

# UW-Madison Digital Library Data Dictionary: Electronic Facsimiles

- **Last Updated:** 2006-02-17
- **Editor:** Peter C. Gorman
- **Version:**5.0
- **Previous Versions:** [v.4.0](#)
- **Current Version:** <http://uwdcc.library.wisc.edu/resources/efacs/EFacsDataDictionary.shtml>
- **Related Documents:**
  - [Entity Relationship Diagram for Electronic Facsimile Collections](#)
  - [Use of electronic facsimile data elements in TEI markup](#)

This document does not specify text structures, record syntax, or display labels. Rather, it defines the core data elements in terms of their internal encoding, semantics, appropriate use, and relationship to other metadata schemes. Crosswalk entries (particularly those for Dublin Core and MARC) are necessarily imprecise, as this document was created for the purpose of mapping relational database fields to TEI SGML structures. They may serve, however, to give readers with a background in those metadata schemes some points of comparison in interpreting the semantics of these data elements. Methods for embedding these data elements in specific applications or transfer formats may be specified in other documents.

**Jump to:** [Contributors](#) | [Key to heading abbreviations](#)

## Objects defined for this application:

### [Collection](#)

Attributes of the collection as a whole.

### [Subcollection](#)

An arbitrary grouping of elements in a Collection.

### [Aggregate](#)

A logical level of organization higher than that of the individual *Issue*. For most serials, this will be a volume. For (single-volume) monographs, this will usually not exist.

### [Issue](#)

Basic unit of distribution. For monographs and some serials this may correspond to a volume.

### [Item](#)

Only unit of organization recognized within an *Issue*. Normally corresponds to chapters, articles, etc.

### [Page](#)

A single page.

### [Standard-Number](#)

Standard identifier according to some published scheme.

--

**Object: Collection**

Attributes of the collection as a whole.

Data Element				Req/Rep		Crosswalk			Comments
Name	Definition	Format	Example	Req	Rep	DC	MARC	TEI	
Collection-ID	Identifier for Collection; unique within scope of all Digital Collections	Text string: characters valid within SGML ID value	JCE	X		DC.Identifier	024 500	<seriesStmt> <idno type="LocalID">  Component of <publicationStmt> <idno type="LocalID">  Component of attributes id (<tei.2>, <div1>, <figure>), target (<ptr>), and entity (<figure>), and of entity names.	May be <a href="#">URN Namespace Specific String</a> .
Collection-Title	Title for the Collection	Text string	The Journal of Chemical Education: electronic facsimile	X		DC.Title	245 440  a	<seriesStmt> <title type="Collection" >	This is the title to be used in field 245 of the catalog record for the collection.
Collection-Title-NFC	Number of non-filing characters at start of Collection-Title	Positive integer	4	X†					In database implementations, this value may be prepended to Collection-Title, delimited by the pipe character ( ).
Collection-Availability	Information about copyright, access rights, etc.	Text string	Copyright © 2000 Board of Regents of the University of	X		DC.Rights	540	<availability>	Standard copyright language should be used when available.

			Wisconsin System						
<p><b>Object: Subcollection</b></p> <p>An arbitrary grouping of elements in a Collection.</p>									
Data Element				Req/Rep		Crosswalk			Comments
Name	Definition	Format	Example	Req	Rep	DC	MARC	TEI	
<i>Collection-ID</i>	Unique identifier for Collection	Text string: characters valid within SGML ID value	JCE	X					Foreign key from <a href="#">Collection</a>
Subcoll-ID	Identifier for Subcollection; unique within scope of Collection	Text string: characters valid within SGML ID value		X		DC.Identifier		Component of attribute id ( <code>&lt;seriesStmt&gt;&lt;title type="Subcollection"&gt;</code> )	
Subcoll-Title	Title for the Subcollection	Text string		X				<code>&lt;seriesStmt&gt; &lt;title type="Subcollection"&gt;</code>	
Subcoll-Title-NFC	Number of non-filing characters at start of Subcoll-Title	Positive integer	<sup>4</sup>	X†					In database implementations, this value may be prepended to Subcoll-Title, delimited by the pipe character ( ).
<p><b>Object: Aggregate</b></p> <p>A logical level of organization higher than that of the individual <i>Issue</i>. For most serials, this will be a volume. For (single-volume) monographs, this will usually not exist.</p>									
Data Element				Req/Rep		Crosswalk			Comments
Name	Definition	Format	Example	Req	Rep	DC	MARC	TEI	
<i>Collection-ID</i>	Unique identifier for	Text string: characters valid	JCE	X					Foreign key from <a href="#">Collection</a>

