What is a Collection?

JAMES CURRALL, MICHAEL MOSS, and SUSAN STUART

RÉSUMÉ Le mot « collection » est monnaie courante dans le monde réel des objets et des événements et a été importé sans trop d'efforts dans notre discours sur le monde numérique. Cependant, il n'en existe pas de définition précise dans aucun de ces domaines. Les auteurs clarifient ce problème en examinant d'abord comment ce mot a été utilisé dans la littérature contemporaine sur les sciences de l'information et ensuite en établissant les critères qui sont employés pour créer une collection. Ils allèguent que l'hypothèse qu'il existe une permanence au sens réaliste du terme ou une fixité dans le monde qui détermine la classification est fausse. Selon eux, la seule approche faisable dans la construction de catégories pour répartir des objets, qu'ils soient numériques ou physiques, est anti-réaliste, et on doit alors porter attention aux intentions et aux décisions de l'institution qui collectionne.

ABSTRACT The word "collection" has been common currency in what we accept as the real world of objects and events, and has been imported with seemingly little effort into our discourse about the digital world, yet there is no clear definition in either domain of what is meant by the term. We clarify this issue by first examining how the term is used in the contemporary information science literature and then by going on to establish the criteria which are employed in bringing a collection about. We will argue that the assumption that there is a realist permanence or fixity in the world that determines taxonomies is false, and that the only feasible approach to the construction of categories to which objects, whether digital or physical, are allocated is an anti-realist one where attention is paid to the intentions and subsequent decisions of the collector.

Introduction

In this paper we intend: to examine what different constituencies mean by the term "collection," discuss how collections come into being, and explore the implications that these have in the world of digital information. This exploration takes us into philosophy and realist and anti-realist approaches to the categorization of knowledge. We then consider the way that archivists and others see collections, and especially the criteria (implicit and explicit) used to define collections. Finally, we discuss how the difficulties that we uncover along the way might be resolved and how the different "information professions" –

archivist, librarian, curator, accountant, statistician, technologist, and so on – might learn from a deeper understanding of each other's worlds.

As a starting point, we find difficulties in current usage with both how collections are defined and how they fail to be defined. For example, the Joint Information System Committee (JISC), which takes a lead in the provision and organization of information for higher and further education in the United Kingdom, has constructed an information environment (JISC IE) which is predicated on an understanding of what collections are and the use of collection-level descriptions "to allow portals, brokers and aggregators to automatically determine what collections are available." JISC IE defines "collection" as "a discrete aggregation of one or more items of content, but will often take the form of a database of one kind or another."[1] We will demonstrate that this is a very narrow and uninformative definition of what constitutes a collection. Earlier work by the UK Office for Library and Information Networking (UKOLN), which provides the collection description service to JISC IE, adopted much broader definitions and recognized that collection-level descriptions will be provided through different distribution channels and media.[2] A much fuller and more useful definition was provided by the RIDING Clump project, which was part of the JISC electronic library program. They pointed out that "a collection may be made up of other collections as well as items, or items and collections together. An item itself may be made up of other items, e.g., a catalogue is actually a collection of catalogue records: a Web page actually comprises text, images, etc. although people think of a Web page as a single item." However, instead of trying to tease out these relationships, "it was left up to the person who probably knows the collection best (i.e., the person in charge of the material) to make the decision about whether it should be described as a collection."[3] The analytical model of collections and their catalogues by Michael Heaney and commissioned by UKOLN with the support of the On-Line Computer Library Centre (OCLC) drew a distinction between collections and the collection of information about such entities. Unfortunately, he neither explored in depth the relationship of surrogate data (however faithfully reproduced) to originals nor provided any adequate definition of what a collection might be.[4] He did claim boldly, however, that his analytical model "should be applicable to physical and digital collections of all kinds, including library, art and museum

1 JISC Information Environment Architecture – Collection Description Service, available at: <http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/collection-description/>.

2 "Creating reusable collection-level descriptions," Collection Description Focus, Guidance Paper 1, available at: <http://www.ukoln.ac.uk/cd-focus/guides/gp1/>.

3 Dr. E.V. Brack, David Palmer, and Bridget Robinson, "Collection Level Description – The RIDING and Agora Experience," *D-Lib Magazine* (September 2000), available at: <http://www.dlib.org/dlib/september00/09brack.html>.

