I. Preliminaries

A. Contents

I. PRELIMINARIES

A. Contents

II. GENERAL STANDARDS

A. Delivery Standards

B. Returning Files

C. Image File Standards

D. File naming conventions

E. Content to be keyboarded

F. Advertisements

III. DOCUMENT STRUCTURE

A. Basic Document Structure

B. The TEI Header + Front Template

C. General guidelines for structure in the body

IV. MARK-UP DIRECTION/CORRECTION OF SPECIFIC ELEMENTS

A. Ambiguities in Keyboarding and Tagging

  1. Foreign Language Tags
NOTES FOR DDD ON KEYBOARDING ERRORS

1. Content still mistaken as being in <p> tags with <hi rend=bold> instead of <head> tag: .................. 37
2. Separate articles often not identified as such: .................................................................................. 37
4. Paragraph tags—almost always in advertisements—are being used incorrectly: ......................... 37
5. Long dashes often replace hyphens in the original page-image: ...................................................... 38
6. End-of-line hyphens breaking words: ................................................................................................. 38
7. Designated content not being keyboarded: ......................................................................................... 38
8. No spaces before/after tags: ................................................................................................................ 38
9. Periods and commas substituted for each other: ................................................................................. 38
10. Replacement of text for what is represented in the original: ............................................................ 38
11. Entities next to number should reflect spacing of original: .............................................................. 38

V. FORMATTING TEXT: ....................................................................................................................... 33
1. Unicode Entities: ................................................................................................................................. 33
2. Standard keyboard punctuation: ......................................................................................................... 34
Unicode Code Charts: ........................................................................................................................... 34
3. Attributes and Values: ......................................................................................................................... 35
4. Consistency of mark-up: ..................................................................................................................... 35
5. Human- and machine readability of files: .......................................................................................... 35

VI. HEAD TAGS: .................................................................................................................................... 15
1. Headlines within paragraphs: ............................................................................................................... 15
2. Head Tag association / new divs: ......................................................................................................... 16

C. MILESTONE TAGS, PAGE-BREAK AND PAGE IMAGE REFERENCES: .......................................... 18

D. ADVERTISEMENTS: .......................................................................................................................... 19
1. Printer’s marks: .................................................................................................................................... 19
2. Advertisements as signed communication: .......................................................................................... 21

E. REFERENCED AND QUOTED MATERIALS: ..................................................................................... 21
1. Opener, Closer, Dateline, Salute, Signed and Address tags: ............................................................... 21
2. Cit, Q and Bibl tags: .......................................................................................................................... 25
3. Reference Notes: ............................................................................................................................... 28

F. TABLES, LISTS AND LINE GROUPS: ................................................................................................ 29
1. Tables: ................................................................................................................................................. 29
2. Lists: ................................................................................................................................................... 30
3. Line Groups: ...................................................................................................................................... 32

B. HEAD TAGS: ...................................................................................................................................... 15
1. Headlines within paragraphs: ............................................................................................................... 15
2. Head Tag association / new divs: ......................................................................................................... 16
B. Purpose

- To repeat information from “Specs_for_Digital_Divide.doc” that and earlier versions of this document that still apply and indicate changes from the specs listed in prior versions
- To call attention to changes we would like to see from the mark-up reflected in the three XML files initially sent to UR and to anticipate correction of errors in future files.
  - rtd18_109_1860_11_06.xml
  - rtd18_109_1860_11_07.xml
  - rtd18_109_1860_11_09.xml

C. Acknowledgements

This document was created originally for the Civil War newspaper digitization project, a cooperative effort between the Perseus Project at Tufts University, and Information Services and the Library at the University of Richmond, made possible by a grant from the Institute for Museum and Library Services (IMLS). This document is based on the “Text Encoding Guidelines for Keyboarding Vendors” created by DLPS Digital Library Production Services (DLPS) at the University of Virginia, and is used as a source by permission of the director, Melinda Baumann. While several specific elements have been retained from the DLPS document, many specifications reflect industry-wide standards and best practices, while other elements, as well as the structure and organization of the document, have been amended for the specific purposes of the University of Richmond and the Perseus Project.

D. Deprecations / Changed Specifications

**Deprecations** are standards that have been promulgated in past editions of this document or which have otherwise been communicated to the keyboarding vendor, but which UR has decided to eliminate. This is distinguished from **Changes**, which retain the element or entity but with modifications. **Additions** are new instructions for tagging.

Deprecations will be listed at the top of the standards document (the present section); descriptions of the aspect of the markup to be eliminated in this document will be retained, but those sections will appear with a light blue background, preceded by the text [**Deprecated:**]. Keyboarding examples used in this document inevitably include examples of other elements and entities than the one being discussed in a given section. In all sections throughout this document, deprecated standards will be changed and such changes will not be noted; deprecated standards will only be noted in those sections which have as their subject the standard in question.

deprecations:
1. `<abbr>` tags (see below, IV-A-2)
2. `<date>` tags (with 3 important exceptions: see below, IV-E-1)
3. additional date, time and name elements formerly specified in this section are withdrawn
4. `id=oclc` attributes of `bibl` tags (see below, IV-E-2)

changes:
1. the only XML entities allowed are 4-digit entities used for the docbook standard (see below, V-2); this is changed from allowing DDD to specify the entities to use.
2. `<bibl>` tags should now take attribute of `n="value"` (see below, IV-E-2)

additions:
1. printers marks, almost exclusively in advertisements, take their own tag: `<term>`
2. advertisements: should take `<signed>` tags; printers notes should always be placed on the last line of a div containing an advertisement, and take the (see below, IV-D)
3. returning files, (see below, II-B)

II. General Standards

All files for the University of Richmond digital library repository are prepared as follows:

- **Transcription**: The electronic text should contain an exact character-for-character transcription of the text of the print source, by industry standard double-keyboarding, with industry-standard computer file analysis, comparison and correction and spell-checking. The industry standard keyboarding error rate is 1 in 10,000.
- **Mark-up**: in XML, according to TEI lite, P4 specifications, and further specifications detailed in this document. The root element should not be repeated in child elements.
- **Entities**: Unicode entities are to be used in most cases (i.e., `&amp;` becomes `&amp`)—see below, “V. Formatting Text” for details; also see changes).
- Files saved with UNIX line-endings (to batch process files from Mac to UNIX line endings, use the command: `perl -pi -e 's/ ?/ /g' *.xml`).
- Lines should be no longer than 100 characters (with the exception of headers and table rows, and entries in lists, which should extend as far as necessary to fit on one line)—see below, “V. Formatting Text” for details.
- Consistency of mark-up in tag usage

A. Delivery Standards

- **Delivery of text files**: text files may be delivered via FTP in compressed archives
- Spreadsheets (Excel) must accompany every delivery of text files as part of the archive; a consistent naming convention for these spreadsheets must be implemented; for DDD, the naming convention for spreadsheets should follow this syntax:

Note that the date elements should be separated by dashes, while all other elements are separated by underscores. Also note that the file name indicates how many files are in a given archive: “file_20” indicates that this archive in question includes 20 XML files.

• These spreadsheets should include a total character count, the figure used to account for the standard Header+Front template provided by UR and appended to each XML file, and the total character count, less the characters from the template. Filenames of each file should be included, and the date the file (batch) was delivered to UR should be indicated.

• Invoices should be delivered on a monthly basis via email (see the most recent SOW) and should be accompanied by a spreadsheet that is simply a compilation of all the separate spreadsheets that accompanied each shipment of XML text files for the month (specifically, for the period and deliveries which the invoice covers). The convention is:

UR_monthly_2004_AUG_file_70.xls

• **Delivery of images:** tiff image files should be delivered on DVDs for archival storage; the DVDs must be accessible on multiple platforms (*i.e.*, Macintosh and Windows) and so the disk should be encoded in the **ISO 9660** format. In theory, however, there is no reason that DVDs encoded in UDF would have been unreadable on OS X.

**NOTE:** In the future, the UDF (Universal Disk Format) 2.01 for DVDs, based on ISO 13346, should be adopted as the standard. Because of problems with DVDs from DDD, we have settled on the ISO 9660 CD format for the present project.

### B. Returning Files

Files not meeting *at least* (but not limited to) the specifications set out in the original version of this document will be returned to DDD for prompt correction. Issues that have arisen include:

• Files with lines significantly exceeding the 100-character limit
• Files using `<p>` tags instead of `<div>` tags
• Files with erroneous or missing content, image declarations, etc.
• Files without keyboarding of “servants for sale or hire” content
• Files with words broken by hyphens (end-of-line hyphenation)
• Files having to be sent back again, due to TEI Header and/or other content being over-written or discarded by DDD.

