

Apples to Apples: Towards a Pan-Canadian Common University Data Set



By
Sean Junor, Miriam Kramer and Alex Usher
The Educational Policy Institute

November 2006

Introduction

In an age of increasing demands for transparency, educational institutions are beset with demands for information about their activities. Policymakers want information so they know that taxpayers' money is being spent in an effective way. Parents and their children want information about institutions so they can make informed decisions prior to spending \$25,000 or more on an education. The demand for information exists and it is increasing relentlessly and a significant part of the increased demand is aimed at more comparative information.

This demand for information places some significant burdens on institutions. In the past, institutions' data collection and analysis capabilities were primarily designed for internal, managerial purposes. The demand for information for external transparency and comparability requires them to collect and report data in very different ways than they have in the past. Thus, we are in a situation where most universities are spending enormous amounts of effort collecting vast amounts of data, but are still – in some quarters – perceived to be unable to report data effectively.

In other countries, there are two ways that this kind of problem has resolved itself. In some countries, governments have stepped in and provided common reporting requirements for institutions. In others, notably the United States, institutions have created a common data set so that governments, publishers, and other interested parties have a single, recognized source of institution-level data. Institutions know what data to produce and the public knows what information to expect. (For more information on the United States' common data set, see www.commondataset.org.)

In Canada, to date, neither has really happened. On the government side, though Statistics Canada has done some excellent work with respect to student surveys, the country's constitutional set-up has precluded it as a federal public agency from gathering much data about institutional affairs (though two significant exceptions are described below in the section below on multi-institutional data sources). With respect to a common data set, on-going friction between institutions and the main media outlet for such data – *Maclean's* magazine's annual *Guide to Universities* – has made institutions reluctant to put too much effort into such a product. Yet paradoxically, despite the widespread institutional discontent with *Maclean's* rankings and its methodology, *Maclean's* annual *Guide* in effect became Canada's *de facto* common data set simply because institutions continued to submit data to *Maclean's* on an annual basis.

So when, in the summer of 2006, two-dozen institutions chose to end their participation in the magazine's annual survey, they were not simply cutting ties with a commercial publisher; they were also effectively destroying Canadian's only common – albeit imperfect – data set. Though the uneven quality of the data set meant that few mourned its passage, the institutions that left *Maclean's* were aware that their actions might be interpreted as a blow against transparency, not as a legitimate protest against poor data quality. Specifically, they said:

*We welcome public assessment of our work, and all our institutions devote significant resources to that end. We already publish a great deal of data about our own institutions on-line and intend to publish more in the future, ideally in the form of standardized datasets that will facilitate valid temporal and institutional comparisons.*¹

¹ From an August 14, 2006 open letter to *Maclean's* magazine declaring their intention to cease participation in the annual rankings exercise signed by Tom Traves (Dalhousie University), Peter George (McMaster University), Michael Stevenson (Simon Fraser University), Indira Samarasekera (University of Alberta), Stephen Toope (University of British Columbia), Harvey Weingarten (University of Calgary), William Cade (University of Lethbridge), Emo Szathmary (University of Manitoba), Luc Vinet (Université de Montreal), Gilles Patry (University of Ottawa) and David Naylor (University of Toronto).

Essentially, the institutions were making two claims. The first was that they *already* published a considerable amount of information about institutional performance on the Internet; the second was that they intended to create data that are more comparable in the future. In short, they argued that transparency is available now, and comparability is in the works.

The purpose of this paper is to shed some light on the current availability of comparative information about Canada's universities and point the way forward for Canadian institutions in terms of collecting data. In Part I, we will examine and comment on what institutions do individually in terms of data collection and publication as well as what they do collectively through national consortia such as the G-13 group and the Canadian Association of University Business Officers (CAUBO). Part II will go on to examine the issue of a Canadian Common Data Set in terms of what such a database would contain as well as the obvious implementation issues that would need to be addressed in order to make it a reality.

Part I. Transparency and Comparability in Canadian University Data

Who needs data on universities?

Broadly speaking, there are three major sets of “clients” for information about institutions. The first is internal: Boards of Governors, university managers, senates, etc. The second is the public at large and the governments through which institutions receive billions of dollars in funding. The third are what might be called “consumers:” students, prospective students and their families. Each of these has different data needs, and institutions respond to their demands for data in different ways.

