

To: Joint Steering Committee for Development of RDA
From: Dave Reser, LC Representative
Subject: Machine-actionable Data Elements for Measurements, Extent of the Carrier, Pagination and Foliation, Dimensions, Extent of the Content, and Duration-Discussion Paper (2015)

Thanks to ALA for the extensive paper on machine-actionable data elements; it is clear that significant thought and analysis has been invested in this topic. Our response is divided into **General Comments**, answers to the **Questions** asked in the paper, and **Specific Comments/Questions** raised by our reading of the paper.

General Comments

As noted in responses to the earlier papers on machine-actionable data elements, the possibility of breaking elements up into machine-actionable sub-elements needs to be weighed against the practicality of doing so. We appreciate that this paper provides a more extensive illustration of what the approach entails in the context of the RDA instructions. As the task group notes, however, the approach “significantly complicates the instructions and examples.” The additional complexity may be better suited to some elements/aspects addressed in the paper than others. For example, it is easier to imagine a machine-actionable need for the measurement techniques applied to attributes such as dimensions and duration than for extent of the carrier and extent of the content. We can imagine use cases such as “retrieve all volumes smaller or wider than 10 centimetres,” or “find all recorded operas lasting 30 minutes,” but have trouble imaging use cases for the extent attributes (find all the carriers that come in 3 units? find all the resources with 12 maps?). Some of the complexity shown by the draft may be more related to technical metadata, or local inventory management.

The task group was correct to “anticipate a fairly long and fraught transition from current practices and methods”. For this reason, we suggest that the JSC proceed with the revisions related to 6JSC/ALA/40 (Revision to RDA 3.1.4, Resources Consisting of More than one Carrier Type), rather than postponing those revisions until the issues in this discussion paper are resolved.

Questions

1. Should RDA contain a super-property for *Measurement* which may apply to all WEMI entities?

Response: We don't think we know enough at this point in time to reach this conclusion. The impact of the new entities in FRBR-LRM, and how they would be actualized in the instruction set is not clear to us, so developing attributes of *Res* seems premature. We are unclear how *Measurement* would be applied to works, or if

that is even a necessity for the solution proposed. The Technical Working Group would be best situated to advise on this recommendation.

a) Should RDA contain a two-path approach, for both machine- and human-generated data?

Response: We're not sure we saw evidence of a machine-generated approach in the discussion paper. The script that was mentioned for calculating extent of a text seems interesting, but to successfully derive a "true extent" would require that all sequences (including "inessential matter" at 3.4.5.3.2, blank pages, advertising, title pages, fly leaves, etc.) would have been recorded in a string, presumably by a human. In other words, most extent statements derived from the current RDA text could not be used to programmatically calculate subunits of carrier for volumes.

If the question was really about machine- and human *actionable* rather than *generated*, then yes, having the two-path approach was quite helpful. It would allow for the additional complexity when use cases, input systems, retrieval techniques warrant that approach, but literal strings may be perfectly adequate in many circumstances.

2. Should the instructions for *Extent (of the Carrier)* be refocused to treat volumes and their subunits in the same manner as other carriers?

Response: There is some value to treating the *carrier extent unit* for volumes to differently than the current instructions if pagination and foliation are retained. The treatment of carrier extent *subunits* of volumes seemed confusing to us (see comments below), and we're not sure that the proposed treatment is an actual improvement over current practice even if it represents a "purer" conceptual approach. Communities that apply the other exceptions to 3.4 for extent of carrier (e.g., cartographic, notated music, still images, three-dimensional objects) would also need to see the refocusing as an adequate replacement.

3. Should a separate set of instructions be developed for *Pagination and Foliation*? If so, should they be placed in Chapter 2 or Chapter 3?

Response: If the proposed solution for carrier extent subunits is used, then yes, instructions for pagination and foliation would be necessary, and frankly, expected in descriptions. While Chapter 2 may be more likely to be used for "identification" attributes over Chapter 3's "selection," we think a similar argument could be made for attributes such as Dimensions (more likely to be used for identification than selection). It seems to behave almost identical to extent, and would be fine with leaving it in Chapter 3.

4. For recording the dimensions of still images and cartographic resources, should the distinction between the dimensions of the sheet and the pictorial area be made using values for *Part Measured*? Or should the dimensions of the pictorial area be treated as an attributed of content (in Chapter 7)?

