Databases which support rather than inhibit Australian Aboriginal ways of knowing and making knowledge

Digital Technologies and the Intergenerational Transmission of Australian Aboriginal Knowledges

ARC Linkage Project: Seminar
Michael Christie
Faculty of Indigenous Research and Education (FIRE) Seminar Room
Charles Darwin University
24th September 2003

In September 2003 the ARC (Australian Research Council) awarded a three year grant for user-friendly database development for indigenous communities. The database development will be aimed at IP owners who wish to explore the uses of digital technology for teaching young people about their land, language and cultures. The grant has been awarded to the Charles Darwin University in collaboration with partner organizations; the Yothu Yindi Foundation, the NT Department of Infrastructure Planning and Environment, and the North Australian Indigenous Land and Sea Management Alliance (through the NLC). This seminar is on the subject of the use of digital technologies for the intergenerational transmission of indigenous knowledge and will raise issues to do with digital technology and Aboriginal knowledge, including technical, philosophical and pedagogical issues.

1. Background

About 18 months ago people from the Centre for Indigenous Natural and Cultural Resource Management (CINCRM) at Charles Darwin University, the Northern Lands Council, the Larrakia Nation (traditional owners of the Darwin area) and other organizations, gathered to discuss digital technology and the protection of Indigenous ecological knowledge. One interesting feature of this meeting was a certain tension between some people emphasizing the need to digitally document traditional knowledge systems, and others saying conversely that Aboriginal knowledge is in the ground, it lives in country, it is embodied in people, relationships and performance. How can you put that sort of knowledge into a computer without destroying that which makes it live? There were Aboriginal and non-Aboriginal supporters for both sides of the argument. We began to tease out issues of politics and epistemology which arise as computers are brought to serve the work of old people teaching new generations of young people.

2. The Original Yolngu Studies Database

We already had some experience using digital technology teaching Yolngu Languages and Culture at Charles Darwin University (CDU). Yolngu Aboriginal people live in Northeast Arnhemland in the Northern Territory of Australia. When we started the program, one of the things that senior Yolngu custodians in Arnhemland emphasized was that students should not study only one Yolngu language. They are required to

study a range of them. All political and religious arrangements in the Yolngu world must always involve people from opposing moieties and thus differing language groups. The weaving together of opposites is fundamental to agreement making. So students need to understand something of the full spectrum of Yolngu culture through the manifold perspectives of the people who own particular languages and associated pieces of land, song, art and ceremonies etc. Therefore, with permission, we started to develop sets of materials that contained a wide range of Yolngu representations of aspects of Yolngu life in a range of languages. We have used an on-line database to make some of these materials available to the public. (###**URL address www.ntu.edu.au/yolngustudies))** After 30 years of bilingual education in Yolngu schools we had access to a vast array of photos, tapes, videos, books and transcribed stories that continues to grow and requires protection and organisation. Particularly pressing is the need for a system of digital archiving that supports rather than inhibits Yolngu ways of speaking, acting, reflecting, representing and nurturing Yolngu knowledge.

We asked the University Teaching and Learning Division to set up a database, and provided them with a set of digital objects (or 'resources') and a series of metadata text files. We structured the metadata as far as possible according to the Dublin Case (spelling?) protocols. (put in some detail here about Software (Jayel) and metadata fields). Together we negotiated suitable search path and interfaces. Over a few years we ironed out bugs and implemented some new ideas, including a key-in text search which searches all the metadata and text data. This is discussed more fully below. Overall, the database, has very usefully served the purposes for which it is designed: making freely available to students and researchers a small range of fully negotiated, authentic representations of Yolngu knowledge in Yolngu languages. The only real frustration has been with the difficulty of uploading from a remote computer.

3. Producing databases for other contexts of Indigenous knowledge production : Some first principles

The Yolngu Studies database owes much of its success to the simplicity of its function and purpose, and the ways in which its design was negotiated in the context of its use. As we begin to look at the much wider range of so-called 'knowledge- management' purposes to which digital technologies are employed in indigenous communities, the relation of purpose to design becomes much more complex.