Owing to the number of files needing to be returned, and in anticipation of needing to return more files for correction, a new procedure must be established for exchanging files:
1. UR will place files being returned on a password-protected web site for download, and inform DDD by email. The site is accessible from the link under Vendor Information at: http://oncampus.richmond.edu/academics/library/digital/reports.htm.

2. The page will include links to a directory from which XML files being returned may be downloaded; it will also include links to the most recent version of the following files:

   - imls_xml_filetracker.xls—corrected list of all files in vol/issue order
   - ddd_batch_deliveries.xls—list of all files in order received from DDD
   - ddd_returned_files.xls—list of all files to be returned, and their status
   - Specs_for_DDD_draft_x.doc—this document

On a separate but critical Q-A issue, the last three months of shipments have included duplicated filenames of those in previous shipments. In some cases these were second instances of original keyboarding of identical content, while in others the duplication was the result of duplication of issue/volume numbers in the original newspaper, but representing different content (from a different date). In either case, it is unacceptable to submit duplicated filenames to UR. Before a shipment of files is posted on the FTP server, all filenames in a shipment MUST be checked against the master list in the imls_xml_filetracker.xls spreadsheet. Any duplicate filenames in this list have already been reconciled, and reflects the 800 files received from DDD by the end of March, 2005.

3. When returned files are downloaded, DDD will send an email with the subject line:

   **UR xml files downloaded: number, date** (number of files, actual date filled in)

   The email body must also list all the files downloaded.

4. DDD will set up a separate directory on its server exclusively for returns called /returns. Any files/archives currently in the /output directory that were returned files will be moved into the /returns directory. As per instructions in the email of 23/2/05, filenames of archives of returned files should have the following syntax:

   **2005_01_28_DDD_returns_4**

   specifying the date of the posting, and the number of files in the archive.

5. DDD will implement changes in the version of the file that is returned by UR when making corrections. It will NOT generate corrections from any other version of the file. The replacement of the TEI Header, or of making changes from a previous version held by DDD is specifically forbidden. This was directed in an email of 2/8/2005. Any file we open to correct will have at least some changes made by UR to the file. These are lost if DDD works from any other version than the one being returned to them. Several files already returned once to DDD are being returned again to DDD because at least TEI Header information was over-written or eliminated. Changes on these files will have to be re-done, from the versions of the files shipped to DDD by UR.
6. All files sent back from UR from this point forward will have some information in the last section of the TEI Header, under the <change> section. This information must be preserved (as in fact any and all changes made by UR are to be preserved when sent back to DDD for correction). Under <item> in this section will be a brief description of the problem(s) for which the file is being returned for correction. Here is an example:

```
<change>
<date>March, 2005</date>
<respStmt>
<resp>Corrector</resp>
<name>Karen Owens, of the University of Richmond Library</name>
</respStmt>
=item>Text validation, generation of authority lists, final tagging.
There are two parts in file that have p tags instead of div tags.
Re-processed and re-corrected by DDD.</item>
</change>
</revisionDesc>
</teiHeader>
```

7. For all files returned to DDD for correction, the following text should be added into the TEI Header (highlighted above) in the <change> section, on a separate line:

```
Re-processed and re-corrected by DDD.
```

To identify files that have been revised by DDD, and distinguish them from any other versions, and certify that changes have been made.

C. Image File Standards

UR makes a distinction between its general standards for archival images, and standards for page-images used for keyboarding. The general standard for archival images is:

- 24-bit color, shot at 600 dpi, tiff images

For projects generating page-images for keyboarding, unless otherwise specified, the standard is:

- 8 bit bi-tonal (black and white) tiff images, shot at 400 dpi

The IMLS Grant project being keyboarded by DDD for UR falls under the latter specification. The rationale for reducing the image quality from 600 to 400 dpi is to reduce image “noise.”
D. File naming conventions

These entity declarations will look slightly different, as the file naming convention is slightly different. The elements are composed of the following:

<table>
<thead>
<tr>
<th>Media Type</th>
<th>Author-Title</th>
<th>Volume</th>
<th>Number</th>
</tr>
</thead>
</table>

For the issue referenced in the previous example, the file name for the XML file would be:

**NW_RichTimesD_018_110.xml**

Where “NW” stands for “newspaper,” “RichTimesD” for “Richmond Times-Dispatch,” and the Volume and Number of the issue, each element separated by underscores. Image files will simply have an additional three-digit number at the end, referencing its page number:

**NW_RichTimesD_018_110_001.jpg**

Therefore, the jpg declarations from the new template look like this:

```xml
<!NOTATION jpg SYSTEM "JPEG">
<!ENTITY NW_RichTimesD_018_110_001 SYSTEM "NW_RichTimesD_018_110_001.jpg" NDATA jpg>
<!ENTITY NW_RichTimesD_018_110_002 SYSTEM "NW_RichTimesD_018_110_002.jpg" NDATA jpg>
<!ENTITY NW_RichTimesD_018_110_003 SYSTEM "NW_RichTimesD_018_110_003.jpg" NDATA jpg>
<!ENTITY NW_RichTimesD_018_110_004 SYSTEM "NW_RichTimesD_018_110_004.jpg" NDATA jpg>
```

E. Content to be keyboarded

All textual data specified must be included in the transcription, and all non-textual data must be included in the mark-up as figures. For the DDD project, content to be keyboarded will be restricted as follows. All content (including advertisements, etc.) already keyboarded by DDD should be brought into conformity with TEI lite and specifications listed here. For page images not yet keyboarded, content should be restricted to **all news stories**, **plus** content under the headings (or headings very similar to):

- “Negro for Sale”
- “Financial & Commercial”
- “Runaways” (runaway slaves)
- “Died” and “Obituraries”
- “Military Notices”

**NOTE—(04-08-2005):** There are many files in which items with a headline of “Servants for Sale or Hire” have *not* been keyboarded. This is the equivalent of “Negro for Sale,” specified in the original version of this spec. This content will need to be added wherever missing.
F. Advertisements

DDD will keyboard the entirety of a given issue of a newspaper, once from a month of the paper’s publication, from the Monday of the second week for all months specified for a job. **NOTE:** For all other newspaper issues, content division placeholders should be keyboarded without content, in a comment but with a generic entity, as in the following example:

```xml
<div1 type="notices">
<head></head>
<div2 type="millinery">
<head>MILLINERY &amp;x26; DRESS-MAKING</head>
<div3 type="ad-blank" n="1">
<head></head><p><!-- &advertid; --></p></div3>
<div3 type="ad-blank" n="2">
<head></head><p><!-- &advertid; --></p></div3>
<div3 type="ad-blank" n="3">
<head></head><p><!-- &advertid; --></p></div3>
</div2>
```

III. Document Structure

A. Basic Document Structure

Follow the basic structural mark-up common to most TEI texts:

```xml
<TEI.2>
  <teiHeader>
    . . .
  </teiHeader>
  <text>
    <front>
      . . . [front matter: masthead etc.]
    </front>
    <body>
      . . . [main body of text]
    </body>
    <back>
      . . . [back matter: final advertisements]
    </back>
  </text>
</TEI.2>
```
B. The TEI Header + Front Template

The TEI Header must be included in all XML files. The TEI Header+Front template for the IMLS Civil War Newspaper Project should be copied and pasted into each document; the only elements needing to be edited are marked in red in the file, and those elements are listed below:

- jpg entity declarations for images
- the `<TEI.2 id>` element
- the `<idno type="URL">` element
- the `<extent>` statement within the `<sourceDesc>` element
- the `<date value>` statement, also within the `<sourceDesc>` element
- the `<date value>` statement, within the `<profileDesc>` element

**NOTE:** DDD did not use the TEI Header from the file `spec_rtd18_112_1860_11_09.xml`, sent 19 June 2004, as instructed. We are providing a new file called `CivilWarHeader+Front.txt` that will be a template for you to copy and paste into all xml files. The assumption is that UR will only be charged for keyboarding changes to this template, not for the characters simply copied and pasted into the file. DDD should account for this in its billing.

**NOTE:** this file will also include the `<front>` division, as well as what is, strictly speaking, the TEI Header. The entire file can be copied and pasted into each XML file that constitutes an issue of the newspaper; portions of the header and `<front>` div must be edited as indicated in the template, including declaration of any foreign language tags used in the file.

**NOTE:** the Header + Front template includes what could be construed as the top of two different columns in the `<front>`. **Be sure not to repeat this as part of the column** in the `<body>`. In terms of its information, it belongs to the masthead and preliminaries.