The first clientele for information – Boards of Governors and management - is in a sense the most demanding. When information is required for local decision-making, it usually receives the highest priority. Most universities in Canada have at least one full-time staff person who does nothing but collect and analyze institutional data; at the larger institutions it can be half-a-dozen people. Historically speaking, this type of data collection is a fairly recent development; as late as the 1980s, most institutions would not have had an office of institutional research *per se*, and indeed the professional body that unites these people in Canada was only born in 1989. Not all of these institutional research offices came into being in the same way. Some grew out of space planning offices; others came from academic planning office while others emerged from budget planning offices. Naturally, each institutional research office therefore has slightly different pre-occupations and responsibilities and produces data in a somewhat different way.

Though this clientele is in a sense responsible for the diversity of data collection practices in Canada, it is also currently a major force pushing for the standardization of data. Governing boards, in particular, are increasingly interested in benchmarking their performances against those of appropriate comparator institutions. For them, the lack of comparable data is becoming if not an obstacle to appropriate governance then at least a serious irritant along the way.

The second client, governments, arguably only really became interested in having institutions report data in the early 1990s, during an age of fiscal restraint.² As money became scarcer, governments and policy analysts began asking what kind of return they were getting on their money. In some cases, notably in Ontario and Alberta, governments began making portions of institutional funding conditional on institutions either publishing certain pieces of data in a common format or meeting certain common benchmark targets. The fact that these benchmarks were frequently trivial, or were based on outputs without reference to inputs, significantly prejudiced institutions against government demands for more data.³ Even when the requests were generally sound and sensible, institutions tended to be suspicious and this has hampered the creation of public, comparable data sets immeasurably. This perceived reluctance to create or share data on a comparable cross-institution basis has sometimes led people in government to bemoan institutions’ “lack of accountability.”

It is worth underlining something at this point: however much one may sympathize with the desire of policymakers to have more comparable inter-institutional data at their fingertips, the fact that institutions do not produce as much such data as policymakers might sometimes like in no way implies that institutions are somehow unaccountable. Every institution has a web of accountability relationships that in no way rely on the creation of comparable data sets. Their legal obligations vary somewhat from province to province, but among the many things they do in various parts of the country are: obey their statutes, report data to Statistics Canada, publish audited financial accounts,

² This is not to say that no data was available prior to that point; various Statistics Canada efforts of varying quality were available prior to that; it is merely to say that it was at this point that governments became significantly more interested in gaining access to comparable data on subjects that were either outside Statistics Canada’s data collection practices or in gaining access to data of a more refined quality than was possible through Statistics Canada surveys.

³ See Finnie and Usher (2005).

appear in front of legislative committees, submit to Freedom of Information legislation, and report staff salaries in conformity with provincial sunshine laws. In addition, Canadian institutions as a whole are getting much better at stating their goals in public reports and subsequently reporting their performance in meeting these goals. A regular international survey of accountability in universities has shown that Canadian institutions – once among the survey’s poorest performers – are now among the best.⁴

We turn now to the data available to the clientele for the third group of clients: students, prospective students and their families. Institutions are currently asking them to invest something in the neighbourhood of \$25,000 in tuition alone for an undergraduate education. Even if this represents only a fraction of the true cost of education, it represents a substantial investment by anyone’s standards. Where consumer durables are concerned – automobiles, for instance - it is completely normal for purchasers to consult various kinds of consumer guides which rate aspects of automotive performance and permit shoppers to compare both price and performance. By doing so, they permit consumers to look “under the hood” of many different types of cars before they proceed to test-drive a few of them.

Given the size of the investment demanded of parents and students, it is completely normal that potential students and parents desire similar kinds of data when making decisions about education. Protestations that universities are more complicated than cars and hence are less amenable to statistical reductionism are true but irrelevant – consumers want comparison data and it is very difficult to make an intellectually-respectable case against giving it to them. There may indeed be a case to make that comparable data on *institutions* is considerably less useful than comparable data on *programs of study* – one we would agree with, incidentally – but that is a very different argument.

Institutions do, of course, strive to give students and prospective students information about their activities. Institutional “view-books” and prospective student websites, for instance, have become much more informative in recent years. However, these are essentially sales brochures and for obvious reasons provide comparative cross-institutional data only in those instances where it is convenient for them to do so. They are not a substitute for comparable cross-institutional data.

