

EAIR Forum Rome 2006

SIG on Exploiting Data Repositories

Marcel Herbst & Mantz Yorke

1 Introduction

This is the second (and presumably last) note (before the Forum) we send to those who have pre-registered to attend the constitutional meeting for a new Special Interest Group (SIG) within EAIR, focusing on exploiting data repositories.

2 Preliminary Agenda of the Session

As of now, we foresee the following preliminary agenda for the session of our SIG:

- Why we are convening this SIG;
- brief introductory statements by those present — essentially covering the ground of our questionnaire — see Section 3 below;
- discussions in separate groups (depending on the areas of interest), in order to discuss what is being done already, what isn't being done, what might be done on a transnational basis, and so on — challenges, problems, difficulties should be addressed (group work);
- reporting and summaries on the group work in the plenum (by group speakers), and collective discussion of the main issues to emerge;
- organizational setup for the upcoming year (2006–07) and formation of a core group in support of the SIG.

3 Survey of Interests

In response to our survey of interests, we have received responses from Nikolaj Burmeister (Denmark), Debra P. Chaparro (USA), and Kurt De Wit (Belgium). In addition, Mantz and Marcel have provided their answers. The answers of the respondents you will find under the respective question listed below:

Motive to register for this SIG “Tell us why you have registered, what you would like to learn or to explore. Refer to problems you will have to solve at your workplace or in which direction you would like to enlarge your professional interests.”

Nikolaj Burmeister: The corner stone of my work is statistics, including setting up a new Danish database for statistics concerning Danish research universities. The Rectors’ Conference often asks us (me) to find some international statistics in order to benchmark Danish universities with universities in other countries. However, we tend to find (even with OECD-statistics) that once you dig down a layer or two, the statistics mean different things in different countries. That is why I am interested in exploring foreign data repositories — and am willing to tell others about the Danish initiatives.

As part of my job I am supposed to supply statistics on all aspects of the Danish universities and, preferably, include international comparisons. However, there are two major questions which I am particularly interested in:

- How much money is actually being spent by the universities on educating, for instance, a medical doctor, a lawyer, an MBA or an engineer (not how much they charge in fees). Statistics on these aspects are very poor.
- What is the status of various countries in achieving the Barcelona-target (3 percent spent on R&D) — and how do the various countries define the target (what is included and what is not included)? In Denmark, we have seen a huge shift in definition.

Debra P. Chaparro: I am a PhD student in Higher Education from the University of Minnesota, Twin Cities in Minneapolis, MN (U.S.). I have just completed my coursework, but I still have very little experience in using data repositories. Any experience I can gain would be very welcome! I have used our CIRP (Cooperative Institutional Research Program) Freshman survey data. I have training in statistics, although I still find it very confusing, to be honest, and I have used SPSS. I recently attended a CIRP data workshop at UCLA and found it to be somewhat helpful.

My motive for signing up for SIG Exploiting Data Repositories is simple: I want to learn. I just moved to Seville, Spain, where I will be finishing the rest of my PhD (now that my coursework is done, I don’t need to be in the Twin Cities). I have a really strong interest in International Higher Education

and I want to see what databases are out there and how to use them.

Kurt De Wit: I registered for this SIG because I carry out institutional educational research at my institution. To this end, I use the student databases my university has developed (formerly in SAS, now in SAP), but also databases on the level of Flanders and Belgium (available as Access and Excel files).

Currently, I largely use these databases for descriptive purposes (the evolution of the number of students in a particular course, and that kind of stuff). From the SIG, I hope to get ideas about how to do more with them and, perhaps, to learn how other databases are organized and built, in order to improve our own databases (and make them more fit for comparative research).

Marcel Herbst: As one of the two initiators of this SIG (with Mantz), I should briefly state why I thought such a group should be formed. I am a retired higher education administrator, formerly in charge of the planning office at the Swiss Federal Institute of Technology (ETH Zürich), and now working as a consultant in the field of higher education management. My motives to form such an SIG are basically the following:

- I have some interest in quantitative analysis (economics and operations research), and I feel that not enough work is being done with information that is made available in data repositories.
- My former and current work suggests that valuable data are lacking. The effect is that issues which appear to be central to higher education research are not (or hardly) explored. Issues drive data collection and exploration, and if there is lacking awareness of issues, analysis and research will suffer.
- There are tendencies to suppress data in higher education research (ethical issue). Data which are being collected are not made public (because of their political implications). In these instances, higher education research would not obey the rules of research in general.

Mantz Yorke: My motive is a belief that there's a lot more data 'out there' than is typically exploited, and that EAIR ought to be in a position to facilitate collaboration across national boundaries.

Data repositories you extract data from “List the data banks — external to your own institution — that you access for your work, and tell us about the context or purpose for accessing these data repositories. State the research questions that guide that work. Tell us about (external) data that you lack, and why.”

