Indicators and Evaluation for Science, Technology and Innovation

Aidan Kane

aidan.kane@nuigalway.ie

Department of Economics
NUI, Galway

www.economics.nuigalway.ie

Background document for
ICSTI Task Force on Metrics and Impact
This draft: May 17th 2001

Contents

Abstract

This paper provides a structured survey of some principal indicators and evaluation methods used internationally to understand science, technology and innovation (STI) activities and inform STI policy. The paper emphasises the variety of such techniques in existence or under development, and offers a discussion of how they may be understood in the context of alternative models of the innovation process, and the distinctive roles in policy formulation and adaptation which they may play. In particular, some techniques, are best understood as reflecting essentially neo-classical economic/linear models of innovation, whereas others reflect the focus of evolutionary/institutional models on knowledge flows, linkages and complex innovative processes.

> He confined the Knowledge of governing within very *narrow Bounds*; to common Sense and Reason, to Justice and Lenity, to the Speedy Determination of Civil and criminal causes; with some other obvious Topicks which are not worth considering. And, he gave for his Opinion; that whoever could make two Ears of Corn, or two Blades of Grass to grow upon a Spot of Ground where only one grew before; would deserve better of Mankind, and do more essential Service to his Country, than the whole Race of Politicians put together.
>
> From "The King's great Ignorance in Politicks", in Chapter VII, Part II, of Jonathan's Swift's *Gulliver's Travels*.

1 Introduction

This paper draws upon an international economic and science policy literature to review science, technology and innovation (STI) indicators and evaluation methods in order to inform the Irish policy debate. The multiplicity of such techniques suggests that apart from operational issues of implementation, they pose many challenges of conceptualisation. How do indicators and evaluation methods fit in to our understanding of the innovation process? How do we select them appropriately? What are their limitations? Who will evaluate the evaluators? This paper raises rather than answers such questions, but aims to do so in a structured way.

We first identify a number of motivations underlying the widespread interest in STI indicators and evaluation methods which has paralleled the increasing importance of STI policy internationally. We explain some of the basic ideas which situate these techniques as distinctive concepts, then outline those indicators and evaluation methods which will be discussed in more detail in the paper. The final part of this introductory section considers how these techniques can differ from one another along a number of dimensions. These distinctions structure the substantive discussion of the principal features of these techniques in Sections 2–3. In Section 4 we conclude by devoting attention to the institutional context within which such techniques are deployed.

1.1 Motivation

The counterpart to STI policy's ascent to the top of the international economic policy agenda is a concern to assess the performance of policy initiatives in this domain and to improve our understanding of the underlying innovation process. This generally motivates the active interest in identifying meaningful indicators and evaluation methods so that STI activities and policy can be judged and improved.

Some techniques reflect their origins in industrial societies structured around the production of tangible goods, government intervention directed at providing physical infrastructure, and economic analyses conditioned by those realities. The emerging reality of knowledge-based economies, where government is a limited influencer of intangible flows, interpreted by less mechanistic economic perspectives, provides an impulse for the development of *new* indicators and evaluation methods.

Governments internationally, having discarded sharp ideological poses, have adopted the language of 'new public management' with which to interpret their responsibilities of efficacy and accountability in sophisticated democratic societies. This is a language presumptively applicable to STI policy, and the STI community views indicators and evaluation methods as a partial response to the agenda it implies. The external political audience addressed by the STI community encompasses not just élite policy makers, but democratic society more generally, so that the contribution which such techniques can make to the public understanding of, and support for, STI activities, is a further motivation.

The internal practitioner audience, apart from regarding indicators and evaluation methods as elements of the rhetoric of public support, appreciates that the STI domain is a complex and changing one, requiring correspondingly nuanced and flexible instruments for learning. This exemplifies a theme which recurs in this discussion, namely that the selection of STI indicators and evaluation methods can be aligned with models of the innovation process—however implicit, partial and contingent those models may be. Differing models offer differing rationales for policy interventions, and suggest differing concepts of policy success, failure and amendment, which our techniques should allow us to illuminate.

1.2 Scope and organisation

1.2.1 Basic concepts

In delineating the scope of this discussion, it is useful to fix some basic ideas now. We have taken for granted up to now what we mean by 'indicators' and 'evaluation'. Are these concepts distinct? While in principle, indicators involve an evaluative exercise, and evaluations provide indicators, we adopt a pragmatic distinction. We can think of indicators as generally being summary statistics of underlying data, intended to report on or measure a distinct aspect of the innovation process. Evaluation involves more by way of judgement against defined criteria, and utilises indicators as the raw material in the application of specific analytical methods—often but not exclusively quantitative in nature. Crucially, evaluation is much more of a *social process*, insofar as the assessment of indicators or other information involves much more by way of explicitly specifying an underlying model of innovation, and implies an institutional framework within with the evaluation takes place. In practice, and in short, indicators are simpler than evaluation.

Indicators and evaluation methods are distinct from conventional economic information in a number of ways. Any policy predicated on the correction of markets failures by definition involves a domain where market prices, quantities and profits are inadequate signals for resource allocation decisions. Even more distant from the market is information relating to the susceptibility of well-intentioned policy interventions to public failures, e.g., incentive structures leading to perverse outcomes, over-burdensome administrative and compliance costs.

These remarks might apply to many areas of policy, but STI policy is an instance of particularly marked challenges in remedying the deficiencies of conventional sources. Keith Smith (1998), as part of a larger guide to STI indicators for policy makers, provides an explanation which is worth quoting in detail:

> It can be strongly argued that in certain respects innovation is incompatible with measurement *tout court*. Measurement is a process of counting or comparison in which we seek to compare entities in terms of some common characteristics, such as weight, dimensions, and so on. In other words, measurement requires an *a priori* dimensional similarity between objects; that is, there is some dimension along which they meaningfully share attributes.
>
> . . .
>
> However, innovation is, by definition, novelty. It is the creation of something qualitatively new, and this leads immediately to problems in measuring and comparing. Innovation is not about extending pre-existing dimensions, but rather changing or replacing technical attributes. In some cases this may mean changing product characteristics, which may certainly be intrinsically measurable in some way—the lift/drag aspects of an aircraft

> wing, for example, or the speed/carrying-capacity of an entire aircraft. However such technical measurement comparisons are only rarely meaningful across products. It is difficult if not impossible, by means of technical measurements of attributes to assess, for example, the degree of innovativeness of a product. More generally, innovation involves multi-dimensional novelty in aspects of activity or knowledge organization which are difficult to measure or intrinsically unmeasurable.
>
> (Smith 1998, pp.11–12)

This is a salutary preliminary statement of the distinctiveness of the indicator (and evaluation) challenge in STI; nevertheless it is apparent, not least from the wider research project of which the above-cited work is a part (the EU-funded IDEA project of the STEP group in Norway) and of which the present paper makes extensive use, that many indicator and evaluation options for STI do exist or are in development, and these provide our substantive agenda.

