The devil is in the details: Describing born-digital records using the *Rules for Archival Description*

Paper for SAA Conference, New Orleans LA, August 2013, by Kat Timms

Slide 1: [Title]

Good morning. I'm very pleased to be here with you today to discuss some ideas of mine concerning describing born-digital archival records using the Canadian descriptive standard, the *Rules for Archival Description*.[1]

I say that "the devil is in the details" because when it comes to describing and managing born-digital records, there are a lot more details for which there are no counterparts in analogue records; I am thinking about metadata, but not exclusively. The fact is also that digital records require management at the item-level, which alone makes them a lot more detailed!

Slide2: Overview

This is a very complex topic that could fill a dissertation or book. Today, I will only be able to share some highlights of the issues as I see them with you.

I started thinking about this topic when I was involved with a project at work a few years back. The project was to transfer, ingest, preserve and make discoverable Government of Canada records originating from an electronic document and records management system (EDRMS). This involved a lot of cross-walking between various standards, and it gave me a chance to wrestle with RAD's Chapter 9 for describing Electronic Records.

So, today, I'm going to split my presentation into two sections. First, I will start by giving you a brief introduction to RAD and its issues which, by the way, extend far beyond those specific to digital records. Actually, RAD's problems is a topic that has gained some momentum in recent months, prompted by the publication of an article in *Archivaria* by fellow Canadian archivist Richard Dancy from Simon Fraser University Archives.[2] I'll introduce you to some of his analysis. He did not, however, address in detail many challenges affecting digital records. Following that, I will share with you a sample of some of the issues I observed while trying to use RAD Ch. 9 in my work project.

[1] Planning Committee on Descriptive Standards, Ottawa: Bureau of Canadian Archivists, 2008.Accessed at http://www.cdncouncilarchives.ca/archdesrules.html on 22 July 2013.

[2] "*RAD* Past, Present and Future", No. 74 (Fall 2012), 7-41.

The second half of my presentation will look beyond RAD into the broader standards landscape. It'll be about questions: which descriptive or metadata standards are needed to support archival management of born-digital records? How does one interpret and manage the various aspects of digital records (especially their metadata) as they enter the archives? What do archivists need to know vs. clients?

Slide 3: *Rules for Archival Description* – Introduction

RAD was first published in 1990 and was based on the framework of AACR2 and ISBD.[3] So, most of RAD's areas and elements of description, punctuation, division into media chapters and the idea of name headings as access points were derived from these models. This said, RAD differed in that it was founded on the archival principles of *respect des fonds*, multi-level description moving from the general to the specific, and the focus on context.

Two ICA standards, for describing records and for describing creators, were developed around the same time[4] but were structured very differently, arguably more in-tune with archival descriptive requirements.

A significant revision of RAD was undertaken about 10 years ago, an output of the Canadian-American Task Force on Archival Description (CUSTARD). While *Describing Archives: A Content Standard* (DACS) was published from those efforts, RAD2 was never fully implemented. Both DACS and RAD2 were more closely aligned with ICA standards, but due to a lack of feedback and consensus in Canada about RAD2, only minor revisions were implemented in 2008.

RAD stands alone in the Canadian context with respect to archival descriptive standards; it is intended to meet all archival descriptive needs. RAD's purpose is explained in its "Statement of Principles": it is to enable discovery, to provide a general understanding of the records through documentation, and to document physical possession/custody after transfer procedures to protect authenticity.

[3] Joint Steering Committee for Revision of AACR, Anglo-American Cataloguing Rules, 2nd ed., 2002 revision (Ottawa: Canadian Library Association, 2002). IFLA Cataloguing Section and ISBD Review Group, eds., ISBD: International Standard Bibliographic Description (Berlin: Walter de Gruyter GmbH, 2007).

[4] ICA Committee on Descriptive Standards, ISAD(G): General International Standard Archival Description, 2nd ed., 2000. ICA Committee on Descriptive Standards, ISAAR(CPF): International Standard Archival Authority Record for Corporate Bodies, Persons and Families, 2nd ed., 2004.

Slide 4: *Rules for Archival Description* – General Issues

So, what does Dancy have to say about what's wrong with RAD?

He discusses how RAD was built on a bibliographic model that has since moved on. FRBR and RDA involve conceptual models defining related entities.[5] Generally speaking in the information community, modeling entities and their relationships is now commonplace. Should not the Canadian archival community investigate and re-think how archives are being modeled, if at all, in the now-dated RAD?

