ISSUES PAPER FOR THE AEG JULY 2005 – SNA UPDATE ISSUE 12

FOLLOW-UP TO THE MEASUREMENT OF DATABASES IN THE NATIONAL ACCOUNTS

Background

1. At the meeting of the AEG in December 2004, an issues paper concerning databases (Ahmad 2004) presented the following recommendations made by the Canberra II Group:

- The reference to " large" databases should be dropped;
- All databases, including those built on own-account, should in principle be treated as fixed capital formation.
- The Group also recommended that the current label in the classification of assets for the category comprising software and databases (AN.1122) should be changed from 'computer software' to 'computer software and databases', with software and databases treated as sub-classes.

2. The paper also described a macro-based procedure by which estimates of own-account production of databases could be derived; (see Annex).

3. The AEG agreed that the reference to "large" should be dropped from the SNA and they tentatively agreed that all databases that meet the standard criteria should be recorded as fixed capital formation. However, at the same time the AEG asked the Canberra II Group to clarify some related issues:

a) to provide a definition of "database" and a definition showing exactly which databases should be included (or excluded) in fixed capital;
b) to consider the distinction between creation and maintenance and the implication for the inclusion in fixed capital;
c) to add precision to the nature of employees to be included in the recommended means of valuing own account databases.

4. Consequently, Nadim Ahmad prepared a paper responding to the AEG's concerns, which received the support of the Canberra II Group at its meeting in March-April, 2005. A slightly modified version of the paper follows. This, in turn, is followed by a questionnaire that you are asked to complete.

THE MEASUREMENT OF DATABASES IN THE NATIONAL ACCOUNTS

Nadim Ahmad, OECD

Introduction

1. Each question is considered in detail in the sections that follow, which is followed by a recommendation. First, however, a preamble.

2. The proposals made in this paper try to be as conceptually robust as possible. That is not to say that other conceptually acceptable solutions don't exist, they do, and it's for the AEG to decide whether the recommendations made at the end of the paper are the most acceptable, or rather, the least undesirable.

3. Much depends on how we define a database. Two definitions present themselves.

4. **The first:** that an electronic database consists of a database management system (DBMS, software) and stored electronic information. The value of which is equal to:

a) the value of the information (with a useful life of more than one year) to be stored on the databases,
b) the costs of converting information from one media/format onto the electronic media and format needed for storage on the database; and
c) the software used to support the database (the database management system).

5. This definition challenges us to think about knowledge and information in surprising ways. Key amongst these concerns the capitalisation, implicit or otherwise of knowledge, data, and information. In other words, if we accept that (electronic) databases embody the value of their operating software and stored information, we indirectly open the door to the capitalisation of knowledge (as long as the knowledge is stored and embodied in the database).

6. Of course not all knowledge is/could be treated in this way, since not all information retains a market (as opposed to social) value after one year. A newspaper's scoop for tomorrow's headlines, priceless today, would be pretty much worthless the day after tomorrow. However, many other forms of knowledge (research papers, time series) may retain some residual value after one year (a statistical office database or directory inquiry database for example).

7. Indirectly therefore, by capitalising databases, we (perhaps inadvertently) capitalise many forms of knowledge that are not currently considered capital in the SNA.

8. Capitalising all knowledge that satisfies the one year rule[1], as long as it's accessible and embodied within a database, is probably going a little too far and probably further than was envisaged when the SNA first included databases; particularly because it suggests that whether knowledge is capitalised or not depends on the medium that stores it: if it's on a database it's an asset, if it's in your grey matter it's not. (The example of a telephone directory database shown later provides a good illustration of this 'inconsistency').

[1] It's interesting to note that this includes the value of all R&D embodied on a database, and, that the definition of R&D used here is potentially significantly larger than the Frascati definition being explored by the Canberra II Group.

9. Given that it might be prudent not to indirectly capitalise expenditures that were not currently within the SNA asset boundary it makes sense to consider an alternative definition for a database.