	Collection	within SGML ID value							
Aggregate-Sequence-No	Sequencer for Aggregate; unique within scope of Collection identified by <a href="#">Collection-ID</a> .	Number: 4 digits, zero-padded	0014	X					May be created during processing. This should start with 0001 for the first Aggregate in the Collection, and continue in unbroken sequence through the last. If Aggregates are added or removed, the sequence must be recalculated. This number has no relationship to any number printed on the source volume. Used to sort files for indexing; not used in TEI data.
Aggregate-ID	Identifier for Aggregate; unique within scope of Collection identified by <a href="#">Collection-ID</a> .	Text string: characters valid within SGML ID value	JCEV23	X†		Component of DC.Identifier	Component of 024 or 500	Component of <code>&lt;publicationStmt&gt;</code> <code>&lt;idno type="LocalID"&gt;</code>	Used for identification and linking. Once assigned, this value should must not change, even as additional Aggregates may be added to the collection.
Aggregate-Author	Author of the Aggregate	Text string	Spenser, Edmund	X†	X	DC.Creator	100 760  a 772  a 773  a	<code>&lt;sourceDesc&gt;</code> <code>&lt;seriesStmt&gt;</code> <code>&lt;respStmt&gt;</code> <code>&lt;resp&gt;Author&lt;/resp&gt;</code> <code>&lt;name&gt;</code>	For monographic works, this may be an author whose collected works comprise the series. Names must be in format compatible with LCNAF (lastname, firstname).
Aggregate-Editor	Editor of the Aggregate	Text string	Child, L. Maria		X	DC.Contributor	245  c 440  c 760  c 772  c 773  c	<code>&lt;sourceDesc&gt;</code> <code>&lt;seriesStmt&gt;</code> <code>&lt;respStmt&gt;</code> <code>&lt;resp&gt;Editor&lt;/resp&gt;</code> <code>&lt;name&gt;</code>	For monographic collections, this may be an editor of a series. Names must be in format compatible with LCNAF (lastname, firstname).
Aggregate-Title	Title of the Aggregate	Text string	The Collected	X†	X	DC.Title	245 440	<code>&lt;sourceDesc&gt;</code> <code>&lt;seriesStmt&gt;</code>	The type attribute of the TEI <code>&lt;title&gt;</code> element should use

			Works of Edmund Spenser.				760   t 772   t 773   t	<title>	the appropriate MARC tag as its value.
Aggregate- Title-NFC	Number of non-filing characters at start of Aggregate- Title	Positive integer	4	X†					In database implementations, this value may be prepended to Aggregate-Title, delimited by the pipe character ( ).
Aggregate- Title-Level	Type of Title of the Aggregate	Text character; one of: m[onographic] j[ournal] s[eries] u[npublished]	m	X†				level attribute (<title>)	
Aggregate- Issue-Sequence- No-List	Range of <a href="#">Issue- Sequence-No</a> included within this Aggregate	Text string: 4- digit sequence numbers separated by a hyphen	0001-0014						Always use the lowest and highest <a href="#">Issue-Sequence-No</a> for this Aggregate. Used for internal processing only.

***Object: Issue***

Basic unit of distribution. For monographs and some serials this may correspond to a volume.

Data Element				Req/Rep		Crosswalk			Comments
Name	Definition	Format	Example	Req	Rep	DC	MARC	TEI	
<i>Collection-ID</i>	Unique identifier for Collection	Text string: characters valid within SGML ID value	JCE	X					Foreign key from <a href="#">Collection</a>
<i>Aggregate-ID</i>	Unique identifier for Aggregate	Text string: characters valid within SGML ID value	JCEV23	X†					Foreign key from <a href="#">Aggregate</a>
<i>Subcoll-ID</i>	Unique identifier for Subcollection	Text string: characters valid within SGML			X				Foreign key from <a href="#">Subcollection</a> . Multiple values may be combined in a

		ID value							single database field as a semicolon-delimited string.
Issue-Sequence-No	Sequencer for Issue; unique within scope of Aggregate identified by <a href="#">Aggregate-Sequence-No.</a>	Number: 4 digits, zero-padded	0003	X					This should start with 0001 for the first Issue in the Aggregate, and continue in unbroken sequence through the last. If Issues are added or removed, the sequence must be recalculated. These values may be created during preprocessing. This number has no relationship to any number printed on the source Issue. Used to sort files for indexing; not used in TEI data.
Issue-ID	Identifier for Issue; unique within scope of Collection identified by <a href="#">Collection-ID.</a>	Text String: characters valid within SGML ID value.	OHehirGaelicLex	X		DC.Identifier	024 or 500	<publicationStmt> <idno type="LocalID">	Used for identification and linking. Once assigned, this value should must not change, even as additional Issues may be added to the collection. May be <a href="#">URN Namespace Specific String</a> .
Issue-Std-No	A standard number or identifier, such as ISSN, ISBN, or URN, associated with the Issue	<a href="#">Standard-Number object</a>			X				
Issue-Printed-No	Sequential Issue (in some cases, Volume) numbering as printed on source's title	Text string	Volume 3, Issue 11	X†		Component of DC.Identifier	Component of 773  g	<publicationStmt> <idno>  Component of n attribute (<tei.2>).	<b>Must</b> include, if present, labels or other non-enumerative text such as "Volume", "Issue", "Number", etc. Include numbering for all levels of aggregation (e.g., volume +