1.2.2 Principal indicators/evaluation methods

We now provide short guides to the principal indicators and evaluation methods reviewed in this paper, while the next section describes some of the ways in which we could organise such a listing along certain dimensions, in order to structure the more detailed discussion later.

We discuss the following:

Indicators of R&D intensity, based on measures of R&D expenditures, which can be distinguished by the sector in which the activity takes place and by sources of finance e.g., government, business, higher education.

Human capital indicators, including employment data, especially on R&D workers, scientists and engineers, data on educational attainments, and indexes of human capital which aim to capture quality improvements, as well as indicators of the mobility of human capital.

Indicators of the technological intensity of production, which scale national and sectoral productive activities according to some predefined notion of high-technology.

Balance of payments indicators, including measures of the technological intensity of exports and imports of good as services, as well as the technology balance of payments, which report on trade in technological knowledge, e.g., by licences sold abroad for domestic patents, and trade in technological services.

Bibliometric analyses, by which the quantity, quality and impact of published scientific literature are assessed, insofar as this literature is a key social mechanism by which scientific knowledge is embodied and transmitted.

Patents analyses, including relatively simple patent counts, but also more elaborate analyses in a bibliometric style aimed at tracing flows of knowledge between science and near-to-market innovations, and a number of techniques which attempt to assess the economic impact/value of patents.

Growth accounting analyses, which distinguish the technologically driven component of economic growth through measures of the productivity of capital and labour from that part of growth which arises from the increased use of capital and labour inputs.

The measurement price and quality changes, which underlies the practice of growth accounting, raises a set of issues of particular importance where, as in STI, the introduction of new and improved products and processes is a defining characteristic.

Cost-benefit analyses/quasi-financial methods, which attempt to capture some of the non-market flows of welfare impacts attributable to STI activities, including models which estimate returns using the concepts of consumers' and producers' surplus, and rates of return, by partial analogy to private sector investment decision making.

Models of productivity spillovers, which draw upon growth accounting analyses to statistically estimate the magnitude and direction of non-appropriable benefits to R&D efforts between and amongst firms, productive sectors, regions and nations.

Models of knowledge flows, which draw upon bibliometric and patent citation data, in order to map the quantum of knowledge of varying types, patterns of knowledge diffusion and interaction, and the strengths of linkages between knowledge-producing and knowledge-using sectors.

Peer review, which has traditionally been a key social process in the scientific community, with varying degrees of adaptation in structuring public policy decisions e.g., some reliance on formal bibliometric analyses and wider selection pools for 'peers'.

Surveys/Interviews, which find wide application in evaluating STI activities, not least because of the opportunity to integrate quantitative and qualitative data.

Case studies, which may draw upon many of the other techniques to provide comprehensive retrospective accounts, especially of well-defined projects with complex evolutionary development paths.

Technology foresight exercises, which similarly can be informed by the eclectic use of information sources, but which also can be enabled by techniques specific to the prospective strategic analysis of STI options.

Implicit in these summary accounts are commonalities among, and differences and linkages between the various indicators and evaluation techniques. We now consider how we might organise such a complex set of relationships.

1.2.3 Dimensions of indicators and evaluation

Some of the dimensions along which we could organise an account of indicators/evaluation methods are as follows:

- By the level of analysis of the system of innovation to which the indicator/evaluation methods is generally addressed e.g., analyses of global, transnational, national, regional systems, or those which focus on specific sectors, industries, firms, or policy programmes, institutions, projects etc.
- Are inputs, outputs or impacts being considered? We take inputs to refer to STI resources or effort (e.g., R&D expenditure) , whereas outputs are intermediate results which *prima facie* contribute to innovation, but are not themselves innovations (e.g., patents), and impacts, the final results of the process, which are the ultimate concern of policy makers (e.g., the production and adoption of welfare-enhancing new goods and services as reflected in economic growth, employment outcomes.) Whether an indicator reports on inputs or outputs can be dependent on the context and the underlying analytical model; for example, patents can be both the output of particular firms, and inputs into the innovative activities of others. Similarly, a neo-classical perspective would view the employment of scientists and engineers as an input, whereas an evolutionary/institutional model suggests that a key output is precisely the knowledge-absorption capacity embodied in highly quality human capital.

- The type of policy intervention, or element of the innovation system to which the indicator/evaluation is most appropriately applied e.g., some models of productivity spillovers have been used in the analysis of the benefits of R&D subsidies to the private sector, whereas bibliometric analyses and peer review might be more natural techniques in assessing grant/subsidies for publicly performed 'basic' research.

- The stage of the policy process being considered, e.g., *the ex ante* prioritisation of policy or selection of projects, (for example, technology foresight) the on-going monitoring of initiatives, or *ex post* judgements of performance against stated objectives (for example, cost-benefit analyses, or historical case studies).

- Whether the indicator/evaluation method is 'summative' i.e., intended to judge, select and/or rank results in a compressed way, often intended to inform a specific policy decision, or 'formative' in being intended more to understand underlying processes and thereby improve performance (Arnold and Guy 1997). Indicators, taken in isolation, might often be summative; at least some evaluation methods can be integrated into a formative process, e.g., feedback arising from client surveys of technology brokerage programmes for firms.

- Who reports an indicator/undertakes an evaluation? STI indicators often arise from specific academic projects, later to mature into mainstream national/international statistical reporting. Evaluation is also implicit in much academic work, while both research institutions and policy makers can draw upon this work, external evaluators, including private sector specialists, and/or internal expert units.

- The intended audience; an evaluation might internal to policy participants and practitioners, or directed towards external sources of policy support such as a funding agency or regulatory authority, or the wider political and social system in which STI is embedded and policy determined.

- Whether a technique is mainly quantitative or qualitative, which may have implications for the degree of expertise needed to implement and interpret the information which it generates. Some indicators are relatively simple counting exercises, others require elaborate econometric tools, while case studies call upon a range of integrative skills to conduct and process.