All the information found in the image below is included in the `<front>` division:

```xml
<front>

DAILY DISPATCH.

VOL VIII.—No. 110
RICHMOND, VA,—WEDNESDAY, NOVEMBER 7, 1860
PRICE ONE CENT.

</-front>
```

**NOTE:** Should the layout or content of the masthead change in future issues, DDD will be responsible to note and reflect those changes in your mark-up!

**NOTE:** Error in DDD files: incorrect jpg declarations:

Two of the files sent by DDD failed to parse. The error was the same in both files, in the following section, near the top of the TEI Header. The DDD file (`rtd18_110_1860_11_07.xml`):
DDD mark-up:

<!NOTATION jpg SYSTEM "JPEG">
<!ENTITY rtd18_110_1860_11_07 SYSTEM "rtd18_110_1860_11_07_01.jpg" NDATA jpg>
<!ENTITY rtd18_110_1860_11_07 SYSTEM "rtd18_110_1860_11_07_02.jpg" NDATA jpg>
<!ENTITY rtd18_110_1860_11_07 SYSTEM "rtd18_110_1860_11_07_03.jpg" NDATA jpg>
<!ENTITY rtd18_110_1860_11_07 SYSTEM "rtd18_110_1860_11_07_04.jpg" NDATA jpg>

Corrected mark-up:

<!NOTATION jpg SYSTEM "JPEG">
<!ENTITY rtd18_110_1860_11_07_01 SYSTEM "rtd18_110_1860_11_07_01.jpg" NDATA jpg>
<!ENTITY rtd18_110_1860_11_07_02 SYSTEM "rtd18_110_1860_11_07_02.jpg" NDATA jpg>
<!ENTITY rtd18_110_1860_11_07_03 SYSTEM "rtd18_110_1860_11_07_03.jpg" NDATA jpg>
<!ENTITY rtd18_110_1860_11_07_04 SYSTEM "rtd18_110_1860_11_07_04.jpg" NDATA jpg>

The information to the right of ENTITY should be identical to that following SYSTEM, with the exception that in the latter case, the information is in quotes, and has a jpg file extension. See the section on the milestone element for errors on referencing the images in the body of the text.

C. General guidelines for structure in the body:

Starting with <div1>: Although TEI Lite allows them, we do not use the <div> or <div0> elements. Instead, top-level structural divisions must be encoded as <div1>.

divs and <head> tags: All div tags must have associated head tags.

type attribute: Each <div#> tag must have a type attribute indicating the kind of division. If the division has no obvious type, the generic term “section” may be used; if “section” has already been used for a higher-level division, use “subsection.” Newspapers usually have information grouped in such a way that an appropriate designation can be found.

Most articles and advertisements will be on div level 3. The highest div level one will be either “news” or “notices.” The latter will include all advertisements, Runaways, Military notices, etc. The former will include “telegraphic news” “news from the north” etc. The “type” designations for level 2 divs should be drawn from the content, but should also be standardized. What appears in the head tag should be what appears in the text (not standardized). If there is not a collective designation for a group of material, it must have empty head tags. An empty div level 2 should be added if needed. See above for changes to markup of empty ad divs.

<back>
All final advertisements (ads at the end of the paper) should be put in the <back> section.

Typical (generic) type values for divisions within <body> are, especially for newspapers are:

- news
- notices — for groups of advertisements
- article
- editorial — for opinion pieces in newspapers
- essay
- letter
- poem
- dialogue
- speech — for a transcript of an oration, not for drama (for which use <sp>)
- entry — for journal entries or encyclopedia/dictionary entries
- plates — one or more full-page illustrations
- part
- section
- subsection

n attribute: If the division is numbered or otherwise labeled in the print source, record that number or label in the n attribute. Otherwise, all level 3 or 4 divs should be numbered by their serial position in the following syntax: <div3 type="article" n="3">.

NOTE: For newspaper mark-up, all articles and advertisements are to be numbered, typically, as they appear in order within a div2 or div3.

IV. Mark-up direction/correction of specific elements:

A. Ambiguities in Keyboarding and Tagging:

1. Foreign Language Tags:

Although it will be rare, part of the standard of TEI lite is the <foreign> tag. It is used to indicate that a language other than English is witnessed. DDD mark-up from the document rtd18_110_1860_11_07.xml has:

DDD mark-up:

seen, simultaneously exclaimed, <q>"<hi rend="italic">Le voila, le voila</hi>"</q>

and swayed to and fro a few moments, eagerly looking up at Brainerd, who

Since “Le voila, le voila” is French, it must be tagged to indicate this, “fre” standing for French:
seen, simultaneously exclaimed, "Le voila, le voila" and swayed to and fro a few moments, eagerly looking up at Brainerd, who

NOTE: If a language is referenced in the within the `<text>` div of an XML document, it must also be declared in the TEI Header. Any file will fail to parse if any language is not so declared.

NOTE: the three-character codes for foreign languages are an ISO 639-2 standard, including:

- fre  French
- ger  German
- grc  Greek, ancient (to 1453)
- gre  Greek, modern (1453- )
- heb  Hebrew
- ita  Italian
- lat  Latin
- rus  Russian
- spa  Spanish

Greek Use the Unicode characters listed alongside the `iso-grk1.ent` character entities, supplemented as needed by the accented characters in `iso-grk2.ent`

Hebrew Use the Hebrew block of Unicode (0590 - 05FF): `&#x05D0;` for aleph.

Russian Use the Cyrillic block of Unicode (0400 - 04FF)

[Deprecated: Vendors should no longer use the `<abbr>` tag]

2. Abbreviations:

19th century newspapers in particular make more extensive use of abbreviations, especially for names, than is currently customary, and it is furthermore not standardized. We would like to capture all these abbreviations with the `<abbr>` tag—attributes or values will be added by UR.
3. Unclear and Sic Tags:

Any character that is currently marked with a # sign to indicate the character was unreadable would instead be marked with the <unclear> element. The “#” sign is not an “industry standard” for_unclears! Also, the entire word must be marked, not just the illegible character.

**DDD mark-up:**

and Shoes. Call at 119 Main street, Richmond, opposite Mitchell & #yler's,

**Corrected mark-up:**

and Shoes. Call at 119 Main street, Richmond, opposite Mitchell & #0026; <unclear>yler's</unclear>,

*Do not key in or otherwise indicate the ambiguous character:* instead, put <unclear> tags around the entire word. When there is a spelling or typographical error in the newspaper (i.e., page-image) original, which is clearly discernable in the image, the error should be marked with <sic> tags—the entire word should be enclosed by the tags—tags should not break up words.

4. Hyphens, Abbreviations and breaking words:

It is more important to have the word intact than to record the information about hyphenation. Therefore, end-of-line hyphens should be eliminated, and words should not break between hyphens. (Note that hyphenated expressions, such as “Sine-American” should retain the hyphen intact.) This is especially true when a tag could potentially break a word. For example:

**DDD mark-up:**

<p><hi rend="bold">FOR THE PURPOSE OF FA</hi> CILITATING THE TRADE of this city,

**Corrected mark-up:**

<p><hi rend="bold">FOR THE PURPOSE OF FACILITATING</hi> THE TRADE of this city,

5. Notes:

Occasionally, decisions have to be made as to how to mark-up a problematic section. It is *industry standard practice* to provide notes in a separate file about such decisions when
significant decisions have been made in keyboarding. The files should be named along the same convention as the XML file, with “DDDnotes” added and a “.txt” file extension. The notes file for NW_RichTimesD_018_110.xml, would be NW_RichTimesD_018_110_DDD_notes.txt.

B. Head Tags:

1. Headlines within paragraphs:

In the following article, the headline is: “DEATH OF THE DUKE OF RICHMOND”—despite the fact that it appears in the paragraph of the article:

When a keyboarder needs to make mark-up decisions, information should always be chosen over appearance, i.e., it is more important to know that “DEATH OF THE DUKE OF RICHMOND” is a headline, than it is to know that it originally appeared within the body of the paragraph. People will want to search headlines, and for that to be possible, they must be designated as such.

DDD mark-up:

```
<div3 type="article" n="3">
<head/>
<p><hi rend="small caps"><hi rend="bold">Death Of The Duke Of Richmond.</hi></hi>

The Lord of the Goodwood races is dead. He was the Duke of Richmond, whose ...
```

Corrected mark-up:

```
<div3 type="article" n="3">
<head rend="smallcaps">Death Of The Duke Of Richmond.</head>
<p rend="nobreak">The Lord of the Goodwood races is dead. He was the Duke of Richmond, whose ...
```
Note that the head element can take a rend attribute; also, we will eliminate <hi rend="bold"> for all head tags, as we will assume they are in bold of some kind. The rend attribute with the <p> tag (<p rend="nobreak">) will indicate that the head does not break from the paragraph. Notice that this will affect many advertisements as well (when those are occasionally tagged).