Of particular concern is the ways in which unexamined assumptions about the political and cultural innocence of digital technologies of the structured, abstract and objective nature of knowledge, lead to uncontested filtering and assimilatory effects. To pin down and address these problems we need to address some fundamental questions: Firstly, What is the range of purposes to which the database will be put? In our particular case the databases will be developed quite specifically to support the intergenerational transmission of traditional knowledge (especially that knowledge which could be referred to as 'ecological'). This is quite different from the development of an archive; for example, a database, which is to be an active participant in the ongoing Aboriginal work of knowledge production, needs to be responsive to the philosophies, contexts and practices of young people learning from their elders. A database is not a repository of knowledge. It is rather a repository of representation or artefacts. How these artefacts become involved in producing new knowledge in specific contexts needs to be studied carefully.

This work of creating a technical object like a database within a specific social/political field involves making decisions, which could be represented relative to two continua on a graph. One continuum relates to specificity of purpose. How contextualised is it? Do we want it to be only for this particular purpose with these particular people with these particular books and records? Or do we want to construct the database so that, no matter what your archiving or educational purpose is, it can be plugged in anywhere, incorporating all sorts of different data sets, by using the same sorts of coding, the same sorts of metadata protocols, and the same file types?

The other continuum represents complexity at one end and simplicity at the other. Complexity allows us to perform advanced searches, relating documents together and finding them quickly. Good metadata structures are especially important when we have very large data sets. Increasing complexity in metadata structuring design, (even or especially those which are coded to reflect the particular sorts of relatedness at work in Aboriginal knowledge systems, among land, language and species of plants and animals etc) necessitates increasing difficulty of input and upload. The question of how- or in fact whether- indigenous Australian knowledge structures can or should be incorporated, as structure into database coding will be discussed further below. But first there is another 'first principle' to be addressed.

Every time we try to determine some aspect of database operation, the determination implies something of how the data and metadata will be uploaded in the first place. If we are providing databases in the first instance for the purposes of indigenous people

who own the information (rather than non Indigenous people who might want to access it for whatever reason), we need to ensure that it is easy for them to use. Easy upload implies simple data structures and vice versa. If they are to support traditional knowledge practices they should also employ and encourage traditional languages. However they tend to favour English language and its structures.

Equally problematically, many young people are not fluent speakers of their own traditional languages, and the older ones are not highly literate. Input and search must become 'friendly' language-specific, purpose-specific, and in negotiated context-specific ways. This is a complex simplification process. The more simple it looks, the more complicated the coding may be, as in the example of the lemma search, where we go to considerable lengths to make the interface simple, requiring behind the scenes, much coding to enable the "fuzzy" search mechanisms to generate and display approximations in spelling or recognize voices.

So we may need to turn our attentions away from the complex, versatile, 'future-proof' databases that will work for a range of purposes, and can plug into other databases elsewhere. We may be better to start with a closer look at how the owners of the information it contains hope to use it, first in teaching their young children, and possibly in their own way and time, in their wider collaborations with students, researchers and scientists. We may best aim for a database system which is able to remain small and local and whose use is intuitive for the people who are the contributors and custodians of the representations they contain.

4. Cultural bias in data structures

We tend to think about databases as theory-neutral and politically innocent, but they naturally carry within them hidden assumptions that may make it easier and more profitable for some people to access than others.

The most fundamental cultural bias at work in a database is epistemological. Western science has long preferred to understand its work through the metaphor of representation: the goal of Western scientists is the production of increasingly accurate representations of the real world. It is in fact this rubric that has prepared the ground for the current flourishing of complex data management technologies in the first place. If however, knowledge from the Aboriginal perspective, is more often understood to be something that people perform, if knowledge is something which is

'in the ground', if it is embedded in the relationships people identify with their land, their totems, and their histories, and which they perform through their narratives, their art, dance and their song, then what exactly is it that is stored in the computer? We may best think of it as information, or data, rather than knowledge. The unexamined relationship between data and knowledge, and unarticulated assumptions about how each produces the other, contribute to the cultural biases at work in databases developed for indigenous knowledge.

Two parallel cultural assumptions: (a) that databases contain knowledge and (b) that education is transferral of knowledge from one head to another, need to be revised if Western scientists are to facilitate the ownership, control, and use of databases by Aboriginal holders of ecological knowledge.

A critical political investment is made in the structuration of the metadata, that is, the ways in which the metadata is distributed into fields and encoded with options as to how the fields may be filled (with pop up lists or keyed in data, for example). Working within the Western scientific tradition it is natural to assume that the world is already intrinsically structured before we arrive to do our work of making knowledge. In this knowledge tradition, we actually assist and expedite the processes of knowledge production by predicting the sorts of work that people are going to do, and encoding appropriate structures into the metadata.