An impetus to re-think RAD has been long-standing, in fact. Consider the ICA suite of standards, which has kept describing records separate from describing creators for several decades. In RAD, authority information describing records' creators is buried within administrative histories or biographical sketches. Further, much information is packed into RAD's "Notes" section, creating a type of catch-all area. Lastly, its organization into media chapters creates a fair amount of overlap and some inconsistency between chapters.

Dancy notes that it's a bit unclear what role RAD is trying to play – it's positioned as a content standard, but at times it also has information belonging to a data value standard (i.e., suggesting required terms, such as forms of material or GMDs).

Dancy also discusses blurring between description of intellectual and physical entities – for example, instructions for physical description of electronic records speak to both records and the carriers they are stored on without relating the two pieces of information together. There is also a lot of variance in how extent information can be formatted.

Slide 5: A Peek at Physical Description

To illustrate my point on mixing intellectual and physical description, here is a peek at the examples given in Chapter 9.

The first two lines show different ways of counting digital extent. Then there are examples for counting physical carriers. If there are more than three types, the information becomes invisible, designated as "other material". This is a carry-over from card catalog days, when there was limited physical space on a card; this is not relevant

[5] IFLA Study Group for Functional Requirements for Bibliographic Records, Functional Requirements for Bibliographic Records (Munich: K.G. Saur Verlag, 1998). Joint Steering Committee for Development of RDA (JSC), Resource Description and Access (Chicago: American Library Association, 2011).

in the digital environment. Following that, there are options to add other physical description details, if desirable.

How do I know **which** of the 6 photographs in the first line are held on a particular CD-ROM or disk? How can I reflect file format information in a heading like the second (6 GB)? Would I tally each format type?

So, this approach conflates entity types, buries information and forces me to eliminate information that I would like to record, but technically can't, according to this standard.

This slide illustrates a few issues:

- How can intellectual items be related to physical carriers? There is often not a 1:1 relationship, although this is the only option promoted by RAD.
- To what level of detail should extent information be provided and how can this be done consistently? Standardized data values could better support automation (e.g., tallying extents or dates up the hierarchy). Given the volume of content, any opportunity to automate is desirable to consider.

Slide 6: What's all the Fuss About? Looking at a Born-Digital Item

I thought it might be a good idea to discuss what all the fuss is about: how are digital records so complicated?

This diagram shows one of many born-digital records scenarios, records originating in an EDRMS. Records management, document management, version, agent and event metadata are structured within this system in relation to the digital objects or records to which they pertain.

All this metadata must be bundled with each digital object, and transferred to the archives, for example by physical carrier. Some of the metadata is pertinent to enabling discovery ("finding aids"), to enabling long-term care (preservation), or for managing access rights or substantiating the authenticity of the records.[6]

Another scenario not depicted here is digital records stored on a shared drive. While they lack metadata infrastructure surrounding them, the folder structure comprises significant information, and there is often metadata embedded in each digital object as well.

[6] The exercise to determine which metadata should be created and acquired to support these archival activities is out of scope of this presentation.

Further, digitized records (with an additional layer of technical metadata documenting the digitization process) could be managed within either an EDRMS or a shared drive. Is there a need to highlight or re-use some of this technical metadata in the archival custodial context? How well are the digitized materials intellectually related with their paper counterparts? Traditional archival descriptive standards are simply not granular enough.

Slide 7: Digital Reality = The Item

The previous slide illustrated how the item level is unavoidably a focal point when describing digital records.

On the other hand, there has been some focus on the opposite end of the spectrum in recent years – with "More Product, Less Process" advocating for more efficiently allocated resources for processing, including more collection-level vs. item-level processing.[7] Further, theorists such as Nesmith have called for broader contextual levels of description,[8] although Yeo cautions that there can be no limit to context.[9] While higher levels of description may be more difficult to definitively box-in or demarcate, the lower levels have become more visible thanks to digital realities. As the EDRMS diagram illustrated, items themselves are types of aggregates, as they are composed of various component parts, such as the objects themselves and their metadata taken from different related tables in a database.

These components, and how they must be archived, consist of bit streams (the binary 1's and 0's), file formats (rules for interpreting the bit stream), and different representations over time (i.e., migrations to new file formats).

Minute details that are unknowable in the paper world are available in the digital world (e.g., event logs). Because this information is present, it cannot be ignored; its value has to be assessed. Some of it may help substantiate certain aspects of "recordness" (i.e., authenticity) as part of the record and its immediate context. Faced with what is largely a sea of metadata as well as technical challenges, archivists need to know what they are dealing with, what metadata has value for what purpose, and how to manage

[7]Mark A. Greene and Dennis Meissner, "More Product, Less Process: Revamping Traditional Archival Processing", American Archivist, (Vol. 68, Fall/Winter 2005), 208-263.