It is assumed that contents of all <head> tags will be bold, so this should not be keyed in; other formatting such as italics, may be indicated by, i.e., <head rend="italics">.

2. Head Tag association / new divs:

When using the <head> tag, it is very important to discern what element the tag belongs to.

In the XML file rtd18_110_1860_11_07, and the image file rtd18_110_1860_11_07_03.jpg, (NOV 7, 1860), we see a continuation of a <div2 type="auction"> from column 5:

**DDD mark-up:**

</div3>
<milestone unit="column" n="6"/>
<div3 type="advert" n="30">
<head>AUCTION SALES.</head>
<p><hi rend="italic"><hi rend="bold">FUTURE DAYS.</hi></hi></p>
<p><hi rend="bold">By Goddin & Apperson, Auct's.</hi></p>
<p><hi rend="bold">VALUABLE FARM IN HENRICO, FIVE MILES ABOVE RICHMOND, SLAVES AND STOCK, FOR SALE.—Having determined to remove to Richmond to reside I will sell at public auction. on the premises, on THURSDAY, the 8th November, 1860, at 11o'clock A M., (if fair; if not on first fair day thereafter,) the Farm on . . .</p>

(Error: please also note the typo: no space between “11” and “o’clock”)
(Error: please note the date should be keyed in with the ISO format; corrected version below)

It is clear from the image itself (below), however, that the category heading “AUCTION SALES” does not belong to the ad, but is a category supplied by the newspaper to organize its ads. The heading “FUTURE DAYS” also is a sub-category supplied by the newspaper, and similarly does not belong to the ad which follows it. Therefore these items should not be included in the head tag associated with the advertisement’s division.
NOTE: Whenever a column heading appears at the top of a column that applies to the whole column (or to more than the ad or article immediately below it) a new division must be started, and the ads or articles within it renumbered starting again from 1:

Corrected mark-up:

</div3>
</div2>

<div2 type="auction">
<head>
AUCTION SALES.</head>
<hi rend="italics">FUTURE DAYS.</hi></head>
<head>
By Goddin & Apperson, Auct's.</head>
<p><hi rend="bold">VALUABLE FARM IN HENRICO, FIVE MILES ABOVE RICHMOND, SLAVES AND STOCK, FOR SALE.—Having determined to remove to Richmond to reside I will sell at public auction, on the premises, on THURSDAY, the 8th November, 1860, at 11 o'clock A M., (if fair; if not, on first fair day thereafter,) the Farm on which I now reside, lying on the Richmond and Fredericksburg Railroad, five miles above Richmond, containing 90 acres of land. After the sale of the land, I will sell, in like manner, my Slaves, Milk Cows, Horses, Mules, Corn, Oats, Poodles, Household and Kitchen Furniture one Hoe-kaway, and two Carts, and whatever else there may be on the farm for sale. Terms.—For the land, a credit of 1, 2, and 3 years, bearing interest, secured by personal and real security. For the slaves, a credit of 6 months' interest added, for negotiable notes and approved endorsers. For the residue of the personal property, cash for all sums under $50, over that amount, 6 months' credit, for bonds with good security. JOS. BERNARD. Sale by GODDIN & APPERSON, Aucts. oo 30</p>

NOTE: the ads at the bottom the prior column (5) for cigars, chewing tobacco, lye, etc., do not fall within the category of “auction sales” in any case, so the division should be changed.

NOTE: see also (E) below on associating <head> tags with tables.
C. Milestone Tags, Page-Break and Page Image References:

For page-breaks as well as column-breaks, the milestone tag is to be used (not the <pb/> tag). Note that some DDD files used the milestone tag for page-breaks and some used <pb/>. Only the milestone tag should be used to indicate page-breaks!

The first milestone tag in each issue (referencing page one) will be in the <front> section, not within <body>. The corresponding page image reference using the <figure entity> tag for the first page will also be in the <front> section. Here is the correct rendering of the <front>:

```xml
<text id="NW_RichTimesD_018_110T">
  <front>
    <milestone unit="page" n="1"/>
    <div1 type="page-image">
      <head></head>
      <p><figure entity="NW_RichTimesD_018_110_001">
        <figDesc>Page image of Daily Dispatch, Volume 18, Number 110, page 1.</figDesc>
      </figure></p>
    </div1>
  </front>
</text>
```

On all subsequent pages (any page 2, 3, 4 etc.) the milestone unit for the page-break, the milestone unit for the column-break, and the figure entity for the page-image will all be clustered together, as follows:

```xml
<text id="NW_RichTimesD_018_110T">
  <milestone unit="page" n="2"/>
  <div1 type="page-image">
    <head></head>
    <p><figure entity="NW_RichTimesD_018_110_002">
      <figDesc>Page image of Daily Dispatch, Volume 18, Number 110, page 2.</figDesc>
    </figure></p>
  </div1>
  <milestone unit="column" n="1"/>
</text>
```

Since the milestone unit for the page-break marks the point at which the page starts, the milestone unit indicating the page should come first, followed by a reference to the page-image (which, is not a representation of the page, because it links to the page-image).

**NOTE:** the entities outside the TEI Header do not take jpg filename extensions!

**NOTE:** the page number should correlate to the image number, and those to the figDesc!

**NOTE:** the formatting of this cluster for any page 2, 3, 4 etc. (anything beyond the <front> section) should always be exactly as in the model above, *i.e.* there should be line breaks after:
Also, the language for the <figDesc> should always follow the format of this example.

Errors:

In the XML file rtd18_110_1860_11_07, the tag was used inconsistently. It correctly referenced the page image for page 2 of the issue near the milestone tag indicating the page-break, but then no page-image was referenced for pages three or four. In the XML file rtd18_112_1860_11_09, the only page-image correctly referenced was the one copied from the <front> section. Some entity references for page-images included the .jpg file name extensions, which is incorrect.

D. Advertisements

1. Printer’s marks:

In advertisements, publishers indicated in shorthand how long a given ad should run.

In the Richmond Daily Dispatch, these notations took the typical form of:

oc 26#x2014;2w*
se 26#x2014;3m

Without specifying the meaning of all such notations, the above are typical forms. The first example seems to indicate that that ad starts running October (oc) 26, and should run for two weeks. The second indicates a start date of September (se) 26, and should run for three months.

Typically, these notations are on the same line as some other text, but (in the original) right- or left-justified. Presumably this was to save column space while setting off the printer’s notation.

NOTE—Spec Addition (01-07-2005): Printer’s notations must be on a separate line, and always on the last line of an ad.

NOTE—Spec Addition (01-07-2005): Printer’s notations should be encoded with their own unique tag:

<term type="printrun"></term>
Current mark-up:

```xml
<div type="advert" n="7">
  <head>BRIDAL AND CHRISTMAS PRESENTS</head>
  <p>
  Elegant Rich WORK BOXES;
  </p>
  WRITING DESKS;
  </p>
  CABAS; MOROCCO BAGS;
  </p>
  MANTEL ORNAMENTS.
  </p>
  TEMPLE OF FANCY,
  </p>
  no 26#x2014;ta(2p) Main street, above 8th.
  </p>
</div>
```

Corrected mark-up:

```xml
<div type="advert" n="7">
  <head>BRIDAL AND CHRISTMAS PRESENTS</head>
  <p>
  Elegant Rich WORK BOXES;
  <lb/>
  WRITING DESKS;
  <lb/>
  CABAS; MOROCCO BAGS;
  <lb/>
  MANTEL ORNAMENTS.
  </p>
  TEMPLE OF FANCY, Main street, above 8th.
  </p>
  <term type="printrun">no 26#x2014;ta(2p)</term>
</div>
```

**NOTE:** The correct usage of the `<p>` tag is used to indicate a logical and grammatical unit; this should be differentiated from the `<lb/>`, which should be used when the break is presentational. If a complete sentence does not occur within the tag, the `<p>` tag by definition should not be used. DDD has usually opted for the `<p>` tag instead of the `<lb/>` tag especially in ads, where the `<lb/>` tag would be preferred, and this should change.