- The underlying model of the innovation process on which an indicator/evaluation method is based. Techniques are model-dependent e.g., growth accounting, and estimates of consumers'/producers' surplus arise from neo-classical perspectives, while at least some indicators and models of knowledge flows can be best understood within an evolutionary/institutional perspective.

- The information gathering and processing costs of indicators and evaluation methods, which may include the effect which they have on the behaviour of STI participants by setting up incentives for compliance.

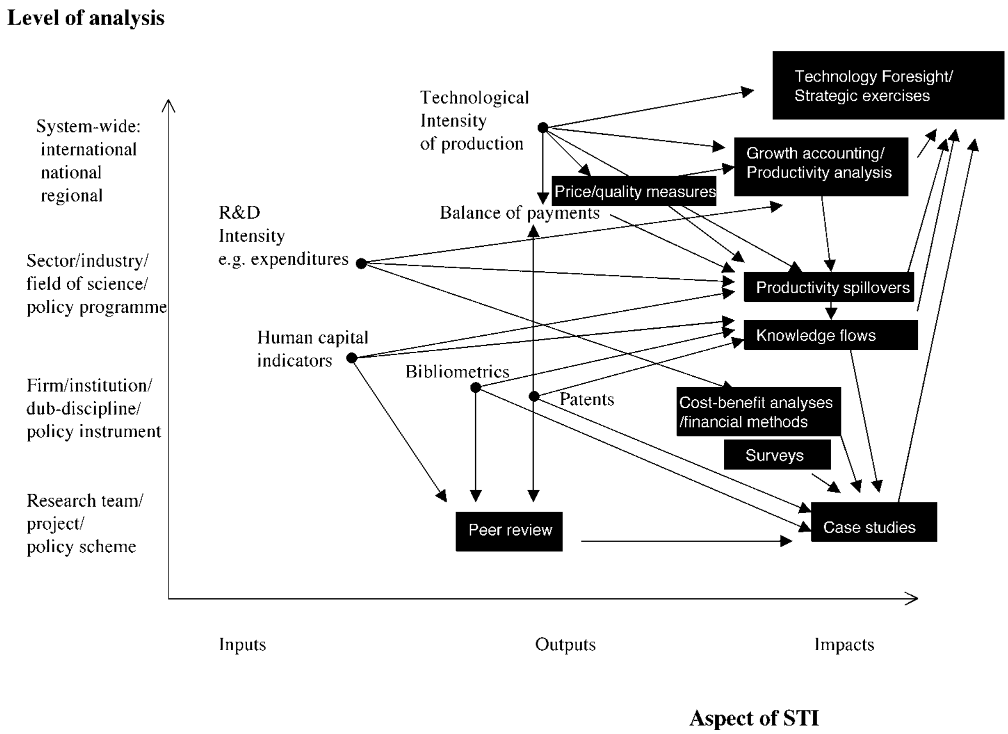

Figure 1: Indicators and Evaluation Methods in STI

A starting point for managing the multiplicity and multi-dimensionality of these techniques is offered in Figure 1. On this diagram, indicators and evaluation methods (in boxes) are mapped according to the dimension relating to the level of analysis to which they are most often applied (from international systems to individual projects), and, on the horizontal axis, according to the extent to which the technique relates to the assessment of inputs, outputs or impacts. The arrows illustrate how techniques can feed into one other e.g., case studies and foresight exercises can draw upon and integrate most other indicators and evaluation methods.

The positioning of techniques in this diagram is meant to be suggestive of broad tendencies only, and not precise, reflecting the earlier observation that some techniques, such as those based on patents, may be fairly simple counts of input or outputs, or more elaborate exercises designed to capture broader 'knowledge impacts'. Much of the following discussion points to such examples of the development of indicators and evaluation methods, which often reflect attempts to capture the dimensions of innovation implied by newer economic models.

The next section discusses those indicators and evaluation methods which can be thought of as reporting on inputs, outputs and impacts, in an essentially neo-classical economics tradition. Section 3 reviews techniques which are closer in spirit to evolutionary/institutional models of innovation, in stressing the importance of knowledge flows, linkages and networks, and in so doing, addressing the distinctive challenges which the complexity of STI poses.

2 Assessing inputs, outputs and impacts

2.1 Inputs

2.1.1 R&D expenditures

The prominence of R&D expenditures as an indicator of innovative effort can be traced to the secular post World War II increase in such activities in advanced industrial economies, in which governments and the private sector interacted closely. The OECD publishes a series of guidelines for the collection of such data, the *Statistical Practice for Surveys of Research and Experimental Development*, referred to as the *Frascati Manual*, starting in 1963, with the current (fifth) version published in 1993.

Sandven's (1998) review of R&D indicators traces the origins of such measures to fairly mechanistic accounts of innovation activities—the 'linear model of innovation' which posits a staged sequence of activities running from basic research towards applied research and the experimental development of existing knowledge. To that extent, and insofar as such R&D expenditure indicators tend to count formal and explicit innovative effort, and thus neglect both innovative outcomes and less formal activity, they have a number of limitations. For example, Sandven discusses a tendency to systematically undercount the R&D activities of smaller firms.

Furthermore, he demonstrates that the common practice of constructing national indexes of innovative activity by e.g., comparing the ratio of Gross Expenditure on Research and Development (GERD) to a measure of national income such as Gross Domestic Product (GNP) across countries, is not straightforward. For example, if larger countries have an intrinsic scale advantage in conducting R&D, reporting the raw ratio will in a sense understate the R&D intensity of smaller countries. Similarly, Sandven argues that to the extent that levels of R&D vary systematically across industries, and industrial structure varies systematically according to the size of countries, further adjustments to the raw indexes are appropriate to better capture underlying R&D effort by taking account of such differences.

A distinction which typically structures the reporting of R&D expenditures is whether the activity is aimed at acquiring new knowledge without particular applications in mind (basic research), or the acquisition of new knowledge with particular practical aims and objectives in view (applied research), or work which draws upon existing knowledge to produce or install new goods, services and processes, or improve existing ones (experimental development). R&D expenditures can also be distinguished by the unit undertaking the activity and the source of finance e.g., governments/supra-national authorities, the private sector, the higher education sector.

