

## The Hunting of the Snark (Looking for Digital “Series”)

Chris Hurley

Recordkeeping Roundtable, Sydney, 25 October, 2011

*They sought it with thimbles, they sought it with care;  
They pursued it with forks and hope;  
They threatened its life with a railway-share;  
They charmed it with smiles and soap.*

*... the usual archival assumption that the arrangement of records reveals their contextual provenance and thus is the key to their description may no longer be adequate in the Information Age.*

Terry Cook, The concept of the archival *fonds* in the post-custodial era: theory, problems and solutions. *Archivaria* 35 (Spring 1993) p.26

**Query :** Is it possible to serialise the digital archives held in the south-west corner of your repository?

That, of course, is the wrong question. If you leave it until records have arrived in the south-west corner of your repository, it is much, much, too late. It is possible to describe electronic records as Series, **if we think that is worth doing**, but only by re-imagining the Australian (“Series”) System – which I shall sometimes be calling, in what follows, the Maclean/Scott System or the Classic View. If any of my six or so regular readers are here today, they will find much that is familiar in what follows and this is because, in one way or another, I have been answering this question for the last twenty years - but not in this form. The irony here is that the Maclean/Scott System was developed to describe records regardless of custody arrangements, so a drive to serialise arising from the needs of an ERA to deal with transfers of digital materials it has in its possession (or thinks it does) is both sad and funny. Underlying what is said here is an abiding respect for the non-custodial vision which defines the Classic Maclean/Scott View, viz. the capacity to deal descriptively with records before they reach an archives (or, to be perfectly correct, **without it making any material difference** to the description whether they have or not). This has the corollary, comforting in a wired world, that records can be described archivally without an archives being involved. Even if the impetus for today’s discussion is custodial, therefore, the conceptual framework must remain focused on the entire recordkeeping process and its environment not just preservation of its archival remnant. Otherwise, it’s just not the Maclean/Scott System.

Moreover, we don’t serialise records. They don’t become Series because we describe them that way. We describe them that way because that’s what they are. Records come to us structured or not at all. Structuration is not something we do to records after we get our hands on them. What we describe is how records are interrelated within the circumstances out of which they were formed. Structure is not imposed by description, it is already part of those circumstances – though it may be lost if description fails. This interrelationship of documentary objects with each other and with context may be the purposeful result of a managed recordkeeping process (a deliberate organisation or depiction of the documentary artifacts belonging to that process) or, more commonly, an unplanned web of meaning discernible in circumstances but surviving by chance and not in consequence of formal arrangement or description. In either case, if records are to be safely kept, the structure of records must be commemorated - as part of their making or of their keeping or of both. The record is preserved as much by the way it is regarded (methodologically) as how it is made and kept (technically). Peter Scott identified Series as basic units of description in purposeful recordkeeping. In one sense, he and Maclean were substituting one level of description for another to manage the component elements in archival description more effectively so as to achieve the same result that more traditional methods aimed at when depicting structure. This continuity of purpose was emphasized to show how faithful their method was to abiding principles of archival management. In another sense, however, they were profoundly changing the purpose of arrangement and description away from depicting the remains of a recordkeeping process and onto documentation of the process itself – moving description from being outside and posterior to the process to being a part of it.

Some practitioners and theoreticians have moved yet further - away from Peter’s own idea of a Series without departing, I think, from the framework that he and Maclean gave us for handling and describing recordkeeping structures. For those of us, struggling with the application of the System to new materials and in new circumstances, reconceptualisation is inevitable and this involves re-examining practices Peter recommended for dealing with the different circumstances that he had to deal with. As our experience of changing circumstances

*Series are used in archival description (and hence in appraisal) regardless of the methodology employed. They are not particular to the Australian (“Series”) System. It is doubtful that we will use them for digital records in the same way as they were used when describing records found in paper-based recordkeeping systems : a problem, you might suppose, for a theory said to have Series as the object of descriptive endeavour. We are looking for entities that best describe how digital records are structured in relation to each other and in relation to context. The requirement is not technological but methodological.*

broadens, the underlying conceptual framework deepens and strengthens. We can now see that there is nothing special about Series and, for the purpose of this discussion, I won't often use that term to refer to the structure of records because it carries too much baggage. Instead, I will call the object of our search a Snark. A Series is a kind of Snark, but only one kind, and not necessarily the kind we will use to describe the digital records in the south-west corner of your repository. And should it transpire, as you must by now suspect it will, that the Snark we find lurking there turns out to be a *Boojum*, it may even then slip through our fingers as we realize that the Snark itself can be any one of a number of ways of understanding how records are structured (how they exist in relation to each other and to context).

## The custodial view

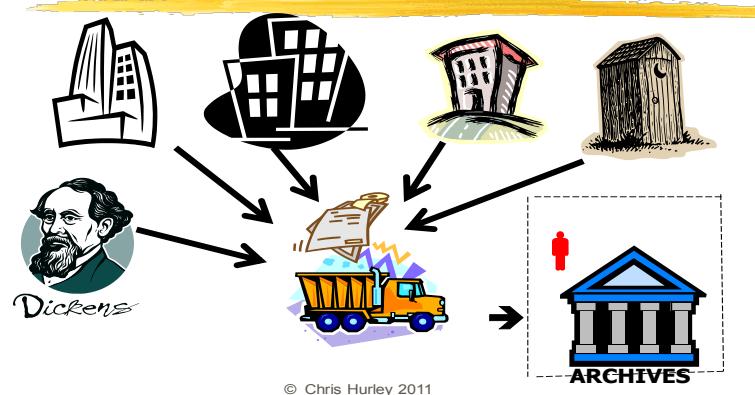


Figure One

Scott proposed that Organisations, Agencies, Series, and Items are the entities that would be used to construct descriptions – along with Functions (a latecomer to the table). These were all, in some form or another, along with Record Groups and *Fonds*, recognizably tools used for description in the traditional custodial view (**Figure One**). Here, the archivist (the figure in red) looks out at the world from which the artifacts have come and describes the holdings from within the boundary established by the repository.

However sophisticated the analysis of context and original order may be, it remains a description of a “collection”, formed, static, dust-encrusted. I do not mean to disparage the work of descriptive archivists in this tradition. In the 1970s - while studying at University College, London - I visited many County Record Offices. I was amazed by the scholarly, detailed, and painstaking work that went into their finding aids. They dealt with family and manorial muniments of great antiquity and of a complexity I had been taught was peculiar to Australian company, trade union, and government administrative arrangements. This disparity between the changeability of Australian context and the stability of British and European circumstances was supposed to account for the development of the System in Australia and resistance to it from elsewhere. But I found that this was not so. If anything, British circumstances were even more unstable and mutable than those I was used to. I later found this to be true of Europe also. Much later, when I found myself dissenting from internationally approved descriptive standards, I was sometimes told by European archivists who nominally implemented those standards – “of course, we do it your way, but we adapt the rules to do it.”

*Each thought he was thinking of nothing but “Snark”  
And the glorious work of the day;  
And each tried to pretend that he did not remark  
That the other was going that way.*

While this was maddening for me, it didn't seem to trouble anyone else much or lead to openly acknowledged dissatisfaction with theory and practice that is manifestly inadequate. All sorts of informal work-arounds were used while the theory remained untouched and to this day one runs the risk of being cast as a heretic for challenging it. All the same, I marveled at how well their materials were described albeit statically and using what I regarded as a primitive and cumbersome methodology. Yet, however well they were done, the perspective, the point of view, was from the inside of a repository looking out. This was a perspective I was accustomed to regard as the wrong one because it is necessarily a partial and incomplete view. Subsequently, it became clear that it also fails to provide a conceptual framework for dealing with digital materials.