For an additional example of both 1 and 2 of this section, see the “Baker’s Bitters” ad in the next section, IV-E-1.
2. Advertisements as signed communication:

NOTE—Spec Addition (01-07-2005): The logic of the following change is that where the business or proprietor name appears at the end of an advertisement, the ad itself may be understood as submitted by the business or proprietor—in other words, as an announcement, and so should take markup similar to that of a letter, especially in its concluding elements. The only elements of the letter syntax that will be incorporated are the <closer> and <signed> elements:

Current mark-up:

```
<div3 type="advert" n="26">
<head type="nobreak">GOODS FOR SERVANTS' WEAR.</head>
<p rend="nobreak">We have on hand a large and varied stock of Goods suitable for male and female servants' wear, manufactured in the South, which we will sell low.
</p>
JOHNSON &\#x0026; HARWOOD.
</p>
</div3>
```

Corrected mark-up:

```
<div3 type="advert" n="26">
<head type="nobreak">GOODS FOR SERVANTS' WEAR.</head>
<p rend="nobreak">We have on hand a large and varied stock of Goods suitable for male and female servants' wear, manufactured in the South, which we will sell low.
</p>
<closer>
<signed>JOHNSON &\#x0026; HARWOOD.</signed>
</closer>
</div3>
```

E. Referenced and Quoted Materials

1. Opener, Closer, Dateline, Salute, Signed and Address tags:

When encoding letters, prefaces, and other personal writings, use the following elements:

- **opener** — groups together dateline, salutation, etc. to start a letter or other division
- **closer** — groups together salutation, signature, etc. to close a letter or other division

Openers and closers, in turn, contain one or more of these elements:

- **dateline** — brief description of the place, date, time, etc., the letter was written
• **salute** — salutation at the start (e.g. “Dear Sir”) or end (e.g. “Yours sincerely”) of a letter
• **signed** — signature at the end of a letter, preface, etc.

**[Deprecated: Vendors should no use the `<date>` tag, with three exceptions]**

• **date** — use the value attribute to provide the date in ISO standard 8601 `yyyy-mm-dd` format. *If part of the date is unknown, omit that part.* All dates should be tagged, except in advertisements.

```xml
<date value="1900-09-22">22d of September, 1900</date>
<date value="09-22">22d of September</date>
```

**Deprecation exceptions:** The exceptions to the deprecation on tagging of dates are in the TEI Header, the `<front>` section of the template, and the top of the first column of the front page, in the `<body>` section. *In those specific places listed below, the full date encoding *must* be used:

a. **The TEI Header**

   The full date within `<publicationStmt>`element *must* be entered:

   ```xml
   <publicationStmt>
   <publisher>BY COWARDIN & HAMMERSLEY.</publisher>
   <pubPlace>Richmond</pubPlace>
   <date value="1860-12-07">December 7, 1860</date>
   <idno type="vol">18.136</idno>
   </publicationStmt>
   
   The *year* must be entered within the `<creation>` element:

   ```xml
   <creation>
   <date value="1860">1860</date>
   </creation>
   ```

b. **The masthead of the `<titlePage>` within the `<front>` div must have a full date:**

   ```xml
   <titlePage>
   <docTitle><titlePart type="main">DAILY DISPATCH</titlePart></docTitle>
   <docImprint>
   VOL. XVIII & #x2014; NO. 136
   <pubPlace>
   RICHMOND, VA.
   </pubPlace>
   <date value="1860-12-07">FRIDAY, DECEMBER 7, 1860</date>
   PRICE ONE CENT
   </docImprint>
   ```
c. The date which is typically (but not always) in the first column of page one:

Here is an example of standard encoding for a letter:

```xml
<div3 type="letter">
  <head>TO MRS. H. LINCOLN.</head>
  <opener>
    Weymouth, 5 October, 1761.
  </opener>
  <salute>MY DEAR FRIEND,</salute>
  <p>
    Does not my friend think me a stupid girl, when she has kindly offered to correspond with me, that I should be so senseless as not to accept the offer?
    ...
  </p>
  <signed>A. S.</signed>
</div3>
```

For letters cited in their entirety within a division, such as a complete poem cited within an essay, or a letter cited in its entirety within an article, the citation should take its own division:
<div3 type="advert" n="4">
<head rend="bold">Baker's Premium Bitters.</head>
<div4 type="letter">
<opener>
<dateline>Henrico County, Oct. 1, 1860.</dateline>
<hi rend="smallcaps">E. Baker, Esq:</hi>
<salute>Dear Sir</salute>
</opener>
<p rend="nobreak">My wife has been suffering with Dyspepsia and Nervous Affection for several years, during which time she ... my family physician's, that BAKER'S PREMIUM BITTERS is the best medicine now before the public for the above-mentioned diseases.
</p>
<closer>
<salute>Yours, most truly, </salute>
<signed>P. W. J. Quarles.</signed>
</closer>
</div4>
<p>
These Bitters can be had of all the Druggists in this city, and at every respectable Drug Store in the State. Orders filled promptly by addressing.
</p>
<closer>
<signed>E. BAKER. Proprietor.</signed>
Richmond, Va.
</closer>
</div3>

[Deprecated: Vendors should no longer use the “additional elements” formerly listed here]

The following example from early DDD files is basically correct in terms of the use of the <opener>, <dateline>, <name>, <closer> and <signed> tags in an article in which some of these elements appear. Please note, however, the keyboarding errors in this article.

The following example from early DDD files is basically correct in terms of the use of the <opener>, <dateline>, <name>, <closer> and <signed> tags in an article in which some of these elements appear. Please note, however, the keyboarding errors in this article.

The following example from early DDD files is basically correct in terms of the use of the <opener>, <dateline>, <name>, <closer> and <signed> tags in an article in which some of these elements appear. Please note, however, the keyboarding errors in this article.
Notwithstanding the extreme inclemency of the weather ...
NOTE—Spec Change: The <q> tag normally indicates a block element typographically offset from surrounding text. Since the majority of quotations are not set off as block elements, the default rendering for the <q> tag will be inline; no attribute will be required. If the quotation is rendered as a block element, the attribute break is used: <q rend="break">.

The general format for (block) citations will be the following:

\[\begin{align*}
&<\textit{cit}> \\
&<q rend="break"> \\
&</q><bibl n="value" id="oclc"></bibl></cit>
\end{align*}\]

The following example of the </cit>, <q> and <bibl> tags from early DDD files is basically correct. Again, this mark-up has also been modified to reflect the changes indicated above concerning headlines occurring within the article paragraph and placement of the <bibl> tag.

(Note that the images above are part of a single column, which has been split for easier viewing.)

\[\begin{align*}
&<\texttt{div1 type="news">} \\
&<\texttt{head}>Richmond Dispatch</texttt{head}> \\
&<\texttt{div2 type="morning">} \\
&<\texttt{head}>\texttt{date value="1860-11-07">WEDNESDAY MORNING . . . . . . . . . NOV 7, 1860</texttt{date}> \\
&</texttt{head}> \\
&<\texttt{div3 type="article" n="1">} \\
&<\texttt{head rend="smallops">}Execution Of A Maticide.</texttt{head}> \\
\end{align*}\]
Ezra Brainard was hung at Three Rivers, C. W., on the 25th ult., for having murdered his mother some months since.

A correspondent of the Montreal Gazette sketches the closing scene of his life. He writes:

Shortly before 11 o'clock, a door leading to the convict's cell was closed, and the hangman proceeded to pinion his victim in the dimly-lighted dungeon, screened by the closed door from the looks of the people in the jail hall. The constables who accompanied him state that while being pinioned Brainerd requested Sheriff Ogden to convey his body, after execution, to Melbourne, in the Townships, for burial, and that the twenty dollars in gold among his effects be appropriated for the purpose. The Sheriff promised to comply with his request. He also said to the hangman, "You are pinching me; you are beginning to murder me too soon."... A few gasps, and in a few moments Ezra Brainerd had expiated his crime.

NOTE: do not use the bibl tag outside the cit tag.

NOTE: In most cases, a paragraph should not close before entering a citation, if it is an inline citation. As an exception, in the example above, since no text from the primary speaker occurs after the citation close, the paragraph closes.

NOTE—Spec Change: double quotation marks are to be encoded as Unicode XML entities (&#x0022;) which is reflected in the “Corrected mark-up” below:

DDD mark-up: (correct according to early spec; no longer correct)

Corrected mark-up:

The boundaries of Virginia have undergone great changes. The charters... he knew not how distant the Pacific Ocean lay from the Atlantic.
Interesting Sketches.

The Alexandria Sentinel has been publishing some sketches of Virginia history for children which are very interesting to grown people. The last of the number contains the following facts about the Old Dominion:

The boundaries of Virginia have undergone great changes. The charters given...

he knew not how distant the Pacific Ocean lay from the Atlantic.

3. Reference Notes

Note reference vs. note body: “Note reference” means the anchor point for the annotation, typically indicated with a superscript number or symbol, using the ref element:

<ref target="n8.1">1</ref>

“Note body” means the content of the annotation, contained in the note id, on this model:

<note id="n8.1" rend="foot"><p>1 See <i>Squier's Monograph ...</i></note>

The most common locations for the note body are:

- the bottom margin of the page (footnotes)
- a separate “Notes” section at the end of a major structural division (endnotes)
- the left or right margin of the page (marginal notes)

Required attributes: Always include the id attribute, which must contain an ID that is unique within the XML document, and the rend attribute, to indicate its placement on the printed page:

- end — note body appears in a separate division containing all the notes for a section
- foot — note body appears in bottom margin
- left — note body appears in left margin
- right — note body appears in right margin
- head — note body appears in top margin
- inline — note body appears within the main body of the text
- above — note body appears between lines of the main text, above the line
- below — note body appears between the lines of the main text, below the line

Note: When creating IDs for notes, use a simple, human-readable numbering scheme. For notes that are already numbered in the print source, include the number in the ID. For example, if the notes are numbered sequentially throughout the entire work, use "n1", "n2", etc. (where “n” is short for “note”):
books, of the like kind, have been published in the extensive provinces of Peru, in South America.<ref target="n8.1">1</ref> See <hi rend="italics">Squier's Monograph of Central American Authors</i>, 1861, pp. 70.&mdash;<hi rend="italics">M</hi>.</p>
<p>An excellent little volume by the learned and reliable bibliographer, Don Joaquin Garcia Icazbalceta, on the subject of books on the American aboriginal languages has lately appeared. . .
</note>

F. Tables, Lists and Line Groups:

1. Tables

Most tables will need only the following elements, with the following attributes (indented):

- **table**
  - **rows**
  - **cols (columns)**
- **head**
- **row**
- **cell**

In some cases (see below) the **cell** element should also take the attribute "role" and value "label". However, this additional mark-up should only be used when clearly indicated. *Note that tables must be contained within <p> tags.* The following DDD example was good:

**DDD mark-up:**