Within the governmental sector, an explicit R&D or science and technology budget may be presented as part of the budgetary process, or more usually, as an *ex post* presentation of activity. Some economists advocate supplementing accounts of government expenditure in general with estimates of taxation revenue implicitly foregone by special taxation measures directed at specific policy objectives—so-called 'tax expenditures'. Warda (1998) discusses measuring the cost of R&D tax expenditures specifically, reviewing a number of approaches and providing some country estimates of the magnitude of taxes foregone in this way.

2.1.2 Human capital indicators

In the neo-classical economics tradition, human capital is thought of as an input to the innovation process, and in more evolutionary/institutional perspectives, as a key output of policy initiatives, insofar as improving human capital is central to an innovation system's capacity to absorb technological and scientific knowledge. More generally, policy makers might naturally think of the consequences of technological change for employment levels and patterns and for wage/skills differentials as important impacts of STI policies. So the aspect of the innovative process captured by such indicators depends in part on the perspective of the client group.

Standard indicators include the stocks and flows of those employed in R&D activities, of graduates in scientific and technological fields, of scientists and engineers, of those employed in specific 'high technology' sectors, or more broadly, of those employed in knowledge-based activities. For international and historical comparisons, such stocks and flows can be scaled against other aggregates, such as the total labour force, the total flow of graduates, the total in employment, the total working age population etc.

An important issue in constructing an overall index of the quality of human capital in an economy is the weighting of different categories. A neo-classical economics perspective suggests weighting different categories of those in employment by a measure of the marginal contribution of a person in such a category to economic production—assumed to be reflected in differential wages. Changes in such indexes may reflect both policy and non-policy influences. For example, all things being equal, the retirement of less qualified workers, given a constant flow of graduates into employment, would automatically raise such an index. However, it may be the case that qualifications do not proportionately reflect increases in human capital, insofar one of the roles of the education system can be to signal pre-existing productivity to the labour market, rather than actually provide it.

Evolutionary/institutional analyses emphasize that tacit knowledge is embodied in human capital, and often transmitted through personal contact. This has motivated much international attention devoted to indicators of the mobility of human capital. We discuss these indicators in Section 3.1 below.

2.2 Outputs

2.2.1 Technological intensity of production

Summary indicators of the technological intensity of production facilitate comparison of industries and countries. As explained in OECD (1992, p. 300), there are two main approaches to the task of categorisation which this involves: by industry and by product. Categorising by industry adopts the criterion of R&D intensity as measured by some index of expenditures (e.g., as a percentage of sales), to place industries into a 'high' 'medium' or 'low' technological band, although the OECD anticipate refinements of these approaches to capture more indirect aspects of technology content. The 'by product' approaches to categorisation inquires as to the R&D content of specific products, rather than whole industries. In this way, mature products in R&D intensive sectors would not receive undue weightings.

2.2.2 Balance of payments indicators

Similar approaches to technological intensity can be applied to that part of production which is exported, and to imports. Such indicators are based on conventional trade statistics, accounting for transactions in tangible goods and in services. To take account of intangible transactions relating to knowledge based trade, and in technological services, the *Technology Balance of Payments* is constructed (OECD 1992, p. 298). This covers for example purchases and sales of patents and licences, models and designs, trade-marks, technical services and the financing of R&D outside the country concerned. As is the case with balance of payments indicators more generally, a deficit or surplus by itself may have little analytical significance, insofar as it reflects voluntary transactions from which one presumes both party benefit, so that these indicators are best appreciated in the context of causal models which might account for their direction and magnitudes.

2.2.3 Bibliometrics

Kaloudis' (1998) guide to bibliometrics for policy makers (another paper from the EU funded IDEA project) introduces the topic as follows:

> Bibliometrics is more than the mere counting of publications. Bibliometrics is a set of analytical concepts and methods for measuring the regularities, the structure, the dynamics, the performance and the institutional setting of written knowledge production in the research system. Bibliometrics is also a tool for situating the research efforts of a country in relation to the world, and the research of an institution in relation to other institutions, and the research of scientific groups in relation to their own communities.
>
> (Kaloudis 1998, p. 92)

Kaloudis distinguishes three types of bibliometric indicators and techniques; absolute and relative counts of publications, analyses of the citation of works, often used as indicators of scientific quality and/or impact, and relational indicators which "... measure and map interaction patterns in the research system" (Kaloudis 1998, p. 94), by e.g., analysing patterns of co-authorship to capture collaborative efforts, co-word analysis to map relationships between documents within a research area, and the citation by patents of scientific literature. As in the case of human capital indicators, we can relate the simpler indicators such as publication counts, to more or less linear innovation models, insofar as they relate to inputs and outputs, while more elaborate techniques such as citation analyses and relational indicators can be thought of as attempts to capture the linkages and relationships stressed by evolutionary/institutional models of innovation. We discuss relational indicators especially in Section 3.1, as part of the review which points to a number of approaches to accounting for knowledge flows.

Bibliometric analyses take advantage of the availability of large bibliographical databases, such as the Science Citation Index, and of software to process the information they contain. Counts of publications and of citations can be used for the assessment of scientific performance at a number of levels, from that of the individual researcher, project team or institution, to the relative standing of national scientific research fields in an international context. Particular scientific fields vary widely in publication and citation practice and journals differ in 'quality', and analysts have developed scale factors to control for such variations. An example is the 'journal impact factor' which is a ratio of the number of citations to items in a journal in a particular year to the number of citable items in that journal.

2.2.4 Patents as indicators

Given the intrinsic difficulties of measurement in innovation adverted to earlier, patent statistics appear attractive to analysts, given their availability, like bibliometric information, in electronic databases, and the apparent directness of their relationship to innovative output. Surveys of the use of patent as indicators, such as those by Griliches (1990) and by Iversen (1998) consider in detail the associated opportunities and limitations.

Patent statistics arise from socio-legal processes not specifically designed to generate indicators *per se*. This has the advantage of utilising criteria for reporting (on the part of patent examiners) which do not involve pre-conceptions of researchers/evaluators, and which are relatively slow to change. On the other hand, not all innovations are patented, or even patentable, and when they are, industries often differ systematically in the extent to which they wish to rely on this mechanism to secure intellectual property rights. Furthermore, even those innovations which are patented are not always commercialised, nor necessarily commercialised successfully.