Maclean and Scott challenged the custodial view and developed descriptive tools to support their challenge. They did not, as the title of Scott's 1966 article misleadingly implied, abandon larger conceptions such as the Record Group or *Fonds* but made use of them in novel and profitable ways. Descriptions of Series and their Creators (Agencies and Persons) were to be systematically processed to produce views of the larger Entities. The more granular descriptions were vehicles for handling the records (in reality data about the records). The larger Entities were used merely to portray recordkeeping rather than as objects of physical storage. The Group/*Fonds* became virtual elements in a process for handling formations of records and were manifest as emanations of that process of handling, rather than as objects of perception. The process in question is not one of describing records after they have been formed into a Group/*Fonds* but rather the process by which they are formed as records and subsequently kept and managed – in short, the process of recordkeeping itself.

The descriptive boundary now encompassed the entire recordkeeping process, not just a collection of the detritus of that process (**Figure Two**). That was the new perspective from which description took place, even when it was not triggered until after the records had been transferred into an archival repository. It is this aspect, rather than a changed focus onto Series or a descriptive separation of records and context, that distinguishes the two views. The change of focus and the disaggregation were simply the means for giving effect to the new paradigm in the particular circumstances in which Maclean/Scott were working in the 1950s and 1960s, not its primary conceptual feature. Maclean was largely responsible for this conception and Scott was largely responsible for giving us the tools with which to accomplish it. In so doing, some measure of reconceptualization was needed but it remained incomplete. The stipulative definitions given to Agency and Series were an understandable but conceptually misleading attempt to retain fixity in a post-record-group-world. I shall call it the Classic View. The little red figure now looks at things from the perspective of the entire recordkeeping process. It is my gloomy view that, while Scott's continuators have gone on using his tools, many have lost Maclean's vision - reverting to description from the collection perspective – while theoreticians and academics have developed the System in more imaginative and productive ways. Anne Picot says I am too pessimistic and that useful work is still being done by practitioners. I hope she is correct.

## Description (classic view)

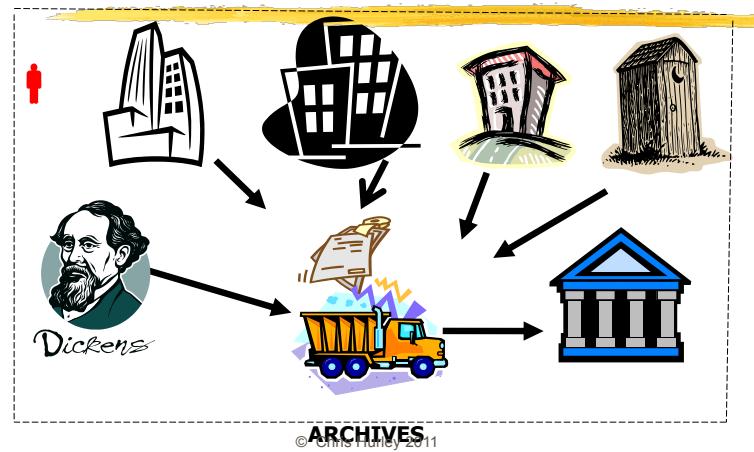


Figure Two

In some ways, both views shared assumptions about the materiality of certain descriptive Entities (notably the Series) that must now be dispelled. The conceptual logic was not followed through. There was no reason why descriptions of Items could not also be systematically processed to produce views of larger Entities (such as

Series) that are vehicles for handling instant records rather than for the arrangement and storage of Items and it turns out that this is what must happen when dealing with digital materials. The Group/Fonds was seen, in the Custodial View, as an actual accumulation of records after completion of the formation process. In the same way, the Series was presented in the Classic View as “the primary level of classification, and the item the secondary level”<sup>1</sup>. We now know, if we follow the axes within Frank Upward’s continuum model, that Series is simply a name we apply to one of many kinds of entity along the recordkeeping axis. Scott replaced the Group with the Series within the “Record Control” area and lifted the Group (as a virtual construct) up into the “Context Control” area. But, while Groups became virtual, Series remained actual. This would now be understood as designing a view of the Group to be a Report from a database while retaining the Series as a Table. We can now see that Series are also virtual and that Items, formulated as Tables, can form themselves into multiple accumulations as reports on Snarks (Series) in a multiplicity of ways in just the same way that Series

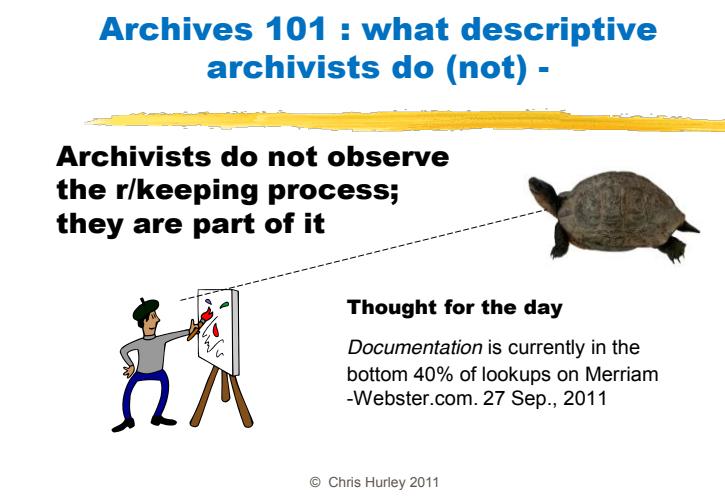
*Series are virtual accumulations of data – to be found in Db Reports not Db Tables. Standardised descriptive practice now allows for Series to relate in a multiplicity of ways to each other and to a multiplicity of contextual frameworks. But the Series itself is still most often regarded as a singularity. When dealing with digital materials, it is necessary to understand that, in exactly the same way, the components within a Series can relate to each other and to a multiplicity of structural frameworks in a multiplicity of ways. Just as a single Fonds or Group could not contain the Series, so we will now find that a single Series cannot contain the digital record.*

(Snarks) can form themselves into multiple accumulations in a multiplicity of ways<sup>2</sup>.

<sup>1</sup> Peter Scott, “The record group concept : a case for abandonment” *American Archivist* (1966)

<sup>2</sup> And data formulated as Tables can, of course, also form itself into multiple accumulations more or less corresponding to Items in a multiplicity of ways.

Sue McKemmish argues that all entities along the recordkeeping axis are virtual. I disputed this at first, arguing that while digital materials are virtual, paper records are real. I could see them, touch them, smell them, and even taste them if I wanted. She would not be moved. "She's being very stubborn", I thought, "I wonder if she could be right". Thinking hard about a paper correspondence file, I tried describing it as an artifact. It's a piece of flat cardboard folded in two, with a fastener in the top left-hand corner, to which papers are attached with writing on the pages. Having said that, you've said about all that is tangible about it. And that would be the most insignificant part of any archivist's description. The things we spend time on – who created the file, what it deals with, what the writing says, how it was used, how it relates to activities of the creator, how that creator fitted into organisational and other contextual frameworks that confer meaning on the contents of the file, and what processes generated the papers it contains – all this is virtual.