```xml
<p><hi rend="bold">WEEKLY PAPERS.</hi></p>
<p><table rows="8" cols="2">
<tr><td>Illustrated London News</td><td>3,393,151</td></tr>
<tr><td>News of the World</td><td>2,885,000</td></tr>
<tr><td>Weekly Times</td><td>1,993,853</td></tr>
<tr><td>Weekly Dispatch</td><td>1,052,450</td></tr>
<tr><td>Bell's Life in London</td><td>466,500</td></tr>
<tr><td>Bell's Weekly Messenger</td><td>34,000</td></tr>
<tr><td>Record</td><td>205,000</td></tr>
<tr><td>Athen&amp;#00E6;um</td><td>81,000</td></tr>
</table></p>
```

The **table** tag can also take a **head** tag, and in this example, the information put in a simple paragraph really bellows to the table, and should be associated with it as a table header. Our specifications require that *all tables must have an associated* **head**, even if it is empty.

**Corrected mark-up:**

```xml
<p>
```
<p><table rows="8" cols="2">
<thead>
<WEEKLY PAPERS.></thead>
<tr><td>Illustrated London News</td><td>3,393,151</td></tr>
<tr><td>News of the World</td><td>2,885,000</td></tr>
<tr><td>Weekly Times</td><td>1,993,853</td></tr>
<tr><td>Weekly Dispatch</td><td>1,052,450</td></tr>
<tr><td>Bell's Life in London</td><td>466,500</td></tr>
<tr><td>Bell's Weekly Messenger</td><td>404,000</td></tr>
<tr><td>Record</td><td>205,000</td></tr>
<tr><td>Athen&amp;#x00E6;um</td><td>81,000</td></tr>
</table></p>

Only in those cases where a clear description of data is being provided should the attribute "role" and value "label" be used. The default value for cells is “data,” so that need never be tagged. In most cases, there will be no need to use the "role" and "label" attributes and values in tables; when in doubt, the keyboarder should refrain from this extended tagging for tables.

<p><table rows="4" cols="6">
<thead>IN THE KINGDOM OF PRUSSIA, AND DUKEDOM OF LITHUANIA.</thead>
<tr><td role="label">Annual Average.</td><td role="label">Births.</td><td role="label">Burials.</td><td role="label">Marriages.</td><td role="label">Proportion of Births to Marriages.</td><td role="label">Proportion of Births to Burials.</td></tr>
<tr><td role="label">10 Yrs to 1702</td><td>21,963</td><td>14,718</td><td>5,928</td><td>150 to 100</td></tr>
<tr><td role="label">5 Yrs to 1716</td><td>21,602</td><td>11,984</td><td>5,599</td><td>148 to 100</td></tr>
<tr><td role="label">5 Yrs to 1756</td><td>28,392</td><td>19,154</td><td>5,599</td><td>50 to 10</td></tr>
</table></p>

2. Lists

- all lists must have associated head tags, or use an empty head if no content
- unordered lists should use the <list> element, without an attribute
- the item element may also take a list, for sub-menus

<list type="ordered">
To choose our own governors.

To cashier them for misconduct.

To frame a government for ourselves.

The following mark-up has the problems of 1) not identifying the list heading, and 2) not identifying the material following the first list, as another, separate list:

**DDD mark-up:**

```
<p>The following is a list of the officers and companies who have reported at the Camp:</p>
<p>ROSTER OF FIELD, STAFF AND OFFICERS OF THE LINE.</p>
<p>Col. Sherwin McRae, Commanding.</p>
<p>Maj. J. J. Werth, 1st Major.</p>
<p>Maj. Thos. G. Armstead, 2d Major.</p>
<p>John F. Wren, Adjutant.</p>
<p>Daniel E. Gardner, Quartermaster.</p>
<p>F. W. Hancock, Assistant Surgeon.</p>
<p>Edmund Fontaine, Sergeant Major.</p>
<p>Walter K. Martin, Paymaster.</p>
<p>Miles C. Selden, Assistant Commissary.</p>
<p>COMPANIES.</p>
<p>Hanover Troop; Capt. Wms. C. Wickham, Lieut. Wm. B. Newton, Lieut. B. H. Bowise.</p>
...    
<p>Goochland Troop; Capt. Julian Harrison, Lieut. T. P. Hobson, Lieut. Geo. F. Harrison.</p>
<p>Powhatan Troop; Lieut. Com'g John F. Lay, Lieut. Chas. Old, Lieut. T. P. Skipwith.</p>
<p>Surry Troop; Capt. T. W. Taylor, Lieut. Wm. Allen. (We regret to learn that Lieut. A. was disabled by a kick from a horse on the way to Richmond.)</p>
<p>A few members of the Essex Troop are here. The officers are; Capt. R. S. Cauthorn, Lieut. Aubrey H. Jones, Lieut. Wm. Oliver.</p>
```

**Error:** mis-keying of “McRae”

**Error:** failure to identify “ROSTER OF FIELD, STAFF AND OFFICERS OF THE LINE.” as a header that needs to be associated with the list that follows.

**Error:** `<p>` tags are used to indicate paragraphs, not line breaks! In this case, what was marked as a `<p>` should be identified as the head of the list, but in such cases, whether in a `<p>` or `<head>` tag, an `<lb>` tag should be used to identify the break.

**NOTE:** in the DDD sample files, many lists were identified with `<list>`, others with `<list type="simple">`. The attribute `type="simple"` need only be added if it is an ordered list, and otherwise should not be added. Additionally, tags must be applied consistently.
The following is a list of the officers and companies who have reported at the Camp:

- **ROSTER OF FIELD, STAFF AND OFFICERS OF THE**
- **LINE.**
  - Col. Sherwin McRac, Commanding.
  - John F. Wren, Adjutant.
  - Daniel E. Gardner, Quartermaster.
  - F. W. Hancock, Assistant Surgeon.
  - Edmund Fontaine, Sergeant Major.
  - Walter K. Martin, Paymaster.
  - Miles C. Selden, Assistant Commissary.

**COMPANIES.**

- Surry Troop—Capt. T. W. Taylor, Lieut. Wm. Allen. (We regret to learn that Lieut. A. was disabled by a kick from a horse on the way to Richmond.)

A few members of the Essex Troop are here. The officers are—Capt. R. S. Cauthorn, Lieut. Aubrey H. Jones, Lieut. Wm. Oliver.

### 3. Line Groups

For more formal block text—typically verse or song—the element `<lg>` should be used instead of `<list>`. The preference is for opening and closing tags to be grouped at the start of a line:

```xml
<p>
<cit><q>
<lg type="stanza">
<l>&#x0022;Know, first, that heaven, and earth's compacted frame,
</l>&#x0022;And flowing waters, and the starry flame,
<l>&#x0022;And both the radiant lights, one common soul
</l>&#x0022;Inspires and feeds;&amp;#x2014;and animates the whole.
<l>&#x0022;This active mind, infused through all the space,
</l>&#x0022;Unites and mingles with the mighty mass:
<l>&#x0022;Hence, men and beasts the breath of life obtain,
</l>&#x0022;And birds of air, and monsters of the main.
<l>&#x0022;Th' ethereal vigour is in all the same,
</l>&#x0022;And every soul is filled with equal flame.&#x0022;
</lg>
</q></cit>
```
Note in this case that the quotation marks do not receive additional <q> tags, since those quotes are identical with the passage as a whole being quoted, and in this case, the <q> tags cannot be placed directly next to the quotation marks. Note the <q> tag does not need the rend attribute, since the default formatting of the <lg> tag will apply.

V. Formatting Text:

1. Unicode Entities:

While TEI Lite allows for standard character entities (i.e., &amp; for the ampersand), UR is adopting the Unicode standard for all entities. Punctuation requiring Unicode encoding are:

**NOTE—Spec Addition (01-07-2005):** Contrary to initial agreement on this issue, XML entities are now restricted only to those listed on the URL below, taken from the docbook specifications, which have a four-character syntax, preceded by “&#x;”. This is a list of possible examples:

<table>
<thead>
<tr>
<th>English description</th>
<th>Character mark</th>
<th>Deprecated</th>
<th>Docbook</th>
</tr>
</thead>
<tbody>
<tr>
<td>double quotation mark</td>
<td>&quot;</td>
<td></td>
<td>&lt;0022;</td>
</tr>
<tr>
<td>single quotation mark*</td>
<td>’</td>
<td></td>
<td>&lt;0027;</td>
</tr>
<tr>
<td>ampersand</td>
<td>&amp;</td>
<td></td>
<td>&lt;0026;</td>
</tr>
<tr>
<td>em dash (long dash)</td>
<td>—</td>
<td>(same)</td>
<td>&lt;2014;</td>
</tr>
<tr>
<td>em space (see below)</td>
<td>(not used)</td>
<td></td>
<td>&lt;2003;</td>
</tr>
<tr>
<td>at sign</td>
<td>@</td>
<td></td>
<td>&lt;0040;</td>
</tr>
<tr>
<td>ligatures (example)</td>
<td>ÀE</td>
<td></td>
<td>&lt;00C6;</td>
</tr>
<tr>
<td>ligatures (example)</td>
<td>æ</td>
<td></td>
<td>&lt;00E6;</td>
</tr>
<tr>
<td>ligatures (example)</td>
<td>ÒE</td>
<td></td>
<td>&lt;0152;</td>
</tr>
<tr>
<td>ligatures (example)</td>
<td>œ</td>
<td></td>
<td>&lt;0153;</td>
</tr>
<tr>
<td>umlaut (example)</td>
<td>ä</td>
<td>(not used?)</td>
<td>&lt;00E4;</td>
</tr>
<tr>
<td>pound sign (English currency)</td>
<td>£</td>
<td></td>
<td>&lt;00A3;</td>
</tr>
<tr>
<td>1/2 fraction</td>
<td>1/2</td>
<td>(same)</td>
<td>&lt;00BD;</td>
</tr>
<tr>
<td>1/4 fraction</td>
<td>1/4</td>
<td></td>
<td>&lt;00BC;</td>
</tr>
<tr>
<td>7/8 fraction</td>
<td>7/8</td>
<td>(not used)</td>
<td>&lt;215E;</td>
</tr>
</tbody>
</table>

For other fractions, other letters with umlauts, circumflexes or any other special characters (see “Standard keyboard punctuation” directly below for exceptions that should be keyed with standard keyboard characters) the XML entities on the link below must be consulted and used.

**NOTE:** *Apostrophes* (see below) are to be marked using a standard keyboard character, as distinct from *single quotation marks*, which are encoded with a Unicode entity. *Strictly
speaking, the XML entity used for single quotation marks, above, is the code for the apostrophe. It is used here as the single quotation mark, because such marks typically are not “curly” but "straight" quotes in the originals. For a list a complete list of entites used by DDD, and the corrected entities which should be used instead, see Entities_used_by_DDD.doc.

The rationale for this change is to reduce confusion in text manipulation and other programming in searching for quotations, to be able to distinguish between double quotes, in particular, that are used in encoding attribute values in TEI tagging, from in-text characters. It additionally allows a means to distinguish quotation marks associated with the <q> element, those in TEI tagging.

2. Standard keyboard punctuation:

Most common punctuation characters can and should be represented using their normal keyboard characters:

- apostrophe     
- exclamation point
- dollar sign
- percent sign
- asterisk
- opening and closing parentheses
- hyphen
- opening and closing square brackets
- opening and closing braces
- colon
- semicolon
- comma
- period
- solidus (forward slash)
- question mark

An ellipse — a series of dots or asterisks indicating deliberately omitted text — is indicated by a series of keyboard-character periods. Use the same number of characters.

If the print source contains a long space that needs to be preserved (for example, to indicate a word deliberately omitted by the author), use a series of \&#x2003; (em space) Unicode entities.

For other punctuation, Unicode entities are required. Entity codes may be found here:

**XML Character Entities**
http://www.oasis-open.org/docbook/specs/wd-docbook-xmlcharent-0.3.html
Conveniently lists the characters in the standard ISO 8879 entity sets, with graphics.

**Unicode Code Charts**
http://www.unicode.org/charts/
Contains links to PDF code charts of Unicode characters, especially for non-Roman characters.

3. Attributes and Values:

Attributes and values of elements should never contain spaces.

NOTE—Spec Change: change in specification for the rend attribute value “small caps”:

Correct:  <hi rend="smallcaps">  Incorrect:  <hi rend="small caps">

Similarly, made-up values for divisions should never contain two words separated by a space; if necessary, two (short) words may be used, connected by a dash.

4. Consistency of mark-up:

Long dashes in headlines: Since these newspapers often use a long dash to distinguish between a headline in an article and the body of the article—while including all in the same paragraph, and UR has specified that headlines should be separated out and identified with head tags, the placement of the long dash should be consistent: long dashes separating headlines from the main text of a division should be placed within the first paragraph, as below:

<head rend="smallcaps">Execution Of A Matricide.</head>
<p rend="nobreak">&amp;#x2014;Ezra Brainard was hung at Three Rivers, C. W., on the 25th <sic>ult.</sic>, for having murdered his mother some months since.

Spacing between sentences: Use one space character between sentences, not two, regardless of the apparent spacing in the print source.

5. Human- and machine readability of files:

A major goal of XML was to create files that are human- and machine-readable. Extra returns at significant points would be helpful in off-setting major breaks in the document structure.

Major divisions should be off-set; closing tags for divs should be on a single line, below the prior div., with an extra line breaking the close of the highest div level.

Milestone units should always have one return above and one return below, for milestone units indicating both page and column-breaks:

</div3>
</div2>

<milestone unit="column" n="6"/>

<div2 type="auction">
AUCTION SALES. <hi rend="italics">FUTURE DAYS.</hi>

By Goddin & Apperson, Auct's.

VALUABLE FARM IN HENRICO, FIVE MILES ABOVE RICHMOND,

Additionally, formatting should be ideal for manipulation by text processors, whether to transform the text into another text file, for output to a different format, or to draw out select information from a file. Since many processors work by reading line-by-line (i.e., PERL), it is ideal to have certain containers grouped together. Here are some general guidelines:

- Content of head tags should be on one line.
- Lines should be no longer than 100 characters (with the exception of headers and table rows, and list items which should fit on one line, within reason).
- Group opening and closing tags together (i.e., <p><p> tags on one line).

**Less preferable:**

```html
<l type="stanza">
<l>&#x0022;Know, first, that heaven, and earth's compacted frame,&lt;l&gt;
<l>And flowing waters, and the starry flame,&lt;l&gt;
<l>And both the radiant lights, one common soul&lt;l&gt;
<l> Inspires and feeds&amp;x2014;and animates the whole.&lt;l&gt;
<l>This active mind, infused through all the space,&lt;l&gt;
<l> Unites and mingles with the mighty mass:&lt;l&gt;
<l>Hence, men and beasts the breath of life obtain,&lt;l&gt;
<l>And birds of air, and monsters of the main.&lt;l&gt;
<l>Th' ethereal vigour is in all the same,&lt;l&gt;
<l>And every soul is filled with equal flame.&lt;x022;&lt;l&gt;
</l>
```

**More preferable:**

