

## Rediscovering the Mystery of Electronic Series (2002)

*We have been spinning coins together since I don't know when, and in all that time (if it is all that time) I don't suppose either of us was more than a couple of gold pieces up or down. I hope that doesn't sound surprising because its very unsurprisingness is something I am trying to keep hold of.*

*God must have created mistakes for their wonderful value in illuminating proper pathways. In all of evolutionary biology, I find no error more starkly instructive, or more frequently repeated, than a line of stunning misreason about apes and humans ... If we evolved from apes, why are apes still around? I label this error instructive because its correction is so transforming: If you accept a false notion of evolution, the statement is a deep puzzle; once you reject this fallacy, the statement is evident nonsense (in the literal sense of unintelligible, not the pejorative sense of foolish)<sup>ii</sup>.*

A coin, thrown into the air, can be expected to land up heads more or less as often as it lands up tails. Yet the chances of its landing up heads next time following a throw of tails always remains 50/50. The reason for this is that throws of a coin are not a series. Those who cannot understand this should not try to document records and they should never play at two-up.

Consider, however, a different set of circumstances – a ballot determined by extracting marbles from a jar – an equal number of black and white totalling 100. When the first marble is drawn, the odds on its being either black or white are 50/50. If the first marble is, in fact, white, then the odds of the next marble drawn out of the jar being black are higher. This is because the boundary established by the jar limits the number of possibilities and each instance of drawing a marble out is accordingly part of a series.

If we had some abstract set of descriptive requirements for acts of this kind, it would be possible to describe each of them using that methodology. If, however, our descriptive tool was designed only to describe instances in series, it would be a misapplication of the descriptive tool to use it for throws of the dice – notwithstanding that all the descriptive fields can be filled in and the resulting description is perfectly comprehensible. We would be describing throws of the coins as a series when in fact they are not – just like the addled gambler who imagines his chances of winning on heads are better immediately after a throw of tails.

*It would be possible to describe a new-born puppy using the documentation required by the Registry of Births, Deaths, and Marriages to register human births. The resulting form could easily be accepted by the system and processed to produce a conventional birth certificate. This would not, however, mean that the thing thus being described was a human child.*

In 1905, the American Museum of Natural History displayed a row of skeletal fossils of horses in ascending order of size demonstrating the “evolution” of the horse. The message conveyed by this description was that the modern horse (the largest in the progression) had evolved through stages and that each stage was earlier than the next. In fact, that message was quite false – because the fossils on display were not in an evolutionary series.

*In conventional charts and museum displays, the evolution of the horse looks like a line of schoolchildren all pointed in one direction and arrayed in what my primary-school drill instructors called “size place” (also stratigraphic order in this case). The most familiar of all illustrations, first drawn early in the century for the American Museum of Natural History's pamphlet on the evolution of horses, by W. D. Matthew, but reproduced hundreds of times since then, shows the whole story: size, toes, and teeth arranged in a row by order of appearance in the fossil record ... But what is so wrong with these evolutionary ladders? Surely we can trace an unbroken continuity from Hyracotherium to modern horses. Yes, but continuity comes in many more potential modes than the lock step of the ladder<sup>iii</sup>.*

Each fossil was, indeed, an example of the evolution of the horse. But they were not instances of a progression or sequence. The fossils displayed did not exist in a definable relationship to each other of which the display was illustrative. A quite false impression of succession was conveyed by the way they were arranged and described.

Records series are evidential records, not because of the manner in which they organised and still less because of the nature of the information they contain. They are evidence because their context and structure forge a link between the information they contain and some event or circumstance. More specifically, the content of a document evidences, in some way, an action or transaction. The defining characteristic of a series is that a sequence (or hierarchy of sequences) is established which contributes to knowledge of context and structure.

