

Towards a Unified Theory of Stuff

By Glenn Sanders BA, Dip Lib, GDDM, MBII, ARMA

Biographical Details

Glenn is one of Australia's leading consultants in document and records management. He has worked on software development for three commercial systems, written several books and articles, and been document manager for Tyndall Australia, Deloitte Touche Tohmatsu and (currently) EnergyAustralia. He is an active supporter of the RMAA and RECMGMT Listservs.

Abstract

The debate about the relativity of 'documents' and 'records' is preventing theoretical advances in our discipline. This note suggests 'stuff' as a single, unifying alternative, and discusses several paths for future research in the light of Quantum theory, info-genic inheritance, automatic indexing and S-space.

There is recurring debate¹ within and around our discipline about the definitions of 'record' and 'document'. Most claim that records are a subset of documents, others the reverse. This is usually expressed as a continuum, though this approach sometimes confuses life cycle with workflow and can have difficulties with short-term, ad hoc status change. Often, the continuum approach reveals that the argument is specious, as the relativities are the same, it is only the terminology that differs.

To introduce more confusion if not debate, we overlap terminologically with related disciplines (eg computing and librarianship), which often have conflicting definitions for terms like 'record', 'file', 'volume', 'document' and 'information'.

Too often, we retreat into specialisation (jargon), talk about statutory responsibilities, and develop highly legalistic definitions which are irrelevant to our customers and further distance us from the immediate core business needs of the organisations we serve.

This debate has gone on long enough. It distracts us from our immediate business needs, divides us from related disciplines we should be working with instead of against, and diverts us from serious research into the real questions and challenges we face. We need a new term, a single unifying term to represent all the things we manage, regardless of whether we call ourselves managers of documents, records, knowledge or whatever.

After all, 'computing' has moved over the last forty years from 'ADP' and 'EDP', through 'information technology' to 'information management'. If you analyse what they mean by the term information, it comes down to 'whatever someone wants to put into a system or run along a bit of copper wire or optical cable'. They avoid being specific, and leave it up to their customers to exploit the infrastructure they provide. Likewise knowledge managers, who use the word knowledge in a variety of ways (though they too are beset by ongoing debate, understandable given the relative immaturity of the discipline).

The answer is 'stuff'. Not in the sense of 'I don't give a . . . ' but in the sense of defining what it is we are putting into our databases, what it is we are managing.

The term is already in widespread use. You will have noticed how conference speakers struggle to define records, documents etc, and then say 'now, so you have all this stuff in your database ...' Stuff is emerging as a shorthand to mean 'whatever things you and those you service decide you need in your system'.

I will admit that 'stuff' perhaps lacks a little elegance, but it is immediately understandable to a wide range of people. It can hardly be accused of being technical jargon. It is certainly no more unusual than the terms used by the scientific community, for example to describe quarks, which now come in a smorgasbord of varieties: up, down, top, bottom, charmed and strange². And I won't even mention that each of these also come in red, green and blue, nor that there exist anti-quarks, in flavours of anti-red, anti-green and anti-blue³.

This terminological irreverence has nevertheless earned Nobel prizes, so I think we are on pretty safe grounds with stuff.

Above all, stuff is cool. It is almost Zen.

Let's play with this idea a bit further. We are putting all this stuff in databases managed by software that is increasingly sophisticated. Most software vendors are working desperately to help our bosses do away with the things too many of us spend too much time on, such as indexing, thesauruses and classification schedules, by coming up with automatic classification and indexing.

Some of this is derived from artificial intelligence, some uses techniques like reverse search analysis. Others use pattern matching or techniques derived from voice recognition. Most benefits from the theoretical work behind the better Web search engines, and pretty well all of it derives directly or indirectly from software developed originally during the cold war era to monitor international phone calls to detect those prone to using naughty words like 'comrade' or 'liberal'⁴.

The point is, this software works, better than we think, and sooner than we think.

So we have databases full of stuff, and software that can do very powerful things in the way of automatic indexing and analysis.

We have also long known of infogenic inheritance - that most stuff (knowledge, documents, records) is based substantially on previously existing information: we all stand on the footnotes of giants⁵.

