The Evaluation of Research Done in Post Secondary Institutions

Cooper H. Langford
Science, Technology, and Society Program
University of Calgary

1. Introduction.

The first issue is to set out the reasons that post secondary institutions (PSIs) do research. A summary list can probably serve present purposes:

0.1 advancing disciplines at the world frontier of emergence;
0.2 maintaining and enhancing professional competence;
0.3 configuration and generation of knowledge for strategic purposes;
0.4 enhancing societal capacity for innovation and critical analysis – research training;
0.5 maintaining an open channel to the world knowledge system;
0.6 informing and enhancing the quality of undergraduate education.

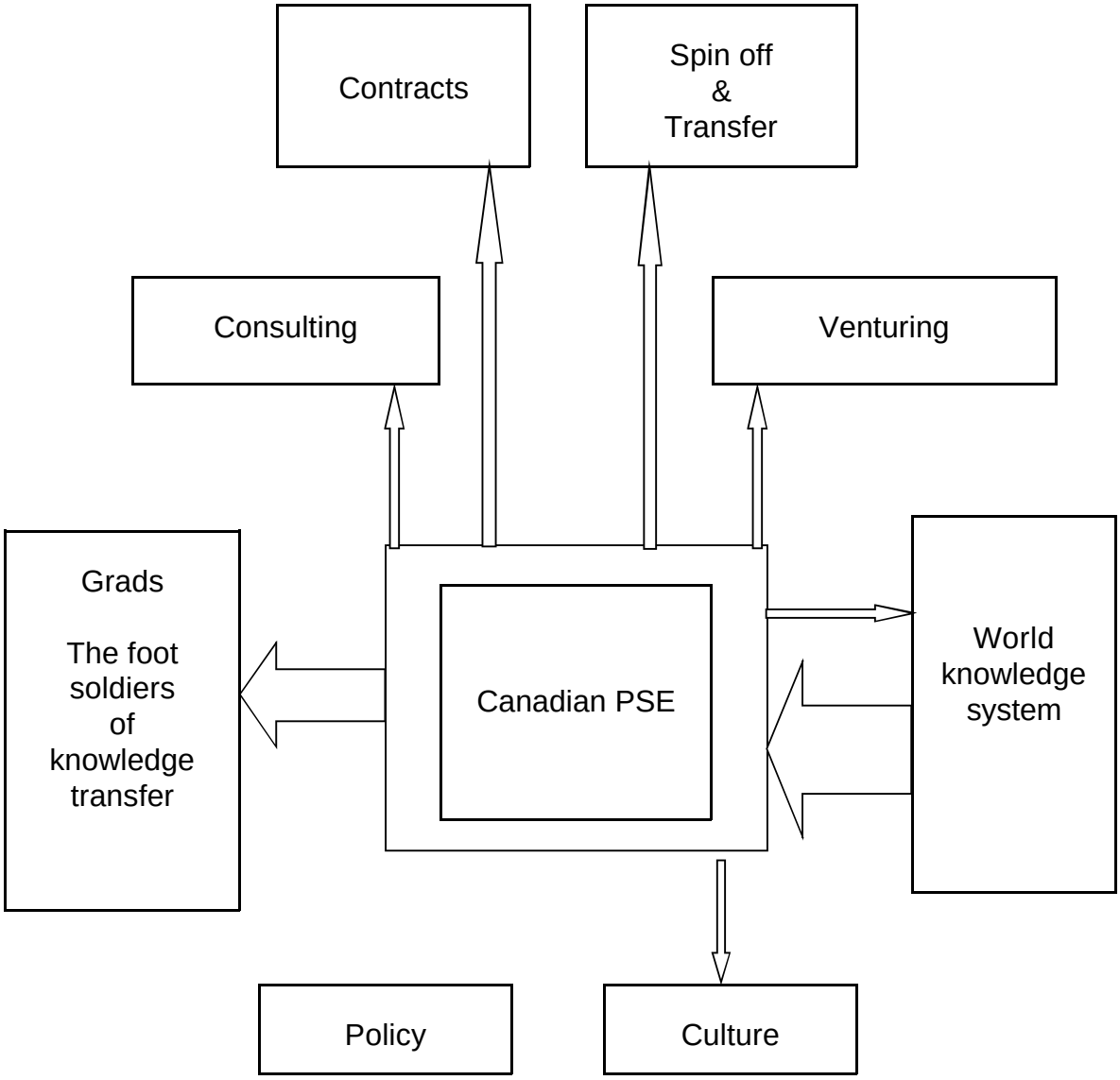
The rationale for public support of these activities can be identified with **outcomes** that flow from these activities. The viewpoint of this paper is that the the outcomes that command the attention of senior policy officials are sufficiently comprehensive that the *focus for development of a framework for quality and accountability can be limited to the issue of outcomes.* One limitation that this view imposes will be that the measures discussed will not include those used to select particular research initiatives (e.g. projects) for sponsorship. That issue has already been well analyzed by various federal and provincial research sponsoring agencies. A second limitation is that only minor attention will be given to input measures. It is not in doubt that these will continue to be collected. The present view is that these measure the scale and scope of activity in a jurisdiction and have little to do with the issues of quality and accountability. Their role in public policy is not negligible. An insufficient level of investment can doom any attempt to achieve quality outcomes, but investment is not in itself a measure of quality.

It should be stated immediately that the interpretation used covers, under the outcomes commanding the attention of senior government policy makers, undergraduate education and cultural development.

A word of warning. In our household, we use the caller identification feature on our telephone to avoid market researchers. The literature on evaluation of research increasingly includes discussion of “evaluation fatigue” on the part of both those evaluated and the people from whom inputs are solicited.

2. The impacts and outcomes of research done in post secondary institutions.

The first step toward analysis of outcomes of research in post secondary institutions (PSIs) is to analyze the profile of the impacts of PSI research so that the desired outcomes can be identified and prioritized. The diagram suggests that PSI research is a node in a network with at least eight important links. The two large boxes on the left and right, development of highly qualified personnel (including undergraduate education) and contribution to the world knowledge system, are commonly recognized. The other six have received less attention. Each deserves some analysis.