4 Michael Heaney, *An Analytical Model of Collections and their Catalogues*, 3rd issue (University of Oxford), p. 4, available at: <http://www.ukoln.ac.uk/metadata/rslf/model/>.

materials, and is by no means applicable only to the resources of large research libraries."[5]

The Open Archival Information System (OAIS), the model for many digital repositories, gives no guidance as to the nature of a collection, except to say: Archival Information Packages (AIPs) "are then aggregated into Archive Information Collections (AICs) using criteria determined by the archivist. Generally AICs are based on the Archive Information Units (AIUs) of interest having common themes or origins and a common set of Associate Descriptions."[6] This is both unhelpful and dangerous, raising wider questions about the role of archives in society expressed tellingly by Jacques Derrida in his book *Archive Fever*: "There is no political power without control of the archive, if not memory. Effective democratization can always be measured by this essential criterion: the participation in and access to the archive, its constitution, and its interpretation."[7] We will return to these issues presently.

While UKOLN, the Scottish Collections Network (SCONE), and OAIS[8] recognize that entities within collections and groups of collections may have many-to-many relations, there is a tendency to endow them with a fixity within a hierarchy that they do not possess.[9] This paper works towards some explanation of what is meant by "collection" and offers a revised notion of how collections come about. To edge this debate forward we will first offer a broad clarification of the realist, anti-realist, and intermediate approaches to the categorization of knowledge.

The Realist and Anti-Realist Approaches to the Categorization of Knowledge

There is a claim that the physical world is real, that it is all there is as the subject of empirical enquiry, and that what exists is mind-independent.[10] Not

---

5 Ibid., p. 3.

6 *Reference Model for an Open Archival Information System (OAIS)*, Blue Book, January 2002, section 4.2.2.7.

7 Jacques Derrida, *Archive Fever – A Freudian Impression* (Chicago, 1966), p. 4, fn. 1.

8 "The AIC organizes a set of AIPs along a thematic hierarchy." *Reference Model for an Open Archival Information System (OAIS)*.

9 "Maintaining collection-level descriptions," Collection Description Focus, Guidance Paper 2, available at: <http://www.ukoln.ac.uk/cd-focus/guides/gp2/> and "Reusing data and avoiding unnecessary duplication," Guide to the SCONE database, available at: <http://scone.strath.ac.uk/service/Guide/gReuse.cfm>.

10 See John Locke, *An Essay Concerning Human Understanding* (1690, Oxford, 1975); I. Newton, *Philosophiae Naturalis Principia Mathematica* (1687), Andrew Motte, trans. (1729), available at: <http:vorlon.ces.cwru.edu/~ames/principia/>; Bertrand Russell, *Our Knowledge of the External World* (London, 1914); Herbert Feigl, "Physicalism, Unity of Science, and the Foundations of Psychology," in P.A. Schilpp, ed., *The Philosophy of Rudolf Carnap* (La Salle, IL, 1963); A.J. Ayer, "My Mental Development," in L.E. Hahn, ed., *The Philosophy of A.J. Ayer* (La Salle, IL, 1992).

unsurprisingly there is an opposing position, that the physical world is not real, or certainly not knowable as real, that all we can ever know are the ideas that correspond to the representations we have of what we perceive to be external, and that whatever we have knowledge of is, therefore, mind-dependent.[11] As Berkeley claimed very elegantly "All the choir of heaven and furniture of the earth, in a word all those bodies which compose the mighty frame of the world, have not any subsistence without a mind."[12] There are also any number of intermediate positions, for example, Kant's Transcendental Idealism which accepts that all we can know, in the sense of experience, are ideas, but that these ideas must arise from a world of physical objects[13]; Dummett's, superficially similar, position that objects exist but they are not mind-independent[14]; and Searle's position that endorses realism about "brute facts" and anti-realism about "social facts."[15] This debate is relevant to our purpose of determining how collections come about because we challenge the realist position that there is, for example, a natural taxonomy existing mind-independently in the universe that is waiting to be discovered and categorized. However, we are not espousing, as some radical alternative, a post-modern, or even deconstructionist perspective.

One systematic categorization with which many of us are very familiar is the biological one of plant and animal phyla, and most important in this field has been Carolus Linnaeus whose work on taxonomies in *Species Plantarum* (1753) and *Systema Naturae* (1758) provided a system for revealing the order that was assumed to be inherent in nature.[16] His system emphasized the need for a universal system of nomenclature, hence binomial naming, and the hierarchical classification of organisms into genera, orders, classes, and kingdoms. Although Linnaeus altered his realist assumption about species being an invariable category without which the revelation of order in nature would be impossible, he did not abandon it completely. He accepted a notion of hybridity in plants and animals, and the production of new species, but he did not accept that this happened outside of the inherent order, or divine scheme, that was there to be uncovered.