10. **The second (proposed) definition** of a database is: a collection of electronic data stored on a DBMS, whose value includes only (b) and (c) above. More precisely, the value of a database is equivalent to the physical, mostly labour, costs of converting/transferring data/knowledge (with an expected working-life of more than year) from one medium/format onto the electronic medium and format needed for storage on the database, including the software application costs (adapting the software for this particular application, including setting up the structure of the database, loading metadata, etc.) plus the costs of the software DBMS. The information that is accessible via the database retains a separate value to the database even if payments are made for both the information content and efficient access.

11. This definition implies no backdoor creation of new knowledge assets within the SNA classification of assets, and, at the same time, ensures that current existing assets continue to be valued and recorded in the accounts, (even if they are accessible via, but not embodied in the value of, databases). As shown below, this greatly simplifies the empirical and conceptual problem.

12. A further problem that affects both definitions, but is empirically of a much lower order of magnitude for the second proposed definition, concerns the depreciation of knowledge/information/data. For the first definition we need to identify: (I) that knowledge that satisfies the one-year rule; (II) its value (including the costs of transferring it onto the database) and (III) how to depreciate it. The second definition requires (Ia) identifying knowledge that satisfies the one year rule; (IIa) the costs/value of physically transferring this data/knowledge onto the electronic format required by the database; and (IIIa) how to depreciate it.

13. Clearly, both definitions require similar levels of mental gymnastics. The key difference however is that, empirically, the value of (IIa) in the second definition will usually be significantly lower than the value of (II). In addition, whereas trying to estimate the value of the many different (and often unique) types of knowledge embodied in databases is a bit like trying to estimate the length of a piece of string, the value of transferring data from one medium and format onto that needed for the database is at least meaningfully measure-able.

14. Consider for example a statistical office database for GDP and assume, for simplicity, that we have been able to establish which data/information is likely to be used for more than one year. A macro-based method seems to be the only practical route for valuation whichever the definition. For the first definition however, it would be necessary (arguably) to include the costs of all staff involved in the collection, preparation, quality-control, survey-design, compilation, balancing etc of GDP. For the second definition it would only be necessary to include the costs of time spent by staff who transfer paper files say onto the format required by the database: (in some countries this may be zero if say survey returns are submitted electronically in the format required by the DBMS). Clearly the second definition is much easier to estimate. Now consider the ubiquitous nature of databases, and the elusive nature of the estimation of knowledge, which again is ubiquitous, and we see that the practical estimation of databases using the first definition is literally impossible.

15. That is not to say that the second proposed definition will necessarily satisfy purists. For example the value of Directory Inquiry databases in some countries may very well embody the value of the information contained with them. Conceptually, and deliberately, definition 2 omits this value. However this is consistent with the idea that if the information is an asset it should be an asset whatever the storage and access medium.

16. What definition 2 provides is an approach that is:

a) Meaningfully measure-able. In practice the physical costs of converting/transferring data from one medium/format onto that needed by the database are, in most cases, likely to prove negligible and, so, can be ignored. Large transfers, such as Google's project (see below), are however captured;

b) Does not create new classes of (ubiquitous) knowledge assets;

c) Does not lead to the capitalisation of data embodied in statistical office databases (excluding the database software which should be capitalised)

d) Largely obviates the need to identify data and information that has a useful expected working life of more than one year and so also reduces the need to estimate depreciation profiles for (often unique) data and information.

What is a database?

17. Although there appears to be some uncertainty concerning the definition of an electronic database amongst national accountants, dictionaries appear to have converged on the following definition:

> *A collection of information organised in such a way that a computer program can quickly select desired pieces of data - an electronic filing system. Information can be accessed or entered via a database management system (DBMS).*

18. And, so an electronic database consists of two parts, the DBMS (software) and the electronic data whose access is facilitated by the DBMS. This seems pretty clear. What is not so clear however is whether the value of the database includes or excludes the value of the data/information stored in electronic format and accessible by the DBMS; that is, definition 1 or 2, above. Before moving on it's interesting to note that some references to databases suggest that they can exist separately to software; e.g. as paper based filing systems. The treatment of these types of databases in the national accounts is considered later.