**Describing a Unicorn (or, “Look, Look! A horse with an arrow in its forehead. It must have been mistaken for a deer”<sup>iv</sup>)**

The arrangement and description of series is fundamental to recordkeeping practice. It is only used by some to deliver information about dead systems to researchers in archival search rooms and on the Internet. Custodialists who wish to use it in this way must have access to the same descriptive tools to describe (as series) electronic records over which they would assume control as the rest of us will need to be recordkeepers in cyberspace. They are, however, unlikely to find them in a continuing fixation on the problems of custody.

The reason for documenting series is not to aid discovery. We document series because the evidential value of a single record depends (in large part) not simply on its context but also upon its connections with other things (chiefly context and other records). At its simplest, the organisation of the records by the recordkeeper into contingent sequences is how connections with other records are made. Records are not subsequently arranged into series so that archivists can describe them for the benefit of researchers. Records exist in series because that is how they are made and kept in the first place.

The question is not, therefore, whether a bunch of electronic stuff can be documented using a series description format; but rather whether electronic recordkeeping is in fact making and keeping electronic stuff in series. The question of how to document the stuff delivered from a records making process into the workings of an archival descriptive process is inextricably tangled up with the question of how electronic stuff is organised within a records making process. The answer may be “as series”, but then again it may not. There is nothing, of course, preventing us from offering our insights on serialisation to those designing and implementing recordkeeping systems.

Not the least of the weaknesses in the custodialist position is that they have failed to develop the descriptive tools they need to achieve custody. To apply series description to electronic records you are trying to take into custody, you must first be reassured that the electronic records you are taking into custody are organised into series as a result of the process of records creation which brought them about.

Sufficient grounds exist for intelligent scepticism that the world of electronic recordkeeping is as yet organising electronic records into describable series. We are probably passing through an intermediate stage, however. It is now possible to outline, theoretically at least, how series can (and probably should) be re-established within the realm of electronic record keeping. Until that happens, so-called transfers of electronic records into a custodial environment will not happily be describable using traditional methodologies.

The descriptive techniques which will be needed to document electronic series in archives custody will most likely come out of the engagement we now have with issues of recordkeeping in cyberspace. First, to paraphrase Mrs Beeton, catch your record.

The custodialists, essentially in denial, must therefore await the results of those whose work regards storage and custody as incidental, not central, to the issue. It will be a sweet irony if the tools the custodialists need to do their work (viz. the descriptive techniques necessary to ensure the survival of electronic records in archival custody) are delivered to them by those of us for whom custody is neither here nor there.

As so often happens when we think outside the square of traditional archival methods, new solutions to new problems arising in cyberspace usually throw light on some hitherto unresolved problems in the world of physical recordkeeping. To anticipate, at this point, the general conclusions of this article, it is likely that electronic series will present us with two problems which will require major surgery to traditional series description:

- a. a much more complicated set of issues around provenance, and
- b. a much higher incidence of records belonging to more than one series.

When we think through the implications of this insight, we find that provenance was just as complicated in the paper world and that sequencing of paper records was also not as unproblematical as it seemed.

Consider a series of book registers. In the paper world, we would describe the volumes as a series according to the system of arrangement given to the books by the records-maker. Thus, we would expect the spines of the books to show the volume numbers or years used to organise the volumes on the shelves. This organising principle or sequencing of the volumes is the basis upon which we would identify and describe the volumes as a series.

Even in the paper world, however, we would have understood that the organising principle for the volumes was not the same principle upon which the data contained in the books is arranged. The system of arrangement of the entries within the register, often organised upon quite a different basis, is also significant. Thus, the entries might be arranged chronologically (for an annual single number system) or according to some classification scheme (for a multiple-number system). Importantly, however, the contents of each book begins afresh (e.g. by returning to the number "1" at the beginning of each volume and prefixing it with a year date). Different sets of numbers might be grouped by prefixes given to certain files (e.g. "P" for personnel files) or in sets of file number blocked out in advance for use in regional offices. Entries in even the most simple register will be some combination of a chronological order and a registration order.