Sometimes raw patent data reveal more about the administrative behaviour in and constraints of patent offices than about the underlying innovative effort we are trying to assess. For example, Griliches relates how reductions in patents granted in certain periods in the United States have been misperceived as indicating productivity slowdowns. The actual causes however, were reductions in the budgets of the patent offices, leading to increases in the lag between applications for patents and grants, a result confirmed by the continued growth in applications themselves.

Elaborations of simple patent counts attempt to deal with these and other difficulties in order to provide meaningful indicators and make valid comparisons. For example, patent grants can be 'filtered', removing those not renewed after a certain period in order to arrive at a sub-set of commercialised innovations. Also, the patterns of renewals and the stock market valuations of firms may be used to estimate the economic value of patents which is not evident from patent counts alone. Similarly, attention can be focussed on those patents which have been licensed internationally, in order to capture those with potential economic impact. Another variation is to focus on 'patent-share' measures, which compare the patenting output of a particular unit (e.g. firm, industry, region, nation, technological field) to the total patenting activity of the relevant population, in order to indicate the relative innovativeness of that unit. An index of 'revealed technological advantage' may be calculated for the unit being evaluated, to indicate whether it engages in more or less patenting activity than its size relative to the total population would predict (Iversen 1998, pp.75–81).

The citation of patents to other patents and to scientific literature, intended by patent examiners and applicants to demonstrate the relationship of innovations to accumulated knowledge, affords analysts an opportunity to map relationships between technological fields and between science and technology and to investigate the determinants of such relationships. Such elaborations of the underlying indicators can be thought to reflect evolutionary/institutional concerns with the mapping of knowledge flows, as so are treated in Section 3.1 below.

2.3 Impacts

2.3.1 Growth accounting

Growth accounting is the principal empirical lens used by macroeconomists to view technological change. In arising from neo-classical economics, it is an attempt to distinguish the extent to which economic growth arises from the increased *use* of inputs, such as labour and capital, as opposed to arising from the increased *efficiency* with which those inputs are used, thought to be mainly—but not exclusively—driven by technological progress. Economists start with given data on the rate of economic growth, the accumulation of capital, the employment of labour, and make a number of assumptions about the relationship between them—as reflected in a 'production function'. These then permit the derivation of that part of growth not explained by the accumulation of capital or labour as a residual, variously referred to as 'multi-factor productivity' (MFP), 'total factor productivity' (TFP), and the 'Solow Residual', reflecting Robert Solow's (1956, 1957) pioneering of this technique. Developments of these techniques are in Boskin and Lau (1992) and Lau (1996).

Solow's work and that of others in the neo-classical tradition identified technological change in the sense of growth in MFP as the key theoretical and empirical force driving long-run increases in living standards. However, such research did not in itself attempt to explain MFP; such aggregates, Solow famously said, were a 'a measure of our ignorance'. Nevertheless, such analyses are important starting points in assessing the relative and historical performance of economies in respect of the impact of innovative activities, and provide in MFP a key aggregate to be explained by further analyses. For example, we consider below models of productivity spillovers which often take MFP or similar measures as the relevant indicator of technological change.

Insofar as growth accounting is based on a quite specific economic model, its validity is vulnerable to departures from the assumptions which underlie that model. These are typically neo-classical in nature, in assuming that resources allocation in economies is via competitive markets, in which scale confers no advantage on firms. Furthermore, the residual nature of the MFP measure means that it may reflect a host of factors other than what we might naturally think of as technological progress, such as increases in the efficiency with which markets work arising from policies which liberalise markets. Empirically, it is easy to mistake productivity increases which arise because employment shifts from low to high productivity sectors for more sustainable productivity increases across all sectors (see the early survey by Kennedy and Thirlwall (1972) on these and a range of related applied concerns).

2.3.2 Price and quality changes

Analyses such as growth accounting (and more basic exercises such as the reporting of raw economic growth rates), depend in part on the construction of appropriate price indexes. Nominal money aggregates are adjusted using such indexes in order to isolate changes in prices from changes in volume—'real changes'. Innovation, in involving both quality improvements in existing good and services and the introduction of new good and services, poses particular challenges for the construction of such indexes, especially when the rate of technological progress is rapid. Quality improvements and novelty are especially evident, and so the analytical challenge is especially acute, when assessing the economic impact of information and communications technologies (ICT).

Conventional economic theory and empirical practice is predicated on a fixed universe of goods and services of constant quality for which prices are comparable over time, and so is not well suited to handling these issues. A seminal alternative approach was that of Lancaster (1966) who thought of consumers as having preferences not over goods, but over the more primitive notions of the features which characterise those goods (e.g., rather than modelling preferences over apples and oranges, one models underlying preferences for qualities of colour, sweetness etc.). This 'characteristics' theory enables the analytic treatment of goods which are improved in respect of certain attributes, and the introduction of new goods, which comprise new combinations of existing attributes. The empirical counterparts to this theory are 'hedonic price indexes' which relate price changes to changes in the characteristics of goods.

For example, in constructing a price index for computer goods, a conventional approach would be to compare the price of personal computers over time. A hedonic price index would track the prices of a number of measurable technical attributes such as processing power, memory, speed, graphical capabilities, and/or network connectivity, and then consider a generic computer good as a weighted bundle of such attributes (Schreyer 2000, Bassanini, Scarpetta, and Visco 2000, Berndt, Griliches, and Rappaport 1993). Much of the 'new economy' debate, in seeking to estimate the impacts of ICT, revolves around the appropriateness of such adjustments, but the central idea is of general relevance to the assessment of innovation. One intriguing example is the work of Manuel Trajtenberg (1989) who adapted this basic approach in order to quantify to benefits owed to improvements in computed tomography (CT scanning) technologies.

2.3.3 Productivity spillovers

A range of econometric techniques start with the analytic framework of growth accounting, often using MFP as the indicator of technological progress, in order to evaluate the magnitude and direction of spillovers, especially those which are expected to arise when the R&D performance of one set of economic agents influences the MFP of another set.

A very large number of studies fall within this category, and there is a corresponding survey literature to provide some navigational help. For example Cameron (1996) has a particular focus on work related to the macroeconomic consequences of such spillovers, while David, Hall, and Toole (2000) review studies which estimates the impact of public R&D on private R&D, Klette, Moen, and Griliches (2000) critique papers on the impact of public subsidies on private R&D, and Hall and Van Reenen (2000) survey the literature on the effectiveness of tax incentives on private R&D[1]. While it is not possible to do justice here to the range of methodologies and results embodied in this literature, we can observe that the econometric sophistication of such exercises reflects both their strength and weaknesses as part of practical evaluation. This is because the need to carefully manage data and implement statistical procedures arises from the non-experimental nature of the data under consideration, so that the econometrician needs to be specific about the underlying model being tested. This means that such exercises require the analyst to explicitly consider the 'no-policy' alternative—the counterfactual—a scenario of direct concern to policy makers, who wish to support activity which would not have happened without intervention.