In other cultures, language is understood less as representational and more as performative, having a strong constitutive power in the production of knowledge and new realities. In that epistemological context, the pre-emptive work that is done sequestering information into metadata fields, is actually going to prevent the creative knowledge-production work which is enabled through the skillful use of words. For example, somebody's name may well also be a place name. It may well be a ceremonial act or object, or it may label a connection between two groups or places. Words, in an Aboriginal language, have their power in the work of knowledge production, by virtue of their potential relatedness to any of the different fields in the metadata structure. In other words they attain their power precisely because they resist that categorization. So whatever the heuristic efficiencies western thinkers may achieve through the distribution of metadata into categories, the same processes may equally inhibit the scientific work of cultures with differing ontologies. Collapsing the structures of metadata and flattening out its content may enable the creative connecting processes upon which Aboriginal knowledge making depends. The philosophy and practice of the production of indigenous knowledge, actually implies

something quite concrete for the coding of software which deals with digital objects in the context of education.

Hence databases are said to have ontologies: assumptions about the fundamental nature of what they contain. Aboriginal philosophy works through particular ontologies that can be encoded directly into database software. Western techno-science working to create useful technologies for Aboriginal people must seek to avoid dragging Aboriginal knowledge through the filter of western objectivist ontology.

Data structures can be seen already to influence the social practices of techno science in the west. Bowker (2001) looks at the relationship between computer databases and biodiversity, demonstrates how different understandings of biodiversity inform and are influenced by the way in which people develop and use databases.

For example, 'grooving' is a process whereby the data structures of databases actually affect the way in which we understand the world. Some things in the world are a lot easier to identify or define than others, they make their way without difficulty into databases, and thereby become constitutive of the theory of reality through which we think. Other things however that may be harder to define, or are contested or have fuzzy boundaries, or are radically singular, (in that they are unlike anything else), will fall through the cracks. They simply fail to make it into the database. After a while, we develop a represented world within our database, which takes on a particular structure or regularity not so much as a reflection of the reality of the world itself, but much more a function of the data structures that we have chosen to depend upon in the first place.

Bowker goes on to identify a subsequent process: 'reverse bootstrapping' where we start to make assumptions about the nature of the world on the basis of the structures of the data that we are using in our database. In research areas, such as biodiversity and language diversity, fundamental (and often unconscious) assumptions about the nature of reality (for example, the ontological status of an ecosystem or species or language) will inevitably become 'hard wired' into the architecture. Their databases go on reflexively to affect the way in which the researchers understand and reinvent the world outside. In the case of Aboriginal knowledge, some things which are perceived to be more charismatic than others, (crocodiles as opposed to algal blooms, for example – both of which are totemically significant in the Yolngu world), find their

way into the database, whereas other things, that are equally important in terms of Aboriginal knowledge don't make it.

There are two directions to search for a solution. One is to encode the complex connectivity of Yolngu knowledge in the data structure. This approach is already at work in the "42-level relational database to catch the way the Yolngu people think about the natural world" at the Galiwin'ku knowledge centre (The Australian, June 10, 2003, pg. 29). The other choice is to do away with the attempt to hard wire the relationality into the database, collapse the metadata categories, and create the conditions whereby indigenous owner-users can learn to invoke the multiple connections in the context of database use, simply by concentrating upon text.

5. Obviating the strictures of text computer literacy

We turn now to a second major issue for indigenous custodians and teachers of ancestral knowledge seeking to use database technology in their work, is the dependence of databases or particular practices of literacy. There are already at work indigenous practices of literacy, already marginalised by text literacy, this marginalisation compounded by the assumptions at work in the logic encoded in everyday interface design. Some of this logic is a direct legacy of the convention of text literacy (eg left-right) and some of it is derived from convergence of design and use convention (buttons etc) specific to history of computer interface design. Graphic user interfaces go some distance obviating dependence on text, but the bottom line is that computers depend upon strings of text (or numbers) for their search strategies. Even with graphic user interfaces, resources must be located by the search function through a process of matching generated strings of text or numbers against texts or numbers in the metadata.

How can this text dependence by harnessed to enhance rather than marginalise the use of endangered languages? We may in fact be better off to call for training for the databases, to make them work in such a way that they are quite intuitive, specifically for the purposes to which their owners are putting them. Train the computers, not the users.