Response: It is easy to imagine measurements of content in Chapter 7, where other attributes exist to indicate the relative size of content (e.g., reduction ration, aspect

ratio, scale), but we can also see the argument for keeping instructions relating the size of the content and the size of the carrier together for ease of use of the instructions by catalogers. There is another similar instruction relating size of the content to carrier at 3.5.1.4.14 (recording the size of the text block in relation to the size of the volume), and there may be other instances of this lurking in Chapter 3. It has been noted by the JSC several times that some examples in Chapter 7 relate supplementary content (e.g., a bibliography in a book) at an expression level to the carrier of a particular manifestation (e.g., the pagination/foliation of a particular manifestation) to help users understand the extent of the supplementary. While this may not pass a “purity” test, the end user is often interested in the “extent” of such content. This issue was also relevant to some of the duration examples changed by the task group. There needs to be some way within RDA to record such information about where “expression” content may be found in a particular manifestation. If this does not happen in Chapter 7, then the concept needs to be accommodated in Chapter 2 (for example, “Colour map of Europe on page 5” or “Bibliography: pages 500-549”).

5. When should *Extent of the Content* be core?

Response: We don’t believe it should be core. Some communities are likely to develop core practices (e.g., for the types of resources treated under the exceptions at 3.4 now).

a) Does it make sense to move Units and Sets of Units with Identical Content (currently RDA 3.4.1.6) to Chapter 7?

Response: Although the concept is inherited from AACR2, it is probably already given more attention than it deserves in the RDA instructions. It has nothing to do with the actual extent, and is only about how a user may use the resource. We think this could be limited to a Note on extent of manifestation.

b) Should a term other than “item” be used in these instructions? If so, what should it be?

Response: It would not be good to use “item” in a manner different than its WEMI definition (our local system uses “item” in a manner suggested by this paper, which led to considerable difficulty of some to grasp the FRBR definition of “item”). Possible replacement is component.

6. Should the instructions for *Duration* make a distinction between the actual duration (for resources with a temporal dimension) and the intended duration (for resources with a stated performance time)?

Response: Lacking any discussion in the paper of the difference between duration and intended duration, or definitions that would illustrate the distinction, we would say “no” at this time. Currently, duration is duration; if the resource gives a duration known to be untrue, this fact is recorded as a “details of duration” at 7.22.1.5. We’re not sure if the paper had some other concept identified as “intended duration.”

7. Is the syntax used for sub-elements in the examples acceptable? Is there a better way to present the examples?

Response: After consultation with the Examples Editor, we have the following thoughts and comments about the examples.

Since ALA explained that the “machine-actionable” examples are not designed to reflect a form displayed to a user, we assume that the ones labelled “as a string” represent both a potential display for public use and a more “human-readable form.” We suggest that the examples boxes could be re-labelled to better reflect that using labels like “*Machine-actionable form*” and “*Display form.*” In order to do that, explanatory text needs to be added to the instructions for those phrases so the examples syntax matches the instructions.

However, the larger issue is that the examples do not follow the instructions as written. For example, according to 3.4.1.3, a “string” should contain the values for the measurement sub-elements from x.y. This means the “string” examples in 3.4.1.3 should contain measurement type, measurement unit, and measurement quantity like this: “carrier extent units, 1 microfilm cassette.” If the intent is for the measurement type not to be part of a “string” then the instructions should say that. Because these examples must be worked into an existing RDA format, there should not be labels like “MEASUREMENT TYPE:” in the examples themselves. Such information is conveyed only in the example explanations. Thus, the appropriate syntax would be:

```

carrier extent units
microfilm cassette
1
Measurement type, measurement unit, and measurement quantity for a microfilm
cassette

```

In addition, the examples should reflect the application of the instructions, and only the application of the instructions in which they are located. For example, there are no instructions on recording measurement type at 3.4.1.5-3.4.1.6, so the examples should not contain measurement type.

8. How great is the concern about the ability to migrate legacy data to the new structure?

Response: This is a significant concern, but one typical for structure standards. We would invite additional discussion on this topic.

Specific Comments/Questions

Section 1: Measurements

The Measurements attribute becomes essentially a “statement” in RDA parlance, whether that statement is made up of the sub-elements for measurement, or those sub-elements expressed as a literal string. We are unsure of how to address any “core” statements in that regard, and question whether there should be anything core in the Measurement

instructions at all, since the “core-ness” in RDA is actually attached to the element that is expressed using Measurement techniques. The “core” statements should also be considered in relation to the Cardinality concept in 0.6.3 (have I satisfied the core requirement with the first measurement attribute vs. subsequent instances?).

We’re somewhat concerned that the element “measurement unit” includes two very different concepts (standard of measurement vs. vocabulary terms for carrier or content terms). Perhaps these should be separated.

The complexity added by “part measured” and “measurement qualifier” do not seem to be warranted. Could the clarifications offered by these sub-elements be better expressed as Notes on/Details of [the element being measured]? We can imagine that “part measured” would only be useful if it were always recorded, otherwise an assumption would need to be made by machines.