**Figure Three**

By persisting in the idea that Series are real, description remains in the Classic View what it was also in the Custodial View – an accurate portrayal of what the eye sees. The descriptive archivist is like an artist committing to canvass a tortoise waddling into view (**Figure Three**). In a recordkeeping world that no longer forms records into Series (imagined as a singular view of the data defined by the circumstances of its creation and use) but instead as multiple views of the data for varied purposes and in manifold combinations, the same objections that Scott raised against using Groups for this purpose apply also to Series. What we now need is a conceptualisation of the recordkeeping Entity that binds both structured and unstructured data into a recordkeeping structuration (interrelation of parts in an organised whole) founded on the circumstances of purpose, formation, and use. This will not be a technological requirement for recordkeeping systems but a methodological requirement for recordkeeping. For that, we need Maclean's vision of our role, one in which we do not simply observe the documentation process, but in which we are instead a part of it.

- Description is not a process applied after records have been formed.
- It is a framework for understanding and managing the recordkeeping process.
- Series are only one kind of descriptive types at our disposal.

For some, all this is peripheral. What does it matter so long as we know how to manage and preserve digital materials? It matters because the Entities used in documenting the structures that make up recordkeeping activity are basic to our work – like atomic structure is to physics. It is simply impossible to undertake recordkeeping without them. They are the language of our endeavor. They are the tools used to make data –

- Time-bound
- Nested
- Context-related.

They provide for an intellectual understanding of content, not for the physical control of the data. The technology of data archiving may be regarded as the digital equivalent of repository management – essential but not sufficient for the making and keeping of records. As of old, a Snark will be supported (but is not defined) by storage arrangements. It is the intellectual controls that sustain it, shaped by the pressure of the business needs that generate it, that makes it a record, lets us appraise it, and facilitates discovery and use.

Some weeks ago, we had a Recordkeeping Round Table event about discovery of records in cyberspace. The lesson I took away was that there has been a discovery paradigm shift :

- We used to construct search pathways based on contextual frameworks. You found the records by you wanted through context.
- On-line, we first find the records you may want, then contextualise them by filtering the results.

I would now like to propose a similar paradigm shift in approach to description :

- We used to organise and preserve records in a form given them by the transactional activity whereby they were created and used.
- On-line, we reassemble data into usable views and invoke description to understand what it means.

In the paper world, the physical arrangement of the documentary objects within a series had great descriptive importance. That was how recordkeepers managed the stuff and we were right to pay attention to it as well as to context and function. In 2002, Canadian archivist, Jim Suderman, described<sup>3</sup> the early stages of the Archives of Ontario's programme for the transfer of electronic records. He looked at defining an electronic series in those circumstances. Noting that considerable "transformation" of physical aspects unavoidably occurs during transfer, he concluded that a continuity between record-making and record-keeping is unavoidable and cited Tom Nesmith with approval :

*... it may well be that the record's existence and characteristics have been shaped far more powerfully by what has happened to it during the custodial and transmissive stages of its life, as well as the archiving process ... But that is not something pre-ordained, but rather an outcome determined by the archivist's interpretive, contextualizing, research into provenance.*

Tom Nesmith, "Still fuzzy, but more accurate : some thoughts on the "ghosts" of archival theory" *Archivaria* 47 (Spring 1999), p.146

Data is simply not managed in a way that presents us with digital objects whose physical structure is all that is meaningful. The content of records is held in tables and summoned up in response to an action demand by a user. Each of these uses will involve the formation of a view of data that is structured in part according to the design of the system (its logic) but also in accordance with the requested view and the circumstances in which it is requested. The data is understood by the purpose for which it was captured and for which it was reconstituted by the use that was made of it. The record trace of an instant event or circumstance was formed not only by the structure of the data engaged but also by -

- Transaction type and the process(es) of which it was part
- The form (diplomatics) conferred by use
- Identification of user, customer/ correspondent, and other involved party(ies)
- Linked records
- Transaction logs and audit trails
- Rules governing data use, access, and security

and, no doubt, much else besides. Such records materialise at the user interface and it is there we must document them<sup>4</sup> but they then disappear again into the data structures in which their substance resides. It is not these data structures alone that we must document; we must also document the structuration of purpose, creation and use.

## How we describe (classic view)

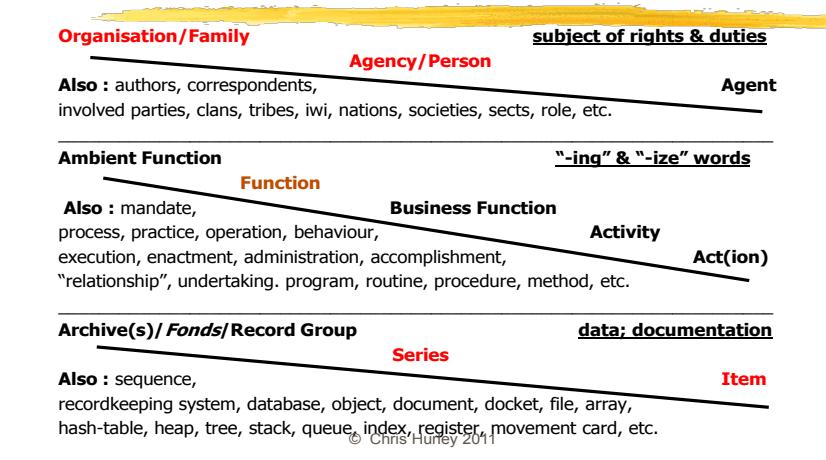


Figure Four

Even before records came to be managed on a data base, Suderman found that old verities were shaken. Describing the transition, he noted :

*With the advent of a computerized tracking system in 1989, the physical arrangement of the records (still paper) changed so as to effectively render items inaccessible (filed numerically) without the tracking system, even though the physical relationship of the incoming and outgoing was preserved. The tracking data was acquired and preserved in part as a finding aid ... (p. 35)*

With the introduction of fully computerised systems, file integrity began to be lost and, a decade after the introduction of the tracking system, much greater reliance was being placed on naming conventions, document profiles, and (case) tracking data. The residual evidence of a recordkeeping process became less significant in relation to residual evidence of business process and use – all of these elements have traditionally gone into

<sup>3</sup> Jim Suderman, "Defining electronic series : a study" *Archivaria* 53 (Spring 2002), pp.31-47.

<sup>4</sup> Because use involves knowledge that is additional to data content.

description but their relative importance and point of origin change when the records are digital. The mix of descriptive elements found in the Classic View is set out in **Figure Four**.

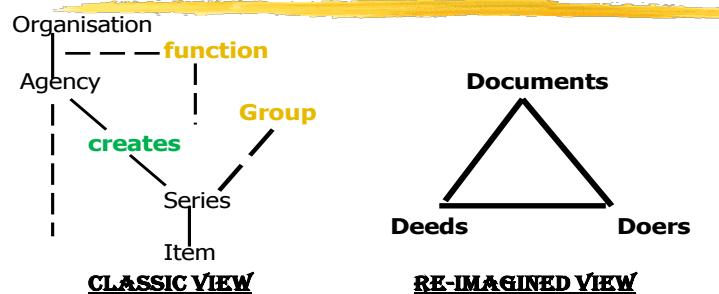
Here we see the elements Scott proposed highlighted in red (Organisation/Family, Agency/Person, Series, and Items). Function is shown in brown because its importance was understood but it was not formally incorporated into the Classic View. But we now understand that these Entities belong to larger families that Frank Upward has identified as populating axes in the Continuum Model – Identity (Organisation/Agency), Transactionality (Function), Recordkeeping (Series, Item). Using this analysis, we can redefine our search. We are no longer looking for Series in the south-west corner of our digital repository, or even for Snarks. We are looking for Recordkeeping Entities – along with the descriptive tools needed to deal with Identity and Transactionality too, of course. This is just as well, because –

- You will look in vain inside your ERA for a Series (defined under the Classic View as a singular meaningful arrangement for the organisation of data reflecting creation and use), and
- In the corporate world of ICT (information and communication technology), Agencies (defined Classically as independent recordkeeping systems) will also be fairly thin on the ground.