2.3.4 Quasi-financial methods/cost-benefit analyses

Cost benefit analyses find general application in evaluation in seeking to coherently account for both the positive impacts of policy initiatives, and for their opportunity costs. Such exercises typically adopt a multi-period perspective, and attempt to appropriately weight benefits/costs in relation to the time at which they accrue, through discounting techniques. By analogy with the financial analysis of private sector investment decisions, evaluators may propose benchmarks such as required rates of returns against which to compare actual rates of return for specific projects. However, the need for specialised techniques in public policy evaluation distinct from such investment appraisal arises from the non-market nature of the impacts being considered, which implies the typical unavailability of markets prices which might be used to value them.

This need for valuation information other than that which markets provide is glaring in the evaluation of STI projects, given the problems of uncertainty and lack of appropriability which characterise innovations, and which may motivate policy in the first place. For example, one standard approach to evaluating the welfare consequences of providing a particular good is to estimate 'consumers' surplus', which in principle requires statistical estimates of market demand functions, based in part on price/quantity data, which

[1] See also Coe and Helpman (1994) and Engelbrecht (1996) on international R&D spillovers.

naturally are not available for innovations which do not yet exist or have not been commercialised. Nevertheless, attempts can be made to remedy such problems, often through survey techniques which probe consumers' (or producers') willingness to pay for new goods and services and/or quality improvements and from which their valuations of outcomes can be inferred. As with other quantitatively intensive techniques, the value of cost-benefit analyses often lies in the structured account of the domain of inquiry they present for consideration, rather than in the supposed precision of specific numerical results they generate (for a discussion of these methods in the context of EU programmes, see PRAGMATA (1999)).

3 Evaluating knowledge flows and processes

This section briefly discusses approaches to evaluation, many of which build on the indicators and methods already outlined, which reflect a model of innovation as an evolutionary, dynamic process, involving flows of often tacit knowledge, embedded in specific institutional frameworks.

3.1 Knowledge flows

The international effort to account for the flow of innovation-related knowledge between participants has developed around patent citation and bibliometric analyses and on indicators of the mobility of human capital. Patent and bibliometric analyses tend to reflect flows of relatively formal, codifiable knowledge, while human capital indicators reflect a concern with the relatively informal, tacit knowledge embodied in people and acquired and transmitted by personal contact and by experience.

It is convenient to consider patent analysis and bibliometric analysis together here, since there are similarities in the information content of patents and of scientific papers; e.g., both have geographically identified and institutionally affiliated authors and/or assignees, papers cite papers, while patents cite scientific papers, as well as other patents, papers and patents can be classified by industrial/technological/scientific field. These similarities can be exploited by comparable analytic techniques, and they permit the interactions between the distinctive innovative activity represented by patents and scientific literature to be examined (see Kaloudis (1998, pp.199–130) and Iversen (1998, 82–86).

Particular applications include:

- Analyses of patterns of scientific collaboration and (through co-authorship analysis),
- analyses of the impact of scientific endeavour in one field on activity in another,
- analyses of interactions between technology and science/and or industry and public research, especially via the patent citation of non-patent literature,
- analyses of the geographical concentration of knowledge, whether in patented innovations or scientific literature and/or the strength of interactions between geographically localised knowledge in these forms.

By way of example, Caballero and Jaffe (1993) undertake an elaborate exercise to represent stocks of public knowledge, drawing upon patent citation patterns to infer both the creation and destruction (through obsolescence) of knowledge. Tijssen and Wijk (1999) consider interactions between scientific research and patented innovations in ICT in Europe, pointing to relatively poor linkages between an underlying strong scientific base and weak commercialising capacity. Zitt, Barr, Sigogneau, and Laville (1999) analyse in this framework territorial patterns of scientific and technological activity in Europe, enabling a taxonomy of regions according to the evolution of science/technology interactions. One especially imaginative application is by Hammermesh and Oster (1998) who assess the impact of ICT on academic endeavour and show that the rise of the internet has, as expected, lead to more international co-authorships, but no increase in academic quality, as measured by the bibliometric impact analyses of those papers.

The development of indicators of the mobility of human capital (especially high quality human capital) is at a relatively early stage. Akerblom (1999) evaluates the possibilities of constructing such indicators from sources such as the European Community Household Panel Survey, country labour force surveys and special country surveys and considers mobility rates in three main contexts[2]:

- mobility between firms and other organisations,
- mobility between research-producing and research-using sectors, and
- the international mobility of highly qualified personnel.

[2]For some early work on international flows of human capital, on the basis of National Science Foundation data, see Grubel and Scott (1966a) and Grubel and Scott (1966b).

The relevance of such indicators is fairly evident to policy initiatives which seek to strengthen linkages between actors in national innovation systems, encourage transnational collaborative efforts, or which are motivated by the immigration of highly skilled people from lagging regions. Furthermore, these data sources might naturally prompt investigations of the nature of codified-tacit knowledge flows, insofar as they can be linked to the patent/bibliometric exercises just discussed.

3.2 Peer review

Science is a social process, and the assent of the community of scholars to the promotion of new knowledge and the discarding of redundant results through peer review is a defining feature of that process. Peer review in the context of scientific publication can be regarded in an economic framework as an institutional development in science which substitutes partially for the unavailability of conventional market reward mechanisms and appropriability conditions, and so which encourages the welfare-enhancing dissemination of new knowledge (Stephan 1996).

Traditional peer review can be modified in the context of public policy evaluation by the availability of some of the data sources already discussed. Formal bibliometric analyses might be considered as particularly relevant in this respect, if only because an expert audience might be thought to be especially conscious of both the strengths and limitations of such approaches. Policy makers may also wish to supplement traditional peer review exercises in evaluating particular programmes with experts who act as advocates for social concerns broader than scientific quality.

Intrinsic difficulties with peer review, not confined to policy evaluation, lie inevitably with mechanisms for choosing the 'peers', to the extent that those currently with the highest prestige may have biases (conscious or unconscious) towards work within current paradigms as opposed to more speculative, but promising, research trajectories. A particular problem for small nation states and regional systems of innovation is that in understandably choosing outside peers in order to avoid the appearance or reality of favouritism, some valuable insights into specifically local institutional factors may be lost.