What this brief survey shows is that there is a considerable market for comparable information about Canadian universities. Institutions themselves need it for management purposes. Students and parents need it to make good decisions about their educational investments. The following two sections will explore the various sources of comparable institutional data available to those who wish to compare the inputs, throughputs and outputs of Canadian universities. The next section will examine sources of data that already cover multiple institutions while the one after will look at the data produced and published by institutions themselves.

⁴ Nelson, M., Banks, W., and Fisher, J., “Improved Accountability Disclosures by Canadian Universities”, *Canadian Accounting Perspectives*, 2(1), (2003), 77-107.

Multi-institutional Data Sources

Canada is unique in the paucity of centrally audited data on post-secondary institutions. Compared to most countries, the amount of publicly available indicators describing the work of public universities and colleges is very small. Part of the problem lies at the level of our national statistics agency, Statistics Canada.⁵ However, there are some pieces of data that Canadian institutions collect individually, but report on collectively (that are and can be used for comparability purposes, but are not published by individual universities). Here we cover some of the data available from these sources, as well as their limitations.

The Enhanced Student Information System (ESIS). This database, designed by Statistics Canada, has existed nationally since 2001. It is a student unit record database that is meant to permit researchers to follow individual students and institutions across time. In theory, ESIS permits full tracking of students from enrolment to graduation, even if they move from one institution to another. The use of individual student identifiers, in theory, also permits the linkage of data from other sources (such as student aid files or income files). ESIS could therefore serve as a source of a tremendous amount of data about institutional completion rates and longer-term graduate income and employment data.

Unfortunately, it is no exaggeration to say that this project, which was cursed both by an over-ambitious design and an under investment in funds and staffing, has not lived up to its original billing. While semi-functional for an important but restricted set of possible research questions, as a source of institutional data its accuracy is patchy at best. The system is plagued by institutional non-responses and inaccurate reporting and it will be at least two to four years before it is properly functional.

The Canadian Association of University Business Officers (CAUBO)'s Financial Information of Universities and Colleges. Produced annually in conjunction with Statistics Canada, this report is the main source of nationally-comparable information on university finances. However, as Snowden (2004) pointed out, the CAUBO data is problematic in a number of ways, not least because it tends to overstate net revenue by not making clear the spending restrictions that go along with particular sources of revenue (notably on tuition fees post-1996 in Ontario institutions). Moreover, because CAUBO covers data for an entire institution (including teaching hospitals), it is not especially helpful in understanding what occurs at a sub-institutional level. However, the data are accurate for accounting purposes, and it is becoming increasingly more comparable over time.

The G-13 (formerly G-10) Data Exchange.⁶ The Group of 13 research institutions (formerly the G-10) has for many years operated a data exchange, which permits institutions to benchmark their performance. The information produced by this data exchange is more finely-tuned than data from any other source. The data have common definitions of staff, students, finances and research dollars, all of which are monitored at a field-of-study level. As a source of data, it is unrivalled in its completeness and in the care that has been taken to ensure comparability. There are, unfortunately, two drawbacks to the G-13 data. Most obviously, its institutional coverage is limited (it covers only 13 institutions). In addition, the G-13 data are not made publicly available. Nevertheless, this data exchange is probably the best model Canada has of a well-constructed common data set.

⁵ To cite only the most troubling example of Statistics Canada's handling of educational data, the agency has been unable to produce an accurate count of college students for the past seven years (i.e. nothing this decade), while counts of specific types of students at the university level (e.g. foreign students) remain highly suspect.

⁶ The G-13 includes McGill University, McMaster University, Queen's University, Université Laval, Université de Montreal, the University of Alberta, the University of Calgary, the University of British Columbia, the University of Ottawa, the University of Toronto, the University of Waterloo, the University of Western Ontario, and York University.

Canadian Undergraduate Survey Consortium (CUSC). CUSC is a co-operative research group involving a collection of Canadian universities. Begun in 1994, CUSC now rotates its survey subject each year between first-year students, graduating students, and all undergraduate students. The Consortium's membership varies from year to year; in 1997, there were nine Canadian universities involved, while in 2001 the number had increased to 26. The survey is coordinated by the University of Manitoba's Department of Housing and Student Life.

Each participating university distributes a survey package consisting of a cover letter, questionnaire, and postage-paid return envelope. Different institutions may use different means to pursue follow-up contact with survey recipients in order to increase the response rate. Technically, survey results are only generalizable to the institutions at which the survey was conducted. Given both the large sample size, and the number and diversity of the institutions taking part, it is probably a very good reflection of the Canadian student body, or at least the portion of it that lies outside the province of Quebec, where coverage is somewhat weak. Also, the data is what might be called "semi-public" – most participating institutions will provide their data if asked, but it is not regularly published and publicized in the way that, for instance, CAUBO's annual report is.