In the Classic View, a single Series comprised one or more items – each of which had its identity from being a component part of that Series and of no other - and it was the Series that provided an object that could be linked contextually and functionally to other entity-types (as well as to other Series in a recordkeeping system). In a data base, this “recordness” can be found in a framework of linkages that is not co-terminous within a single Series but is to be found (as well, if not instead) in linkages from each record with other records and with context. Suderman notes some of the implications if “in the electronic environment it is more likely that a record may belong in more than one series at the same time” (p.42) – what we would now call simultaneous multiple provenance.

Suderman has also applied this reasoning to appraisal<sup>5</sup>. If I understand him correctly, he argues that appraisal involves both a determination of which record characteristics survive (itself an act of appraisal) and an evaluation of those characteristics. But the digital record (whatever its technological peculiarities) remains methodologically what records were in the paper world – viz. instances of a Series, part of a collectivity of some kind, i.e. of a Snark. The characteristics of the Series (or Snark) are at least as important (if not more so) in determining an appraisal outcome. Which instances to keep within the appraisal outcome determined upon may require assessment of the characteristics of the instant record, but that is another question. Concede that the same data may manifest itself as part of more than one instant record and the same instant record may manifest itself as part of more than one Series (or Snark) and it follows that the same instant record may have two or more different disposal outcomes depending upon which view of it is taken. Appraisal, on this view of the matter, must involve looking the characteristics of the Series (or Snark).

## New typology of description



© Chris Hurley 2011

Figure Five

What we can expect to find is something that supports recordkeeping as we have always understood it to be – something that can be described (or documented) according to a re-imagined understanding of the Classic View. A new typology of description (**Figure Five**) is needed to replace that of the Classic View. In the absence of adequate descriptive solutions, relatively too much attention has been paid (I believe) to standards and metadata. I do not under-value their inherent importance, but rather the relative importance they have acquired. The re-imagined descriptive view deals with three entity-types (set out below) :

- **Documents (recordkeeping)** : an object or accumulation of documentary objects containing or conveying meaning and satisfying the requirements for recordkeeping;

<sup>5</sup> Jim Suderman, “Context, Structure, and content : new criteria for appraising electronic records” available at [mybestdocs.com](http://mybestdocs.com).

- **Deeds (transactionality)** : a documented event, activity, or action or the invocation of an action (e.g. the process or step that is being documented in business or in recordkeeping);
- **Doers (identity)** : corporate/natural persons or originators/initiators of action who undertake, authorise or are otherwise involved in documented activity – including roles they adopt in that regard.

This new typology reflects a revised understanding of what the System means, beyond the formulations of those who have developed and applied it heretofore. This process of revision makes it difficult for those new to it to disentangle what the System now means from how it has been articulated in earlier expositions (based on an unrevised understanding of its purpose) – expositions now being overtaken by fresh thinking, like learning Newtonian physics in a post-Einstein world. The limitations of the Classic View were always there but could not be clearly seen from the perspective within which it evolved. Re-imagination is needed to remove limitations that impede our progress in defining the descriptive tools we need to operate with digital materials, the structure of which involves manifold views of data in manifold combinations. Maclean and Scott dealt with what they saw before them –

- Organisations and Agencies are not conceptual, they belong to the world Maclean/Scott set out to describe, but they are merely instances of the **Doer** type;
- Series and Items are not conceptual, they are **Documents**;
- Functions and Activities are **Deeds**.

If we want to leave their world and operate in the one that has since taken its place, we must conceptualise where they stipulated. We don't disavow their work, but we re-build it.

Our quest is not, therefore, for a digital equivalent to the Series as it was used to describe records in the Classic View. Trying to organise digital materials into Series of that sort is fairly pointless; just as applying descriptive techniques developed to handle paper records (irrespective of the methodology) is a bit silly<sup>6</sup>. Our quest is for an Entity (or Entities) that is (or are), at one and the same time, both wholly new and wholly traditional. Use of the data will enfold it within nested structures along axes formed around the actual –

- activities, processes, and functions (Deeds) in which it is involved,
- persons, families, groups, and corporations (Doers) involved, and/or
- other data in singular or aggregated form (Documents).

These structures will be formed as describable entities, but whether we will want to term any accumulation of data a "series" or any other entity stipulated in the paper world is open to question. We will want some kind of Document or Documents, but whether Series, Snarks, Boojums, or something else remains to be seen.

Scott's Entities were defined stipulatively (setting out observable characteristics of instances of the entity-type) to pin them down within the CRS application he advocated. He didn't provide conceptual definition (hypothetical constructs involving the application to particular cases of a theory of description) at any rate, not definitions that we can apply to digital materials and, it must therefore be inferred, not based on a comprehensive theory. He wasn't wrong but, like Darwin trying to explain evolution before Mendel's work on genetics (or Newton before Einstein), he didn't have all the pieces. Scott's formulations derive from a world of paper-based recordkeeping in a registry-based environment. He dreamed of testing his ideas by applying them to examples of recordkeeping that pre-dated the eighteenth century when systematic paper-based recordkeeping began in Australia (thereby broadening the conceptual base). Circumstances now require that it be broadened stipulatively within the framework of digital recordkeeping and, in so doing, we must re-imagine it conceptually so that it can support archival description irrespective of the technological application.

*The Beaver brought paper, portfolio, pens,  
And ink in unfailing supplies:  
While strange creepy creatures came out of their dens,  
And watched them with wondering eyes.*

We can't just go on applying the old formulations, though we do not invalidate them, and we must not simply replace them with new formulations particular to the digital environment. We must re-imagine them in ways that will comprehend both and would have astonished our forebears when they first developed the descriptive theory we now struggle to apply in a larger framework.

In a re-imagined world, we've only just begun. We should stop stipulating (until concepts are clear) –

- Series and Items are just labels we attach to instances of the **Document** type.
- We should think of them as **big Documents** swallowing **little Documents** (nesting).

We must ask questions, e.g. –

- Can a **Deed** swallow a **Doer**? Some relationships must cross the barrier between types. Are there any which do not?
- SPIRT concluded that recordkeeping itself is a sub-species of **Deed**. Is there any justification for this distinction, or is a recordkeeping Deed just a variant like any other? Are sub-types even allowed? If so, what is their significance?
- Can an entity belong to two types? Hybrids: **half-Document** and **half-Doer**?

---

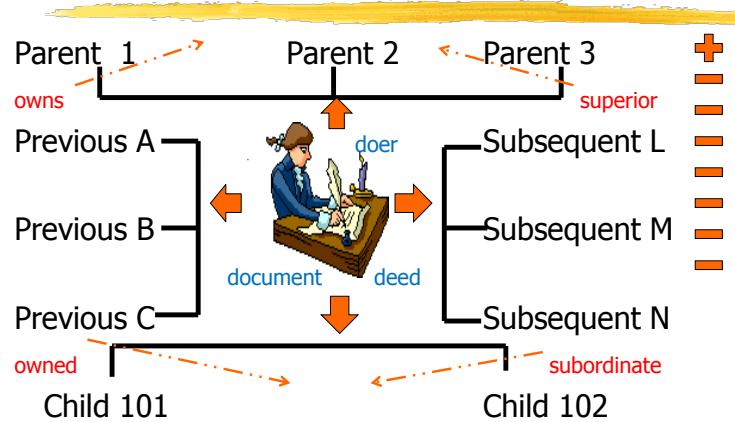
<sup>6</sup> We need a descriptive methodology that encompasses varying techniques and, without relitigating my objections to current international standards, let me say again here that we do not yet have one.