Section 2 : Extent of the Carrier

(3.4.1.1) Scope of Extent of the carrier: we assume that “... measurement of the number and type of carrier ...” should be “...measurement consisting of the number and type of carrier...”.

(3.4.1.3) The instructions for measurement type say to record **either** the carrier extent unit **or** the carrier extent subunits, but examples presented show both, and there is an instruction to “specify the number of subunits, if applicable” later in the same instruction implying that both are allowed. This needs clarification.

(3.4.1.5) If “pieces” would always need to be qualified by “various”, then it would be far simpler to indicate that the measurement unit is “various pieces.”

(3.4.1.7) Not sure the logic is right vis a vis the replacement wording on formats that parallel another type of resource. This will require further study.

(3.4.1.7.1) Following the principles of separating carrier and content, it would seem that the measurement units of “audio files, video files, data files” seem more related to content, and are the terms specified at 3.19.2.3. The presence in both places needs to be evaluated.

(3.4.1.7.2-3.4.1.10) The instructions merely say to specify the number of something; we believe the intent is to “record the set of measurement sub-elements for a [flipchart, etc.]”. As indicated in the introduction, a vocabulary of carrier extent subunits still needs to be developed.

(3.4.1.7.11) The instructions for recording the subunits for Volumes was difficult for us to understand. Those who didn’t know in advance (from a CC:DA discussion at ALA Midwinter) that it was only supposed to represent the exact true extent of the subunits did not notice that the instruction was any different than the current practice. Even the true

intent (expressed on page 4) does not play out in the instructions—if the intent is to record simply the carrier extent subunits without regard to whether there is numbering, blank or printed, in a sequence or not, a plate vs. page or leaf, then the instruction should merely be to record the number/measurement quantity of leaves in the volume and a term for the measurement unit (that would have to be “leaves” by default). Every leaf is by definition 2 pages, so if we just care about the number of pieces of paper bound in the book without regard to the content on them, leaves seems to be the only applicable term. Furthermore, the concept of folded leaves does not represent extent—it is merely a leaf. Indication that some or all leaves are folded is not a measurement qualifier because it doesn’t modify the number. The fact of folding would be a matter for a note on extent, pagination and foliation, or perhaps an expansion of the Book Format element which covers other ‘folding’ issues for volumes.

(3.4.1.9) The concept of “each” really isn’t appropriate to simply stating the extent, and seems to add an unnecessary complexity even for machine-readability. The fact that each unit contains the same number of subunits should be a “note on/details of” treatment, if mentioned at all.

(3.4.1.10) The examples showing *not yet complete* or *unknown*, do not follow the instruction to record those terms as measurement quantity. The concept of “loose-leaf” is not a measurement qualifier; it perhaps needs to be incorporated into a re-developed “Book Format” element.

Section 3: Pagination and Foliation

(z.1.1) The scope needs to indicate that more than numbers are included—the terms for the type of unit (leaves, pages, plates). The final paragraph (reference to 3.4) should make clear that it is for extent of carrier, and refer only to volumes (sheets do not have carrier subunits other than panels, portfolios and cases only have carrier subunits that are leaves or sheets).

(z.1.2) The sources of information should have another sentence: “Take additional evidence from any source.” This way you can get pagination information from a bibliography that describes the manifestation if your resource is imperfect.

(z.1.4) RDA editorial style issue: the references seem to be to non-existent instructions called “Single Unit” and “Multiple Units”; this can be resolved editorially (not shown here).

(z.1.4.1) The instructions do not indicate that the number recorded is for the last numbered leaf or page of the sequence (this is covered for leaves or pages of plates at z.1.4.8.1). The information about “loose-leaf” is not a Pagination and Foliation issue but a Book Format issue (if that element is further developed). If Pagination and Foliation becomes more of an identification element, it may be useful to transcribe the words used as numbering rather than to convert them to numeric equivalents (e.g., thirty-two pages rather than 32 pages).

(z.1.4.2) If the resource is completely unnumbered, then the term “unpaged” should be supplemented by “unfoliated” to reflect that the leaves are printed on one or both sides as is done in DCRM(B).

Section 4: Dimensions

The use of the Measurement technique is useful to express the particular dimension being measured (e.g., height, width, depth, etc.). This could also be part of the “string” representation of the dimensions (e.g., 3 cm high x 5 cm wide). This is an internationalization issue. There may be many places in the world where height x width isn’t the usual order of measurements so recording “3 x 5 cm” may not be understandable. The task group suggested changing the order of the dimension aspects for some units, but clearly expressing what is being measured even in a string may be a better substitute.