[8]"Reopening Archives: Bringing New Contextualities into Archival Theory and Practice", Archivaria, (Vo. 60, Fall 2005), 259-274.

[9]Geoffrey Yeo, "Debates About Description", in Currents of Archival Thinking, Terry Eastwood and Heather MacNeil, eds., Westport CT: Libraries Unlimited, 2010, p. 98.

digital records (the objects and their metadata) over the long-term, as well as rendering them accessible.

A good start for making sense of this complexity would be to develop a way to view all the component parts as one related whole: devise a conceptual model for archival records, including at the item-level. Dancy called for this, and conceptual modeling of archives is on others' radar, too. Archives New Zealand introduced their approach to modeling archival items at the 2012 ICA Congress.[10] Also, ICA's Experts Group on Archival Description has announced their intent to develop a conceptual model for archives in the next four years.[11]

Slide 8: *Rules for Archival Description* – Chapter 9: Records in Electronic Form

The peek at physical description illuminated some of the issues I see with Chapter 9. To take a step back, some might question the need for separate rules for digital records at all. Textual records, graphic records, sound and moving image records can all be in digital or analogue form, so why have a separate chapter? Having worked with digital records, I maintain that there is a need for separate models, rules, and instructions for managing them, intellectually and physically.

I cannot provide my full analysis on Chapter 9 and the broader issues it raises in this short talk, but I will highlight three particular points.

First, thinking of the EDRMS diagram, how does one interpret "date of creation", a staple descriptive element, in this context? The document's metadata profile is a separate entity related to each distinct document version. Digitized materials are "created" in a sense when they are scanned. Is the date of creation of this physical surrogate relevant along with the date of creation of the original analogue record? Each virtual folder is a "thing" that is created and has a discernible date of creation. So, there are, in fact, many dates of creation, one for each component of a single digital item. RAD's Chapter 9 does not provide any insight or answers on this issue. It does not deconstruct an item in this way and theoretical clarification of the concept of "date of creation" is missing. Perhaps theoretical explanations belong outside of RAD, but RAD

[10] Hywel Gwynn Williams and Anna Henry, "Building a Digital Archive: Integrating Theory and Implementation", paper presented at International Council on Archives Congress 2012, Brisbane, Australia, 22 August 2012. Accessed at http://www.ica2012.com/files/pdf/Full%20papers%20upload/ica12Final00280.pdf on 22 July 2012.

[11] "EGAD Activities and Projects: EGAD Strategic Work Plan", accessed at http://www.ica.org/13845/egad-activities-and-projects/egad-strategic-work-plan.html on 22 July 2013.

should provide adequate rules or implementation guidance for the situations it is intended to addresses.

Second, RAD tips its hat to physical management, including conservation treatment. If file format migration were considered as a conservation activity, metadata for this should be captured at the item-level. Further, information about file-formats and digitization such as one would find in the NISO data dictionary for digital still images[12] is not comparable to anything in RAD. There is not enough granularity in RAD to document either of these activities sufficiently.

Third, regarding arrangement, in addition to needing clarification on how to relate items and their component parts, there is a need for clearer instruction for describing different groupings of aggregates. In the current scheme, there is the item, file, series, and *fonds* or collection. This does not reflect digital realities such as items being comprised of component pieces, or of (1:many) relationships between a file classification and a folder structure, (i.e., thinking of folder arrangements in shared drives or EDRM systems). If RAD speaks to archival description, it pre-supposes that archival arrangement rules have been identified, which is not the case. This is clearly a gap that leads to confusion, inconsistency and idiosyncrasy when processing digital records comprised of a web of interconnected parts, with more levels of aggregation than are reflected in the traditional four-level hierarchy.

Slide 9: What Do We Need?

It's clear that RAD alone cannot illuminate all aspects of archival management of digital records, but it's also clear that it's insufficient in guiding the areas that it attempts to address – mainly intellectual management, but with aspects of physical management intermixed.

Various efforts in the archival, recordkeeping, digital curation, and bibliographic communities seem to be saying the same thing: devise a conceptual model that includes an analysis of the item-level, and identify and codify relationships between archival components.

These different communities have produced a body of standards that can be used together to cover all the required bases – for example, PREMIS for digital preservation

[12]NISO, Data Dictionary – Technical Metadata for Digital Still Images. ANSI/NISO Z39.87-2006 (R2011). Accessed at http://www.niso.org/apps/group_public/download.php/6502/Data%20Dictionary%20-%20Technical%20Metadata%20for%20Digital%20Still%20Images.pdf on 22 July 2013.

management,[13] and EAD for structured finding aids.[14] This doesn't negate the need, however, to revise standards such as RAD that fall short of their intended targets, including providing instructions for describing digital archival material.