19. For now the discussion is restricted to electronic databases and, according to this definition, virtually all types of data stored electronically and accessible through a software system form part of an electronic database. This includes customer banking records, electoral registers, directory inquiry databases, statistical office databases, personnel administrative databases, company registration databases, and even the electronic discussion group database set up for the Canberra Group papers. That's not to mention the multitude of databases available on the internet, such as greeting cards, encyclopaedias, screensavers etc.

Which databases should be considered as economic assets?

20. This issue has been discussed on a number of occasions by the Canberra Group. We have generally always agreed that all databases that satisfied the one-year rule qualify as economic assets. Proposals to limit the scope of databases to data providers or commercial industries were considered but not pursued because they were not satisfactory (conceptually incomplete).

21. I've listed a number of databases below going from Yellow Pages to administrative (personnel) and statistical office databases, including databases where access is free or paid for (directly or via advertising revenue). I can see no good reason for the Group to change its earlier recommendations that virtually all databases that satisfy the standard criteria should be considered, in principle, as economic assets. Certainly the software component will nearly always be considered as fixed capital formation.

I. Databases used purely for internal administrative purposes: such as personnel databases.

II. Databases developed for internal administrative purposes but with dual applications: such as databases of customer records and credit-ratings held by catalogue companies.

III. Databases used to deliver market services: for example Directory Inquiry databases.

IV. Databases sold in their entirety as reproductions, such as Microsoft's Encarta or the Encyclopaedia Britannica.

V. Databases that provide free access but where advertisers pay fees: for example the forthcoming Google encyclopaedia[2], Yellow Pages, and FriendsReunited.com

VI. Databases run by public administrations, where access is free: for example national accounts databases and databases on births, deaths and marriages (and which generally provide internal and external services).

VII. Database libraries of research papers/books/documents held at universities, or any institution, including the papers on the UNSD website: access to which may be free or via payment.

VIII. Databases of information held for less than one year and not used continuously, such as company account databases for tax purposes.

22. Clearly many electronic databases satisfy the economic asset criteria set out in SNA 10.2 and 10.3. Most obviously Type (III) databases, for example databases held by market research companies that sell data and directory inquiry databases owned by telecommunications companies. But does the definition of economic assets extend to other electronic databases for example Statistical Office databases or databases developed purely for in-house administrative purposes. Unfortunately the SNA provides little further guidance. However it is clear that even these databases satisfy most, if not all, of the criteria set out in the SNA: they provide economic benefits to their owners (why else have them) and even if, in some cases, they can be copied and acquired or used freely by others, the versions retained by the (original) owners are still owned by them. The fact that copies can be made or that access is provided free of charge merely affects the value of the asset held.

23. At the recent OECD Working Party on National Accounts meeting a number of delegates expressed a view that statistical office databases should not be recorded as fixed capital formation on the grounds that the information derived from the databases was publicly available. Although this is partly true (some information is not publicly available) the case of Google shows that this does not necessarily present good grounds for the exclusion of these databases from the asset boundary (roads are another example). Moreover it is indisputable that the productivity of statistical offices has been greatly enhanced by the development of these databases – in other words, they provide statistical offices with clear economic benefits.

24. So, it would appear that, with the exception of those databases that can be deemed 'small tools' and those whose data do not satisfy the one-year rule, ***all databases used continuously for more that one year should be treated as fixed capital formation***.

25. As shown below, using the first definition of databases, this becomes an almost irresolvable measurement problem. But not if the second definition is used. Since, in this case and in practice, the value of electronic data stored on databases, excluding the value of the content of the data, will usually be negligible; even if the electronic data is able to satisfy the one-year rule. And so, although most databases should be capitalised, with definition 2, this amounts, in most cases, to only capitalising the software

[2] In December last year Google announced its plans to convert the holdings of many US national research libraries and Oxford University into digital files that would be freely available on the Web. Librarians have predicted that the project could take at least decade, at a cost estimated at $10 for each of more than 15 million books and other documents.

component (DBMS), which is already the case. (Later, I argue for practical reasons that the value of databases should only reflect the access costs and not the value of the DBMS which will normally be recorded as software).