National Survey on Student Engagement (NSSE). The NSSE started as an initiative of the Pew Charitable Trusts in 1998. Drawing on decades of research on student outcomes in the United States, this instrument was designed to measure, via student surveys, the presence of institutional activities or practices, which are known to be related to important post-secondary education outcomes. Because it deals with actual educational activities and outcomes rather than financial inputs, the NSSE is widely seen within institutions as an attractive alternative to an ordinary rankings project. The first Canadian institutions started to use it in 2004, and by 2006, 30 Canadian institutions participated in the survey. While widely liked as a managerial tool to help identify areas of educational experience in need of improvement, NSSE is not very useful as a transparency tool. NSSE actively discourages its use as a rankings measure. In addition, while many Canadian institutions laudably choose to publish their NSSE data, in their raw form they are a daunting, highly technical and not at all a useful tool for parents and prospective students.

Maclean's Magazine. Though it never set out to achieve this, the magazine has essentially become the source of common information about universities in Canada since it was first published in 1991. It contains 21 indicators – most of which measure inputs in one way or another. However, not all of *Maclean's* 21 indicators actually contain real data; on some indicators, *Maclean's* has chosen to print ordinal rankings without actually revealing the data behind the indicator. Integral data on student retention and success rates are posted at an institutional level, as are certain data about incoming students and class sizes

There are some legitimate questions about the accuracy of various pieces of data reported to *Maclean's*. There is no shortage of reports or theories about how particular institutions have taken liberties in the definitional interpretation of the numerators and denominators that make up certain indicators. Regardless of the quality and scope of the data, *Maclean's* was, until this summer, one of the few accessible sources of comparable inter-institutional data in Canada.

It is evident that each of these various multi-institutional data sources is an imperfect source of good, comparable information about institutions. ESIS is incomplete. CAUBO is too general to be of much use and in addition is somewhat misleading with respect to revenues. CUSC lacks a truly national scope and, for those included, only measures current student satisfaction. NSSE was not designed to

be a public reporting vehicle. The G-13 data scores high in terms of accuracy and comparability, but gets no points for transparency. Conversely, *Maclean's* gets high marks for transparency, but lower marks for comparability and accuracy. That being said, these are only one source of data about institutions; another, far more important source is the institutions themselves.

Institutional Data Comparability

As the "*Maclean's 11*" said in their August 14 letter, institutions do indeed – for the most part – publish a great deal of data about themselves on their websites. However, while it is true that virtually all institutions publish *some* data in some fashion, the value of this data for students and parents on the one hand, and institutions for benchmarking purposes on the other, is somewhat limited if the data are neither easily accessible nor easily comparable from one institution to the next. It is to these two issues that we now turn.

Comparability is crucial to prospective students and parents who are looking for the right university. It is all very well, for instance, to know that 80% of students at a particular institution graduate within four years, but without knowing how similar institutions fare on this measure, this figure is nothing more than a factoid, from which one cannot infer anything about institutional quality. The ability to compare one institution with another is, largely, what makes traditional rankings exercises such as *U.S. News and World Report's America's Best Colleges* and *Maclean's Guide to Canadian Universities* so attractive to parents. The ability to identify truly comparable institutions and measure one's performance against that of comparable data is also of immense value to managers and Governors.

The fact is that very little data produced by institutions for purposes of their own performance measurement or public accountability – other than that mandated by government or Statistics Canada – is actually comparable. At the broadest level, most institutions do report on similar areas of activity: most produce and publish some indicators about their students, their faculty members, their finances (both financial and physical) and their research activities. But they frequently use different types of indicators to describe their institutional activities, and even when they use the same type of indicator, they frequently collect the data on a different basis, and use different metrics when displaying the data. For example, the NSSE information, which is data derived from a central survey administered by one source, is incomparable for the simple reason that different universities use different inputs within each indicator to measure their performance. For instance, while Queen's University and the Universities of Calgary, Saskatchewan and Toronto, all present information on scholarship and bursaries as a percentage of operating expenses, they use different figures and sources to derive this outcome.

