research-and-Development-Budget-Priorities-and-Cross-Cutting-Actions.pdf

Wirkierman, A., Ciarli, T. and Savona, M. (2018) *Varieties of European National Innovation Systems*. Brussels: European Commission Horizon 2020 Project.

Wood, J. (2025) 'Why America's innovation ecosystem will survive Trump's cuts to R&D', *LSE United States Politics and Policy Blog*, 17 December. Available at: https://blogs.lse.ac.uk/usappblog/2025/12/17/why-americas-innovation-ecosystem-will-survive-trumps-cuts-to-rd/

Zimmermann, A. (2025) *U.S. R&D and innovation in a global context: the 2025 data update*. Washington, DC: American Association for the Advancement of Science, 30 April. Available at: https://www.aaas.org/news/us-rd-and-innovation-global-context-2025-data-update

Annex B: Abbreviations and acronyms

Abbreviation	Meaning
AMED	Japan Agency for Medical Research and Development
ARC	Australian Research Council
ARIA	Advanced Research and Invention Agency
BERD	Business Enterprise Research and Development
BRICS nations	Brazil, Russia, India, China and South Africa
CBA	Cost-Benefit Analysis
CEA	Cost-Effectiveness Analysis
CSIRO	Commonwealth Scientific and Industrial Research Organisation (Australia)
CSO	Central Statistical Office (Ireland)
CSTI	Council for Science, Technology and Innovation (Japan)
CMO	Context-Mechanism-Outcome
DARPA	Defence Advanced Research Projects Agency (US)
DID	Difference-in-Difference
DOD	Department of Defense (US)
DOE	The U.S. Department of Energy
DOI	Digital Object Identifier
DSIT	Department of Science, Innovation, and Technology (UK)
EBPM	Evidence-Based Policymaking
ECOOM	The Centre for Research & Development Monitoring (Belgium)
EIB	European Investment Bank
EIF	European Investment Fund
EIPM	Evidence-Informed Policymaking
EIS	European Innovation Scoreboard
EPO	European Patent Office
ERC	Expenditure Review Committee (Australia)
FPB	The Federal Planning Bureau (Belgium)

Abbreviation	Meaning
FPS BOSA	Federal Public Service Strategy and Support in Belgium
GBARD	Government budget allocations for R&D
GERD	Gross Expenditure on Research and Development
HMRC	HM Revenue and Customs (UK)
IAC	Impact Assessment Committee
ICT	Information and Communication Technology
IA	Impact Analysis
IP	Intellectual Property
JST	Japan Science and Technology Agency
KIPs	Key Impact Pathways
METI	Ministry of Economy, Trade and Industry (Japan)
MEXT	Ministry of Education, Culture, Sports, Science and Technology (Japan)
NCSES	National Center for Science and Engineering Statistics (US)
NAO	National Audit Office
NEDO	New Energy and Industrial Technology Development Organization (Japan)
NHMRC	National Health and Medical Research Council (Australia)
NIH	The National Institutes of Health (US)
NIS	National Innovation Systems
NPV	Net Present Value
NSF	The National Science Foundation (US)
OBPR	Office of Best Practice Regulation (Australia)
OIA	Office of Impact Analysis (Australia)
OMB	The Office of Management and Budget (US)
OSTP	The Office of Science and Technology Policy (US)
PCDA	Plan-Do-Check-Action
PCT	Patent Cooperation Treaty
PDCA	Plan-Do-Check-Action

Abbreviation	Meaning
PIRs	Post-Implementation Reviews
PM&C	Department of the Prime Minister and Cabinet (Australia)
PSC	Public Spending Code (Ireland)
R&D	Research and Development
RDI	Research, Development and Innovation
RDTI	R&D tax incentive
RE	Realist Evaluation
RIA	Regulatory Impact Assessment
RIETI	Japanese Research Institute of Economy, Trade and Industry
RIS	Regulatory Impact Analysis
RIS3	Research and innovation Strategies for Smart Specialisation
ROA	Real Options Analysis
ROAMEF	Rationale, Objectives, Appraisal, Monitoring, Evaluation, and Feedback
ROR	The Research Organization Registry
RSB	Regulatory Scrutiny Board (European Commission)
SBIR	Small Business Innovation Research
SFI	Science Foundation Ireland
SIPs	Strategic Innovation Programmes (Sweden)
SIPP	Strategic Innovation Promotion Programme (Japan)
SMART	Specific, Measurable, Achievable, Relevant, and Time-bound
SRI	Science, Research and Innovation
STTR	Small Business Technology Transfer
SWD	Staff Working Document (European Commission)
TIP	Transformative Innovation Policy
TIS	Technology Innovation System
TOC	Theory of Change
UKRI	UK Research and Innovation

Abbreviation	Meaning
VfM	Value for Money
WP	With-project
WOP	Without-project

Source: Case study authors

Annex C: Summary of insights from country cases

See glossary in Annex B for definitions.