- What rules, procedures, routines are needed to make this work? Is “role” a **Doer** or a **Deed**?

If the conceptualisation of entities is primitive, the understanding of relationships is dismal. In the Classic View, there are three kinds of Relationship :

- **Generation (create, control)** : make, embrace, author, own, comprehend, include, contain, perform, authorise, develop, apply, etc.
- **Succession (precede, succeed)** : post-date, pre-date, ante-date, pre-date, supervene, displace, replace, supersede, supplant, etc.
- **Inter-action (engagement)** : participate, involvement, interest, commerce, dealings, intercourse, fertilisation, etc.

## How entity-types are used



© Chris Hurley 2011

**Figure Six**

We still have much to do in this area – we must break these categories down<sup>7</sup> and add to their number based on better observation of what happens in the recordkeeping world. We need to clarify the meanings of the relationships we use. Does a father generate a son, for example, or does a son succeed to the father?

*The method employed I would gladly explain,  
While I have it so clear in my head,  
If I had but the time and you had but the brain –  
But much yet remains to be said.*

In **Figure Six**, the re-imagined elements of description come together in ways that remain faithful to the Maclean/Scott paradigm. Here we see the target resource – a construct made up of instances of the three entity-types - at the epicentre of a network of relationships with other entities that, in combination, give the structuration that is the essence of recordkeeping. This is a view that exists in time – you can go forward and back. The arrows also allow you to navigate to higher and lower levels of aggregation. But the most significant navigation tool is on the right-hand side. Like an inter-active map on the Internet, this allows you to scale up or down, moving in and out as well as up and down the map. This is the scalability axis (see below). Descriptive and relationship rules will be formulated, in the first instance, by type rather than by instance, enabling recordkeeping to be supported by appropriate design and adequate metadata.

We must now clarify the kind of record-keeping processes within which our re-imagined descriptive entities are to operate. Digital archiving is simply a sub-set of the process of migration that is a common feature of all data management. No new problems arise at the point of transfer into a digital archive that have not already arisen when the data was first migrated from any legacy system in which it has been supported. But not all movement of data counts as migration. It is useful to identify three kinds (**Figure Seven**).

Simple migration needs to be differentiated from Transformation and Mortification. In the latter, recordkeeping has stopped and dead records are preserved in a digital mortuary. Many ERAs will mortify the data. Ideally, a digital archive would be a vehicle for migration, but this could be expensive and impractical while digital recordkeeping remains so primitive. It will remain so until we can recreate in the digital world the kind of methods for preserving the recordkeeping system, through description *inter alia*, that Maclean/Scott provided for us in the Classic View. That is what we should be trying to do here today. Transformation involves copying the data for re-use in alternative ways. That too may describe what is happening in some digital archives. I like to find old-world examples of new concepts and I would suggest that the mediaeval cartularies are examples of transformed data. Historians and diplomatics sometimes dismiss cartularies as imperfect copies of the originals but a correct

<sup>7</sup> Or, possibly, combine them. I have argued that succession and generation are the same thing from different perspectives, but “engagement” certainly hides a multitude of connection-types.

view of them recognises that they had their own authentic recordkeeping purpose that was simply different from the originals from which the abbreviated summaries were extracted.

## Migration / Transformation

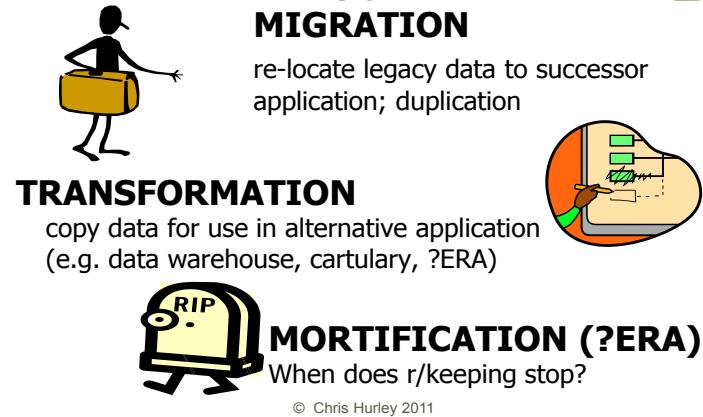


Figure Seven

A re-imagining of the Classic View (Figure Eight) retains the boundary around what is going on, but what happens inside that boundary has now altered. The Archives – an ERA, off-line digital storage device, data warehouse, or body of legacy data, whatever form it takes – is now on the left. It is simply part of the mix for dealing with digital materials rather than an ultimate destination. All the arrows are now two-way indicating easy movement back and forth from one state to another – showing that storage boundaries have a certain irrelevance, descriptively speaking. A new component has been added on the right through a link to the world-wide-web where data is accessed and maybe even stored in the cloud – certainly a place where our data may be linked with someone else's. In the middle, in a turmoil of data churn, the processes we have identified are still going on, driving the continuing process of making and keeping records – records formation, migration, transformation, and (yes) even mortification.

## What we do (re-imagined)

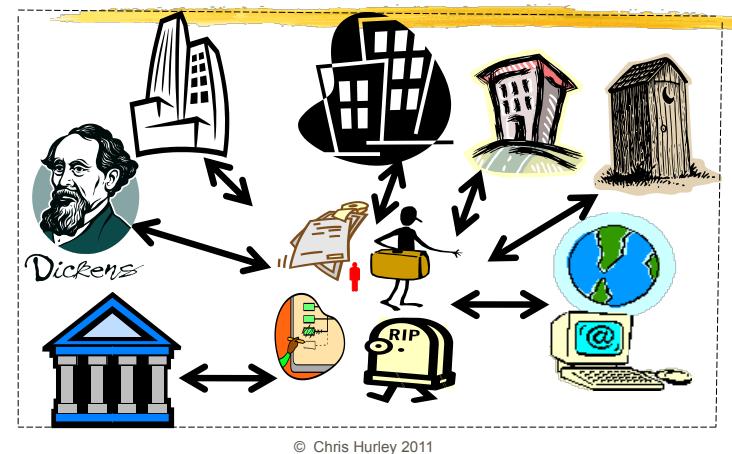


Figure Eight

The Classic View of Context is not immune from re-imagination either. Scott gave us the notion of multiple-provenance, by which is meant the capacity to view the successive stages in recordkeeping within a single descriptive framework. It is now necessary to deal with the concept of parallel provenance or multiplicity of view. Parallel provenance describes the idea that context (at all levels) can be found in places outside the one that has been documented.

- **Multiple provenance** : 2+ "creators" in different time periods in the same context.
- **Simultaneous multiple provenance** : 2+ "creators" in the same time period and the same context.
- **Parallel provenance** : 2+ "creators" in the same time period but in different contexts.