11 George Berkeley, *Essay Toward a New Theory of Vision and Other Writings* (London, 1910); George Berkeley, *The Principles of Human Knowledge* (London, 1937).

12 Berkeley, *Principles of Human Knowledge*, p. 6.

13 Immanuel Kant, *The Critique of Pure Reason*, Norman Kemp Smith, trans. (London, 1929).

14 Michael Dummett, *Truth and Other Enigmas* (London, 1978); Michael Dummett, *The Seas of Language* (Oxford, 1993).

15 John Searle, *The Construction of Social Reality* (New York, 1995).

16 Carolus Linnaeus, *Species Plantarum: exhibentes plantas rite cognitas, ad genera relatas, cum differentiis specificis, nominibus trivialibus, synonymis selectis, locis natalibus, secundum systema sexuale digestas* (Stockholm, 1762); and *Systema naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis* (Vol. 1) (Stockholm, 1758).

Linnaeus's search for a natural system continues, with some revisions, to be in use today with biologists now concentrating on the evolutionary relationships between taxonomic groups or species. There are, and will continue to be, frequent revisions to the categorization of such a system; but even these will be insufficient to demonstrate the scarcity of well-defined categories in the natural world for the determined realist; for the determined realist can always claim that the definitive taxonomy is always just around the next taxonomic corner.

We hope to demonstrate that many information scientists, and undoubtedly some biologists, are in the inconsistent position of being realists about a mind-independent taxonomy into which every object or event will ultimately fall, whilst using language which is not only mind-dependent, but which circumscribes the limits of our knowledge. We will also argue that in doing this they manipulate and determine these categories intentionally, and in this intentional determination of categories they express preferences that are based on a range of choices, personal, professional, cultural, historical, religious, political, and so on. This is precisely the trap into which the two most cited treatments of the question in information sciences fall.[17] While Berman accepts terminology will change to reflect societal expectations, he infers that we are simply travelling down a road of revelation to the true (American) taxonomy.[18] Bowker and Star raise some interesting issues, but argue misguidedly that standards impose classification schemes and that "algorithms for codification do not resolve moral questions" when there is no reason that they should.[19] This is the confusion that is evident in archival literature among authors, for example, Duff and Harris who likewise confuse standards with taxonomies and the activity of placing objects into categories with the creation of the taxonomies themselves.[20]

Our position is simply this: it is human beings, with their language and intentions, who determine the categories and, thus, the collections into which things are placed. They might do this with some feeling of there being a correspondence with how things are in the world, but ultimately this correspondence remains unknowable since what is perceived and described is determined by the limits of senses, thought, and language. The fact is that the process is iterative, categories are redefined to reflect new knowledge or to allow for greater subdivision of a mass. We would argue that collections rarely

---

17 Sanford Berman, *Prejudices and Antipathies: A Tract on the LC Subject Heads Concerning People* (Metuchen, NJ, 1971) and Geoffrey Bowker and S.L. Star, *Sorting Things Out: Classification and its Consequences* (Cambridge, MA, 1999).

18 Berman, *Prejudices and Antipathies*.

19 Bowker and Star, *Sorting Things Out*, p. 15.

20 Wendy M. Duff and Verne Harris, "Stories and Names: Archival Description," *Archival Science* 2, nos. 3–4 (2002), pp. 263–85.

come about as a result of chance or happenstance, though an element of fortune might play some part; more often they are the result of choices made by individuals or groups of individuals, who define and determine what they know by their cultural-specific use of language, and these categorization choices privilege, possibly inadvertently, some kinds of information or knowledge over others. The categorization of artefacts makes this claim clear. Artefacts, with the exception of natural artefacts like driftwood or a lump of agate, whether physical objects or digital entities, are the product of some agency and their allocation into categories and collections is, there can be no doubt, the consequence of someone's relationship with the object and, ultimately, their choice. This is the contextual information archivists make much of, but more often than not must be inferred from the object and its relationship with others. It must not be confused with the descriptive categories (libertine or not) into which a custodian might place it.[21]

The categories that result from these choices and decisions are the social facts that Searle has claimed come into existence through human construction.[22] In creating our social reality we assign functions to objects and information, and classify them into groups, so that the things that we are sitting on are chairs, the paper we use in financial transactions is currency of agreed denominations, and Wagner's *Tristan and Isolde* is, we have decided, a wonderful expression of Late Romantic grand opera. Agreement about how to classify these things derives from the fact that we share a common social intention to treat those things in a particular way, and they will remain in those categories until we change the rules of classification, or *en masse* we fail to conform to those rules, or some *deus ex machina* intervenes.