2.1 Links to the world knowledge system – the intelligence function.

About 4.5% of the publication output of research institutions worldwide is contributed from Canadian institutions. Publication, with associated informal communication, is the bloodstream of the scholarly dialogues and critiques that convert research into knowledge through the emergence of consensus.[1] It is clear that Canadian PSIs are not a dominant voice in the world system. Nevertheless, participation by Canadian researchers is quite important to other outcomes of PSI research in Canada. Through contribution to the world dialogue, Canadians earn a place at the tables where consensus emerges and gain a deep appreciation of current knowledge. PSIs become a storehouse of contemporary knowledge that can be transferred into other sectors in Canada (commonly by people movement) to fuel a knowledge economy and enrich culture. Metaphorically, the PSI researchers function as an **intelligence agency** for

[1] Ziman, John, *Reliable Knowledge*, Cambridge University Press, 1978. P.6-8.

Canada, gathering information from around the world and making it systematically available.

2.2 Highly qualified personnel.

Knowledge transfer is a contact sport. The most effective path for knowledge to flow from PSIs is movement of people. Both research students and undergraduates (who participate mainly in classroom, not laboratory, experiences) can be influenced by the PSI's research mission. Flows of graduates are easily identified, but it is less clear how to assess their impact in private and government sectors or the degree to which their post-secondary experience has included contact with research[2]. As well, there is a research-training component in the experience of research associates, temporary faculty, and technicians that needs to be accounted but does not show in graduation statistics.

2.3 Venturing.

A recent study of economic impacts of research at the University of Calgary[3] indicates that the launch of new commercial or not for profit ventures by university faculty is relatively common. (>200 in twenty years at Calgary). Most of these are not what is usually meant by "spin off" companies or "technology transfer". More often, ventures flow from an opportunity arising out of a faculty member's accumulated professional and research experience rather than from one particular discovery or technology package. The venture results from a growing recognition of a gap in the economy and/or society that can be filled by an activity based on the professional competencies that a faculty member has developed over time. The emergence of the venture cannot be effectively influenced by the usual agencies for "technology transfer" because the process that leads toward the venture is "sub-terranean". This is a point that has received very little attention despite its major role. Indeed it may be one of the most important outcomes of university research activity.

2.4 Consulting.

The Calgary studies[3] emphasize the large role played by expert consulting, both paid and unpaid, in the transfer of university knowledge. There is high consistency between the view of this process from university faculty and from their clients. In the job creation dimension of the economic impact study, clients of university consultants attributed more than 15% of job creation captured by the study to consulting. If the impact of formal "on campus" clinics such as Calgary's *Venture development Clinic* is added, consulting in one form or another accounts for almost half of identified job creation.

---

[2] A very recent OECD appraisal of university education in Denmark remarks, once again, that it is difficult to detect the influence of research on the experience of first year students.

[3] Chrisman, James, *Economic benefits to Alberta from the Faculty of the University of Calgary*, University of Calgary, Calgary, 1994. This study and two others: Unrau, Yvonne, and McDonald, Jack, *The Frequency, Nature, and Impact of Faculty Influence on Policy External to the University*, 1995; and Park, Elaine V. *Connected Energies, Calgary and its University* (on cultural impact), 1998, provide the basis for sections 2.3 through 2.8.

2.5 Contract research.

Contract research is a significant factor. It promotes direct communication between researchers in PSIs and researchers in government and industry sectors. The extent of knowledge transfer (both ways) is sensitively dependent upon the nature of the relationship. Longer-term relationships and ones of a program rather than short project character have greater impact. Thus, activities including industrial research chairs and research consortia are usually the most productive.

2.6 Spin-off and technology transfer.

The terms spin-off and technology transfer are used here in the limited sense of formation of a new enterprise, or licensing to an established firm, based on specific outputs of a research program such as a patentable technology or a focused technology package. The development exploited must be based on activity carried out formally in the PSI. This is an area that has received substantial recent attention. The Chrisman study[3] suggests that it is a factor comparable to the previous two*, but not more important,* in economic impact.

2.7 Policy analysis and formation.

A major impact of research done in PSIs is felt in policy analysis and formation in both public and private organizations. The Calgary study[3] found that more than 70% of faculty members participated in external policy formation over the two-year period surveyed. A similar survey conducted in Quebec reached parallel conclusions. In general, clients shared the same perception of the impact of faculty contributions that the faculty members held.

2.8 Artistic culture.

Most PSIs would wish to include creative activity of an artistic or similar character within the domain of research. Some Quebec universities use the title “Vice recteur a la récherche et a la création” for their senior officer. In most communities, this activity of PSIs plays a quite significant role in the cultural life of the community, an area of concern for policy makers.

3. Definitional problems surrounding quality and accountability measurement.

The eight-fold framework set out in section 2 identifies impacts and outcomes sufficiently broadly that the clear focus on outcomes proposed above seems justified.

3.1 Speciation of quality measures.

Quality is not a homogeneous concept. No statement about quality is value free. Within PSIs there is continual debate about relative quality of different elements of the institution based as much on the different value systems as on the different styles of presentation of the outcomes of research (publication cultures). At least three perceptions of university roles determine the value systems that pre-condition any quality assessment[4].

[4] These three represent an elaboration of a classification proposed by Robert Birnbaum.

1. *The discipline/professional community view*: Institutional conformity to professional and scholarly norms[5]. The outcomes of the university most related to this view are the linkages to the world knowledge system (2.1) and the character and capacities of graduates (2.2).

2. *The community interest view:* The degree to which important collective constituencies are satisfied. The outcomes most related to this view include the consulting, contract, policy impact, cultural impact, (2.4 –2.6) and (again) character and capacities of graduates (2.2). Where the collective constituency is government, venturing and spin-off impacts (2.3 and 2.7) may become importantly related to government economic development policy.