Countries	(a) Uses Logic models — Appraisal	(b) Uses Logic models — Evaluation	(c) R&D-specific guidance — Appraisal	(d) R&D-specific guidance — Evaluation	(e) Centralised vs decentralised	(f) Appraisal/evaluation clearly distinguished
EU	**YES** Intervention logic required in all Impact Assessments	**YES** Required via Toolbox; KIPs map full causal chain	**PARTIAL** Guidelines for all policy area (Chapter IV for ex-ante Impact Assessment); Tool #22 (Research and Innovation) Tool#9 (Spending programmes) no standalone guide	**YES** Horizon Europe Monitoring and Evaluation Evidence Framework SWD (2023) 132 -KIPs Framework, ERA-LEARN toolkit	Hybrid Central rules + RSB; distributed execution European Partnerships	**YES** Separate chapters, procedures, RSB tracks
Japan	**PARTIAL** Functional equivalents (PDCA cycles, output/outcome distinctions, roadmaps); no formal TOC	**PARTIAL** PDCA cycle; flowcharts recommended but not mandated	**YES** National Guidelines for Evaluating Government Funded R&D set by Cabinet Office/CSTI) + ministry guidelines (MEXT guidelines) cover ex ante	**YES** National Guidelines for Evaluating Government Funded R&D set by Cabinet Office/CSTI + ministry guidelines (MEXT) + Funding agency guidelines (JST, NEDO, AMED)	Hybrid-hierarchical PM guidelines → ministries → agencies	**YES** Four explicit temporal stages in guidelines (ex-ante, mid-term, ex-post, follow-up); but "evaluation" language for all stages.
Ireland	**PARTIAL** Structured CBA, CEA, but no formal TOC mandate	**YES** Programme Logic Models required in PSC Quality Assurance in-depth checks and VfM Reviews	**PARTIAL** Parallel Ecosystem of sector-specific tools. Enterprise Agencies Economic Appraisal Model provide CBA for grants not always used for R&D	**PARTIAL** Forfás framework (for Enterprise supports) and SFI (now Research Ireland) impact framework pillars; not PSC-integrated	Hybrid PSC central; agencies autonomous operationally	**YES** PSC across lifecycle explicitly separates expenditure into three phase
USA	**PARTIAL** Peer review dominates; logic models in some NSF programmes	**YES** Encouraged by Office of OMB; used in DOE, NSF evaluation	**PARTIAL** Agency-level only (NSF, NIH peer review criteria)	**PARTIAL** DOE methods booklet; Evidence Act is generic	Highly decentralised OMB sets generic standards; agencies diverge	**PARTIAL** Functional separation; not formally branded
Australia	**PARTIAL** CBA-driven; TOC not mandated for appraisal	**YES** TOC templates in Evaluation Toolkit (RMG 130)	**NO** Generic OIA framework for all policy	**NO** Generic evaluation policy; ARC's Excellence in Research for Australia and Engagement and Impact assessment discontinued 2023	Hybrid Increasingly centralised framework; distributed execution	**PARTIAL** Institutional separation (PM&C vs Treasury) but less formalised
Belgium	**PARTIAL** Some agencies require impact linkage informally	**PARTIAL** ECOOM uses causal methods; no published guidance	**NO** Generic RIA only	**PARTIAL** FPB and ECOOM practice; no published guidance	Highly decentralised Federal/regional split; no coordination	**NO** No formal distinction at any level
Sweden	**PARTIAL** Vinnova requires impact logic; no central mandate	**PARTIAL** Framework exists; NAO found quality deficient	**NO** No central R&D appraisal guidance	**PARTIAL** Stonge agency independence; Vinnova "Focus on Impact" VA 2008:01 ex post Evaluation and long-term impact analysis; no central guidance	Strongly decentralised; Each agency autonomous	**NO** System-wide; **YES** at Vinnova

Source: Innovation Research Caucus, based on literature

SQW



Contact

For more information:

Tim Vorley
Co-Director,
Innovation & Research Caucus
T: 01865 485839
E: tvorley@brookes.ac.uk

Oxford Brookes Business School,
Headington Campus,
Oxford, OX3 0BP

Rebecca Pates
Director, SQW
T: 0161 475 2105
E: rpates@sqw.co.uk

3rd Floor
1WR
1 Wellington Road North
Stockport
SK4 1AF

www.sqw.co.uk

About us

SQW Group

SQW and Oxford Innovation are part of SQW Group.
www.sqwgroup.com

SQW

SQW is a leading provider of research, analysis and advice on sustainable economic and social development for public, private and voluntary sector organisations across the UK and internationally. Core services include appraisal, economic impact assessment, and evaluation; demand assessment, feasibility and business planning; economic, social and environmental research and analysis; organisation and partnership development; policy development, strategy, and action planning.
www.sqw.co.uk

Oxford Innovation

Oxford Innovation supports economic growth by providing businesses with Advice, Finance and Space. Oxford Innovation Advice is the UK's outstanding provider of business support for SMEs, providing specialist expertise to help businesses accelerate their growth and fulfil their potential. Oxford Innovation Finance is a specialist in providing growth and development capital for ambitious and innovative businesses. Oxford Innovation Space helps to transform local economies by creating environments where dynamic and innovative businesses come together and thrive.
www.oxfordinnovation.co.uk