



### Session 3: Tools and policies for collecting, managing and disseminating data and metadata (1)

## The Path towards Improving Quality in the International Education Database

UNESCO Institute for Statistics

### **Introduction**

This document looks at the International Education Database - the greater part of which (at least in terms of the number of countries covered) is hosted by the UNESCO Institute for Statistics (UIS) - and the manner in which the UIS is striving to improve data quality through strategic initiatives targeted at both data providers and data users as well as better internal processing and systems. It describes work started during the Institute's first year of operation that is still on-going.

The UIS was established in 1999 with the mission of:

*"providing statistical information on education, science, culture and communication which helps decision-making in Member States and facilitates democratic debate in UNESCO's areas of competence, employing to that end the highest professional standards and intellectual independence in data collection and analysis".*

There are many institutions collecting international data. The challenges faced vary substantially from institution to institution and from domain to domain. The UIS has the lead-role and responsibility amongst the UN agencies for collecting data on education and hence for establishing and maintaining a global database of education statistics. The Institute is required and expected to have internationally comparable data, which are as complete as possible for as many countries as possible in as many years as possible. We are, of course, not unique in this respect in the UN system but in the field of education we are the only organization with such a wide remit. Our main partner organizations in the education data collection field - OECD and Eurostat - are both focussed principally on much smaller numbers of countries that are more homogeneous economically and politically and have more similar education policy aims.

While most international organizations deal with highly aggregated data, this is specifically sensitive in the field of education. The International Standard Classification for Educational Statistics, ISCED, is per se somewhat unhandy and difficult to adapt to national education systems. Further, contacting countries, sending/receiving materials electronically is difficult in several regions, but offers significant benefits in terms of timeliness.

This paper illustrates some of the issues the Institute has encountered in striving to improve the quality of data by focussing on the UIS's annual Education Survey. The survey complements two other similar but not identical surveys - UOE<sup>1</sup> and WEI<sup>2</sup> - that are run jointly with OECD and, in the case of the former, Eurostat as well. Almost 60 countries are involved in either UOE or WEI - both of which are designed to meet the data needs of all participating organizations and countries and thus to minimize both duplication

<sup>1</sup> The joint UNESCO/UIS-OECD-Eurostat survey completed by all OECD Member States, Israel and approximately 10 additional countries in Europe of special interest to the European Union.

<sup>2</sup> The World Bank-funded *World Education Indicators* survey jointly administered by OECD and the UIS which is completed by 19 countries around the world including most of the most populous countries not covered by the UOE survey.

of data collection and duplication of follow-up queries. The remaining 140 or so countries are asked to complete the UIS's own Education Survey.

All three surveys are typically completed by national Governments (either Ministries of Education or National Statistical Institutes – or, often, a combination of both). Data are consequently largely drawn from administrative sources<sup>3</sup>.

When the Institute was first set up, one of the early quality decisions made in relation to the education database, was to defer the UIS survey for one year, allowing time for a review of the data collection strategy and a redesign of the existing UIS survey.

Although not stated explicitly at that time, the review took account of many of the characteristics generally regarded as necessary for good data quality:

- Relevance
- Accuracy
- Timeliness
- Accessibility

### ***Back to basics: relevance***

The first step was to identify the information needs to be met by the UIS survey instrument. The questionnaires were redesigned with the resulting indicators and hence uses (and users) in mind. Ideas were taken from the UOE and WEI survey development and adapted to the UNESCO situation.

Account was taken of the likely needs of key users – including countries – by examining a number of basic topics relevant to education-policymaking and taking ideas for policy-relevant statistics and indicators from those commonly used or commonly requested by UIS's main customers: international organizations (including UNESCO), global initiatives (in particular Education For All) and countries themselves. Some limited consultation with a group of users also took place.