3. *The personal development view*. The degree to which programs contribute to the personal development of students and other clients. The outcome most related to this view is, once again, character and capacity of graduates and others who have gained experience within PSIs (2.2).

Each of the value systems that drive each of these three perspectives will be important to *senior policy makers* concerned with establishment of a quality framework. For example, it is clear that the value systems of the international invisible colleges of the disciplines strongly influence the market for human resources for research. As well, these value systems are critical to the "intelligence function" associated with the transfer of knowledge into specific institutions so that impact can arise through the distribution of knowledge. The value systems of the various constituencies in the community, which define "community interest", are important because those constituencies play social roles that influence *government* agendas. The issues surrounding personal development are important to any government with a human resource policy.

3.2 Assessment agencies

The first agency of quality assessment is **the institution**. An effective post secondary institution will construct a quality accounting procedure with the goals of supporting resource management decisions and motivating unit performance. At the institutional level, the first step in such a process will be the identification of areas of excellence and areas of opportunity. Neither of these need correspond to categories shared by other institutions. Indeed, a good "competitive" strategy may suggest identification of configurations specific to the institution, just as effective corporations try to decide "what business are **we** in". *It can be advantageous that other related institutions would classify and focus differently*. For example, a university with a chemistry program may wish to focus on the presence of a group of internationally leading inorganic chemists precisely because its neighbour can identify a comparable strength in organic chemistry. It will be wise to ignore a tendency to lump the two into a broad category labeled "chemistry" (even though the Science Citation Index lumps "chemistry" in the tables it wishes to sell).

---

[5] These are most often the norms of "academic science" identified by Robert K. Merton (Merton, R.K. *J. Legal and Political Sociology*, **1942**, *1*, 115-26.) and summarized by John Ziman under the powerfully suggestive acronym "CUDOS" (*Prometheus Bound* , Cambridge University Press, Cambridge, 1994. P. 175-9.)

Given the tremendous variety of activity encompassed by a modern PSI, the first step in an institutional program must be *a request to research units to define clearly their separate goals*. Having defined those goals, the expertise of the units can be exploited to try to define the measures of outcomes that would meet these goals. This done, it becomes possible to identify research units against which the particular one can benchmark. A wise final step in the sequence is establishment of a process for validation of opinions generated within the research unit that exploits peer review. The process can be summarized:

- research units define goals;
- identify measures of goal achievement appropriate to the unit;
- benchmark against relevant, parallel (goal sharing), research units;
- peer review the validity of the products of the process.

This is a process focused on maximization of performance in one institution. In its focus on benchmarking the individual research units, it pays little direct attention to the question of benchmarking the institution overall. Indeed, the focus is *differentiation*. If the process is followed in a number of institutions in a jurisdiction (a province or even Canada) it creates, quite self-consciously, a problem of comparing apples and oranges for those who would compare the quality of the institutions within the jurisdiction. At the same time, it meets objectives of senior policy makers. *It supports institutional focus, avoidance of unnecessary duplication, and effective internal resource distribution*.

The second agency of quality assessment, is the **funding agency** for institutions. In Canada, this is predominately the provincial department responsible for post-secondary education. At first sight, this level has rather different concerns. It is charged with resolving issues of resource allocation and seeks more often to benchmark institutions rather than research units. It will focus attention on aggregate information. It will, within the several value systems it recognizes, seek to compare institutions within its jurisdiction to each other and to standards set by institutions in relevant benchmark jurisdictions. There will be a drive to define categories that can be applied across institutions in a uniform ("fair") manner. *This creates a disjunction between institution and funding agency approaches.* The institution will be oriented to more idiosyncratic metrics. The agency will seek general ones. This is a disjunction that merits very careful management. The agency shares the institution's interest in maximizing the effectiveness and productivity of the individual institution. It must be concerned not to create general metrics for evaluation that constrain institutional improvement and specialization. If agency measures were to work against institutional differentiation, the final goal of the agency, namely the overall effectiveness and efficiency of activity in the jurisdiction, would be compromised. There is no doubt that one of the goals of senior policy makers is to encourage effective management in the institutions. As the framework paper for this exercise put it, a motivation for assessment is: to "influence continuous improvement in [institutional] management".

3.3 The nature of research activity.

A critical feature of the emergence of the contemporary "knowledge economy" is the rise of a new style of knowledge creation. This style is a problem rather than discipline oriented mode. It is interdisciplinary. It often gives pride of place to configuration of knowledge over discovery. Yet it clearly reflects knowledge creation. It has been called

"mode 2"[6]. The better understood "mode 1" responds to the dynamics of disciplines and focuses on discovery at the surface of emergence of the discipline. Both modes are essential to the complex of research outcomes, but most of the indicators that are currently favoured in quality assessment are much more responsive to mode 1 than to mode 2. In part this arises from the fact that mode 2 knowledge production relates to the fact that no single corporate body in our society can any longer satisfy its knowledge need internally. Knowledge creation is becoming a more collaborative activity. In academia, there is a strong tradition of sharing knowledge. The product of research is seen as common property, but great credit goes to the originators. The owner of knowledge is not identified in mode 1, but the creator is celebrated[7]. In mode 2, ownership may find a locale, but creation is hard to place. Objectives are institutional more than individual[8]. The team is central. This implies a measurement asymmetry. In mode 1, creation is easily measured. In mode 2, it is less easily localized. A danger arises as mode 2 becomes an equal partner in knowledge creation. An academic institution that does not participate actively in mode 2 runs a risk of becoming irrelevant to 50% of contemporary research. Measures of mode 2 activity are badly needed.

3.4 Accountability: for whom?

It has been said that a business is accountable to its suppliers, its customers, its employees, its shareholders, and the general public. PSIs are accountable for their research quality to their sponsors (governments and private), their suppliers in the world knowledge system, their students and employees, and their communities. The outcomes enumerated do affect all of these constituencies and it is a reasonable role for senior policy makers to ensure a system of evaluation that answers to all these dimensions of accountability. Sharing, with institutions, this broad responsibility does not create conflict with the concerns of government.