Definition 1

If databases embody the value of the information stored within them how should we value them?

26. A key question here is what information provides services for more than one year, since it is length of the expected working life of the data that determines whether it is GFCF. For some databases, there's no definitive and singular answer to this. It depends on how we view the stored information. For example one could argue that the personnel records stored within an office are updated every year and, so, the database (excluding the software component) never provides services for more than one year. On the other hand one could just as credibly argue that most of the records and details (staff number, home address, health details, career history) continue to last for more than a year, assuming normal levels of staff turnover, and that the updating of the database merely reflects increased investment or maintenance (discussed below). Although, clearly, some records such as remaining annual vacation entitlement are almost certainly not investment.

27. Better and perhaps more contentious examples are statistical office economic accounts databases. These clearly provide economic benefits, and, so, could be considered as economic assets. The fact that payments are rarely received for the services they provide is not relevant, since the same argument could be applied to other assets used to provide non-market services (for example, computers, software, roads etc). However, do they provide services for more than one year? Yes, if we think of these databases as providing historical and recent data, where revisions are seen as improving quality. But, not if we think of these databases as being completely revised each year, or where almost all of the value in the database is embodied in the statistics for the very latest year.

28. It's certainly true that a disproportionate amount of value is probably embodied in the latest year's estimates, for example 2004 statistics are probably worth more than 2003's and a lot more than 1963's, but it's unlikely that any single year's data is worth more than half, say, the total value of the database. And so, on these grounds, it seems more likely than not that statistical office databases, including the value of embodied data, should be viewed as economic assets.

29. Some statistical office databases are less affected by these arguments, for example, registry databases on births, deaths and marriages (where it could even be argued that data from later years is just as valuable, perhaps more so, than data in earlier years, certainly for genealogists this is the case).

30. For other databases the position may be a little clearer, for example, the information embodied in telephone directories generally lasts more than one year (although even here it's not immediately obvious that a 2004 database is not completely superseded by a 2005 database).

31. Google's planned new encyclopaedia site, which is estimated to cost in the order of $150 million (people scanning books being the main cost) is another contentious example. If users pay to access some premium content information such as books, (after Google acquires the rights to electronically reproduce and charge for them), how should these books be valued within the Google database?

32. Some databases however clearly do not satisfy the rules of economic assets; in particular the one year rule. For example, databases of financial statements in some companies recording invoices and outgoings within a calendar year for tax purposes, or databases maintained by Ministries of Finance to monitor Departmental spending during the year.

33. As can be seen, identifying data that meets the one-year requirement is relatively elusive and very difficult. But estimating the value is even more complicated. For example, for a statistical office, the costs of producing data are potentially comparable to those of running the office and could include the costs of employees working on quality (such as national accounts' GDP balancers) as well as data in-putters. This is a challenging issue, just as challenging however is the fact that definition 1 appears to create a new class of knowledge assets, dependent only on the fact that they exist within a database.

Does definition 1 create an 'inconsistency': Information stored on databases is (part of) an asset, the same information stored elsewhere is not.

34. At this stage it might be useful to consider a relatively surreal example. Let's consider the Luddite telephone company that keeps all customers' phone details (name, address and telephone number) on paper files, with the files organised by name only. When a request for a phone number is made somebody in the data warehouse has to leaf through the thousands of records. Although income is generated as a result of providing this service, via directory inquiries, my reading of the SNA is that these files are not gross fixed capital formation. To improve accessibility and to speed up its service the company subsequently decides to store all of the records in a three dimensional repository sorting the records by three fields: name, region and postal-code. The re-organisation of these files will clearly lead to an improvement in accessibility and timeliness but again (my interpretation) the files are still not treated as gross fixed capital formation, leading to improvements in TFP. Slowly accepting the winds of change, the company decides to automate the search process buying in a fleet of robots that can scan the records automatically. The acquisition of the robots is clearly fixed capital formation, but the paper records remain unchanged. With a Force 10 gale blowing up the company gives in to the inevitable and converts all of the paper files into electronic format and places them onto an electronic database. We know the software (DBMS) is fixed capital formation. The key question is: is the value of the content of the electronic data also fixed capital formation. If so, what makes the electronic data especially different to the paper copies that it should be treated differently? Should the storage media matter?