Understanding digital records is still an evolving process as well. Theory and practice need to advance so that we can more consistently, and with stronger conviction, manage these records over time in a standardized way.

For government records, too, there needs to be a solid connection between active and archival records; archival standards for description, at least at the lower-levels of aggregation, should not differ from recordkeeping metadata devised and created upstream. Records continuum thinking needs to be more deeply embedded, ensuring that the infrastructure is built to support more re-use and channeling of existing metadata along with ensuring that digital objects are initially created in sustainable formats with long-term preservation in mind.

Slide 10: What's Out There?

The good news is that we are hardly starting from scratch. Various parties in the archival community have either started or have voiced the need to develop conceptual models for archives.

A focus on the item level in the bibliographic community is something that archives are leaning on in their use of Dublin Core or MODS for digitized items,[15] and the NISO data dictionary for still digital images can also be useful in this area with respect to technical metadata. PREMIS, too, is geared toward item-level management.

I think the point here is that many of the issues are shared among communities: archivists, records managers, digital curators, and librarians. Working together on common problems is the most efficient way of moving forward. But where does this leave RAD?

Slide 11: Who Needs to Know What?

[13]PREMIS Editorial Committee, PREMIS Data Dictionary for Preservation Metadata, v.2.2, July 2012. Accessed at http://www.loc.gov/standards/premis/v2/premis-2-2.pdf on 22 July 2013.

[14]Library of Congress, Encoded Archival Description (EAD), Version 2002 Official Site, accessed at http://www.loc.gov/ead/ on 22 July 2013.

[15]Kelsey Shepherd and Jenn Riley, "A Brave New World: Archivists and Shareable Descriptive Metadata", American Archivist, (Vol. 72, No. 1, Spring/Summer 2009), 91-112.

Perhaps a revised RAD has no place in speaking to describing lower-level(s) of archival records, particularly the item. Following the creation of a conceptual model for archives, perhaps a framework of standards should be codified, each standard with its particular purpose and place.

New RAD in this world could be geared toward describing records in accordance with the broad conceptual model, and likely this would flow from the suite of ICA standards for records, creators, functions and repositories.

As digital archivists, though, we'll still need to know how to manage and describe born-digital content, and for some of us, greater collaboration with records managers and systems designers will be necessary. We must build our systems acknowledging the reality that upstream systems feed into archival systems, including digital preservation systems.

New RAD should define how it should be interpreted and applied within the conceptual model and the framework of archival standards, and companion guidelines or tools may be required.

There is a need for ongoing theoretical engagement, too, to be able to better understand and interpret complicated record structures found in database and business systems to better understand what to acquire and how to preserve it. I think that we need more explicit guidance with respect to documenting the physical aspects of collections separate from but related to their intellectual contents. We need to conceptually deconstruct digital records while keeping them in context.

In addition to our own curatorial needs, we need to consider what our clients need and want to know about our collections and how to better enable access to them. Some archival theorists and authors, such as Nesmith, are calling for more high-level contextual description. Others, such as McNeil are calling for more transparency regarding custodial actions in public descriptions.[16] On the whole, what do users want to know about records before accessing them? Supplying a complete metadata record for a digital object, including PREMIS metadata, could be considered as being transparent, but would a user be interested?

How we form our standards has two purposes: facing inwards (archival management) and facing outwards (serving clients). I think this issue needs to be considered, too,

[16]"Picking Our Text: Archival Description, Authenticity, and the Archivist as Editor", American Archivist (Vol. 68, Fall/Winter 2005), 272.

when working on the new conceptual model and using the framework of standards, as well as when revising standards like RAD. Who needs to know what, and for what purpose?

Slide 12: Final Thoughts

I hope my presentation has gotten you thinking about what we need to manage and make accessible our born-digital archival holdings. I think the community is moving in the right direction with respect to developing a conceptual model, and I think that a standards framework would be a logical companion to that. I think that there is a need for separate guidance for digital records, but I think it needs to be very different than what we currently find in Chapter 9 of RAD.

I also think that RAD has a future, in a revised form. Its place may no longer be to address digital records specifically (at a lower level), but there remains a need for an over-arching content standard in Canada that better aligns with a modern understanding of archival entities, their component parts, and their relationships, along with being geared toward current systems environments and tools.

Slide 13: Thank you!