4. What is evaluated?

Over time, evaluations of quality and accountability regimes have evolved from a focus on inputs and processes to embrace outcomes. *The present paper advocates a clear decision that the evaluations under discussion (which do not face the need to predict the future in the way a grant selection committee must) are about* ***outcomes***. The Framework paper, which set the terms of reference for this study, notes that "many measures of outcomes are indicative only". This can be read as if it were an excuse to continue to measure inputs and outputs. The argument here is that the wrong "excuse" is adduced. During the work of the Alberta task force that proposed the current research "key performance indicators", the greatest problem confronted by the members was identification of **databases** which would allow measurement of the aspects of performance sought. It rapidly became clear to the task force that **surrogates** would be central to any final system of indicators[9]. The excuse for continuing to use indicators of inputs and outputs must be that they serve as **surrogates** for **outcomes**. If this point has been little discussed in the literature of

---

[6] M. Gibbons, C. Limoges, H. Nowotny, S. Schwartzman, P. Scott, and M. Trow, *The New Production of Knowledge*, Sage, London, 1994.
[7] Here, the values noted in footnote 5 are spelled out, to some degree.
[8] This is a value system that Ziman, *op cit* has labeled with the suggestive acronym PLACE and described as the value system of industrial research.
[9] I thank Mr. Robert W. Martin, chair of the task force, for these crucial insights.

research evaluation, it is not novel. Econometricians know well that GDP is not an end in itself but a surrogate for economic goals that are hard to quantify. It is, then, necessary to identify and justify the surrogate relationships sought. In some cases this is fairly straightforward. In others it is subtle in the extreme and great caution is required.

Some examples of the inputs and outputs that are popular for measurement include the following targets:

Inputs
- sponsored research funding
- research complement (number, qualification, discipline structure)
- research trainees (graduate students and PDFs)
- research infrastructure investments.

Outputs
- publications, etc.
- patents, designs, and copyrights – including their realization in licensing agreements, etc.
- numbers and target destinations of graduates at M.Sc. and Ph.D. levels.

If any of these measures are to be incorporated into indicator schemes, it will be necessary to be clear about the **outcomes** for which they function as statistically "robust" **surrogates**. For example, the Alberta task force judged that success rates in peer reviewed national granting council competitions were a reasonable surrogate for the degree to which Alberta institutions were connected to the world knowledge network and could deliver the outcomes dependent on that linkage (see sections 2.1 and 3.1.1).

The identification of surrogates when using statistics at the level of an institution has a critical dimension. Research cultures vary radically across each institution. The presence of a medical faculty will, for example, profoundly skew any institutional statistics, which involve dollars of research sponsorship. On the other end of the scale, dollar based statistics will under represent activity in the social sciences and, more profoundly, the humanities and fine arts. If dollar denominated statistics are to be used in inter-institutional comparison, it will be necessary to show the **structural similarity** (as to distribution of effort among fields) of the institutions so compared[9]. Fortunately, it is possible to classify Canadian universities into fairly homogeneous structural groupings so that there is some hope for use of readily accessible dollar based statistics as surrogates[9]. (It is relevant that the Alberta key performance indicators extend the recommended intra-institutional practice of self-defining benchmark groups to the institutional level. Each institution has some opportunity to propose its benchmark group, thereby addressing the need for dollar comparison to institutions with parallel structure.)

One of the continuing foci of evaluations has been *peer review* to an international standard. From the UK quadrennial research evaluation exercise to Finnish efforts, international standard peer review has been a central element of what is measured[10]. What outcome does it represent? Clearly, it is most often, given the peers chosen, a surrogate for the effectiveness of the connections to the world knowledge system. It

[10] OECD, *The Evaluation of Research: Selected Experiences*, OECD, Paris, document OECD/GD(97)194, 1997.

measures the capacity of the domestic research institutions to contribute to disciplinary discovery and its associated outcomes in both the human resource and knowledge flow areas (See sections 2.1 and 2.2). These are important but the emphasis on peer review does reflect the CUDOS value system of academic science, which has an Achilles heel in the degree to which the traditions of science celebrate the "heroic" individual discoverer. As the sociology of science has known for some years[11], multiple discovery is much more the rule than is usually acknowledged. The reason for the heroic tradition probably has more to do with a socialization process that promises to entrants to the scientific community "fame" as the reward for originality. Thus, peer review tends to be skewed to recognition of individual accomplishment in contrast to some of the collective outcomes that may be favoured by policy makers.

Without denying the value of the expertise brought by peer review, it is necessary to ask who is a peer. There is a growing tendency to open the peer review process to "experts" from outside the value system of academic science. This is an important development. It may be possible to construct peer processes much more closely related to "mode 2" knowledge production if means can be found to characterize the key components in interdisciplinary/intersectoral collaborations and enlist cooperation in identifying the factors of successful outcomes.

Behind these general remarks lies the whole literature of the social studies of science. In the narrow framework of the present paper, the above brief, sometimes oracular, remarks must serve to set the frame of reference. Therefore, we now turn to analysis of indicators available for each outcome category.

5. Measures of exploitation of the world knowledge system

Ensuring a strong connection of Canada and the several provinces to the world knowledge system through PSI research effort is the locus of most of the well-established measures of research performance. The key assumption is that the opinion that other members of this system have of Canadian contributions reflects, and is a robust surrogate for, the stock available to Canada. The measures of the opinion of Canadian contributions held by the larger community in the knowledge system depend on the role played by the academic science value of communal ownership of knowledge. This generates the emphasis on publication and the value placed on criticism and debate as the route to reliable knowledge. This makes bibliometrics a major tool for assessing a Canadian group against world standards. Its major competitor is analysis of outcomes of peer review.