This idea has been significantly extended by Pratchett, whose work adds multi-dimensionality to infogenic inheritance, and frees it from linear time constraints. He originally suggested that all books, in libraries everywhere, are linked in 'L-space'⁶. He later hypothesised that all appointments, including those that are yet to be made, exist somewhere in appointment phase space (A-space)⁷. However, he has disappointingly failed to realise that appointments are a subset of stuff (S-space), just as books are a quaintly old fashioned subset of documents (D-space), in turn a subset of stuff⁸.

All this is, of course, a variation on Ted Nelson's Xanadu project⁹, now realised to an extent in a highly uncoordinated way in the World Wide Web.

We now have three stuff-related factors to consider:

- Infogenic inheritance (linear)
- S-space, including its subsets L-, D- and R-space (multi-dimensional, time-independent)
- The analytical ability of modern and sometimes still classified software

I suggest that these three factors, logically extended, now give us the potential to predict, with considerable accuracy, and entirely automatically, the content of the majority of the stuff *which does not yet exist*¹⁰.

We can thus eliminate the time we now spend on classification and indexing (not to mention the time we spend on discussing the meaning of document and record), and get on with the serious issues facing our discipline.

At a practical level, for example, will we need financial adviser licenses because of our ability to predict intellectual property income? Will the scope and integrity of our operations be constrained by their potential impact on share prices? At an ethical and moral level, should we attempt to censor potential stuff? What are the implications for business intelligence and freedom of (potential) information? How can corporate counsel and external lawyers extend privilege to potential stuff?

Even more interesting are the theoretical issues. For example, Heisenberg's uncertainty principle¹¹, applied to S-space, means that as we increase the accuracy of our content prediction, we decrease the accuracy of our time prediction, and vice versa. So if we know exactly what some potential stuff will contain, we cannot know exactly when it will exist; if we know when it will exist, we cannot predict accurately what it will contain.

And does stuff in S-space exist only if and when someone uses it? When we open a box, will the stuff inside be information, or knowledge¹²? How can we allow for all this in our disposal schedules? Can they remain concise and useful while extending to the multiplicity of dimensions (some curled up) required by current string theory¹³, or will they become even more prescriptive, enumerative and voluminously unusable than they are already?

Much theoretical and experimental work is needed yet before we have a unified theory of stuff. The challenge lies before us.

¹ Eg a search on 1 November 2000 for 'record or document' on the RECMGMT list archives (www.lists.ufl.edu/archives/recmgmt-1.html) covering only eight months of activity, found 29 hits.

² Davies, Paul. *God and the New Physics*. Penguin, 1984, p.153.

³ Davies, Paul. *Superforce*. Penguin, 1995, p.125.

⁴ I know, I know, but you've got to fit at least one conspiracy theory in somewhere.

⁵ With apologies to Newton: *If I have seen further it is by standing on ye shoulders of giants* (Newton to Hooke, 5 February 1676). Newton was being recursive: the phrase had been around for 500 or so years prior to his use of it.

⁶ Pratchett, T. *Guards, Guards*. Corgi, 1990. An allegorical analysis of the relationship between knowledge, will and reality.

⁷ Pratchett, T. *Jingo*. Corgi, 1998. An allegorical analysis of the impact of parallel phase spaces on politico-military reality.

⁸ There is, for example, no mention of S-space in *The Science of Discworld* (Pratchett T, Stewart, I, and Cohen, J. Ebury Press, 1999), incorrectly classified by most bookshops under Sci-Fi / Fantasy but actually a serious work on the philosophy of science, in a style reminiscent of Gaarder's *Sophie's World* (Phoenix, 1995).

⁹ Nelson, Ted. "A New Home for the Mind?" *Datamation*, March 1989.

¹⁰ Pratchett has discussed this, specifically as far as L-space is concerned (*Lords and Ladies* Corgi, 1992, p.55) but has not generalised the concept to S-space.

¹¹ Greene, Brian. *The Elegant Universe*. Vintage, 2000, p.112.

¹² In some dimensions, the box could instead contain Schrodinger's cat: see Davies, Paul. *God and the New Physics*. p.114.

¹³ Greene *op cit*. Serious dimensionality here folks: if you struggle with four, how about nine, or sixteen, or more?