It will, by now, have become clear that we disagree strongly with any claim that collections are the result of a pre-existing system, a realism into which objects, information, and events will slot regardless of determinants like language, culture, and an individual's wishes and desires, let alone "standards." In the sections that follow we will examine how this has influenced what we mean by "collection" and how we ought to determine collections in the digital world. But it should be noted that in writing this paper we have restricted our consideration of contemporary information science to the theoretical questions underpinning the nature of collections, that is, to collecting and collections *per se*.[23]

---

21 Ibid., p. 284.

22 Searle, *The Construction of Social Reality.*

23 More practical considerations of how this thinking affects both traditional and digital collections will be the subject of a paper entitled "Why Privileging Information is Inevitable" forthcoming in *Archives and Manuscripts*, 2005.

A Contemporary Understanding of Collections

In the physical world the term "collection" is employed variously by archivists, museum curators, and librarians. Sometimes it is used to describe all the holdings of a repository which have been "collected," as it were, by successive curators. In these circumstances the contents of the repository are the result of a "deliberate" policy of collecting. For example, we might only collect Roman artefacts or political papers or books about rabbits and, although we might imagine that we can easily classify members of a particular set or collection, with a little reflection we will realize that the content of any collection is usually dynamic; they may not be the same even from day to day, as objects are added or deleted and the boundaries modified to take account of the changes.[24]

The term "collection" is also used as a sort of *post hoc* shorthand for a bunch of objects that have been deposited in a particular location and, perhaps now**,** form an identifiable subset of, or are simply part of, a larger collection. In these instances the act of collecting or assembling the collection is shared between the original owner and the repository**,** with the possibility that the guardians of the repository might be unwilling to accept, for whatever reason, an entire collection from a benefactor even when offered it. In determining what will and will not be accepted, the guardians will be making enormous assumptions about what the boundaries of a particular collection should be. Such collections may subsequently be de-accessioned or parts even destroyed, nevertheless they are considered to have fixed boundaries and cannot usually be added to by curators or depositors except explicitly; further accretions are usually referred to as "additional deposits" in catalogues.[25]

Although Hilary Jenkinson,[26] in his germinal book on archive administration, made no reference to "collections," manuscript curators in the United States were by then interested in them.[27] According to T.R. Schellenberg, as

---

24 One of the best comic examples of the perpetual flux of boundaries is Polonius' attempt to categorize the types of acting the traveling players perform: "The best actors in the world, either for tragedy, comedy, history, pastoral, pastoral-comical, historical-pastoral, tragical-historical, tragical-comical-historical-pastoral; scene individable, or poem unlimited. Seneca cannot be too heavy, nor Plautus too light. For the law of writ and the liberty, these are the only men." W. Shakespeare, *Hamlet* (Middlesex, 1980), Act II, Scene 2, lines 395–401.

25 This issue is addressed by Terry Eastwood, "Putting the Parts of the Whole Together: Systematic Arrangement of Archives," *Archivaria* 50 (Fall 2000), p. 7.

26 Hilary Jenkinson, *A Manual of Archive Administration* (London, 1922).

27 Jenkinson was later dismissive of Schellenberg's concept of artificial collections in his famously dismissive review of *Modern Archives: Principles and Techniques* (1956), reprinted in *Selected Writings of Sir Hilary Jenkinson* (Gloucester, 1980), pp. 339–42, in which he set the tone by using the latin tag "ex America semper alaquid novi": "to make the fact that Archives have been subject to selection ... an essential part of Archives quality is to mask the

early as 1904 "Worthington Ford was, perhaps, the first manuscript curator to concern himself with the nature of collections received by manuscript reposi- tories."[28] Schellenberg, critical of the lack of definition of the term, identified three types of collections: "(a) organic collections, (b) artificial collections, and (c) collections of miscellany."[29] He developed this typology by explaining in what circumstances each term should be applied, but failed to draw a satis- factory distinction between artificial collections and collections of miscel- lany.[30] The Canadian *Rules for Archival Description* (*RAD*) narrowly define a collection in terms of "b" and "c" in Schellenberg's list:

An artificial accumulation of documents of any provenance brought together on the basis of some common characteristic, e.g. way of acquisition, subject, language, medium, type of document, name of collector, which may be treated for descriptive purposes as a unit under a common title.[31]

Organic collections are described as fonds, "the whole of the documents, regardless of form or medium, automatically created and/or accumulated and used by a particular individual, family, or corporate body in the course of that creator's activities or functions."[32] The difficulty with this distinction is that many institutions continue to use the term collection or series to describe what *RAD* terms "fonds" and there seems no possibility of resolving this ambiguity.