Put simply, though institutions are making data available, very little of it is common and therefore very little of it is comparable, which limits its usefulness to those wishing to make appropriate in-depth comparisons between institutions. Although this situation is challenging and disappointing, there is no reason to suspect a sinister motive behind this situation. Institutions collect and report data for a variety of reasons, but most of them, traditionally, have been managerial in nature. Thus, to take but one banal example, institutions tend to have very little data on their staff because the computer systems that deal with personnel focus on payroll issues rather than human resources. Also, for the most part, data has historically been collected for the purpose of specific internal audiences. As a result, the manner in which they are displayed varies from one institution to another depending on local preferences. It is only relatively recently that institutions have been asked to use their data to create meaningful inter-institutional comparisons. It should therefore not be a surprise that these disparate systems do not, as of yet, produce results which successfully meet the test of comparability.

Institutional Data Accessibility

Even though institutions are indeed publishing a reasonable amount of data, very few of them are doing so in a manner that the general public would be likely to read. While some institutions put their major indicator data into reasonably attractive, easy-to-understand documents (l'Université de Montreal's *Annual Report* deserves special praise in this area, others choose to put much of their reporting in so-called fact books or data books, which are exceedingly technical in their orientation, occasionally phone book-like in their readability, and constitute public reporting only in the narrowest sense of the term.

Similarly, while some institutions (notably the University of Toronto and the University of Western Ontario) put their data in a single, comprehensive document, most split their reporting between at least two, and in some cases three or more, major documents including fact books, annual reports, business reports, university statistics reports, and student surveys. This means that what is available in a single spot on a website at one institution could be located in two or three different spots on another institution's website. Even where indicators were clearly presented in a single document, that document itself was not always easy to find as performance indicator information is sometimes scattered across several documents located on different web pages managed by different units within a university (e.g. Vice-Presidents, Provosts, Offices of Institutional Analysis, etc).

Again, there is nothing sinister in this practice and there is absolutely nothing to suggest that institutions are trying to hide information. It is simply to say that because the data has been produced with internal purposes in mind, it tends to be published in ways that are not always conducive to helping students and parents find it. However, this finding does tend to weaken the point made by the "Maclean's 11" that they do not need to work with a commercial media outlet because they already publish much of the data. Clearly, institutions are making data available. However, how useful is data if it can only be accessed by individuals who are familiar with the information and where it might already be located on the university website?

Conclusion to Part I

There is a tremendous public appetite for information about universities. The annual sales of *Maclean's Guide to Canadian Universities* are, if nothing else, clear evidence of this fact. And universities themselves are not hiding from this demand; by and large, they do publish a great deal of information about themselves. However, institutional reporting is hampered by poor accessibility, dispersion of reporting among multiple publications and, above all, lack of inter-institutional comparability.

The short supply of information relative to the demand means that people wanting to meet the need for information will make do with inferior data. For instance, *Maclean's* will publish old data where no new data is available, and try to create indicators from published institutional reports. Predictably, this will upset institutions who will correctly claim that their data is being misrepresented. *Maclean's*, equally correctly, will reply that they are doing the best they can with available data. Mutual suspicion and recrimination will no doubt be the result, and Canadian students and parents will continue to lack the accurate and comparable data they need and deserve to make informed decisions.

There are only two ways out of this situation. The first is for some kind of third party – presumably governments – to get involved and start mandating institutions to publish certain types of data as part of a “transparency” agenda. This might be acceptable to some, but “transparency” agendas, in the hands of governments have a way of turning into “accountability” agendas and, as we have seen with the implementation Key Performance Indicator initiatives in Ontario and Alberta, governments tend to focus on simple, but not particularly relevant, output measures. This outcome is probably not in universities’ interests.

The other way out of this situation is for institutions themselves to come together and start producing more data, and to make this data more publicly available and set to a common standard. Universities, despite their concerns and problems with the methodology of magazine rankings have, by virtue of their participation over the years, effectively signaled their belief in the value of such a tool; the open letter to *Maclean's* dated August 14 makes it clear that they still believe in the need for common data.

In short, what Canada needs is a common data set. Part II of this document describes what this common data set might look like.

Part II. The Need for Common Data Sets

As noted in Part I, there is an increasing demand by a number of stakeholders to increase institutional transparency. It is appropriate for different stakeholders to find different ways to compare the inputs, activities and outputs of institutions, and may require different indicators to make their comparisons. Governments may legitimately want to know whether some institutions are more efficient than others at turning research into patents or scientific publications. Prospective students may want to know how different institutions compare on issues of cost, student engagement and long-term job prospects. This increasing demand for data is creating increasing pressure on university administrations to devote scarce resources to fulfilling an ever-growing number of data requests.