The organisation of the contents of a series and of the physical packages in which the content is held, while it may sometimes be identical, is frequently different. Except in very problematic cases, we have (in the paper world) preferred the organisation of the physical packages as the basis for "serialisation" over the organisation of the data contents. Thus we will describe a series of book registers based on the sequencing of the volumes, rather than the contents within and across volumes.

We are choosing to emphasise one sequence at the expense of another in order to fit the data into a preconceived idea of what a series is like. A more systematic approach would be to regard the contents of the series (both the data and the packages of data) as an organisation of records-stuff and observe (if the facts require it) that this assemblage involves several sequences, not one.

It is clear, however, that we could have also made a series out of the contents - in preference to the physical packages. Thus, identifying as a series the entries in a body of book registers (rather than the registers themselves) we now see to be not only possible, but arguably more sensible.

In the world of physical sets, we would, of course, have to register the physical packets. We can now see, however, that just as the context of the *fonds* is virtual vis

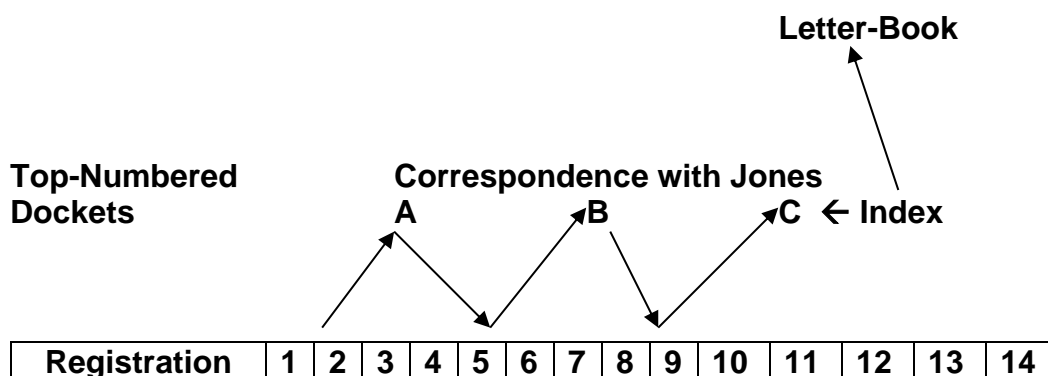
a vis the physical series, so too the physical series could be treated as virtual vis a vis the data contained therein.

A nineteenth century docketing system employing top-numbering provides an even more obvious example. Each piece of incoming correspondence is registered and docketed. Outgoing replies would typically be copied into a letter book cross-referenced to the dockets using registers and indexes. The control records (registers and indexes) would be a guide to the physical location of each docket and the whereabouts of replies in the letter-book.

More importantly, however, these control records document sequencing of the records (for recordkeeping purposes) which is different from the order of the registration numbers. If you want to “assemble” a transactional record within this series it is necessary to use the control records to guide you to the disparate documents which make up the whole record.

New correspondence would be freshly registered and previous papers often attached (top-numbered) into the new docket. The removal of the old docket from its place in the sequence to be filed under the new number used to register subsequent business would be recorded in the register. Not all old dockets were top-numbered, however, some were simply cross-referenced and, in some cases, there is no evidence that a connection was ever made between two pieces of business. A complete transactional record does not exist in these systems except as a “view” provided by the operation of the system as a whole.

**Figure One**  
**Nineteenth Century Docketing System (with Top-Numbering)**



Archivists have the choice of registering as a series the sequence of registration numbers (1, 2, 3, etc) - many of which no longer exist because they have been top-numbered - or the sequence of top-numbered dockets (A, B, C, etc) – which has unexplainable gaps to someone unfamiliar with the process. Whichever is chosen, it will be seen that both sequences are relevant to construct both these and alternative views of the data.