```html
<l type="stanza">
<l>&#x0022;Know, first, that heaven, and earth's compacted frame,
<l>And flowing waters, and the starry flame,
<l>And both the radiant lights, one common soul
<l> Inspires and feeds&amp;x2014;and animates the whole.
<l>This active mind, infused through all the space,
<l> Unites and mingles with the mighty mass:
<l>Hence, men and beasts the breath of life obtain,
<l>And birds of air, and monsters of the main.
<l>Th' ethereal vigour is in all the same,
<l>And every soul is filled with equal flame.&lt;x022;
</l>
```
VI. Notes for DDD on keyboarding errors

1. Content still mistaken as being in \texttt{<p>} tags with \texttt{<hi rend=bold>} instead of \texttt{<head>} tag;

There are many instances of this in, \textit{i.e.}, NW_RichTimesD_018_112.xml.

2. Separate articles often not identified as such.

Articles are not being properly identified, as in these multiple articles, designated as only one article, from NW_RichTimesD_018_130.xml (Also note the missing “T” keyboarding error):

```xml
<div3 type="article" n="9">
<head>FRANCE.</head>
<p>
There were rumors of ministerial changes rife. It was reported that M. Laihy, who figured in the Savoy negotiations, was to be Minister of the Interior.
</p>
<p>
It was said that at the Cabinet Council a plan for the army reserve was finally settled.
</p>
<p>
A project was on foot to abolish the present system of venerable furloughs, and to establish a reserve on the plan of the Prussian laudereche, combined with that of the English militia.
</p>
<p>
The commercial treaty of France and England was completed at Paris on the 16th, and received final signatures. Mr. Cobden was said to be completely prostrated by his labors.
</p>
<p>
The Bourse on the 16th was firmer and $\frac{1}{4}$ higher, closing at 69f. 95c.
</p>
</div3>
<div3 type="article" n="10">
<head>ITALY.</head>
<p>
There was no movement of importance.
</p>
</div3>
```

\textbf{NOTE}: the entity representing $1/4$ above is not an XML entity! The form would have to be $\&\#x00BC;$; and the correct entity (as of version j of the specs) is $\&\#x00BC;.$

3. Paragraph tags—almost always in advertisements—are being used incorrectly.

Where there is only a \textit{break}, not a logical or grammatical separation, the \texttt{<lb/>} should and \textit{must} be used, not the \texttt{<p>} tag!
5. Long dashes often replace hyphens in the original page-image.

Long dashes must be distinguished from hyphens. Hyphens are not encoded as XML entities.


The original spec indicated that words split by end-of-line hyphens should NOT be split in the key-boarding, and the hyphen should be eliminated. The entire word should instead appear on the next line (i.e., where the second half of the originally split word was).

7. Designated content not being keyboarded.

- There are many files in which items with a headline of “Servants for Sale or Hire” have not been keyboarded. This is the equivalent of “Negro for Sale,” specified in the original version of this spec. This content will need to be added wherever missing.

- There are many files in which advertisements not otherwise identified as content to be keyboarded (i.e., ads for consumer products, real estate, etc.) are included in one part of the paper, but not another. Inclusion of these ads should either be complete or absent. Files keyboarded part-way will be returned for completion.

8. No spaces before/after tags

In some cases, there are no spaces keyed in before or after a tag, which is the equivalent of typing like this. This occurs especially next to (the now deprecated) <abbr> tag.

9. Periods and commas substituted for each other.

While the page-image may not be clear, there are many instances where periods clearly have been substituted for what are commas, as easily determined by the context.

10. Replacement of text for what is represented in the original

Keyboarders tend to type quotation marks, especially in tables and lists, where those are intended to mean “the term above is repeated,” with the original (i.e. “bbls”)—this is INCORRECT.

11. Entities next to number should reflect spacing of original.
Especially in the stock market reports, entities for numeric fractions are found in originals. Typically, in the original, *there are no spaces between the whole number and the fraction.* However, your keyboarders have been inserting spaces between the whole number and the fraction: this is incorrect. Also, many of the entities—even beyond the fact that they are not XML entities—are incorrect. If a keyboarder cannot read a fraction well enough to have a fair degree of certainty, these should be marked with `<unclear>` instead. Clearly, some of these were guesses. Notice also that an entire sentence was missing from the end of the citation!

For ligatures, there have been many instances where the lowercase entity is used when it should use the upper case entity, and vice-versa. Better Q-A is needed here (count as mis-keys).

**NW_RichTimesD_018_121.xml:**

```
<div3 type="article" n="3">
  <head></head>
  <p>The Baltimore American of yesterday, writing of commercial affairs, under date of Saturday, says:</p>
  <cit><q>
    <p>Financial matters were rather gloomy to-day, and the market was evidently more feverish that it has been for a day or two past. The money market continues very stringent, and it is with great difficulty that even strictly first-class paper can be disposed of. Our banks continue to discount to a moderate extent. Gilt-edged first class paper may be quoted at 1 ½ to 1 ¾ per cent; but capitalists. as a general thing, are not disposed to invest, even at these high rates. The market for breadstuffs was again very heavy, and in order to effect sales, a further decline was submitted to on the part of holders. Grain of all kinds was also dull and drooping, and there seemed to be less demand for export to-day than has been the case any day for two or three weeks past. Nothing whatever done in any other branch of trade.
    </p><p>At the Stock Board to day there was quite a large business done, the sales adding up nearly $125,000 Railroad stocks were pretty active, but the market was decidedly weaker. The sales of Baltimore and Ohio foot up 1,470 shares at $56 ½; down to $55 ½; cash and the opening, but they closed at $56 bid, $56 ½; asked, a decline of $1 on yesterday's closing rates. Since the Board closed, we learn that several hundred shares were sold at $56 Northern Central gave way $1, with sales of 1,100 shares at $14 regular way and buyer 3 days, and closed at $14 bid, $15 asked Canton Company declined ½; with sales of 100 shares at $14 ½; and $15, and left off at $14 ½; bid, $15 asked. All of the Mining stocks show more firmness, and there was less disposition to force sales. Loans were pretty active, some $10,500 Baltimore city 6's having sold at $96 ½; down to $96 ½; for 1890's, and $95 for 1875's. Railroad Bons were quiet, and rather heavy. The New York stock market was very heavy to-day, and most of the leading stocks gave way. Virginia 6's sold at a decline of 7; Missouri 6's at 2 ½; Erie Railroad at 1 ½; New York Central & ½; closing heavy.
    </p></q><bibl n="The Baltimore American">The Baltimore American</bibl></cit>
</div3>
```
The Baltimore American of yesterday, writing of commercial affairs, under date of Saturday, says:

Financial matters were rather gloomy to-day, and the market was evidently more feverish than it has been for a day or two past. The money market continues very stringent, and it is with great difficulty that even strictly first-class paper can be disposed of. Our banks continue to discount at a moderate extent. Gilt-edged first-class paper may be quoted at 1¼ to 1½ per cent.; but capitalists, as a general thing, are not disposed to invest, even at these high rates. The market for breadstuffs was again very heavy, and in order to effect sales, a further decline was submitted to on the part of holders. Grain of all kinds was also dull and drooping, and there seemed to be less demand for export to-day than has been the case any day for two or three weeks past. Nothing whatever done in any other branch of trade.

At the Stock Board to-day there was quite a large business done, the sales adding up nearly $125,000. Railroad stocks were pretty active, but the market was decidedy weaker. The sales of Baltimore and Ohio foot up 1,470 shares at $55½ down to $55¼ cash and the opening, but they closed at $55 bid, $56¼ asked, a decline of $1 on yesterday's closing rate. Since the Board closed, we learn that several hundred shares were sold at $55 Northern Central gave way $1, with sales of 1,100 shares at $14 regular way and buyer 3 days, and closed at $14 bid, $15 asked. Canton Company declined ¼ with sales of 100 shares at $14½ and $15, and left off at $14½ bid, $15 asked. All of the Mining stocks show more firmness, and there was less disposition to force sales. Loans were pretty active, some $10,500 Baltimore city 6's having sold at 90½ down to 90¼ for 1890's, and 95 for 1875's. Railroad Bonds were quiet, and rather heavy. The New York stock market was very heavy to-day, and most of the leading stocks gave way. Virginia 6's sold at a decline of 7; Missouri 6's at 2½; Erie Railroad at 1½; New York Central ½; Reading ½; Galena and Chicago 1½, and Harlem ½, closing heavy.