A similar phenomenon occurred in the United States over a decade ago as other publishers and government agencies followed *U.S. News and World Report's* lead in asking for more data in the name of transparency. As the requests for data multiplied, institutions in the United States took a proactive position and began to create a single, common set of data for interested parties to use. This set, known as the Common Data Set (CDS), is now the main way American institutions report on their internal activities to the broader public. Members of the CDS consortium may take data from the set and put it into whatever kind of publication (or, presumably, ranking) they wish.

Moreover, while it is entirely legitimate to ask institutions to provide data that permit these kinds of comparisons on individual indicators, it is also entirely legitimate for institutions to note that many inter-institutional comparisons – even on ostensibly uniform indicators – are not actually apples-to-apples. To take only the most banal example, institutional graduate income levels and student loan default rates are highly correlated with the set of program offered by an institution; the presence of law and medical schools, for instance, can exert a highly distorting effect on the data. As a result, it is not always best to make comparisons at the institutional level.

If comparisons should not take place at the institutional level, then where should they be made? Worldwide, it is increasingly the case that universities are compared at the sub-institutional level. University rankings in Germany and Italy are being done at the field-of-study level, while both the Shanghai Jiao Tong and Times Higher Education Supplement (THES) rankings are now able to disaggregate by broad fields of study as well. Within Canada, almost no data is publicly available at the field-of-study level, but *individually*, the G-13 group of institutions shares an astonishing wealth of indicators at the field-of-study level. Indeed, the effort that the G-13 has made to ensure the complete comparability of the data is remarkable; for that reason its efforts – and indeed its indicators – would make an excellent starting point for a Canadian common data set.

What indicators should be in a Common Data Set?

When thinking about what kinds of indicators should go into a common data set, it is important to cast a broad net and look at what kinds of indicators might potentially be useful in terms of informing either public policy or consumer choice. We should definitely not be restricting ourselves to output indicators, but should be considering how various forms of input indicators (students, money, staff) condition the output indicators as well. A reasonably fair portrait of institutional activity be drawn only when we have data on both sides of the input-output equation.

In any common data set, we would certainly want to include something about *students' beginning characteristics*, such as gender, ethnicity, aboriginal status, mother tongue, presence of a disability, age, marital status, presence of dependents, family educational background, family financial background, secondary school marks, etc. At a bare minimum, we should have summary statistics at an institutional level on all these fields. Better yet, we should have this data disaggregated to the field-of-study level.

We would also know something about *teaching staff* at each institution. At a bare minimum, we need acceptable head-counts of full and part-time staff (including temporary faculty or graduate students acting as class instructors), on a field-of-study basis. Ideally, though, we would also know something about salary bands, average years of teaching experience, as well as information on research grants, teaching awards received, various bibliometric and citation data, and the number of classes taught.

There are variables related *institutional resources* that need to be included as well. At an institutional level, we need to know the amounts and sources of various revenue streams, and we need to know amounts and types of expenditures. Much of this is available from the annual CAUBO survey, but the validity of this source for comparative institutional purposes has been rightly questioned (see Snowdon 2004), and some work would need to be done in order to make this data set fit for this purpose.

We would also want to know certain things about capital, including available square footage, and estimates of deferred maintenance costs; we should know various things about the state of Information or Computer Technology (ICT) resources on campus; in addition, we should know about expenditures on student services and financial aid; and we should certainly know more about library resources than we currently do.⁷

As they move on through an institution, we would ideally like to know a number of things about *student progression*. Cohort drop-out rates (i.e. aggregate data) would be useful information, though not as interesting as the knowing the progression or failure of each individual students in the system and then linking that to unit record data about course selection and the experience of individual instructors along the way. We should be interested in whatever data is available through NSSE or similar instruments with respect to contacts with teachers, participation in collaborative learning environments, or any of the other activities associated with “engagement.”