35. One could argue that the paper copies of records might be considered fixed capital formation but this is an issue the Group should consider in its deliberations over the classification and nomenclature of assets. It is certainly true that users may occasionally pay for the knowledge/information embodied in databases, and that this knowledge contains asset characteristics. But the knowledge displays these characteristics whether it's embodied in the database or not. And the SNA explicitly recognises, for good reasons, that not everything that displays some asset characteristics should be recorded as fixed capital formation. For consistent treatment, therefore, this suggests that databases should not embody the value of the information stored within them (and that they should be valued as set out in definition 2).

36. But this is not the only problem, where valuation is concerned, as shown below; where we consider the distinction between investment and maintenance. Although this also concerns definition 2 the scale of the problem is significantly smaller for definition 2. Since, for definition 1 we're trying to establish the circumstances under which the value of new data/information should be added to the database whereas, for definition 2, we're only trying to establish the circumstances under which the value of the physical costs of transferring new data from one medium/format onto the electronic format needed by the DBMS should be added to the DBMS.

How can we distinguish between creation and maintenance?

37. Let's stay with statistical office databases. If we view the addition of a new year of data as maintenance, then, a logical argument can be given to show that, apart from the DBMS, no economic asset exists, since every year that was added in the past also represented maintenance. The sole (possible) exception being the first time all paper records were converted onto an electronic format, and, because, in

most countries, this would have happened a long time ago, it can be assumed that the value of this effort has become negligible due to obsolescence.

38. If, on the other hand, we view the database as an asset then the addition of a new year's worth of data reflects a quality change and, so, new investment.

39. In fact, strictly speaking, because maintenance merely returns an asset that has depreciated through wear and tear back to its original state, there is no such thing as maintenance of the data embodied in a database. Since, no wear and tear occurs (depreciation is through obsolescence) and the addition of new items improves the quality of the database, taking it beyond its original state. So, in theory, these changes should be viewed as fixed capital formation if the data that are added are expected to be used for more than one year and as expenses otherwise.

Conclusion (definition 1)

40. Definition 1 therefore implies the following: The software component of all databases should be capitalised. All information that is stored on the database and that is expected to be used for more than one year should also be included in the value of the database. Care should be taken to ensure that reductions in the values of other recognised assets, for example originals, are reduced by the amount of value embodied in the database. Where it is difficult to estimate the value of information stored within the database a macro-based method should be used. This would need to include the costs of all individuals involved in information creation (including an estimate of the proportion of the value of the information that is embodied in the database[3]).

41. For statistical office databases the implication is that much of the output of the office is database creation. Indeed one could argue that excluding support services (who may have their own databases to manage), the work of most people in the office is involved in producing the database, since, this group includes not only those individuals responsible for manually inputting returns into electronic format but also those responsible for ensuring the quality and coherence of the data.

42. Providing a precise definition of these individuals as well as those that work on updating and managing personnel systems, directory inquiry systems, is arguably not practicable. But it would result in a definition along the following lines:

43. Macro based estimates of fixed capital formation in databases should include the time spent by all individuals involved in the updating, processing, quality-assurance, creation (including scanning and inputting) of data used in databases that provide services for more than one year. This includes statistical office databases and personnel record databases. (Although, in the latter case, the costs may not be significant and, so, can be excluded for convenience.) Costs should include relevant labour costs as well as other input costs; in particular, any costs involved in acquiring the data.