This is, in one sense, the Peter Scott argument re-imagined. Peter said it would be better to ascribe records that have passed through successive phases of generation to all successive Doers (or Deeds) involved and not

simply to one Doer (or Deed). Showing successive creators in this way is multiple provenance. But in Peter's writings there was still only one provenance at any given time. This derives from traditional archival ideas of the uniqueness of provenance which emphasised the singularity of origination. We can now see that traditional archival ideas of origination are themselves flawed and that context may be conferred on records in manifold ways not simply the ones implicit in archival theories about "creation". Moreover, in computerised environments records are enfolded into a multiplicity of virtual contextual frameworks that are not constrained by predetermined organisational or functional arrangements. The multiplicity that Scott identified was successive, not simultaneous. But he and I did experiment with simultaneous multiple provenance in cases I have described in the literature, starting with the papers of Ministers of State. Since then I have come to see that those early stirrings of multiplicity need to go much further than he or I imagined back in the 1970s.

The meaning of a Document (and hence its appraised value) comes from relationships it has with other entities – within a landscape that has features and particularity. The instant record cannot stand alone. All provenance statements are parallel and hence flawed, until contextualised, and all descriptive views are partial until they are set within a framework that establishes a point of view. Such points of view must be exclusive in order to confer meaning but they must also be diverse in order to be accurate and coordinated in order to be complete. The doctrine of impartiality governs how we construct the descriptive view; but all description is, by its nature, a partial representation because it must place what is being described into a landscape that supports one understanding to the exclusion of others. A sent email differs from the same email held by the recipient unless a single copy is displayed within the same descriptive boundary showing at least two different views of it nested inside the same system that allows for at least two different sequences. There will always be parallel provenance because there is no single cosmic descriptive view – although I once suggested that if such a framework existed it might be registered as BIG 1. The BIG 1 was a dream of Peter Scott's when he handed me two handbooks to populate way back in 1971 – that would give common context to all government recordkeeping in Australasia, government and private, since 1788. These Handbooks, alas, remained largely unpopulated although I developed the idea in my 1986 proposals (never acted on) for establishing a single, common contextual framework for Australia's archives programmes as a mechanism for standardisation.

*He had brought a large map representing the sea, Without the least vestige of land;  
And the crew were much pleased when they found it to be A map they could all understand.*

\*\*\*\*\*

*"Other maps are such shapes, with their islands and capes! But we've got our brave captain to thank;  
(So the crew would protest) that he's brought us the best – A perfect and absolute blank."*

The point of reference in special description is called a datum. When the site of the Sydney Opera House was ceremonially launched, the big wigs unveiled a stone that was to serve as the datum from which all engineering measurements would be made. In mid-project, a new architectural firm took over and their first concern was to find the location of this datum. The builders laughed and showed them a shed into which the stone had been moved for safekeeping. This was when the building was already half constructed! The architects were horrified

until they were shown a rivet blasted into the concrete apron that the builders had put there to use as a datum while they got on with the work. My favourite datum is a sandstone obelisk in Macquarie Park – just south of Circular Quay in Sydney. Someone cares for it enough to keep the black lettering freshly painted. It says that Governor Lachlan Macquarie erected this obelisk in 1816 to mark the point from which all road distances in New South Wales would be measured. One use for a datum in archival description is recontextualisation (the alternative path to standardisation, the road not taken) but we lack the datum that Peter's Handbooks, if realised, or my 1986 proposals, if acted upon, would have afforded. In a minor key every descriptive programme must establish a datum of its own, like the builder's rivet in the concrete apron, in order for us to get on with our work.

The Snark is not, as we once supposed the Series to be, a "primary" mechanism for organising "items" or anything else. Nor is it fixed in such a way that we can build a system on predictable placement of sub-entities (instances of each of the three types) within different applications. The reasons for having more than one

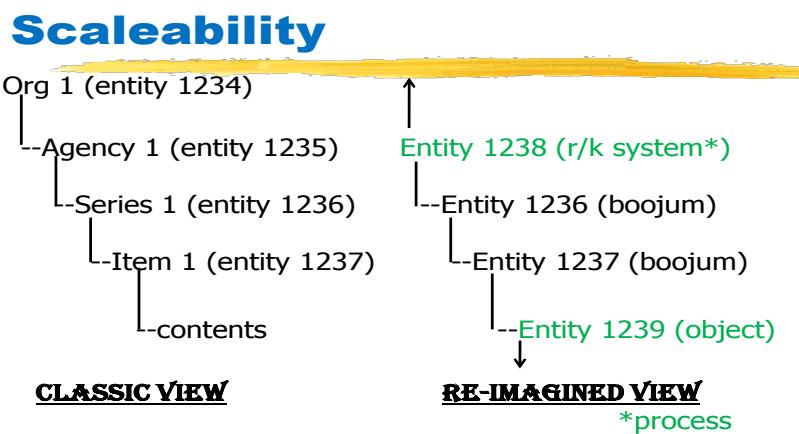
instance per type will be particular to the application concerned – purely local – customisation, if you like, of the out-of-the-box product that the System delivers to us all. We will not have Series, or Snarks, or Items, or Boojums because our descriptive theory dictates that we do. We will have instances of the Document type and those entities must behave in ways appropriate to the making and keeping of records. That is all that need now be said.

The three types may be regarded properly as sub-types because there will be some attributes and relationship rules that apply regardless of which type is involved. I call this super-type the URO (Universal Recordkeeping Object) having been persuaded that its former name, the HERO (Hurley's Enduring Recordkeeping Object), was immodest. The URO will have the mandatory metadata that must be present for all sub-types, cascading down

*The concept of Series in the Australian System was developed stipulatively within the framework of paper-bound recordkeeping systems. That does not provide an adequate a basis for analysing structure within a digital environment and earlier formulations of the System may now actually impede understanding unless a conceptual model is found. Existing descriptive concepts can be subsumed into three entity types : Document, Deed, and Doer. Entities needed in particular applications belong to one or other of these three. More work is needed on standardised methodologies for dealing with these entity- types and especially on the relationship rules that link them in recordkeeping processes.*

through the data structures of every child of the URO. (I think “Child of the URO” has a surreal ring to it, don’t you?) Each sub-type (the Deed, the Document, and the Doer) will, similarly, confer, upon instances of the sub-type, additional metadata particular to that sub-type which will, in turn, be cascaded down to sub-sub-entities (the instances) along with the URO metadata inherited through the sub-types.

In terms of universal descriptive theory, these sub-sub-entities will not be sub-sub-types but instances of the sub-type. Reasons for having more than one instance of the same sub-type include having to deal with significant variation in characteristics (metadata) between instances of the same sub-type and the requirement that different instances behave in different ways in your application (according to sub-rules, desirably, of the common relationship rules for that type). It is also possible to deal effectively with super-types (an example of which is the recordkeeping system composed of two or more series). In most archives that I have seen, the description of Items is sufficiently different from descriptions of Series and their behaviour is sufficiently different for it to make sense to handle them differently – at least when dealing with paper-based records. But the emphasis traditionally given to these differences has also obscured the similarities between descriptive and behavioural characteristics of – for instance – Items and Series. It may well be that, in the digital world, these differences are seen as less significant and you might then have Series and Items to deal with paper records and a different kind of Snark, different instances (one or more), to deal with digital records – unless, of course, they turn out to be *Boojums*. So long as they all fit within the Document sub-type and simply operate as diverse instances according to rules that are universal for the sub-type along with additional local rules for the instances utilised, I see no problem with this. It will not be necessary for two descriptions to agree about what instances to use, provided a common datum is employed, only in the stipulation of the entity-type and of the number levels used.