3.3 Surveys

Surveys in general provide the raw material for many conventional reported economic aggregates, not least in the analysis of STI (e.g., R&D expenditures, as discussed earlier). STI policy makers and researchers have undertaken considerable work in refining this central tool of economic information gathering to map and understand innovative processes, as is evident from the detailed account and evaluation by Arundel, Smith, Patel, and Sirilli (1998). They consider a number of international innovation surveys and propose models for the design of surveys, in a framework which explicitly addresses concerns raised by models of innovation along evolutionary/institutional lines. Thus the authors, while adopting an eclectic approach, generally adopt a framework for the development of survey-based innovation indicators which is structured around not only knowledge creation and dissemination, but is also concerned with the capacity of actors in an innovation system to absorb knowledge. This framework is set out in Table 1. The authors, amongst many other issues, consider best practice in the detail of survey design, taking account for example of the trade-off between the value of extra information gathered against the costs imposed on questionnaire respondents.

Table 1: The framework for the development of innovation indicators of Arundel *et al.* (1998)

	Knowledge creation	**Knowledge dissemination**	**Knowledge use and absorption**
Who	Who creates the knowledge?	Who is the carrier of the knowledge–scientists, patents, journal articles, new equipment?	Who is using the knowledge?
What (How)	What type of knowledge is created?	What specific knowledge is disseminated?	What absorptive capacities must be present to be able to use the knowledge?
Where	Where is this activity located?	Where are the linkages—where do they start from?	Where is it being used?
Why	What purpose does the knowledge serve?		Why is it being used? What purpose does it serve?
External factors	What external factors influence this activity? How do financial factors influence knowledge creation?	What external factors influence this activity?	What external factors influence this activity?

3.4 Case studies

Case studies of particular policy programmes, research efforts or national systems of innovation, as Figure 1 suggests, have the capacity to integrate information derived from many other indicators and evaluation methods, and especially to combine quantitative and qualitative information (see Georghiou and Rossner (2000) for an assessment of some particular case study evaluations). For example, survey questionnaires may contain open-ended request for observations or identification of relevant influences and impacts, while interview techniques may be especially important in eliciting localised, tacit and institutional knowledge. Similarly, survey techniques can be applied to over time to cohorts of firms/research projects which can be partly assessed in a case study framework. Furthermore, the systemic and interdependent nature of innovation adds to the case for the case study approach, in enabling the consideration of long term impacts and interactions not well captured or even anticipated by numerical measures.

The necessity for such approaches is underlined by the lack (and probable impossibility) of comprehensive formal quantitative models which could reliably capture inputs, outputs and impacts and the causal mechanisms which link them in a manner useful to policy makers and STI participants. The ability of case studies to situate innovative activity in a variety of socio-economic, political, cultural, personal, scientific and technological contexts is well illustrated by quasi-historical accounts of the origins of specific technologies, such as the consideration of the origins of the internet by Abbate (2000) and that of the development of radar technologies by Buderi (1998).

3.5 Technology foresight

Technology foresight exercises, in seeking to match the current capacities of an innovation system to future societal needs and scientific/technological opportunities, can synthesise information from a diversity of sources in a strategic framework. For example, comparative accounts of conventional innovative effort and outcomes as surveyed earlier, might enable the benchmarking of a national system against its peers and/or competitors, providing relatively transparent targets for policy action. Bibliometric, patent-based, and human capital analyses could inform such processes, in pointing especially to areas where the linkages and absorptive capacities of a national system have comparative advantages, or are especially weak. Equally, the information embodied in peer review exercises and case studies can usefully inform strategic attempts to assess current activity against futures needs.

Of course, the intrinsic difficulties of future-oriented prioritisation in a domain characterised by deep complexity and pervasive uncertainty are acute in such potentially broad exercises, so that specific techniques such as scenario analysis and the Delphi method have been developed in order to structure decision making. An obvious in-built constraint might be the reluctance of participants to reveal private information given the intended publicness of such exercises and the associated appropriability problem.

4 Conclusion: evaluation, processes and institutions

Evolutionary/institutional models of innovation emphasise that policy making in STI is best conceived of not as the application of algorithmic optimisation techniques in well-defined situations, but as a dynamic process of learning within institutions and technological environments characterised by change, complexity and uncertainty. This is taken to imply a move away from naïve notions of indicators and evaluation methods as providing one-shot summative numerical measures which can mechanically determine policy decisions, towards a view of them as part of an experimental and learning process for STI policy makers, researchers and participants alike. Furthermore, such perspectives on STI tend to make a case for a portfolio of evaluative techniques and measures, given that particular techniques capture only certain dimensions of innovative activity, whereas the over-arching goal is to address systemic performance in a holistic manner.

These general considerations are reflected in the experiences of evaluation exercises internationally, as evidenced for example in the collections of 'best practice' case studies and analyses in OECD (1997a) and OECD (1997b). These practitioners, while acknowledging the motivations underlying evaluative exercises, point to the dangers of 'evaluation fatigue', especially in contexts where such exercises appear to participants to be unrelated to the development of policy measures, or do not assist in the improvement of research programmes and/or the technological activities of firms. Such concerns inform the embedding of evaluation as a continual part of policy development, as discussed by Guy and Arnold (1998). Some of the tensions which emerge here are rooted in perceived tensions between the essentially open-ended nature of the search for new knowledge, and public management agendas which privilege demonstrable and quantifiable indicators of accountability.

From an economic point of view, any 'evaluation of evaluation' would acknowledge that such exercises themselves utilise resources otherwise applicable to the substantive concerns of STI, that evaluation exercises can implicitly set up more or less desirable incentives, and that among these is the creation of academic, public sector and private sector constituencies—an 'evaluation industry'— whose interests may not be precisely aligned with the broadest public policy goals. It is against such costs that the potential for such techniques to inform policy, and more generally to illuminate the fascinating terrain of STI activities, should be assessed.

References

Abbate, J. (2000): *Inventing the Internet.* MIT Press, Cambridge, Massachusetts and London.

Akerblom, M. (1999): "Mobility of highly qualified manpower: a feasibility study on the possibilities to construct internationally comparable indicators," Report for Camire (Eurostat) and the OECD (Bluesky project on S&T indicators).