5.1 Citation analysis

The Science Citation Index provides the opportunity to measure the linkages in published knowledge development. The literature is now large and has become quite sophisticated[12]. This discussion will present several conclusions without detailed development of their rationale. Citation analysis has two components, the interpretation of total citations and interpretation of citation impact. Total citations are

---

[11] Merton, R.K. *Singletons and Multiples in Scientific Discovery*, Proc. Amer. Phil. Soc. **1961**, *105 (5)*, 470-486.

[12] See, for example, Leydesdorff, L. *The Challenge of Scientometrics; The Development, Measurement, and Self-organization of Scientific Communication*, DSWO Press, Leiden, 1995.

an extensive measure of the overall involvement of the unit measured in the knowledge system under study. Citation impact (the ratio, citations to a unit / publication from a unit) is an intensive measure which seeks to identify the influence of the work. Both are useful.

Citation analysis is probably most useful in evaluations conducted at a *meso* level. There is general agreement that it is not too useful at the *micro* level of the individual researcher or individual research unit (e.g. professor, research students, and research associates). At the same time, it has severe limits when applied to the *macro* level of institutions. Citation cultures differ widely among disciplines and the available data are concentrated in the natural sciences with much less coverage of social sciences and humanities. Consequently, citation analysis at the institutional level will magnify the role of certain disciplines. *If it to be used at the institutional or jurisdiction levels, it will be critical to address the issue of benchmarking against other institutions or jurisdictions of comparable structure.* This principle is, for example, imbedded in the Alberta KPI system by allowing institutions to nominate the benchmark group of non-Alberta institutions. Magnification of the role of natural sciences is not, however, avoided.

At the *meso* level of a discipline, the value of citation analysis can be maximized. The publication cultures are reasonably homogeneous and groupings are frequently large enough to support statistical robustness. The only major problem is the definition of a discipline. Of necessity, citation analysis defines a discipline by constructing a list of journals. The contribution and influence of an institution will be measured by the publication and citation of work with author addresses at the institution that appears in the listed journals. Ideally, an individual evaluation exercise should exploit the raw citation database with the opportunity to tailor the list of journals to specific purposes. Unfortunately, this is expensive. An extreme (and amusing) illustration of the problem arose recently at the University of Calgary. A citation impact study produced the result that dentistry is a strong discipline at the university. Calgary lacks a dental faculty. Some researchers in the department of physiology had published important and controversial work on the toxicology of dental amalgams.

There is one area in which citation analysis is an almost unique tool. It is only in the international literature that readily quantifiable aspects of comparisons among countries can emerge. Within the Canadian system, other processes (see peer review below) can provide can provide comparative data. Few such processes work across national boundaries. A sub-set of the international issues that can be addressed by bibliometrics is the pattern of international interaction and international knowledge flows. This issue is susceptible to analysis by study of patterns of co-authorship. Within Canada co-authorship anlysis can be extended to study of relationships among PSI research effort in the different provinces.

Two technical matters must be mentioned. First, the Citation indexes are highly flawed. Names, addresses, and other pertinent information are often confused. To achieve reliability, especially as the size of the unit under analysis decreases, it is very important to carry out painstaking and time-consuming data clean up. In Canada, an important initiative in this area has been undertaken by *L'observatoire des sciences et des technologies* at UQAM. (For a methodological analysis of bibliometrics, see É. Gautier, *Bibliometric Analysis of Scientific and Technological Research*, Statistics Canada working paper # 88F0006XPB No. 8, Ottawa, 1998.) Second, the time sequence of development of citation information must be respected. Citation analysis has little to say about this year's work. Given the inevitable delays in conducting any

evaluation process, the time lags in citation analysis make for quite late delivery of evaluation results.

5.2 Peer review

At the *micro* level of individual researcher and individual research unit, peer review is the main tool of evaluation. It is most often used in a predictive mode to select projects for support. However, over time, the accumulated results of decisions made on support of research initiatives can become robust surrogates for standing in and contribution to the world knowledge system. Analysis of patterns of peer review decisions can become a quantitative tool. Especially in the Canadian granting council system, where track record plays a large role in decisions, patterns of success with proposals becomes a valuable measure. There are two measurables. Total dollars awarded convolutes issues of research quality, research cost, and national research priorities. Measures of success rate focus more sharply on quality, but must be analyzed in the context of demand levels and impact of resource decisions made in funding the councils themselves.

In broadly based evaluations, as for example the institutional level, council success rates are the most useful tool. They can provide information about a broad spectrum of disciplines. However, they provide little or no information about the relationships among disciplines that fall within the mandate of different councils. There are significant differences in procedure and resource environment among the councils. Not all awards made by councils are of equal interest. The awards which best reflect perceptions of impact of research in the world knowledge system are those of the core programs that have an orientation to individual researchers. It is success rates in these programs that are the best surrogate for the outcome under discussion. The main problem to keep in mind is the composition of the review panels. In the main, they are constructed along lines of traditional disciplines and can undervalue research conducted in boundary areas. There is ample circumstantial evidence that researchers in boundary and interdisciplinary areas are disadvantaged in the competitions.

The dollar values of council awards are not irrelevant. In addition to reflecting costs of research, they do include a factor indicating priority attached to completion of an initiative and its importance. It is only necessary to guard against reducing this factor to a simple input measure.

Recommended indicators:

Citation impacts – meso level, institutional level (caution), national level.
Total citations – meso level.
Co-authorship analysis of knowledge flows.
Core program federal council success rates – meso level, institutional level (desegregated by council), provincial level.