The information needs were then translated into data items to be collected and new questionnaires were designed. The extent to which the information needs could be met was limited by what could reasonably fit on a paper questionnaire; by what could reasonably be expected to be available in countries and what could be defined and hence compiled in an internationally-comparable manner. Given such a wide-range of potential uses and users, it was inevitable that information needs seen as key for some could not be fully addressed.

The new UIS survey is more relevant than its predecessor over time it has become apparent that there are still a number of shortcomings and gaps and that the survey is not relevant for all purposes. Given the limited amount of consultation during the survey design, even if it is relevant for many purposes it is not always seen as such. The expectations of users were not addressed or managed as effectively as they might have been.

### ***Starting at the source: accuracy (1)***

Having redesigned the UIS survey, the next step in improving the International Education Database was to try to improve the accuracy and completeness of the data coming from the countries themselves.

---

<sup>3</sup> The UIS has not so far run household surveys of its own and is only just about to launch a sample survey of schools (jointly with OECD) in countries participating in the World Bank-funded *World Education Indicators* project. The UIS analytical team does make some use of household survey data from UNICEF. Surveys assessing educational achievements of individuals such as PISA or TIMSS (for school pupils) or IALS (for adults) have also so far been the remit of other organizations although in 2004 the UIS will launch a pilot-survey (LAMP) in up to 7 countries to assess the literacy skills of adults.

The new survey was accompanied by an improved Manual of Guidance for completion of the questionnaire: a conscious decision was taken to keep the Manual fairly short – around 20 pages – to be understandable without being overwhelming and to make it (and the questionnaires themselves) available in 5 of the 6 official UN languages<sup>4</sup>. It was felt that short notes of guidance were more likely to be read than long notes – especially if they were being read by a non-native speaker of the language in which they were written. On the other hand, short notes may over-simplify the situation and provide insufficient advice and guidance. There is some evidence to suggest that this is now the case and therefore that the Manual needs to be reviewed and improved.

The new survey also collected more metadata than in the past – in particular on the structure of the national education system and the mapping of national programmes of education to the International Standard Classification of Education (ISCED) - the framework used for collecting internationally comparable data on education.

But perhaps the most direct attempt to improve the accuracy of data coming from countries was to launch the new survey in a series of Regional Workshops between March and September 2000 to which all 140 countries expected to complete the UIS survey were invited. The workshops allowed the Institute to explain the purpose of the survey and its key elements; to teach participants about the ISCED and help them to map their national programmes to the ISCED before attempting to complete the questionnaires; it allowed us to strengthen our contacts with individual countries and motivated participants to do their best to provide the data requested.

The Regional Workshops were repeated annually during the following three years but became gradually less effective and thus have been cancelled for this year whilst we review this aspect of our data collection strategy and seek to integrate this work more closely with the Institute's statistical capacity-building programme which has grown enormously since the Institute was first established.

## ***Improving data processing: accuracy (2)***

Another crucial aspect of accuracy of data is the manner in which data are handled once they are received by the data requester. This is probably the area that has been the weakest or at least has often appeared to be the weakest.

Information systems to assist in effective and efficient data processing were not available when the UIS was first created and until recently, there was no capacity to put these systems in place. The expertise and knowledge of statistical assistants and other staff responsible for processing the data was heavily relied upon to ensure the accuracy of the data – but such ad hoc methods are not conducive to quality results. Various actions were taken to minimize these risks – including training; the development of guidance notes and other documentation to try to ensure the application of similar procedures across different individuals; the implementation of some automatic cross-checks between data points in the database and a more systematic review of resulting indicators.

Even though the use of electronic questionnaires is increasing, capturing paper questionnaires will be required for the foreseeable future. For the data capture that is taking place in 2004 and onwards, the UIS will be introducing double capture of data received as a data reliability check.