It will be seen that the first physical manifestation of the correspondence with Jones will not be the first registration item (no. 2) nor the first index entry (linking nos 2 and 6) but rather the last docket into which letters from Jones were top-numbered (no. 9). As a recordkeeping system, it is necessary to document the organisation of data into four sequences (or series):

1. Registration order (1, 2, 3, etc.)
2. Transaction order (A, B, C, etc.)
3. Physical order (arrangement of surviving dockets)
4. Arrangement of entries in the letter books.

In addition, this system presents us with at least three other arrangements of data:

1. Arrangement of the registers themselves
2. Arrangement of the letter books
3. Arrangement of the index.

It is also possible that the arrangement of entries within the Letter Books and the Index provide alternative arrangements. In all, this makes for a possible identification of 9 different sequences (or series) within this one recordkeeping system where, in the physical world, we would ordinarily only identify three (viz. the surviving dockets, the letter books, and the Index). I have long argued that the world of electronic recordkeeping should look to nineteenth century docketing systems for its model of how to organise electronic records, rather than the silly attempt to emulate files in cyberspace using “folders”.

What files are good at, however, is representing business processes. Records are essentially the intersection of recordkeeping processes (e.g. filing rules) and business processes (e.g. classification rules). In Part 3, we will explore how this traditional exception of recordkeeping can be built upon to develop recordkeeping systems in cyberspace out of which electronic series really do emerge so that even a custodial archivist could describe them. In the meantime, however, it is necessary to dwell upon the limitations of current electronic recordkeeping systems in organising data into recognisable series.

Why is any of this important? If we did describe as series any convenient collection of data, what would be lost? The answer to that question, once appreciated, is shattering. The answer is: everything.

The rules of arrangement and description do not support data discovery. They support recordkeeping. Making and keeping evidence is the task of the recordkeeper. We do that by keeping the arrangement of the records as evidence of transaction and use. In order to describe an electronic series, even a custodialist must first find (in order to preserve) what the record-maker made. If we simply impose “series order” on records which never had that order to start with, then we are no better than the systematisers in reaction to whom our very profession was formed in the first place.

### **Behind the Looking Glass (or, Alice Doesn't Live Here Any More<sup>v</sup>)**

A series is identifiable by its structure and context. A recordkeeping system is compounded of series within a controlled workspace and is comprised of the elements of control and process. Two kinds of process are relevant : recordkeeping process(es) and business process(es).

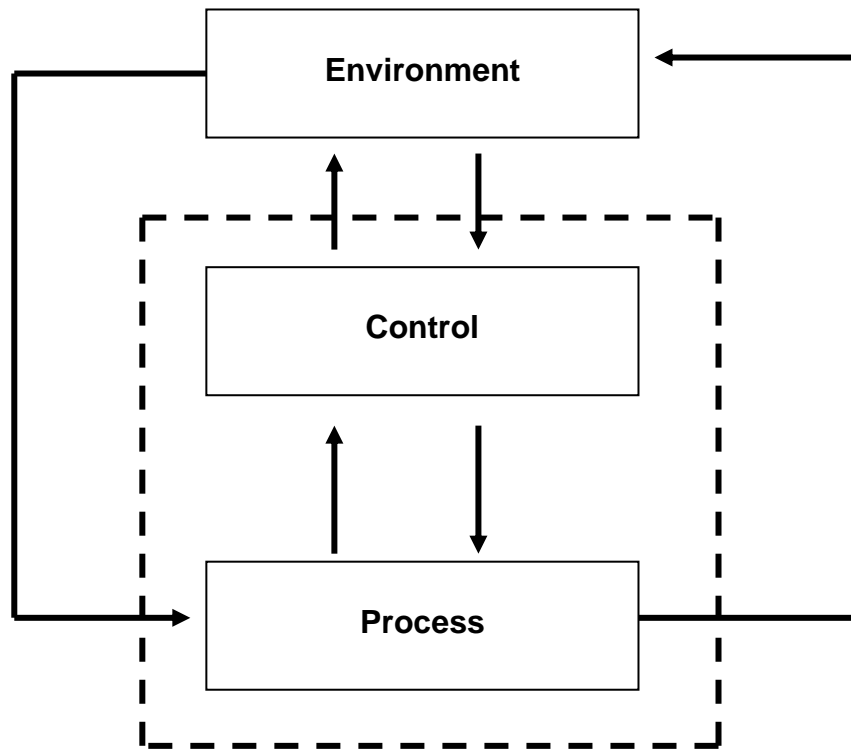
In the traditional description, the provenance statement usually embodies the principal statement about context and is more or less synonymous with control (see Figure Two). There may be many actors in a complex process (including actors from outside the controlled workspace (i.e. from the environment) but it is the controller (“creator”) of records of the process to which/whom primacy is given. In the physical world of paper records control and custody were closely linked. Recorderkeepers have been accustomed to dealing with situations in which control of recordkeeping processes and control of business processes have been more or less synonymous.