© Chris Hurley 2011

**Figure Nine**

We should now think of recordkeeping entities as instances nested within a sub-type of the URO, within which the Document, Deed, and Doer sub-types are themselves nested. Despite his insistence on the primacy of the Series as a focus of description, Peter Scott was well on the way to this insight when I was working with him. I have written up several times how he tried to organise Series into Recordkeeping Systems, appearing as super-entities on the Inventories of Series. This was to solve the problem that related Series often ended up pages apart in a chronologically arranged list or inventory. Peter tried using a kind of double-entry (sometimes triple- or quadruple- entry) to group related series on the Inventory while preserving the over-all chronological arrangement as well by describing the Series a second time in their correct place in the chronological order – to give what we would now call duplicate or alternative views of the same data. Each view represents, of course, a different contextualisation of the object of description and this would now be recognised as a variant of simultaneous multiple provenance. Neither the entity nor its description changes but its meaning and significance do as the perspective shifts. In Classic Theory, physical evidence of original order at the Group level is un-picked in descriptions to show the discernible stages of organisation through which stuff has passed. This recreates, on sound principles of observation, what the creator actually did. It does not, in Peter’s theory, impose a view on the records but restores it. In dealing with digital materials, we will always be handling descriptive “views” of the data that may not represent ones that were ever actually formed. So far from representing a “problem” for Maclean/Scott, I cannot understand how this can be dealt with using any other method. How the conceptual model relates to evidence-in-need-of-preservation is, I grant you, still unclear. But this is the result of not having freed ourselves of stipulated definitions of Series etc. and not having explored with sufficient rigour the descriptive objects we will need for recordkeeping purposes.

In the Classic View, entities are identified by type and code (**Figure Nine**). Series 1 is different to Item 1 and it is sufficient to say what kind of entity it is to avoid any possible confusion occasioned by the fact that they both bear the same identifier – “1”. But they will almost certainly be assigned different identifiers from a single sequence for the purposes of entity management within whatever application is used to support the descriptions – Series 1 being entity 1236 while Item 1 is entity 1237. This is how we must rebadge our descriptive entities (at least to the level of the sub-type) once we have annihilated the special conceptual distinction that instances of the sub-type have traditionally been given in the Classic View. Any distinction between a Series and an Item, for example, will be a local variation of the universal rule and they will each be viewed as variant instances of Snark, coded in the

same sequence. It is in implementation only that one decides whether or not to have one or more instances of each sub-type. The choice will be determined by the needs and preferences of the owner of the application and presented to the world as instances of the sub-type that should be universally recognisable in any recordkeeping system. All Document entities will relate in predictable ways to other Document entities and in equally predictable ways to the other two entity sub-types. Those rules remain constant regardless of how many instances of each sub-type are chosen. The rules as to how Documents relate to each other will apply to all instances of Document and all instances you have of that type will relate to each other in that way irrespective of whether they are a Series, an Item, or a Snark. And they will relate in equally predictable ways to all instances of the Deed and Doer sub-types.

Much of my own writing has been at the macro-level, but from time to time I use email to demonstrate more granular digital application of the theory at the micro- level (e.g. "Relationships in Records"). This is useful because emails are accessible (not too different from paper-bound stuff) and pose issues we are used to handling (viz. records of communication). As a simple demonstration, consider emails passing between Kate Cumming and me in the development of this session. Each of us has a separate in-box and an out-box - two sequences each of digital objects (four sequences in all). These look like Series because they are the organisation of documentary objects for use – a classic definition. But they are not. Kate and I will, each of us (regardless of the functionality of our application), view them from as one sequence (frequently incorporating the last message, or a chain of previous messages, as a "tail" to the next email we send). It is the logical view of the email chain that is the record of transactional Deeds between two Doers. Unlike the Series, which we found in the physical arrangement of the documents by a system, we find the Snark in the logical view of the data defined by the actions of the parties involved. These same emails can be sorted into numerous other views, some of which may involve other Doers (Barbara, Cassie, and Anne, for example) and many Deeds (inviting the speakers, arranging the venue, finalising the program). Any logical possibility for organising the digital objects (either functionally possible or actually realised) becomes a series.

Over the years, I have chided developers for not imposing more registration control over emails so we could manage them better. In fact they do, but clients seldom want to make use of it and so it invisible to most users. Within the control of every system administrator is the power to reconfigure email so that there is only one copy of any email in storage - to which as many links as users decide all point but all opening onto the same digital object when it is needed. Lotus Notes (many years ago) actually had an email application that worked just like this. The functionality specifications for such a system would, in fact, be the Register of Snarks. Each Snark would be a sub-set of the global accumulation. Subsidiary Snarks would be created by Doers as they conducted their affairs. The e-mail Snark would nest within a Super-Snark that encompassed all the data, not just emails, within the corporate/domain.

Now imagine that, instead of such a boundary being established around the activity of Doers tied together within a corporate entity, the boundary exists virtually as a shared workspace (probably one of dozens to which each of us will belong) both inside and outside the boundaries of the entity that supports the application we are using. Views based on organisation of the documentary objects will be far less interesting than those based on the activity within the shared workspace – mainly, I imagine, around descriptions of the Deeds undertaken and the Doers involved. I have no doubt that series theory can deal with this but it remains a matter of faith and imagination until we work it out in application. It needs to work for both digital objects and structured data (data-bases). One possibility (uncharacteristically optimistic for me) is that structured data will be easier to deal with because there will be less of the physical getting in the way.

When James Boswell sends an email to Dr Johnson, inviting him to dinner at Stone's Chop House, we have a single digital object. In its primitive state, it belongs in a "sent" box and a "received" box. But in an application of any sophistication, both will want to "see" and "use" the object on a variety of ways – to chronicle their correspondence, to manage invitations between each other and with other parties, to diarise appointments, and to make dinner reservations at Stone's and other eateries. Each of these views entails linkages between the documentary object and the Doers (the correspondents, Boswell and Johnson, and an involved party, Stone's) and the Deeds (correspondence, invitation, diarisation, reservation – to name only those documented as such and saying nothing of events that can be inferred, such as gustation). These are all possible, and many will be realised, views of the data. Which ones we choose to document and/or preserve as Snarks is another story.

Lest there be any confusion, let me be clear that this does not involve the archivist imposing a view onto the data but rather identifying and documenting a structure that is already there in the circumstances of its creation and use. This is not to say that the archivist may not also be involved in opening up the record to new and additional varied uses that the data supports (in effect, becoming a participant in the making, as well as the keeping, of the record – in an archival after-life or, if you prefer, in data re-use). In the pre-digital world, we were accustomed to

*The cycles through which digital records pass are no longer linear. Were they ever? The conduit between the source system and the archives is not one-way. Data re-use invalidates many custodial and life-cycle assumptions. Stability of the descriptive object is also affected by data management processes – e.g. migration and transformation. The value of archival description is the perspective it provides from outside the making and keeping process but the only feasible posture for the descriptive archivist is the one proposed by Maclean and Scott – not describing records after they have been formed into a Group or Fonds but rather within the process by which they are formed as records and subsequently kept and managed : in short, describing the process of recordkeeping itself.*

finding a singularity and preserving that as untouched evidence of past events. In the digital world, we do not impose a multiplicity, but that is what we are going to find and a multiplication of views (not the duplication of data) may well continue long after the point where in the pre-digital world the data would have been consigned immutably to archival storage and preservation. This multiplication is the result of data-sharing and the diversity of uses to which data can be put (re-use). Under classic theory, an immutable singularity was delivered to us by transfer for preservation and explanation and we selected how much of it to preserve. In a re-imagined view the same thing happens but it arrives in the form of possibilities from among which we must now select ones that both document and preserve (giving full weight to the actual uses to which the data has been put) and support data re-use from which new additional structuration may derive that was not there at the outset.