Over ten years teaching Yolngu languages and culture at CDU, we have developed a language teaching CD, a dictionary CD, a library CD and an on-line database. In the collaborations between Yolngu and non-Yolngu, much effort has been expended

upon making searching easier. So, all successful database searches depend upon a correct text string. The basic bottleneck has always been making the Aboriginal languages central to the process, and creating useable search strings in these languages. Aboriginal languages can be very difficult to spell for a range of phonological and orthographic reasons. They are also very different from each other so that implementing language-friendly searches in a variety of contexts would require a range of specialist work. Furthermore, some of the older people who are the most important knowers and potentially the most creative configurers of the information, are most often the least able to generate good strings of text for the search because their text literacy skills may not be very good. Their grandchildren on the other hand may have better text literacy but less facility with the language and conceptual systems they require.

Herein lies the most debilitating bottleneck in the work of a database. The database user approaches the software through a particular search path with its own series of interfaces. The only way he or she has of navigating through to a successful 'hit', is by generating a string of text, which can match, element for element, a string in the metadata (or the data). The text string must be engaged by the search to call up all resources whose data contain that text string. Creating the correct string is the bottle-neck. The solution is to find canny ways of enabling poor spellers, or even, good speakers with little spelling ability to generate a workable string.

One way of doing this is to use voice recognition. Another is to use a 'fuzzy find'. Both these processes depend upon a 'lemma' form which could be generated as follows. As soon as text is put into the database, it is glossarised. There is a 'machine' that strips out hyphens, commas, spaces etc. This list of words is then processed by an algorithm which generates a lemma form for every word. A lemma form is a basic approximation of all the complexities of a word. For example in Yolngu languages there are a number of different consonants which sound a bit like an English 'd'. They are written d, ḏ, dh and dj. In some positions is a word they can be confused with t, ṯ, th, and tj. There is also a vowel 'a' which can be under some circumstances lengthened phonemically and spelt 'ä'. There is also in some contexts a phonemic distinction between g and k. These differences of course are unconsciously 'known' by native speakers because they use them to generate differences in meaning. But they may not be easy for people to spell. In the list below we see a number of lemma forms generated by algorithm from the real forms: (for example "for 'dh', and 'ṯ', and 'd' write 'd'", "for 'g' and 'k' write 'g'", "for 'a' and 'ä' write 'a'")

dhiku diku

deku diku
degu diku
dhigi diki

Once you have a list of lemma forms, for all words in your data, you are able to perform a fuzzy search. This can be by voice recognition, keyboard entries, or by clicking on graphics. Your input will be matched against the lemma list, and from that, generate (backwards, as it were) a list of possible words you might be looking for.

There are other ways to simplify the approach to finding a text string. For example, there is the 'filter find' where all the possibilities for the first letter (or digraph) of a word are found in a pop-up menu, (which selects all words starting with that letter) then a selection from the possibilities for the second letter (presenting a list of all words in the glossary starting with these two letters). Then the third, and so on. This is a way of narrowing down the corpus without the use of a lemma form. Decisions about how to generate 'smart search' processes like these depend upon the phonological morphological and orthographic conventions in particular languages.

6. How databases can be used by the owners for the intergenerational transmission of knowledge.

The priority for contextually useful digital technologies must lead us away from delivering pre-packaged systems with metadata, and search protocols in place. They need to be negotiated in situ. The first basic principle for developing ownership and usefulness would be to start with very limited data that has been produced by the people on the ground which they have decided will be useful for their own purposes in teaching younger generations, making collective memories, or celebrating connections. The most useful interface solutions need originally to be negotiated without reference to constraints of programming.

We need to find ways of using vernacular languages in digital technology that support their use, and in some cases their revitalization, using plenty of sound aid, canny searching, and processes that encourage people to depend on what limited language they may have, while strengthening it. Currently, computers tend actually to prevent people from using their own languages.

Secondly, we need to conceive of a database in terms of the potential for its use. The way in which a database is structured needs to be understood in terms of the uses to

which it will be put, rather than what we understand to be the nature of the knowledge that we think it might contain. Metadata should not reflect what we think adequately describes the object, but rather how it could possibly be retrieved.

Thirdly, we need to radically minimise the structuration of the metadata. Complex metadata both pre-empts the complex connectivity upon which Aboriginal knowledge depends, and also renders the processing and upload of new objects very difficult.

The database itself is a text. Not only does the data in the database need to be read through some sort of a literacy practice, but the database itself needs to be read both materially and discursively.

Finally, the database is not a repository of knowledge; it is a technology whose nature and function is determined by the context of its use.