Similarly, we should be interested in students' levels of satisfaction with the general campus experience as well as with specific aspects of the campus experience (e.g. availability of professors, library resources, etc.). It would be important also to track students' financial data, specifically to look at the evolution of aspects like student income and student debt. Finally, student grades would be of great importance. Many of these things are actually currently available at the aggregate

⁷ In regard to library resources, the Association of Research Libraries (ARL) recently has been conducting some excellent research on library performance indicators for the digital age which should certainly be adopted in Canada as quickly as possible. To the extent possible, all the data should be cut to permit field-of-study level comparisons, though it is clear that in some areas this simply may not be possible.

institutional level, either through administrative data or student surveys, but are not necessarily made public. At a minimum, this data should go public. The next step should be to collect the data in sufficient quantity that could provide accurate field-of-study data and ensure the data could be published.

Learning Outcomes are one area where institutions regularly report some kind of performance indicator, even if the indicators themselves are not well developed. Graduation rates are usually available at an aggregate rate, but ESIS should be collecting graduation data at an individual level. Data is usually aggregated at an institutional level, but this could easily (and should) be extended to a field-of-study level. We can also look at final grade at a unit level, but since grades are highly subjective, they are not the best way of measuring knowledge acquisition or cognitive development. In future, as products such as the College Learning Assessment (CLA) become more widely available, we may be able to perform and record some more direct measurements of learning outcomes. The OECD is also currently looking at this area and considering development of instruments similar to the Programme for International Student Assessment (PISA) to gather comparative cognitive data about university-enrolled students.

Finally, there are students' *final outcomes*. Employment rates and income levels as collected through various kinds of graduate surveys are also often available (and sometimes reported) at the institutional level, but would be more appropriately reported at a field-of-study level. The same would also be true of rates of students returning to further education or student satisfaction after a number of years in the labour force (currently collected through Statistics Canada National Graduate Survey (NGS) and a number of graduate surveys). An almost limitless number of other variables could also be collected via special surveys (e.g. civic participation, happiness surveys, etc.), and reported at either an institutional, field-of-study or unit level.

Table 1 – Table of Possible Common Data Set Indicators

Broad Indicator Set	Specific Indicator	Level of Desired Reporting	Source of Data
Beginning Characteristics	Gender composition	Institution, Field-of-study	Admin data
	Mother Tongue	Institution, Field-of-study	Survey
	Aboriginal status	Institution, Field-of-study	Survey
	Disability Status	Institution, Field-of-study	Survey
	Age	Institution, Field-of-study	Admin Data
	Socio-economic backgrounds	Institution, Field-of-study	Survey
Teaching Staff	Average Entering Marks	Institution, Field-of-study	Admin data
	Full-time staff	Institution, Field-of-study	Admin data
	Part-time staff	Institution, Field-of-study	Admin data
	Bibliometric and research award data	Institution, Field-of-study	Thomson ISI, Granting Councils
	Teaching awards received	Institution	Awarding bodies
	Salary bands	Institution	Survey
Institutional Resources	Average Years of Experience	Institution, Field-of-study	Institutional data
	Revenue sources and amounts	Institution	Modified CAUBO
	Types of Expenditures	Institution	Modified CAUBO
	Space Indicators	Institution	TBD
	IT Resources	Institution	TBD
	Library Resources	Institution	Modified CARL
Student Persistence and Engagement	Student Services and Financial Aid	Institution	Modified CAUBO
	Student Persistence	Institution, Field-of-Study	Admin Data
	Student Engagement data	Institution, Field-of-Study	Survey
Learning Outcomes	Student Satisfaction data	Institution, Field-of-Study	Survey
	Graduation rates	Institution, Field-of-Study	Admin Data
Final Outcomes	Cognitive outcomes	Institution, Field-of-Study	Survey ⁸
	Return for further study	Institution	Survey
	Labour market outcomes	Institution	Survey
	Well-being outcomes	Institution	Survey

The foregoing is, of course, only a potential list of indicators in a common data set. Other indicators could be imagined; equally, some could be dropped if they proved too difficult or too costly to collect in a standard, comparable manner.

⁸ Such a survey does not functionally exist in Canada – though the OECD is considering such an instrument and the Council for Aid to Education has a pilot in the field.

Collecting Common Data

An accurate common data set requires one or more common data collection platforms. In the case of the kinds of common data sets we have been describing here, these could take one of two forms.

With respect to indicators which are based upon administrative data either directly or indirectly (e.g. CAUBO), a common data platform would simply take the form of filling in an agreed-upon survey with data according to agreed-upon definitions. The key obstacle here tends to be getting institutions to agree to a common set of reporting standards – a problem which is no doubt magnified in situations where institutions have got used to their own ways of collecting and reporting things.