In cataloguing their holdings, librarians and, to lesser extent, museum cura- tors have tended to start with individual objects and attribute them to either received or constructed collections; for example, the collection presented by Colonel *X* or the collection of French plays. The OAIS model is then in a sense a library-oriented approach. Archivists usually start with collections, for example, the papers of Sir Winston Churchill and, if resources allow or signif- icance demands, catalogue to the individual piece or item level, described somewhat opaquely in *RAD* as the "lowest level of description and the small-

---

sad conclusion that our generation is bringing Archives a long step nearer to the status of those artificial "Collections" to which Dr. Schellenberg a little later assigns, in agreement with me, an inferior quality as evidence; and that in doing so it surrenders one of the most valuable Archival Characteristics – their impartiality." This seems to us more a dig at the American obsession with breaking up important European archives in the post-war than an attack on the concept of a collection itself.

28 T.R. Schellenberg, *The Management of Archives* (New York, 1965).

29 Ibid., p. 172.

30 Ibid., pp. 173–74.

31 Bureau of Canadian Archivists, *Rules for Archival Description* (1990, revised 2003), Appendix D, Glossary, available at: <http://www.cdncouncilarchives.ca/rad_ part1_revised_Dec2003. pdf>.

32 Ibid.

est intellectual entity within a fonds useful for descriptive purposes."[33] Museum curators fall somewhere between the two; they tend to concentrate on describing the contents and sometimes on the provenance and history of individual objects or the collection itself. But on the whole, curators have avoided addressing questions of "intention" in describing collections, even though they have arrogated to themselves the right to select and dispense objects for preservation within their stewardship. Terms such as "artificial collection" (collections in the Canadian sense) "usually done by bibliophiles, historians, or dealers for commercial purposes" militate against such enquiry except in the crudest terms.[34] To condemn the motive of such collecting as "commercial" fails to recognize that such collecting can be just as intellectually coherent as that of public repositories.[35]

The Criteria for Collections

To date, the information professions have failed to come to grips with the "what-ness" or perhaps better the ontology of the objects in their stewardship within the physical let alone the digital environment.[36] Although curators are aware that objects within their collection or collections have been selected for inclusion, it rarely occurs to them that the process of selection is both dynamic and constructed. It is more obvious in the case of a taxonomic reference collection, such as a herbarium or a stamp album, where specimens are added either to fill gaps or because new plants have been discovered or new stamps issued. Michael Buckland has argued that much collection development in libraries is "value-laden" – "The books placed on the shelves, in the reader's face, so to speak, carry an implicit endorsement: These you should read; these are good books for you; or these are books you will like. Other material, those not selected for (or weeded from) the collection, are actively (though implicitly) treated as less suitable for readers."[37] Although in some instances the process of building collections "is more haphazard than constructed," the winnowing hand of time is rarely as Buckland suggests only by happenstance, even if chance has played a part, for example, in the form of a leaking roof or a long forgotten drawer. The very act of placing documents in a drawer which

---

33 *Rules for Archival Description*; Eastwood, "Putting the Parts of the Whole Together," p. 97; Duff and Harris, "Stories and Names," pp. 272–73.

34 Duff and Harris, "Stories and Names," p. 274.

35 An excellent example of the intellectual coherence of commercial collecting is given in Miles Harvey's 2001 interview with the collector Gilbert Bland. See Miles Harvey, *The Island of Lost Maps – A True Story of Cartographic Crime* (London, 2001).

36 For an interesting discussion of this particular issue, see Sarah Tyacke, "Archives in a Wider World: The Culture and Politics of Archives," *Archivaria* 52 (Fall 2001), pp. 1–25.