Studies have revealed that error rates can vary greatly, between 0.44 and 4.83%<sup>5</sup>, where variables defined as "key variables" tend to have a lower error rate than others. It is in the nature of the UIS Education Questionnaire, that most variables are thought to be key variables. This could possibly lower the overall wariness but there is no empirical data with which to judge. Therefore, taking the overall number of data cells in the UIS Education Questionnaire (approximately 2200), would therefore result in between 9 and 100 incorrect values. Given the high level of aggregation in the data, and the way errors in one data cell can lead to multiplied errors in indicators calculated, error rates should near 0%. Even

---

<sup>4</sup> Chinese was not used as China itself participates in the WEI project and so does not complete the UIS's survey.

<sup>5</sup> Statistical Policy Office: Statistical Policy. Working Paper 31. July 2001

though most of these errors should be found during data processing, the additional resource costs and time investments are substantial which justifies the use of double capture at least until empirical data exists suggesting otherwise.

Another area of the statistical system where new technical capabilities could help the process greatly was thought to be in the area of batch edits. During 2003 a proof of concept was undertaken to determine how effective a batch editing and reporting system would be in processing the education data. The results were overwhelming in support of taking this project to the next stage. While this may not be surprising, one of the additional side benefits of the proof of concept is that there is now staff buy-in and commitment to put a more formal system in place and commit time and energy to identifying the necessary hard and soft edits that need to be included.

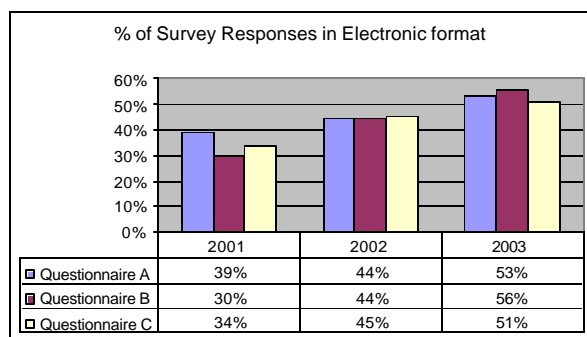
The system is currently under development and the hard edits will be fully implemented over the next month. The system will be self-documenting so that the edits as articulated in the system will be included in the error reports and the documentation in a consistent and accurate manner thus avoiding the need to maintain separate documentation system that is rarely up to date. The reporting structure will capture changes over time in order to measure progress and improvements. Error reports will also be available in various levels of detail for managers and statistical officers. A more complete examination of this system and its impact on data processing can only be accurately addressed at a future point in time. Suffice it to say that the system should lead to timely and accurate error identification that will greatly increase the effectiveness of staff carrying out survey processing.

## ***Timeliness***

Although the UIS survey has always been designed as a paper questionnaire, an electronic version was also developed and introduced in 2000 in the first round of the new survey. The electronic version is currently available in three languages: English, French, and Spanish. Internal discussions have led to the decision that electronic questionnaires should be made available in the same languages as the paper questionnaires. This will result in the electronic questionnaire also being made available in both Arabic and Russian in upcoming years.

Electronic survey instruments will be a significant component to increasing timeliness along with quality of responses. Paper questionnaires can take several weeks if not months to reach their intended destinations whether they are being sent to a statistician in a Ministry of Education or the Institute itself. Electronic questionnaires are one of the easiest ways to speed up this process although appropriate steps have to be taken to ensure that the data submitted via electronic questionnaires are vetted by the relevant national authorities and hence represent the official response from that country.

The electronic survey capability has been enhanced with the adoption rates in countries increasing year over year <sup>Figure 1, Figure 3.</sup>



**Figure 1**

There are two primary shortcomings of the existing electronic method. The first is that the application must be installed on the respondent's computer as opposed to following a zero-footprint model. This can create usage barriers in countries depending as well as increased demands to quality assure the application. This issue will be taken into consideration as part of the redesign of the instrument for Survey 2006. Secondly, responses to questionnaires must be returned via email to individual UIS staff members (or via postal mail on a diskette) and then subjected to additional handling before capture/processing can be undertaken. While only affecting timeliness to a limited degree, this process does not provide the necessary delivery guarantees and controls that are desired.