	page or cover.								issue) in this value.
Issue-Author	Author of the Issue	Text string	Shakespeare, William	X†	X	DC.Creator	100 773  a	<titleStmt> <author>	For monographic works, the main author. Names must be in format compatible with LCNAF (lastname, firstname).
Issue-Editor	Editor of the Issue	Text string	Haugen, Einar		X	DC.Contributor	245 773  c	<editor>	Names must be in format compatible with LCNAF (lastname, firstname).
Issue-Submitter	Submitter of the (source) Issue	Text string	Walden, Barbara. University of Wisconsin-Madison. Libraries	X	X		720  a  e	<respStmt> <resp>Submitter</resp> <name>	Names (personal or corporate) must be in format compatible with LCNAF. Affiliation must be included when applicable.
Issue-Title	Title of the Issue	Text string	The homes of the New world; impressions of America	X†	X	DC.Title	245 773  t	<titleStmt> <title>	For monographic works, the main title as found in subfields  a,  b,  n,  p of field 245 in the MARC catalog record.
Issue-Title-NFC	Number of non-filing characters at start of Issue-Title	Positive integer	4	X†					In database implementations, this value may be prepended to Issue-Title, delimited by the pipe character ( ).
Issue-Title-Level	Type of Title of the Issue	Text character; one of: m[onographic] j[ournal] a[nalytic] u[npublished]	m	X†				level attribute (<title>)	
Issue-PubPlace	Place of publication of the Issue	Text string	Reykjavík	X†			260 773  a  d	<pubPlace>	
Issue-Publisher	Publisher of the Issue	Text string	Mál og menning	X†		DC.Publisher	260 773  b  d	<publisher>	
Issue-Chron	Period of time represented by	Text string	March 1932			DC.Date	260 773  c  d	<publicationStmt> <date>	For periodicals, this will normally be a month or

	source Issue or date of publication of source Issue, as printed in the source.								quarter and year. For monographs and some serials, this will normally be a year.
Issue-Extent	The physical characteristics of the Issue.	Text string	168 p. : ill. (part fold.) ; 27 cm.			DC.Format	300  a	<sourceDesc> <extent>	Whenever possible, this should be copied from the catalog record for the Issue.
Issue-Page-Sequence-No-List	Range of <a href="#">Page-Sequence-No</a> included within this Issue	Text string: 4-digit sequence numbers separated by a hyphen	0001-0385						Always use the lowest and highest <a href="#">Page-Sequence-No</a> for this Issue. Used for internal processing only.
Issue-Text	Whether this Issue has text available.	Boolean: one of {y n}	y	X					Default is n.  Null values will be treated as n.
Issue-Abstract	A textual summary of the content and significance of the Issue.	Text string	Reminiscences of a pioneer settler in Milwaukee, Wisconsin, who left his home in Vermont in 1831 [...]			DC.Description	520 3_  a	<notesStmt> <note type="Abstract"> <p>	No line breaks or markup may be included in the value.
Issue-Availability	Information about copyright, access rights, etc.	Text string	Copyright © 2002 Board of Regents of the University of Wisconsin System.	X	X	DC.Rights	540	<availability>	Standard copyright language should be used when available. If a value is not present in the metadata, it will be copied from <a href="#">Collection-Availability</a> .
Issue-Production-	Whether the Issue has been	Boolean: one of {y n}	y						All and only Issues with a value of y in this field will



Ready	released for production								built in the staging environment. All Issues will be built in the test environment, regardless of the value of Issue-Production-Ready.
Issue-Last-Update	The date the Issue was created or last updated.	YYYY(-MM(-DD)) format specified in the <a href="#">W3C profile of ISO 8601</a>	2006-02-14	X					This will include only the date of the most recent update. The full revision history will be maintained within the TEI file.
Issue-Last-Update-Reason	The reason the Issue was last updated.	Text string	Issue created.  Inserted three missing page images.	X					This will include only the reason for the most recent update. The full revision history will be maintained within the TEI file.

**Object: Item**

Only unit of organization recognized within an *Issue*. Normally corresponds to chapters, articles, etc.