37 Michael Buckland, *Redesigning Library Services: A Manifesto*, available at: <http://sunsite. berkeley.edu/Literature/Library//Redesigning/supp lement.html>.

was subsequently forgotten or plant specimens under a roof which subse-
quently leaked was "intended," just as much as their creation in the first place,
even if the consequences were not. It is impossible to maintain that the weed-
ing of books from a library, or documents from an archive, or objects from a
museum is not intended. We are not yet interested in the nature of the inten-
tion, that is, whether it be a moral, social, or political judgement – something
that troubles Bowker and Star, and Duff and Harris – but simply in the inten-
tionality expressed within the act itself.[38]

Thus, any collection must be approached with both happenstance and inten-
tionality in mind since the reason for those collections having come about will
lie somewhere along that scale. Although it is difficult to imagine, all but a
few will survive by happenstance alone (the letters which fell behind the desk,
or apples that fell off a lorry or the people frozen in time at Herculaneum), it is
quite possible to imagine many instances of entirely intentional survival; I
keep all the letters my mother wrote to me, and so on. Intention can be subjec-
tive (conforming to internal rules as in this case), or objective (conforming to
external rules) or a combination of both. Rules for selection have different
weight, from all the terrors of the law – the need to keep your car insurance
and registration documents handy – to entirely private self-referential criteria
– I only collect yellow flowers or piebald horses. As we all know rules, both
objective and subjective, can and do change to conform to changing social
norms and differing cultural perspectives, and Marilyn Strathern would argue
that in our increasingly audit society they become one and the same.[39] We
only have to think of the way the rules for proof of personal identity have
changed over the last twenty years or the way in which interest in the history
of women and gardening has resulted in changes in collecting and cataloguing
procedures. External rules are usually explicit but may not necessarily be
recorded. Internal rules are rarely recorded and have to be inferred from the
content of a collection; only the letters from my mother survive even though it
is unlikely that I received no other letters except from my mother. Moreover,
even if it was my intention to keep letters from my mother, it may well be that
I failed to do this consistently or I passed some on to my sister. Even where
external rules apply, such lapses in intention can occur. But in both cases the
modification of intentionality can only be inferred unless explicitly recorded.

If it is accepted that all collections are in some ways constructed simply to
make it easier to handle or describe the objects contained (a bottle full of flies,
or a bowl of fruit, or a flock of sheep), then it follows that we are dealing with
relationships which can be dis-aggregated and re-aggregated to form other

---

38 Bowker and Star, *Sorting Things Out*; Duff and Harris, "Stories and Names."
39 Marilyn Strathern, *Virtual Society? Get Real!*, available at: <http://virtualsociety.sbs.ox.ac.uk/
    GRpapers/strathern.htm>.

collections.[40] The items need not be taxonomically related. I grow bog plants such as primula, irises, gunnera, caltha, and so on. I could just grow irises but not all irises grow in bogs even though they are related taxonomically, or at least they are at present. This applies equally to books, archives, and museum objects. A book by author *X* about the French Revolution belongs to a library's whole collection, as well as to its collection of books by author *X*, to its collection of books about French history and, perhaps, to the collection gifted by Madame Pompadour. Since it is most unlikely that a library will have all the books written about French history or necessarily even all the books written by the same author, its collections will be arbitrary in the very limited sense of being dependent on such things as acquisition and disposal policies, purchase budgets, and availability.

The allocation of category membership is far from straightforward. In collections of manuscripts there can often be doubt or uncertainty about the content of a document. The signature of the correspondent may be missing, the nature of the content may be ambiguous and so on. The same can be true of books. Where a book about French agriculture may in fact be as much about agricultural history, the librarian may, in a capricious moment, decide to shelve it under agriculture rather than history. The issues here are well rehearsed in Benoit.[41]

Cataloguers in the information professions are uncomfortable with such collections even though they regularly encounter them and prefer, for convenience, to allocate individual objects and the collections to which they belong to hierarchies with explicit taxonomies. The world of physical cataloguing reinforces this approach but it does not necessarily improve resource discovery. Take, for example, a collection of family papers belonging to the Duke of *X*. Accepted practice is to arrange the papers under the various Dukes of *X* and perhaps their duchesses with sub-divisions under each heading, such as political papers, military service, family correspondence, and so on. This is only helpful if the user knows to look amongst the papers of the Dukes of *X* to satisfy a search and it remains arbitrary since a letter on a political subject might include details of how to grow potatoes, and so on. Moreover most of the correspondence in the papers will not be of the Dukes of *X* at all but of those who corresponded with them. Their letters will invariably be held in other collections. As a result no matter the size of the collection, the papers of the 3rd Duke of *X* will always be incomplete. Although there can be no doubt that the 3rd Duke of *X* lived from such and such a date to such and such a date, documents ascribed to him at one moment in time may subsequently be discovered

40 This is precisely the reason given in the OAIS model.
41 Gerald Benoit, "Toward a Critical Theoretic Perspective in Information Systems," *Library Quarterly* 72, no. 4 (2002), pp. 441–71.

to have belonged to someone else. This problem is particularly acute in what Schellenberg called artificial collections where provenance can be obscure, the identity of the object uncertain, and the relationship of one object to another may simply reflect the caprice of the collector.[42] One has only to think of paintings in collections, which were acquired in good faith in the belief they were by Titian and are now described euphemistically as "after Titian" or "school of Titian."