Nikolaj Burmeister: I most frequently use: OECD; Eurostat; Statistics Denmark; Statistics Sweden; Nordic Council; Ministry of Research, Denmark; Ministry of Education, Denmark; Ministry of Finance, Denmark; The Danish Center for Research Studies and Research Policy; THES; Chronicle of Higher

Education; CERD (sic?), Switzerland; The Danish Confederation of Professional Associations.

Kurt De Wit: The Flemish government has a database on all students in higher education in Flanders (Database Tertiary Education). I use this for comparisons between my university and other institutions). It also publishes a yearbook on education in Flanders (available as data files). The University Foundation provides figures about students in higher education in Belgium, but with a time lapse of two to three years.

Marcel Herbst: Since my retirement, I have used data of OECD, the US National Research Council, and CEST (www.cest.ch), among others.

Mantz Yorke: Repositories I use:

- Data from the UK Higher Education Statistics Agency, but the trouble with these data are that they are generally not available free of charge. This is irritating since it is 'our' (i.e. UK HEIs') data which we (institutionally) have to send in and pay for to get back.
- Internal institutional datasets, largely for analyzing assessment patterns of varying kinds.

I have published papers based on data from both types of source.

Data repositories you supply with data “List the data repositories — external to your own institution — that you supply with data, and tell us about the context or legal framework which guides this data exchange.”

Nikolaj Burmeister: None. We only gather and publicize data.

Kurt De Wit: I provide data for the Database Tertiary Education. Every higher education institution in Flanders has to provide data for this database. This was contractually agreed between the government, the Flemish Inter-University Council, and the Flemish Colleges of higher education Council.

Mantz Yorke: Repositories I supply — nothing personally, but institutions do as indicated above, and also supply to agencies like the funding councils and the Universities and Colleges Admissions Service (UCAS).

Focus of your work “Tell us of the particular foci which underly this data exchange:

- uploading data:
 - reporting requirements, i.e. supply supervisory boards or national agency with data and reports (e.g. quarterly or annual reports);

- voluntary data exchange with peer institutions.
- downloading data from government or quasi-government agencies in support of:
 - comparative benchmarking regarding academic achievement or educational outcomes (SAT, drop-out rates, transfer rates, graduation rates, employment or promotion rates, etc.);
 - comparative benchmarking in fields like performance measurement of research output (e.g. publications) or research impact (e.g. citations).”

Kurt De Wit: Uploading:

- Each higher education institution in Flanders has to report annually about student numbers, pass-rates, drop-out rates, etc.;
- each HEI also has to provide data for the Database Tertiary Education. This is an ongoing process, the (academic) year round;
- within the framework of quality assurance, a detailed analysis of student numbers, pass-rates, drop-out rates, market shares, etc., has to be made for a number of courses each year.

Downloading:

- The data we have to upload for our own institution, are downloaded afterward for all institutions, to be able to compare educational achievement and outcomes (for internal use only).

Marcel Herbst: I am interested in comparative analyses (internationally) of higher education systems. In particular, I am interested in a phenomenon which I called the “production-morphology nexus”, namely the linking of output (of an institution or institutional unit) to its morphological (or structural) setup (an issue which is practically unexplored and for which it is very difficult to find the necessary data). There are a range of associated or unrelated questions which would interest me (to name a few):

- productivity of research teams in terms of their setup (using econometric analyses);
- employment/replacement patterns of faculty (using Markov-analysis);
- effectiveness issues (using data-envelopment analysis);
- assessment of the speed with which new research results (of particular fields) propagate, in function of the higher education system.

Mantz Yorke: Focus of my work — assessment issues mainly, that are amenable to ‘macro-level’ analyses.

Issues you have to grapple with “Tell us of problems, including ethical issues, which are associated with this data exchange and which ought to be addressed in our SIG.”

Kurt De Wit: Ensuring the interface between our databases and the Database Tertiary Education; cope with changes to this government-managed Database (changes that are imposed by government; or changes we want and therefore have to negotiate with government).

Ethical issues: except for obligatory data exchange with governmental agencies, we keep most data for internal use only. When external researchers want to use data, we ask guarantees for ethical use.

Marcel Herbst: I have mentioned a grave ethical problem above.

Software and platforms “List the software (and associated platforms like Windows, Linux, Macintosh, Unix, et cetera) your institution uses. State the software (and associated platforms) you yourself use at your workplace (if that differs from your institutional setup). State your own competencies in this regard (rough estimate) and your needs.”

Kurt De Wit: Institution: mainframe software: SAP (and SAP Business Warehouse); PC software: Microsoft Office; database management software: Access; statistical software: Excel, SAS;

My competencies: all fairly good, but limited with regard to Access and SAP.

Marcel Herbst: Now, I mainly use R (as a statistics and graphics package) and \LaTeX (and associated software) as a document preparatory system. I also have some competences in the field of optimization software.