44. This is not realistically achievable. Literally everybody in information industries works on developing information and so, indirectly, database creation, and many others in non information industries also spend part of their working day working directly or indirectly on databases. To achieve the AEG's request therefore of "adding precision to the nature of employees to be included in the recommended means of valuing own account databases" is to, all extents and purposes, impossible if definition 1 is used.

[3] Another elusive entity.

Definition 2

45. ***If however we decide that databases do not embody the value of the data/information stored on them*** the empirical problem becomes significantly easier and solvable. One would only need to include the costs of converting data onto the electronic format needed for database management, and, only then, for data that had an expected working life of more than one-year; since it is the expected working life of the data that determines the life of the costs incurred in entering the data. In many offices, particularly those that collect information electronically these costs would probably be negligible and ignorable, particularly if the value of this work incurred fairly rapid depreciation, which is likely to be true in many databases.

46. On balance therefore, all databases whose data provide economic benefits for more than one year are in scope, including those used to support ancillary activities and statistical office databases but only the costs involved in converting data/information into the electronic format needed by the database (and that has a working life of more than one year) and the DBMS should be capitalised. For the databases of statistical offices, for example, these might be quite small and may largely comprise of establishing the metadata and structure of databases. Where significant activities of this type occur however (for example, scanning or data in-putting) the costs cannot be ignored (for example Google).

Recommendations

47. And so, definition 2 implies that a database is a tool that provides access to information. The value of the database reflects the value of the DBMS (software), which may already be capitalised in the accounts, plus the costs of converting data from one medium/format to the medium/format required for access by the DBMS. The value of the database does not however include the value of the information that can be accessed by it. The data/information may already be recorded in the accounts as assets for example originals or they may not be, for example paper records.

48. Some payments for using databases may include payments for the information accessed by the database as well as the access costs. The services provided by the database are for access services, which are provided by the DBMS, and for the fact that information has been made accessible by the database (the value of which is equal to the costs of converting information into this format, making sure not to include the costs of the information itself). Only the loading costs for data and metadata and DBMS application costs that have an expected service life of more than one year should be recorded as fixed capital formation. All updates to a database that satisfy the one year rule should also be treated as fixed capital formation.

49. Most creation of databases occurs on own-account. Values of theses databases could be estimated using a sum of costs approach. As part of the costs, all employees involved in converting data from one format/medium to the electronic medium and format required by the DBMS (for example the costs of data-inputting or data scanning, where the data is to be used continuously for more than one year), should be included; ensuring that only the time spent on these activities is included.

50. **In practice the software component of databases will already be recorded elsewhere as software**. Quite often this will have been estimated using a macro-based method, meaning that the software component of databases is not easily identifiable. **As such it seems simpler <u>(and is recommended)</u> to adopt the convention that own-account databases only includes the costs involved in converting data from one medium/format to that required by the DBMS, including the application costs (adapting the software for this particular application, including setting up the structure of the database, loading metadata, etc.).** This allows for macro-based estimates of databases to be derived as follows:

*Total number of employees working on converting data with an expected working life of more than one year from one medium/ format onto that required by the database and on the DBMS application **

*Average remuneration **

Proportion of time spent on these activities+

Other intermediate costs used in these activities (not including any costs associated with the acquisition of data) +

Notional operating surplus related to these activities (costs of capital services, for example capital services of scanning machines and computers)

51. Where the ownership of a database changes hands the payment may reflect three components – the value of the information contained within the database, the costs of converting information from their original medium/format onto that used by the database; and the software (DBMS). Each of these transactions should be considered on a case by case basis in determining the asset class of the transfer. Where database reproductions are purchased/produced the rules on originals and copies already formulated by the Canberra II Group apply.

Points for discussion:

1. Do you agree that all databases holding data with a useful life of more than one year are a fixed asset?

2. Do you agree that the valuation of a database should be that of Definition 2, but exclude the value of the DBMS, which should continue to be recorded as a software asset?

3. Do you agree there is no maintenance entailed with databases and that all set-up and updating costs should be recorded as capital formation?

4. Do you agree with recommended method for deriving estimates of own account database capital formation?