Arnold, E., and K. Guy (1997): "Technology diffusion programmes and the challenge for evaluation," in *Policy Evaluation in Innovation and Technology: Towards Best Practice*. OECD, Paris.

Arundel, A., K. Smith, P. Patel, and G. Sirilli (1998): "The Future of Innovation Measurement in Europe," IDEA Paper No.3 STEP Group.

Bassanini, A., S. Scarpetta, and I. Visco (2000): "Knowledge, technology and economic growth: recent evidence from OECD countries," OECD STI working papers.

Berndt, E. R., Z. Griliches, and N. Rappaport (1993): "Econometric estimates of price indexes for personal computers in the 1990s," NBER working paper No. 4549.

Boskin, M. J., and L. J. Lau (1992): "Capital, technology and economic growth," in *Technology and the Wealth of Nations*, ed. by N. Rosenberg, R. Landau, and D. C. Mowery. Stanford University Press, Stanford.

Buderi, R. (1998): *The Invention that Changed the World: the Story of Radar from War to Peace.* Abacus, London, 3 edn.

Caballero, R. J., and A. B. Jaffe (1993): "How high are the giants shoulders? An empirical assessment of knowledge spillovers and creative destruction in a model of economic growth," in *NBER Macroeconomics Annual*, ed. by O. J. Blanchard, and S. Fischer. MIT Press, Cambridge, Massachusetts and London.

Cameron, G. (1996): "Innovation and economic growth," Centre for Economic Performance Discussion Paper No. 277.

Coe, D. T., and E. Helpman (1994): "International R&D spillovers," *European Economic Review*, 39(859-887).

David, P. A., B. H. Hall, and A. A. Toole (2000): "Is public R&D a complement or substitute for private R&D? A review of the econometric evidence," *Research Policy*, 29, 497–529.

ENGELBRECHT, H.-J. (1996): "International R&D spillovers, human capital and productivity in OECD economies: an empirical investigation," *European Economic Review*, 41, 1479–1488.

GEORGHIOU, L., AND D. ROSSNER (2000): "Evaluating technology programs: tools and methods," *Research Policy*, 29, 657–678.

GRILICHES, Z. (1990): "Patent statistics as economic indicators: a survey," *Journal of Economic Literature*, 28(4), 1661–1707.

GRUBEL, H. B., AND A. D. SCOTT (1966a): "The immigration of scientists and engineers to the United States 1949–61," *Journal of Political Economy*, 74(4), 368–378.

——— (1966b): "The international flow of human capital," *American Economic Review*, 56(1/2), 268–274.

GUY, K., AND E. ARNOLD (1998): "Strategic Options for the Evaluation of R&D Programmes of the European Union," Report for Scientific and Technological Options Assessment (STOA) unit, European Parliament.

HALL, B. H., AND J. VAN REENEN (2000): "How effective are fiscal incentives for R&D? A review of the evidence," *Research Policy*, 29, 449–469.

HAMMERMESH, D. S., AND S. M. OSTER (1998): "Tools or toys? The impact of high technology on scholarly productivity," NBER Working paper No. 6761.

IVERSEN, E. J. (1998): "Understanding innovation indicators based on patents," in *Science, Technology and Innovation Indicators: a Guide for Policy Makers*, ed. by K. Smith. IDEA, STEP Group, Oslo.

KALOUDIS, A. (1998): "Bibliometrics," in *Science, Technology and Innovation Indicators: a Guide for Policy Makers*, ed. by K. Smith. IDEA, STEP Group, Oslo.

KENNEDY, C., AND A. THIRLWALL (1972): "Surveys in applied economics: technical progress," *Economic Journal*, 82(325), 11–72.

KLETTE, T. J., J. MOEN, AND Z. GRILICHES (2000): "Do subsidies to commercial R&D reduce market failures? Microeconometric evaluation studies," *Research Policy*, 29, 471–495.

LANCASTER, K. J. (1966): "A new approach to consumer theory," *Journal of Political Economy*, 74(2), 132–157.

LAU, L. J. (1996): "The sources of long-term economic growth: observations from the experience of developed and developing countries," in *The*

Mosaic of Economic Growth, ed. by R. Landau, T. Taylor, and G. Wright. Stanford University Press, Stanford.

OECD (1992): "Consequences of the Technology/Economy Programme for the Development of Indicators," in *Technology and the Economy: the Key Relationships*. OECD, Paris.

——— (1997a): *The Evaluation of Scientific Research: Selected Experiences*. OECD, Paris.

——— (1997b): *Policy Evaluation in Innovation and Technology: Towards Best Practice*. OECD, Paris.

PRAGMATA (1999): "Cost-benefit analysis of EU research and technological development: methods for the assessment of the impact of such programmes," Working document for Scientific and Technological Options Assessment (STOA), European Parliament.

Sandven, T. (1998): "R&D Data and R&D Indicators," in *Science, Technology and Innovation Indicators: a Guide for Policy Makers*, ed. by K. Smith. IDEA, STEP Group, Oslo.

Schreyer, P. (2000): "The contribution of information and communication technology to output growth: a study of the G7 countries," OECD Directorate for STI working paper series.

Smith, K. (1998): "Innovation indicators," in *Science, Technology and Innovation Indicators: a Guide for Policy Makers*. IDEA, STEP Group, Oslo.

Solow, R. M. (1956): "A contribution to the theory of economic growth," *Quarterly Journal of Economics*, 70(1), 65–94.

——— (1957): "Technical change and the aggregate production function," *Review of Economics and Statistics*, 39(3), 312–320.

Stephan, P. E. (1996): "The economics of science," *Journal of Economic Literature*, 34, 1199–1235.

Tijssen, R. J., and v. Wijk (1999): "In search of the European Paradox: an international comparison of Europe's scientific performance and knowledge flows in information and communications technologies research," *Research Policy*, 28, 519–543.

Trajtenberg, M. (1989): "The welfare analysis of product innovations, with an application to computed tomography scanners," *Journal of Political Economy*, 97(2), 444–479.

Warda, J. (1998): "Measuring the cost of R&D tax incentives," OECD Room Document No.4 (Bluesky project on S&T indicators).

Zitt, M., R. Barr, A. Sigogneau, and A. Laville (1999): "Territorial concentration and evolution of science and technology activities in the European Union: a descriptive analysis," *Research Policy*, 28, 545–562.