Most of these traditional assumptions are under threat in cyberspace. We can no longer assume that control of recordkeeping processes and control of business processes will be synonymous. At the very least, statements of series creation/control will need to become more sophisticated to deal with this complexity. More profoundly, it is likely that documents will belong to more than one series within shared workspace.

Moreover, the possibilities which automated systems have for sequencing documentary traces in series reflecting simultaneously more than one evidential thread of action, almost certainly mean that a single documentary trace will need to be organised into more than one series even within the same span of control. This will be because the physical sequence no longer limits the links which can be established between one documentary trace and another within either recordkeeping or business processes.

These are *new* problems, in the sense that descriptive practice has not hitherto dealt with them very well. They are new problems because electronic recordkeeping has not yet developed the implementation strategies for dealing with the consequences.

**Figure Two**



In this regard, it is of paramount importance to understand that series description is not imposed upon records (electronic or otherwise). We do not organise records into series, rather they are records because they are being managed in series. If electronic recordkeeping is not organising documentary materials into series which capture and maintain their evidential value as records then we are simply confronted by a reality which defies us to apply our descriptive tools to it. Imply, if there are no series, there is nothing to describe.

Since electronic recordkeeping has so far developed only the most crude attempts to organise materials into sequences which fully satisfy the recordkeeping need for evidence, we can really only guess at what, in practice, will eventually be delivered to us for description. Many different implementations will be possible and we should rightly refrain from trying to predict what the market will eventually decide is the most effective and efficient way of going about it. We know, at least, that it will be materially differently from implementation strategies in the paper world.

What we can know now, in advance of a fully developed implementation model, is the theoretical basis upon which any successful implementation will have to be justified. A series of marbles drawn from a barrel is governed by probability theory; a series of species is governed by evolutionary theory; and a series of records is governed by recordkeeping theory.

A second key insight is that these problems are not really new. Compound processes in shared workspace may look like a novel situation create by the possibilities opened up by IT, but in fact it has been with us always. What we have hitherto regarded as a single workspace within a corporation has, in reality, been identical to the situation created by shared workspace in cyberspace. For “agency” in cyberspace read “business unit” or “actor” within a corporation and all of the issues are the same.

That we have not dealt with these issues and the complexity they give rise to is the product, truly, of the limitations resulting from managing records physically. Nevertheless, they already exist. Within the records of a single corporation, the role of business units and officers definitely alters the provenance of records and dictates how documentary traces are organised. Traditional descriptive practice has ignored this complexity by assigning primacy to the corporation and assuming that the corporation exercised perfect control over the activity of business units and officers.

Our context statements have sometimes acknowledged and documented the different roles and activities of actors within the corporation but seldom to the extent of elevating them to the role of provenance. Even traditional provenance entities cannot be said to have been unproblematic. In order for a provenance statement to be effective, assumptions must be made about the role that provenance entity has in relation to the records in question. I have explored this issue in a preliminary way elsewhere<sup>vi</sup>. Almost completely unexplored is the related problem of the different ways in which a provenance statement (once correctly documented) can relate to a series (i.e. in how many different ways can a series be created and controlled?).