In an attempt to overcome some of these problems, catalogues of objects are sometimes elaborately cross-referenced, with users employing their own catalogues and cross-references in the construction of their own "collections." We have only to think of the books and articles "collected" to write this article. The collection has a temporary fixity in that it is bound by this article but that is all. The digital environment makes the construction of such transactional sets of objects very easy. A good example is an Internet search that yields a set of hits that may or may not satisfy the subject of an enquiry and will vary from day to day, notwithstanding the use of identical search terms. The only way to make the search less arbitrary is to add metadata to objects describing their content.

Within accounting systems it is essential that every transaction is not only unambiguously recorded but properly allocated to different headings (collections, if you will). In the physical world this problem has been resolved by generating complex indices to transactions with accompanying summary information in the shape of journals, ledgers, and so on. This approach has been preserved in the digital realm by replacing this summary information with metadata attached to the individual objects. This permits collections of information to be generated at will providing the existing metadata rules allow for it. Gone are the ledgers and journals, only to be replaced by box or byte-full upon box or byte-full of receipts and invoices. Such an approach would be possible across the digital world and this is what the so-called "archival information package" of the OAIS model attempts to do, but the individual object would then become the core of a preservation system rather than the collection which would, then, only exist as a transaction. Although this is a library approach, the digital world in some senses equates to a library in that items enjoy an independence that is analogous to a book.[43] Where Web pages are created dynamically what the user views is collections of objects which only exist as a particular transaction in response to a particular request made at a particular instance. In most cases there is little point in preserving or recording that "collection" as it can be repeated providing the underlying datasets and

---

42 Schellenberg, *The Management of Archives*.

43 In other senses it equates more to an archive in that the items are unique objects rather than an "edition."

interfaces are preserved as a "collection" – the JISC-IE *modus operandi*.[44] Such an approach is less straightforward than it might seem; if the Web pages are delivered through a portal, they may be drawing, unbeknown to the user, on several collections which have different creators whose intentions and methodologies will, in all probability, vary. This problem can easily be seen from union library catalogues where the same book will appear variously described depending on individual cataloguing rules. In the electronic environment, much more than in the physical world, the accurate description of the provenance and history of the collection to which an individual object belongs is essential to guarantee integrity and authenticity and to avoid the dangers to which Buckland alludes.[45] Unlike with a physical collection, it is almost impossible to obtain a view of the whole collection and then to infer the rules of construction, which would presumably be catered for by the administrative history entities in the creator, producer, collector and, owner sections of the Heaney model, although not explicitly discussed therein.[46]

There is some merit in such an approach since one document (object) never relates simply to one subject or activity and has, as archivists (curators) remind us, to be understood in the context of other documents which themselves may relate to yet other activities. Holding documents as independent objects will require additional metadata but this will avoid the need in the physical world of filling in all the cross-references in ledgers or in catalogues. In such a model the function defines the set, and the morphology of the function the strength of the definition. For example, to board a plane we need strong rules to define which documents are required, ticket and passport or international driving license; but for some functions we do not need binding rules to define the set, for example, the members of an evening class on the molluscs that inhabit the seas of the Western Isles of Scotland. The significance of the function and the strength of the definition of the set is a matter for its owners which will inevitably involve some assessment of risk and utility. Such assessments may not be entirely rational but they will be intentional. In other words the relationship between the document and the function is at one removed. The committee does not define the minute, it defines the need for agendas, minutes, papers, and so on. The botanist does not define the herbarium but the need for a collection of types as a reference against which to identify further discoveries.

Having now established that collections are brought about through a combination of the collector's or curator's intentional states and a modicum of happenstance, it follows that they are dynamic, changing in response to individual

44 See footnote 1.
45 Buckland, *Redesigning Library Services*.
46 Heaney, *An Analytical Model of Collections*, pp. 12–13.

choice, moral, social, political and fiscal pressure, and occasionally, acts of God.

How Might We Progress?