Before undertaking any significant changes to the current electronic collection model, an assessment of the state of technology in countries was performed. This was done in conjunction with the annual series of regional workshops in 2003.

A summary of the findings indicates that approximately half (47%) of the computers are using relatively current operating systems and MS Office products from Microsoft. Approximately half (49%) of computers are using older generation operating systems and MS Office products from Microsoft. Many staff have access to the internet but in many cases access is provided in a Managers office and access must be shared and scheduled with the Manager. Internet browser products and versions along with email client software products and versions cover the majority of products available on the market. Almost all (95%) of computers have cdrom readers. A small majority (65%) of system hade PDF reader software installed.

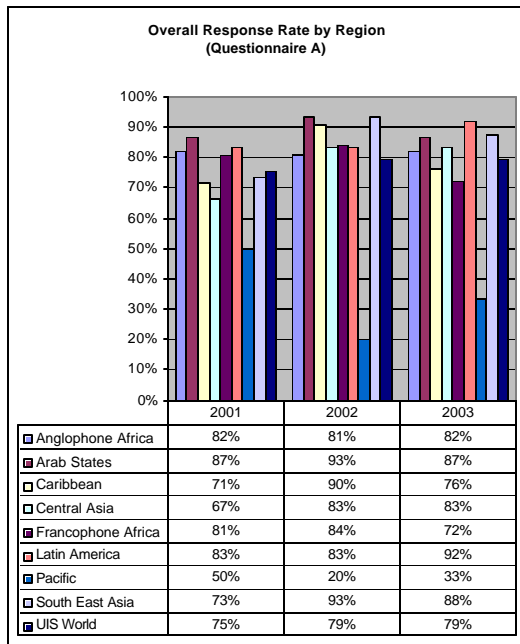
These findings contradicted some popular beliefs held up to that point in time as to what technologies were available. The findings were used to immediately change the delivery of data and products to data users and data providers by:

- Changing electronic questionnaire distribution methods to use Cdroms instead of multiple diskettes as had previously been used.
- Introduce a companion Cdrom to the launch of the 2003 Global Education Digest. The Cdrom contains data tables in electronic format along with electronic copies of many UIS publications. This secondary distribution of data via cdrom has provided users with direct access to the education data in electronic format without the need for internet access. It has also broadened access to UIS publications in electronic format.

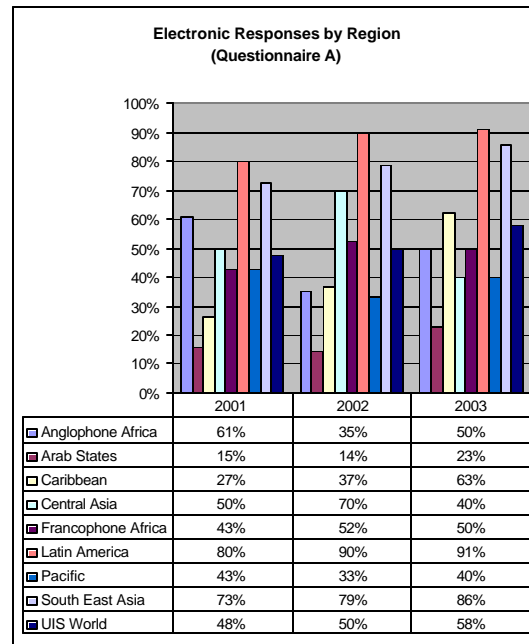
These findings along with feedback from various sources will be integrated into the redesign of the electronic questionnaire for Survey 2006.