6. Highly qualified personnel

The largest group of individuals who are influenced by the PSI research activity is the cohort of university graduates. The ones for whom the influence of research activity is most readily traceable are those receiving graduate degrees. Consequently, important indicators can be constructed based on the experience of recipients of Masters and

Doctoral degrees. The National Graduates Survey[13], conducted by Statistics Canada with HRDC sponsorship is the key existing tool. It currently explores employment rate, employment type, job satisfaction, and satisfaction with educational experiences. The survey is conducted approximately every four years and surveys graduates two years after graduation. (The study of the class of 1990 was supplemented by a follow-up survey of 1990 graduates in 1995 to determine changes between the second and fifth career year.) The results are analyzed in detail, for many specific areas of study, in the "Job Futures" publication of HRDC. At present, the graduate survey cannot be analyzed at the individual institution level. However, it provides the model of a first step toward quantitative indicators. The employment rate and job satisfaction items are important indicators of successful transfer of skills formed in PSI research environments. The linkage between PSI program and "discipline" of employment is more problematical. It is not obvious that the optimal outcomes are obtained by avoidance of migration across "discipline" boundaries.

A number of institutions conduct exit surveys of their own graduates. These can become useful. They may offer an opportunity to develop information on the impact of research on the broad student population, not just thesis students. Research is needed on the nature of questions that would elicit valuable information. A critical problem is the degree to which assessments can evaluate value added. (There is a cynical observation that: "Harvard specializes in making silk purses from silk purses".) The Centre for Educational Research and Assessment at Guelph has reviewed outcome measures identified with change in knowledge, values, and skills[14]. All of the institutions that attempt to measure these outcomes use self-assessed questions in exit surveys.

The Canadian Association for Graduate Study (graduate deans) collects information on graduate study. At present this is done manually. More useful information might be generated if the project could be undertaken by a statistical agency, e.g. Statistics Canada. Certain discipline groups have collected data of a more longitudinal character as a part of an effort to understand the human resource dynamics of their disciplines. The survey conducted by the Council of Canadian University Chemistry Chairs is an example.

Recommended indicators (the first two require enhancement of the graduate survey):

Graduate degree holder employment.
Graduate degree holder satisfaction.
Employment outcomes from the National Graduates Survey;
Institutional and regional survey research.

7. Venturing

Venturing is an outcome that is hard to measure, and about which available statistics are hard to interpret. Where reporting has been attempted, completeness has been extremely variable and definitions have differed. A part of what is here included in

[13] See http://www.hrdc-drhc.gc.ca/hrdc/stratpol/arbsite/research/class90
[14] See http://www.css.uoguelph.ca/cera/PSE/OUTCOMES

venturing is included as a subject in section 5 of the Statistics Canada survey *Intellectual Property Commercialization in the Higher Education Sector, 1998.* It follows from the distinction that is made in the survey between start-up and spin-off companies. The term start-up is limited to new firms that are dependent upon licensing the institution's intellectual property. New firms that sponsor further research in the university with a view to ultimate commercialization or that offer services previously offered in institutional departments are classed as spin-off but not start-up. This curious usage does not exhaust the reasons that faculty launch ventures (either for profit or non-profit)[3]. It does, however, serve to clarify the point that one term (e.g. spin off) is insufficient to characterize the phenomena.

As the term venturing is used here, the recommended approach to quantification is survey research based on carefully designed sampling. Venturing is sufficiently widespread, diverse, and random in incidence that it is unlikely to be reliably represented in systematic collection of institutional data.

8. Consulting

The Statistics Canada survey cited above asks institutions to report how much information on consulting activity they collect from their faculty. At present, it seems difficult to collect more detailed information on a comprehensive basis. As well, activity is not an entirely satisfactory surrogate for desired outcomes. If only activity is measured, at least statistics on repeat activity should be collected to give some sense of client valuation.

The most satisfactory methods to evaluate the outcomes of consulting are based on survey research. Surveys need to be based on careful sampling plans and must include some validation of consultant self appraisal using client interviews.

Recommended indicator:

survey research.

9. Contract research

Contract research is most commonly measured in terms of the input metric of dollars of contracting by an institution or a unit. This metric does stand as a surrogate for an outcome if it is measured over time and used to indicate not the level of activity but its acceptability to clients. In this respect, the numbers may be interpreted in a fashion analogous to interpretation of sales of a firm. An extremely valuable supplement to input measures are surveys which assess client response over several dimensions as well as provider self appraisal.

Use of the income-input measure does skew reporting to give greater prominence to high cost efforts in engineering or medical clinical trials. Important efforts in policy research areas may only slightly perturb overall income statistics and artistic commissions are likely to go unreported. The problem of structural comparability is quite severe if measurement is made at the institutional level, across disciplines. It is interesting that government agencies, even those with an interest in university performance such as CMEC, have a propensity to contract with individual scholars rather than with the institutions. This inhibits reporting.

Reporting of contract research income is one part of the collection of data on research which is presently organized for Canadian universities through the Canadian Association of University Business Officers (CAUBO). There is an increasing recognition that not all institutions use the same definitions in the data reported to CAUBO. For several years, the Canadian Association of University Research Administrators (CAURA) has called for standardization of reporting. Recently, the problem of non-standard reporting has caused problems for the Canada Foundation for Innovation. This problem is not limited only to reporting on contracts.

A special problem, which is here classed with contracting, is the work of research consortia. Consortia range from the programs that arise around industrial chairs to multi-university collaboratives illustrated by TRlabs (telecommunications research) in the west. To some degree the provincial centres of excellence and the Networks of Centres of Excellence present the consortium problem.

A consortium pursues a research program of value to groups of clients. The shared activity often belongs to the "precompetitive" domain. It is consequently difficult to trace the ultimate use of the research product. Some larger organizations have developed sophisticated tools to evaluate economic impact, but these are not practical for smaller or more localized units. One area where information from the consortia is readily obtained and of considerable significance for assessment of impact is information about the level of transfer to consortium sponsor employment by graduate students working in research projects of the consortium.

In the areas of outcome which contractual research (especially through consortia) encompass, a very interesting issue is the measurement of knowledge flows among sectors. Bibliometrics has an important contribution to make. To the extent that researchers from other sectors join with university researchers in publication, bibliometric analysis can be applied to identify linkages between university research and collaborators by analysis of co-authorship in the publication databases[15]. The subtle flow paths of knowledge can reveal themselves in patterns of co-authorship. Unfortunately, knowledge on the move leaves the printed page of the scientific journal early in its journey through society. Much that is important to knowledge flows is not recorded by formal publication.