Similarly, in the paper world a single documentary trace can be seen to belong to more than one evidential sequence. Photocopying the same document to many files is one way of crudely doing this in the paper world. More sophisticated approaches employ relationships between documents and/or files to transfix the documentary trace of a single (transaction into more than one sequence. The entries in a letter book arrange the outgoing correspondence in a single evidential sequence (chronological) within a system which also incorporates the letter-copy into the (trans)actional record of which it is part.

We can analyse (and could have described) these and other complexities of context and structure already based simply on our experience of what has actually happened in the paper world in which we are already familiar. It is possible, to foresee that the electronic world is going to raise similar complexities in ways which we will be unable to avoid.

If we had dealt with these issues in physical space, we would now have some solutions. We would see, for example, that a file is both a sub-series and an item. We would recognise that a single documentary trace on a file belongs simultaneously to at least two series : the file itself and the series to which the file belongs.

Until the world of electronic recordkeeping actually delivers robust series of electronic records for us to describe, we will have to keep guessing what they will look like. We can, however, as an exercise in pure theory, articulate how they will need to be described.

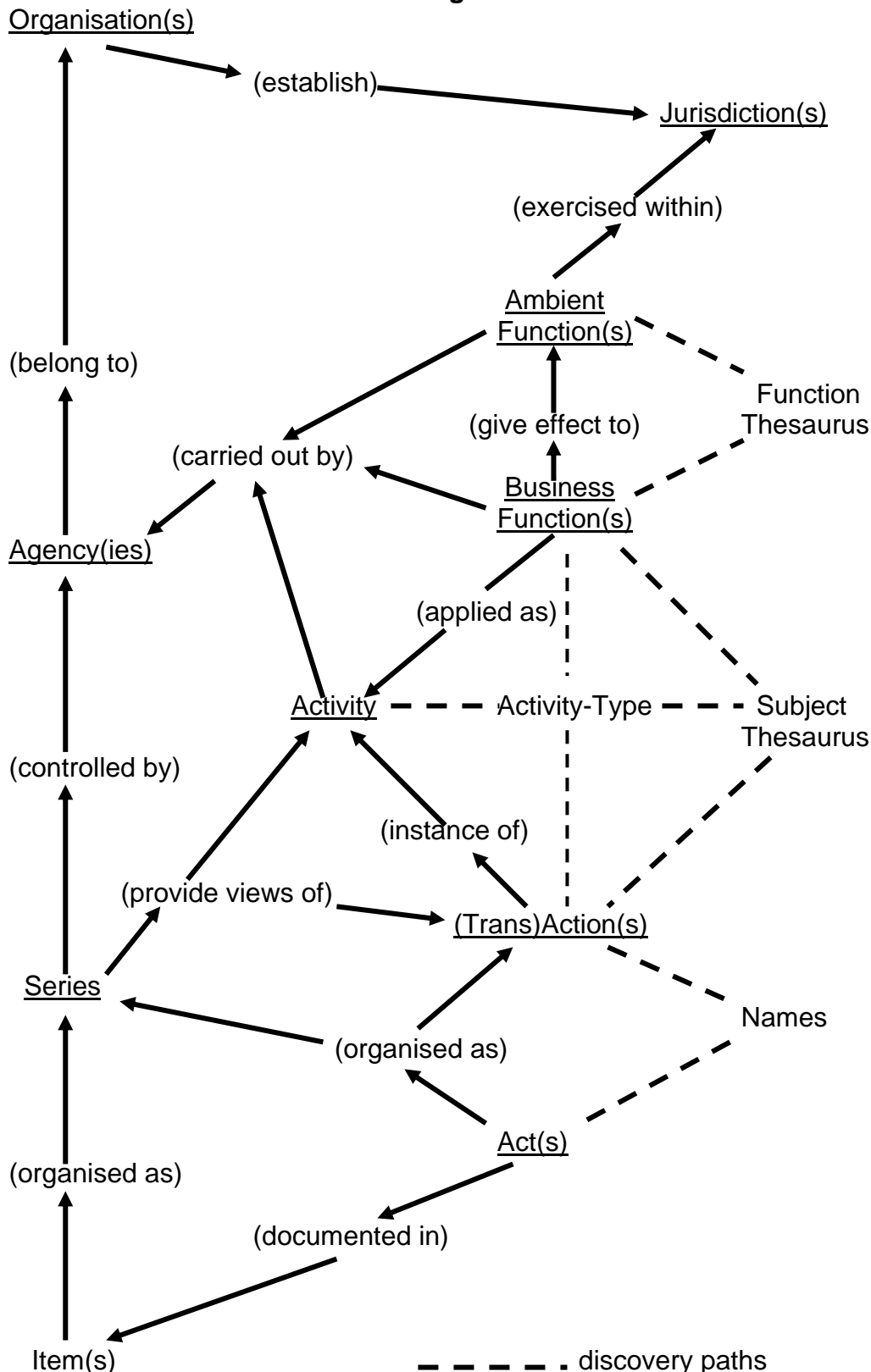
### **Some Modest Proposals (or, You Can Get There from Here<sup>vii</sup>)**