In the Custodial and Classic Views, much attention was given to developing agreed definitions of the descriptive entities and to applying them uniformly. In fact, an examination of the actual work performed by archivists operating under those preconceptions reveals that there was very little uniformity. In the re-imagined View, common descriptive rules will be applied (so far as possible) at the level of the conceptual entity-type, leaving much room for individual variation in the application to particular instances. The Document I call a *series* might be called an *item* by you. You may distinguish between items and series while I may treat them both as variants of the same kind of entity. It will not matter because both descriptions will conform to the rules for Documents. In the Classic View, descriptions of sequences appeared on Inventories such that it was difficult to determine whether you were looking at a list of documents, files, cards, maps, volumes, or the "series" components of a recordkeeping system. The description was regular and systematic to support its organisation and display of the underlying coherence within the inter-action of Documents, Deeds, and Doers at all levels of aggregation. We

now understand that standardising the view we take of relationships between entity-types within that interaction is more significant than precision in depicting the entities that participate according to an agreed formula. Provided entity-types are reasonably stable, it is the relationships, rather than the entity-definition, that is critical to the archival view of data.

*Data can be viewed in a multiplicity of ways in relation to other data and to context. The same aggregation of data (the digital equivalent of a document) can be read, compounded, and used in a variety of ways within the integrated or related business processes - necessitating fresh understanding of how recordkeeping takes place in a digital environment.*

*Simultaneous multiple provenance was developed to deal with inadequacies in the attribution of provenance over Series. Applying the methodology at the level of data management, it is now possible to hypothesize a descriptive methodology (simultaneous multiple provenance on steroids) that accommodates a multiplicity of views of the same data in as many sequences as may be required. The recordkeeping boundary will seldom coincide with the domain of a particular application so techniques based on ideas of parallel provenance will also be apposite.*

The development of an adequate response to the challenge of electronic recordkeeping will necessarily (in my view) involve admitting simultaneous and parallel provenance to our descriptive lexicon - because "creation" in e/systems invariably involves a multiplicity of formative processes rather than a singularity. It is inconceivable that this multiplicity can be made to conform to a standardised notion of entity definition outside of a single application and most e/records will have to survive across (rather than within) a single application. Since we document the formative process and do not impose our view on it, we must be able to deal with diversity. To do that, we must lose our fear of describing a Snark that may turn out to be a *Boojum*. "Is it an *item* or a *series*?" will turn out to be a less significant question than "If it's a *Document*, how is it related to instances of the same or other entity-types?"

Potential views of the data (not just realised ones) are possible Snarks and we don't yet have a rationale that I am aware of to determine which views

(potential or realised) we should preserve and document. Those who say lower storage costs makes appraisal irrelevant don't yet understand, I think, how selective we are going to have to be. When emails are joined to other kinds of digital object, sitting on other applications within the same domain, for example, this model becomes very complex - rich, rewarding, and very powerful for recordkeeping - but very challenging to manage. As defined in the Classic View, the Series was a single, authentic, formed view of documentary materials delivered to us as evidence. In a re-imagined world, a Snark may be any of the views taken (or capable of being taken) of the data alongside some log or audit of uses actually made. There won't be just one view, and therefore there will not just be one Series, there will be manifold views of the same data in the same context. It will be simultaneous multiple provenance on steroids:

- Which views of the data will qualify as Snarks?
  - ▶ all or only some of the logically possible views;
  - ▶ all or only some of the functionally possible views;
  - ▶ all or only some executed/implemented views;
  - ▶ all or only some utilised/ accessed views?
  - ▶ and on what basis shall a selection be made?
- How are reconstituted records to be "captured"?
  - ▶ must the data, use logs, and system functionality be preserved;
  - ▶ or, can we formulate records based on use - relying on metadata (supported by adequate description) to preserve them as system-independent entities?
  - ▶ whatever happened to the metadata encapsulated object?

- Relationships must be identified and documented
  - ▶ recordkeeping requires at least one link between entities;
  - ▶ yet a taxonomy of recordkeeping relationships is still lacking.
- An appropriate datum must be established that does not limit the variety of views and yet anchors the record to its correct context -
  - ▶ some description, therefore, will always be required, but
  - ▶ how much can be embedded through good system design/functionality, or
  - ▶ adequate metadata?
- What metadata is mandatory? At the descriptive level for the URO – cascading then to all sub-entities, I propose only three -
  - ▶ identity, because all records are unique;
  - ▶ dates, because all records are timebound;
  - ▶ relationships, because no record stands alone.
- A solution must be adopted for preserving the record so that authenticity is not compromised -
  - ▶ migrating (or mortifying) data as archives have traditionally done, or
  - ▶ transforming it (as some ERAs appear to do).

Once we deal more effectively with these matters (and others besides), we may be close to the end of our search; unless the Snark turns out to be a *Boojum* after all. Justice Potter Stewart famously declared that he could not define pornography but he knew it when he saw it. When we can define, with some confidence and justification, and know better how to manage, which views of our data satisfy the requirements of structuration for recordkeeping purposes (how they exist in relation to each other and in relation to context including the context of purpose, use, and re-use), then we will have found our digital Series. But perhaps ...

*In the midst of the word [we are] trying to say,  
 In the midst of [our] laughter and glee,  
 [It will] softly and suddenly vanish away —  
 For the Snark was a Boojum, you see.*

© Chris Hurley 2011

## The Deed (in application)

**Purpose** : end rather than the means by which it is accomplished

- we assist low income earners to get affordable housing
- we improve the welfare of Aborigines

**Enablement** : outcome to be predicted in a specific instance.

- we build and manage public housing units
- we provide rent subsidies
- we maintain an Aboriginal Health Service
- we are going to steal your children

**Action** : a step taken in a specific instance.

- we are building a block of flats here and you will have one
- I am releasing a tender now
- I am signing a contract now
- I am signing your lease agreement

© Chris Hurley 2011

---

Figure Ten

## The Document (applied)

**Fonds** defined by provenance

- ▶ Elvis Presley Archive
- ▶ James Boswell Papers
- ▶ State Archives of Ruritania
- ▶ Ruritanian State Department Fonds

**Sequence** defined by structure

- ▶ Scores and Recordings
- ▶ Invitations Sent and Received
- ▶ Archives of the Presidency
- ▶ Overseas Cables : Outwards

**Object** defined by purpose

- ▶ Love Me Tender (a score)
- ▶ Inauguration Ceremony Records
- ▶ 1946, August 13 : Offer of Aid to Transylvania
- ▶ 1754, Oct 19 : To Dr J. – invitation to dine at Stone's Chop House

© Chris Hurley 2011

---

Figure Eleven

# The Doer (applied)

**Corporation or Family** organises action and bestows authority/unity

- ▶ United Nations Organisation
- ▶ Government of the Commonwealth of Australia
- ▶ The Tudors
- ▶ Ministry of Foreign Affairs
- ▶ British Embassy, Washington

**Actor, Agent, or Person** undertakes action

- ▶ Secretary-General, UNO
- ▶ Henry VIII
- ▶ Fred Nerk (Minister of Foreign Affairs)
- ▶ Kim Philby

**Role or Function** mission or responsibility undertaken

by Arthur Wellesley, Duke of Wellington (1769-1852)

- ▶ C-in-C Allied Forces, Belgium (June 1815)
- ▶ Prime Minister, Great Britain (1828-1839; 1834)

© Chris Hurley 2011

---

## Figure Twelve