Recommended indicators.

> contract income reported according to standardized definitions and subdivided by types of contractors, problem areas, and types of providers (individual labs, or consortia);
> bibliometric analysis of knowledge flows;
> survey research.

10. Spin off and technology transfer

As noted above, the target in this category is the impact of creation of new firms from, or transfer to established firms or institutions of, the results of specific PSI projects. It is

[15] Godin, B., Gingras, Y., and Davignon, L., *Knowledge Flows in Canada as Measured by Bibliometrics*, Science and Technology redesign Project, Statistics Canada, ST 98-10, Ottawa, 1998.

more focused than venturing on activities for which institutions can keep records. *The Statistics Canada Survey of Intellectual Property Commercialization in the Higher Education Sector, 1998* attempts to collect information on both aspects. The licensing area is also the subject of annual reports prepared from data supplied by members to the Association of University Technology Managers (AUTM).

The Statistics Canada effort measures a number of important quantities. These include numbers of firms, survival of firms, university equity holdings in firms, licensing revenues, numbers of licenses and domestic vs. foreign revenues. If the data can be made complete and reliable, the Statistics Canada approach will provide information on all but one of the obvious quantitative issues. The data do not address job creation. Some surveys have attempted estimation of job creation[3]. There is a fairly robust methodology, an example of which is the methodology used in the assessment of the U.S. Small Business Centers Program.

Unfortunately, there is serious concern that the Statistics Canada study is not yet complete or consistent. AUTM data on licensing and royalty revenue suggest that the Statistics Canada survey may have missed significant information. Similarly, there are claims to identify spin off companies back to the 1970s or before. Records are notoriously incomplete over this time frame, and since the fate of spin of companies over a number of years, as well as trends in activity, are important to assessment, relatively complete older data are very desirable. Finally, the estimates of total royalty revenue and total equity holdings by Canadian universities seem low in comparison to figures released by some leading universities. There are, of course, definitional problems. Is an equity holding reported at book value or market value? I know of one case personally where a single company holding might be valued at a sum larger than the total reported for Canada.

The Statistics Canada questionnaire does raise one vexing question – the impact of production of educational materials. There is no doubt that these have considerable economic and cultural value. However, they have normally arisen in ways that make it difficult to achieve an accounting. PSIs have tended to regard educational materials as goods that arise autonomously. Although professors are authors, this was usually thought of as an exercise accomplished while burning the midnight oil. Also, such materials are very commonly inter-institutional efforts following no fixed pattern of collaborations. (My own textbook, *Inorganic Chemistry* – a world market leader - is the joint production of an American, a British, and a Canadian author. There is little doubt that it owes much of its success to the critical readings and advice provided by international authorities for whom acknowledgement required a two page single spaced list.) At the same time, production of multi-media and other novel materials has tended to occur in the framework communication units on campus who have little connection to the research administration. Separate administration of IP policy is common.

Recommended indicators:

the Statistics Canada survey after refinement;
survey research bearing on job creation.

11. Policy analysis and formation

A major difficulty confronting measurement in this important area is that so much of the formal research done in support of policy formation is sponsored in relatively informal ways and is not easily extracted from institutional records. Often, sponsors prefer a direct relationship with a scholar and seek to leave the institution out. Some institutions refuse to count activities so sponsored. Moreover, some of the important contributions are volunteer efforts that escape even individual faculty member annual reports. Data from indicators for contract research may contain some of the desired information. Similarly, studies of consulting will expose parts of this activity. In several institutions, survey research has proved revealing.

12. Artistic culture

Records of exhibitions, artistic commissions, major performances, productions, concerts, works of fiction, and other creative activity are frequently available from the annual reports of faculty. Summaries of these activities can form a valuable adjunct to assessment. Artistic creation is hard to quantify. However, highlight summaries prepared by the units involved can aid assessors to gain an appreciation of outcomes.

13. Trade publications and media contributions

In both the policy and culture areas the work of “popularization” carried out by PSI faculty is important. Just as with artistic creations, quantification is difficult. However, summaries of institutional highlights in trade book production, media presentations and journalism can serve as a valuable aid to assessment.

14. Research intensity

This discussion cannot end without an evaluation of the most enduringly popular of indicators for research activity in Canadian universities. This is the research intensity metric – the ratio of income from sponsored research to income from operating grants and tuition. It is essentially an input measure that indicates the intensity of involvement in sponsored research. If appropriate assumptions can be justified about similarity of institutional structure, it can serve as a relative measure of involvement by different institutions from the same structural classes. In gross terms it does, itself, help to classify institutions in terms of their “operationally” identified mission, whatever may be the language of mission statements. Its linkage to outcomes is at one remove. Nevertheless it will probably remain a part of well-developed indicator packages as a sort of pre-screen which helps to define the expected roles of institutions. It will help to determine which outcome indicators are most relevant.

15. A basket of indicators

How is a practical basket of indicators to be constructed in the face of all the uncertainties that confront us about both the significance of indicators and the reliability of accessible data? Clearly, the approach must be experimental and evolutionary with a continuing process of critical analysis applied to the indicators introduced. As well, it should be carefully monitored to determine whether it is skewing performance in

undesirable directions. (A clear example is system wide indicators that distract an institution from building a good internal quality system.) A good way to start on this difficult question is to look at experiments that are underway.