Two of the chief characteristics of a series (whether in a traditional or an electronic environment) will be provenance and structure. The fifth file opened in 1990 by Agency X dealing with “general correspondence” can be usefully described as Item 1990/5 of the General Correspondence Files of Agency X. It is perfectly possible to imagine, within the data model we are developing for implementation of the Australian (“series”) System, how series of electronic records could be derived. The model we will use to demonstrate this has (or will be) outlined in greater detail elsewhere (**Figure Three**).

The entities outlined in **Figure Three** can all fit the SPIRT Project’s RKMS Model for metadata. In the electronic world, the provenance of a series will be less likely to be a structural entity (agency, business unit, section, officer) as a span of control or responsibility for an activity - a set of (trans)actions. This responsibility may correspond neatly with the organisational divisions within which they were traditionally

assigned in the paper world, but increasingly they are likely to take the form of so-called compound-transactions within “shared workspace”.

**Figure Three**



Moreover, a single act may well form part of several (trans)actional sequences – just as a single (trans)action may belong to several activity sequences. Thus, a single record item may be simultaneously part of more than one “series”. To deal with this, it will be necessary to document recordkeeping systems as an entity at the series level (perhaps treating activity- and action-based sequences as *sous-series*).



We should begin by recognising that the issues of sequencing which give rise to the “problem” of describing electronic series always existed in the paper world, but that they were suppressed by the limitations of the physical world.

Records are the result of the intersection of two related processes – business processes and recordkeeping processes. Physical records represent a triumph of recordkeeping process over business process. Records are organised systematically into sequences of filing which approximate or relate to concepts in the business process. Thus, a file represents an instance of a series of transactions belonging to a class of similar transactions. A class of similar transactions is equivalent to an “activity” in **Figure Three**.

In the paper world an “activity” component of a filing system represented the recordkeeper’s attempt to classify records into surrogates for business processes. Accordingly, a recordkeeper maintains a severe distinction between the classification scheme (recordkeeping metadata) and the subject index (discovery metadata). The classification organises material into recordkeeping structure and the subject index enables retrieval of information from the records.

Information about “buildings”, for example, might be found throughout the records.