Against the rather fluid background just outlined one might imagine that we are advocating a radical agenda to abandon the approach that has been built up over millennia to deal with the physical world[47]; on the contrary, there is an even more pressing need to create taxonomies and associated rules and standards to allocate names to objects or sets of objects for the simple reason that "privileging of the better and, by default, the non-privileging of the rest, remains a significant needed service."[48] However, as with all taxonomies, they will function merely as useful constructs with no guaranteed fixity over time. Some will be stronger than others but that does not of itself endow them with permanence. We will need to cater for this lack of fixity in dynamic systems by recording "histories" of functions and accompanying definitions of the supporting sets of documents independently of the objects and their accompanying metadata. It would be worthless to know that this is the correspondence of *X* or the papers of *Y*, if there is no means of knowing what *X* or *Y* did. We delude ourselves if we believe these descriptions have fixity as they themselves are subject to changes in cultural norms and perspective, and language use. Similarly it is no good knowing that this is the record of the documents which allowed *Z* to board a plane, if there is no evidence of the rules which defined the set of documents necessary at that time. Where documents are destroyed and only some are selected for preservation, the rules and the mechanism for legitimizing them must be preserved if for no other reason than to protect the custodian. We do this already in the paper world even if not explicitly but, like everything in the digital world, it has to be made more explicit. This, in turn, will facilitate further processing either automatically or manually.

The digital world places us all in very deep water and it is no wonder that the advocates of collection-level descriptions do not wish to ask too many questions about the ontology of collections. If we accept that all objects, whether in the physical or digital world, were created and survive as a result of some underlying intentionality, then we will more readily face up to the challenges of how the criteria for collecting are both devised and implemented, and how far they contribute to problematizing the preserved objects. If all objects can be defined as "texts" laden with cultural and political baggage rendering them ambiguous and of no more value than personal experience, then the water gets deeper and we should respond swiftly by

---

47 Benoit, "Toward A Critical Theoretic Perspective."
48 Jenkinson, *A Manual of Archive Administration*, "Supplement to Chapter 6," p. 6.

surrounding the object with processes which render its ontology unambiguous. In the physical world this was achieved in part by means of a "strong box" where access was controlled in a beneficial fiduciary fashion with, as far as was practically possible, an absence of any intention of serving the dominant discourse unless at one remove.[49] Objects of proven provenance – purporting to be what they said they were – were placed in the strong box and held there securely on behalf of the community or some third party. Although the criteria for deposit were mediated by the guardians of the box, criteria of this sort were on the whole determined by an interplay of the needs of their creators/collectors and of potential users, although this could be perverted as Buckland suggests.[50] Trust in the security of the box was fundamental to attracting deposit. A collector would be unlikely to give his collection of rare books to a library if he thought they might be stolen or despoiled. A person in authority would hardly be likely to hand over confidential papers if there was any possibility of unauthorized release or alteration. This is all much more difficult in the digital world and there has been too little discussion of the crucial fiduciary nature of the service provider. This may be because much of the thinking has been informed by the library profession where, given that the stock-in-trade is multiple copies, the problems are less acute. As we saw at the outset this is reflected in the role ascribed to the archivist in the OAIS model, but the more power archivists and curators have to allocate objects to collections or within collections, the more they are exposed to the strictures of Buckland[51] and Derrida.[52]

So, how far have we come? If collections are to be simply transient transactions, then they cease in any useful sense to be collections as we have known them in the physical world. This can apply as much to the physical as to the digital. It is just that the digital has made it so much easier to generate and abandon such collections in museums, libraries, and archives. The collection, which must continue to be described, is the bunch of objects bound together by a defined activity or institutional framework. As Schellenberg proposed, they may have different typologies and these need to be explored without the confusion of transactional collections.[53] Whatever the collection type the user in the digital world will need to have trust in their content. Much more than in the physical world, where failure in trust once suspected can be identified, in the digital world confidence in the collection being complete could easily evaporate unless processes are put in place that make such detection possible.

---

49 See for example W.E. Tate, *The Parish Chest, A Study of the Records of Parochial Administration in England* (Cambridge, 1969), pp. 35–42.
50 Buckland, *Redesigning Library Services*.
51 Ibid.
52 Derrida, *Archive Fever*.
53 Schellenberg, *The Management of Archives*.

Trust will extend, more explicitly than in the physical world, to the navigational devices provided as they will often be the only means of accessing the object, unlike a book which can be picked off a shelf. How entities are described in the metadata will govern the outcome of the generation of transactional sets and, as in the physical world, the entities will inevitably be culturally determined. In the United Kingdom we have trousers and braces; in North America, colleagues have pants and suspenders, but if the inquirer knows this then it will be reflected in the search strategy.[54]

54 For a lengthy discussion on this very issue, see Berman, *Prejudices and Antipathies*.