Assuming the hurdle of agreeing to common definitions can be overcome – which it certainly can be, given a reasonable dose of patience and goodwill – a common data platform is simply a matter of institutions filling in boxes on one or more surveys. In the case of the indicators we list in the table, there would be at least one to look at staffing issues, a second to look at financial data, a third for physical space indicators and a fourth for libraries and ICT data. Collecting data for these surveys would no doubt be time consuming but not necessarily difficult. Currently, most of the data mentioned here (with the exception of some of the library metrics) is already regularly reported within institutions; reporting a slightly different set of numbers according to common definitions will probably change more in terms of collection practices than in staff work loads.

Developing a common data collection platform for the collection of student survey data would be a more challenging process. In our brief survey, we looked at several types of student data, including certain types of socio-economic data, satisfaction and experiential/engagement data, and some educational outcome data. Now, as noted earlier, most institutions have at least one kind of student survey that covers at least some of these topics; most also have some kind of graduate follow-up survey that covers other related topics. The challenge in this area is not to re-invent the wheel, for this would lead to a substantial duplication of effort. Rather, the real challenge is to create a meta-survey that covers all the subject areas without creating survey fatigue among students or overly onerous burdens on institutional research offices.

Canadian universities already have a number of common data collection platform data, most notably the NSSE and CUSC surveys. NSSE deals primarily with satisfaction and engagement; CUSC deals primarily with satisfaction, student backgrounds and student finances. Combining the data from the two at a unit level could reveal some very significant relationships and, theoretically, it should be possible to create a hybrid instrument that covers both while eliminating the overlap relating to satisfaction. Such an instrument could also contain several additional modules to respond to local survey needs, which would further increase its utility to institutions. If the instrument were electronic (i.e. web-based) with a reasonably large sample drawn from a local database (i.e. the institution's own list of student emails), then both survey costs and processing time could be significantly reduced from traditional surveys.

Some institutions use both of these instruments, others one or the other, and some use neither. Statistics Canada surveys such as the NGS have the potential to be common data platforms even though they are not currently used as such. With respect to administrative data, ESIS and the annual CAUBO surveys are also common data platforms that are currently used.

In addition, similar revision could be done with respect to graduate surveys, which most institutions still conduct periodically. Like with the undergraduate survey, a common graduate survey need not mean the end of institution-specific questions. The trick is to mix a common set of questions with modules that vary by institution so that the benefits of local surveys are maintained while the benefits of commonality are obtained.

Conclusion – Building a Pan-Canadian Common Data Set

As we saw in Part I of this report, many universities spend a great deal of time and energy collecting and generating statistical reports on their activities. And yet, in part due to inadequate presentation, and in part due to a lack of comparability, their work receives far less attention than it should. More importantly, the current situation deprives a variety of stakeholders of fairly important information presented in a comparison-enabling format.

In Part II of the report, we have examined some potential elements of a pan-Canadian common data set and noted that many of the individual data fields are already being collected in one form or another. What is holding back the creation of a common indicator set is simply the absence of common definitions and common data collection platforms. Overcoming these obstacles will not necessarily be easy, but neither is it rocket science – all it takes is institutional good will, patience and hard work. However, the fact that a common data set has yet to come into existence suggests that at least one of those elements has been missing to date.

We conclude here simply by noting that the events of the summer of 2006 with respect to *Maclean's* magazine, if nothing else, changes the situation somewhat. It is no longer possible to pretend that institutions collectively provide Canadians with a comparable set of data about their activities. To their credit, the “*Maclean's* 11” promised in their letter to *Maclean's* publishers that they would be taking steps to create more comparable data. This is a laudable aim that now needs to be translated into action not just by these 11 institutions but by all Canadian universities.

It is our hope that this paper has succeeded in detailing the nature of the challenges that lie ahead in the construction of a common data set, and in providing a possible model for the final form such a common data set might take. There is hard work ahead, and we look forward to providing what assistance we may to Canadian universities involved in this task.

Glossary

CARL	Canadian Association of Research Libraries
CAUBO	Canadian Association of University Business Officers
CLA	College Learning Assessment
CUSC	Canadian Undergraduate Survey Consortium
ESIS	The Enhanced Student Information System
G-13	The Group of 13 Canadian research universities
ICT	Information or Computer Technology
ISI	The (Thomson) Index of Scientific Information
NSSE	National Survey on Student Engagement
NGS	National Graduate Survey
PISA	Programme for International Student Assessment
THES	The Times Higher Education Supplement