Other actions of a more traditional nature have also been taken to try to improve on timeliness – both by issuing questionnaires earlier, setting earlier deadlines and by taking more vigorous reminder actions to follow-up on non-response - all of which have had some success. The shortening of the survey cycle, which has reduced the number of months that Education Ministries have for filling out the questionnaires <sup>Figure 4</sup>, has put additional pressure on the competing goal of increasing response rates. Over the last three

years the amount of time allowed to respondents has been cut in half without negatively affecting overall response rates <sup>Figure 2</sup> or electronic response rates <sup>Figure 3</sup>.

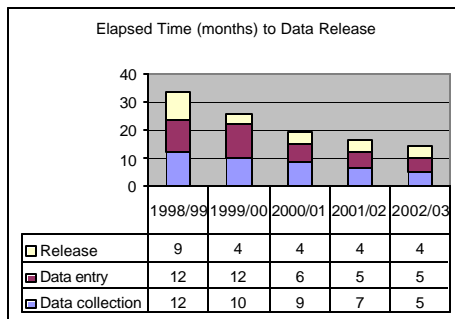


**Figure 2**



**Figure 3**

Improvements in timeliness have resulted in first preliminary results of the survey now being available in the calendar year following the end of the school year for which data are being collected <sup>Figure 4</sup> (although actual publication of data – a few months later – is still nearly two years after the end of the school year in question). Various efforts are underway to continue to improve timeliness so that effective reporting can be done for countries whose education systems are changing rapidly and often unpredictably - as is the case for many of the least developed countries or countries in transition.



**Figure 4**

## ***Accessibility***

Another of the early quality decisions made in relation to the education database, while possibly not immediately evident and not necessarily popular with all data users, was the decision to stop production of the annual UNESCO Statistical Yearbook and focus instead on a wider range of targeted publications containing data as well as some analysis. These include a series of Regional Reports providing analysis for a smaller group of countries covering issues of interest to the given region but also allowing read across between regions on key topics or indicators. A new Global Education Digest together with a Cdrom of data replaces in part the role of the Yearbook but contains an analytical chapter focusing on different topics in different editions not included previously in the Yearbook. The presentation of new analyses of familiar education indicators is intended to make our data more accessible to the average reader as well as to provide greater insights and ideas for further analysis. This will help to ensure the data available are used often and well. The feedback we inevitably receive as a result of frequent and better use of available data help to further improve the quality of data and accessibility.

The UIS has committed to making the UIS web site ([www.uis.unesco.org](http://www.uis.unesco.org)) the primary dissemination vehicle for all UIS data. Furthermore, access to all data is provided free of charge. In order to provide a richer data access environment, the data dissemination was implemented in 2003 using Web Data Server from Beyond 20/20. The initial implementation had the majority of the necessary features but does not provide the most intuitive user interface. However, there has been a great improvement in the quality and consistency of dissemination activities as a result of this technical implementation. The UIS is now able to provide consistently accurate data tables, metadata, and an improved mechanism for managing both footnotes (at cell, variable, and data table levels) and bilingual text.

A common complaint from the user community is that the historic data series is not readily available and it is not comparable over time with the new series. The main cause of the break in the time series was the introduction of a revised ISCED in 1997 that was implemented for 1998 and later reference years but, so far, the past series has not been adjusted retroactively. This is not, in fact, a simple task and when completed is likely to result in some loss of data as, to ensure comparability over time, it will be necessary to present results at more aggregate levels than were collected originally.

## ***Conclusions***

The activities outlined in this paper have touched many of the dimensions of data quality (Relevance, Accuracy, Timeliness, and Accessibility). Some dimensions have received more focus than others. This paper has also outlined some deficiencies in the system or survey that needs to be addressed in the future.

There are many opportunities and activities that could not be adequately covered in this paper. The surest way to improve data quality is to improve the data being submitted by countries. The UIS has been investing enormous resources into Statistical Capacity Building (SCB) activities and these investments are sowing the seeds for the future.

A wide range of activities have been undertaken in efforts to improve the quality of the international education database. The actions have had varying degrees of impact but taken together, significant improvements are being realized.