Documents filed in sequence on a file thus classified :

5.0.0	Accommodation
5.1.0	----: Buildings
5.1.1	----: ----: Leases
5.1.1/1	----: ----: ----: 24 Reilly St, Auckland

provide an evidential record of a sequence of transactions relating to the leasing of the property in question from which meaning is derived not simply from the informational content of each document, but also from their accumulation together. A sequence of documents relating to transactions documenting leasing arrangements for that property would, typically, be found on file 5.1.1/1, in a chronological sequence which would be expected to place documents in the successive order in which they were created. The documents would be fastened by a split-pin (or tie-tight).

This sequence of documents constituted a sub-series. The file itself would ordinarily be placed inside classification system which would facilitate discovery and make connections with other files representing other views of the activities of the business more or less closely related to the transaction(s) documented on the file. Although the sub-series (the file) and the series each reflect important aspects of business process, their primary orientation is towards a recordkeeping process (i.e. classification, registering, filing, and indexing).

The first response of the electronic world has been to replicate within electronic records management systems a virtual equivalent of the file” – i.e. the ubiquitous virtual folders. As argued above, a more appropriate model would be found in docketing systems. Both folders and files, however, involve thinking about electronic records as if they were still constrained by the limitations of physical space. We seem unable to imagine the sequencing requirements for electronic records (upon which the reality of electronic series will be founded) – like our forefathers were unable to think of motor cars except as horseless carriages.

Although the real world does not yet provide us with example of sequencing solutions (electronic series) which properly utilise the potentialities of the technology, it is easy enough to imagine what they will be like. It is already clear that recordkeeping will demand the documentation of Acts (see Figure Three). These Acts will belong in sequences reflecting the manifold views of that Act made possible by its location within a contextual network of relationships with (trans)actions and the activities from which they derive.

Each electronic trace of an Act will this be held together, not by a single split-pin (or tie-tight) but by manifold ribbons of contextuality derived from the functional context from which it derives. It is these manifold sequences (or views) to which each document will belong that will be the electronic series we must document. An electronic item (or document) will belong to many series (or views) not just one. Each series will share its content with any other series. We have yet to develop the descriptive tools to deal with this.

The documentary traces of Acts will this be joined not in a physical sequence dictated by the requirements of filing, but in virtual sequences based on an analysis of business functions. Acts should, desirably, be documented entities within recordkeeping systems. In applying workflow, these documented Acts will be much ore stable and enduring than the work-flow analysis which they will populate. Business need will determine whether or not a record will be needed of each Act. A typical work-flow might consist of dozens or hundreds of Acts, of which a few will be documented and a business rule established that a record be created of each instance of that Act.

A record of the workflow(s) within which a documented record of the Act is made (and of the version of the workflow involved) will need to be kept – as part f the knowledge informing the record. The sequence (series) will not, however, be imposed by the work-flow sequence. The sequence will reflect the business process(es) embodied in (Trans)actions and activities related to the Act.

This will move the emphasis we once gave in the paper world to recordkeeping process over business process, and replace it with an emphasis on business process over recordkeeping process. We will then be left with the problem of how to introduce recordkeeping process back into recordereeping. This can be done by using the control boundary (Figure Two) to define the recordkeeping system to which the manifold views (series to which each item belongs. Each series will this provide a view of a sequence of evidential records within a documented recordkeeping system.

---

## END-NOTES

<sup>i</sup> Tom Stoppard, *Rosencrantz and Guildenstern Are Dead*, Act 1 (London, Faber, 1968)

<sup>ii</sup> Stephen Jay Gould, “The Declining Empire of Apes” in *Eight Little Piggies* (London, Jonathan Cape, 1993), p. 284.

<sup>iii</sup> Stephen Jay Gould, “Life’s Little Joke” in *Bully for Brontosaurus* (New York, Norton, 1991), pp.174-175.

<sup>iv</sup> With apologies to Tom Stoppard.

<sup>v</sup> With apologies to Bobby Goldsboro.

<sup>vi</sup> Chris Hurley, “Problems with Provenance”

<sup>vii</sup> With apologies to Shirley Maclaine.