The Alberta system of key performance indicators (KPIs) includes a basket of research measures. The first is **research intensity**. It is recognized as an input indicator. The Universities of Alberta and Calgary choose to define a reference group of institutions that have a structure drawing large dollar sponsorship per faculty member. (All the "peer" institutions have medical faculties.) The University of Lethbridge chooses a different benchmark group. The indicator serves to measure the membership in an institutional structural class and reflects the level of success in implementing a mission consistent with this structure. Three-year rolling averages are used and changes in the indicator are regarded as significant at the same level as the value itself. (Is the institution continuing on a path of development consistent with the implicit mission it structure class indicates.) The risk in the use of this indicator is that it will be seen as a driver toward excellence in only those areas with high dollar cost of research. Especially, it may skew appraisal against areas of great economic significance (e.g. management, economics, and political science) where the relation of the institution to sponsors tends to flow through individuals in preference to institutional contracting. Similarly, the lone scholar makes little contribution to this indicator, however brilliant her contribution, and the artistic commission is likely to be invisible. A small problem with this indicator is that data are not yet reported to CAUBO on a consistent basis across the country, or even in one province.

The second Alberta indicator is **citation impact**. Five-year rolling averages are used to deal with the problem of citation pattern over time. As suggested above, this indicator is seen as a measure of involvement with the world knowledge system. Again, change is valued to respond to institutional development. The benchmark groups are different for the Universities of Alberta and Calgary on the one hand and Lethbridge on the other, recognizing that structure influences this indicator. Citation cultures differ by discipline. The outcome measured is undoubtedly linkage to the world knowledge system. Publications by student authors are included which is sensible. An indication of direct involvement of students in research is thereby derived. It would be desirable to find a way to disaggregate student contributions to make the indicator more useful to analysis of human resources outcomes. The indicator biases evaluation of institutions toward those disciplines with an ISI oriented publication culture. Gingras[16] has shown that weighting citation impact by citation frequency of the discipline can reverse the order of impact of institutions! This does not even require moving outside the science disciplines where citation data are most complete.

Citation impact is usually used in a "mode 1" oriented fashion. The citation impact indicator might well be supplemented by an analysis of the citation of scientific publications by patents. This would have bearing on the technology transfer and consulting outcomes. Narin[17] has documented the interesting result that there is geographical correlation between the addresses of patentors and the addresses of authors of basic science publications cited in the patents. This probably measures the

[16] Gingras, Y. *Performance Indicators, Keeping the Black Box Open*, in Measuring the Impact, Natural Sciences and Engineering research Council, Ottawa, 1995. P. 45-49.
[17] Narin, F, *Indicators of R&D Impact based on Citations from U.S. Patents to Research Journal Papers., op cit* p. 50-52.

"intelligence" function of the connection of national basic science to the world knowledge system. The publication patent relationship is a stronger mode 2 indicator.

The third Alberta indicator is **Council success rates**. The use of the indicator is based on the notion that peer review panels use relatively consistent criteria over time so that the trends are meaningful. Two elements are tracked: 1) the number of awards in relation to number of applications; and, 2) the average dollar value of an award received. The first requires modulation because the application rate and success rate are strongly discipline dependent. The second biases attention to expensive areas (e.g. medicine). The two do not compensate the weaknesses of each other. An important feature is the option open to universities to report data disaggregated by council. This improves the interpretability of statistics. However, with all care in data disaggregation, the bias in council deliberations against interdisciplinary work remains a problem. Mode 2 research is not well served. This is the area where it is interesting to consider distinguishing among types of awards. Mode 1 and links to the world knowledge system are best represented by core program award trends over time as discussed above. Other classes of awards shade into "contract" research (especially if consortia are included in contract research).

The fourth Alberta indicator set is related to **graduate students**. The indicators include; graduate student enrollment in thesis (i.e. research based) programs relative to academic staff; total graduate enrolment per academic staff member; and the number of competitive national graduate scholarships and fellowships in relation to full time academic staff. All of these parameters are chosen because graduate students have greater mobility than undergraduates do and their institutional choices reflect student appraisal of quality. The first is, at the moment, strongly linked to Mode 1 research styles, as is the third. The outcomes relate immediately to the world knowledge system. In contrast, the second indicator fully reflects the strong growth in professional graduate programs where an integrated problem solving style using case studies from practice is celebrated. It is more Mode 2 responsive. The use of graduate student judgement as a surrogate for judgements of the larger community is an interesting device that deserves monitoring.

A fifth Alberta measure attempts to synthesize knowledge system, contract research, and technology transfer outcome aspects. This is the **research impacts** indicator group. It includes:

1) council support ratio (granting council income to total research sponsorship);
2) community and industry support ratio (non-governmental income to total research support);
3) industry sponsored research, which relates industry sponsorship to total community and industry revenue;
4) total annual licensing revenue; and
5) rate of disclosure of inventions.

This group expands assessment into the contract research, technology transfer, and spin off fields. In each case this is done by using an input or output measure as a surrogate for an outcome. Well-designed survey research could serve as a valuable supplement. It might even justify the use of the surrogates. This group represents an attempt to get at some of the Mode 2 issues. As noted above, there are serious data collection issues.

Finally, there is a collection of indicators aimed at measurement of **community service and economic impact.** This is a basket of conventional econometric indicators. It reflects a useful picture of short-term effects of research spending. The disadvantage is that it draws attention away from the more profound outcomes that occur only with time lags.

The present Alberta indicator set represents a brave experiment. The indicators are broad gage, reaching for a wide range of outcome aspects. They are limited by the problem of data, especially in the context of the recognition that indicators would not be meaningful if they could not be benchmarked beyond Alberta. There is more use of surrogates than is desirable. Areas that could be enriched by survey research, a tool of growing importance in Europe[18], have not yet benefited from this approach. Finally, they are more focused on the indicators of Mode 1 research than even the current balance would recommend. A very concrete illustration of the dangers of an imbalanced set of measures does exist. Over the years of this century that new professional faculties (education, management, and to a certain degree engineering) have come into the universities, these faculties have confronted the Mode 1 academic research values of their colleagues. This has distorted research in these areas toward an artificial mode 1 style and undermined efforts of some of the most creative in the professions.

[18] According to statistics complied by Ugur Muldur of the European Commission, the use of survey research has grown faster in the decade than the use of either bibliometric or "econometric" style statistics. The rationale for this trend has been given above.