Data Element				Req/Rep		Crosswalk			Comments
Name	Definition	Format	Example	Req	Rep	DC	MARC	TEI	
<i>Collection-ID</i>	Unique identifier for Collection	Text string: characters valid within SGML ID value	JCE	X					Foreign key from <a href="#">Collection</a>
<i>Issue-ID</i>	Unique identifier for Issue	Text string: characters valid within SGML ID value	00120003	X					Foreign key from <a href="#">Issue</a>
Item-ID	Identifier for Item; unique within scope of Issue identified by	Text string: characters valid within SGML ID value	WTDesmond	X		DC.Identifier	024 or 500		Used for identification and linking. Once assigned, this value should must not change. <b>Must</b> be used if Item-Type is "Article" or

	<a href="#">Issue-ID.</a>								"Work". May be used in a <a href="#">URN Namespace Specific String</a> .
Item-Sequence-No	Identifier for Item; unique within scope of Issue identified by <a href="#">Issue-ID</a> .	Number: 4 digits, zero-padded	0023	X					This should start with 0001 for the first Item in the Issue, and continue in unbroken sequence through the last. If Items are added or removed, the sequence must be recalculated. The division of the Issues into Items must be complete (every Page must occur within at least one Item), but may be somewhat arbitrary when there is no explicit internal structure or when the internal structure is hierarchical and must be flattened into a single Item layer. Used to sort Items within the TEI file; not explicitly used in TEI data.
Item-Std-No	Standard identifier for Item	<a href="#">Standard-Number object</a>			X				In relational database, may be entered as a semicolon-delimited string.
Item-Type	Type of Item	Text String; one of the <a href="#">values for &lt;div1&gt; element types</a> defined in section 5.2.1 of <i>Guidelines for Markup of Electronic Texts</i> . To this list may be added: Title Page.	Article	X		DC.Type		type attribute (<div1>)	Default value is "Section".
Item-Author	Author of the	Text string	Steinfeldt		X	DC.Creator	100	<notesStmt>	Names must be in format

	Item		, Harry				Compon ent of 505	<bibl> <author>  <div1> <docAuthor>	compatible with LCNAF (lastname, firstname).
Item-Title	Title of the Item	Text string	The Use of Chemicals in Education: an Experiment		X	DC.Title	245 505	<notesStmt> <bibl> <title>  <div1> <head> <title>	A default string may be supplied by the application when no title is present, e.g. to provide clickable text in a contents list.
Item-Title-NFC	Number of non-filing characters at start of Item- Title	Positive integer	4	X†					In database implementations, this value may be prepended to Item-Title, delimited by the pipe character ( ).
Item-Abstract	A textual summary of the content and significance of the Item.	Text string	[...]			DC.Description	520 3_  a		No line breaks or markup may be included in the value.
Item-First- Printed-Page- No	Page number printed on first Page of this Item	Text String	A19						Used for internal processing only.
Item-Page- Sequence-No- List	Range of <a href="#">Page- Sequence-No</a> included within this Item	Text string: 4- digit sequence numbers separated by a hyphen	0023-0031	X					Always use the lowest and highest <a href="#">Page-Sequence-No</a> for this Item. Used for internal processing only.

**Object: Page**

A single page.

Data Element				Req/Rep		Crosswalk			Comments
Name	Definition	Format	Example	Req	Rep	DC	MARC	TEI	

<i>Collection-ID</i>	Unique identifier for Collection	Text string: characters valid within SGML ID value	JCE	X					Foreign key from <a href="#">Collection</a>
<i>Issue-ID</i>	Unique identifier for Issue	Text string: characters valid within SGML ID value	00120003	X					Foreign key from <a href="#">Issue</a>
Page-ID	Identifier for Page; unique within scope of Issue identified by <a href="#">Issue-ID</a> .	Text string: characters valid within SGML ID value	p0023	X		Component of DC.Identifier	Component of 024 or 500	Component of attributes id (<figure>) and entity (<figure>), and of entity names.	Used for identification and linking. Once assigned, this value should not change, even as additional Issues may be added to the collection. May be <a href="#">URN Namespace Specific String</a> .
Page-Sequence-No	Identifier for Page; unique within scope of Issue identified by <a href="#">Issue-ID</a> .	Number: 4 digits, zero-padded	0029						This should start with 0001 for the first Page in the Issue, and continue in unbroken sequence through the last. If Pages are added or removed, the sequence must be recalculated. This number has no relationship to any number printed on the source page. Used to sort Pages within the TEI file; not explicitly used in TEI data.
Page-Printed-No	Sequential Page number as printed on source page.	Text string	A20	X†		DC.Identifier		n attribute (<figure>) n attribute (<pb>)	Roman numerals, etc., should be transcribed exactly as printed on the source page.
Page-Description	Textual description of Page content.	Text string	Chairs; Tables; Furniture			DC.Description	6xx	<figure> <figDesc>	May be used to provide keyword access to page images containing figures or illustrations. No markup should be included in text.
Page-Text	ISO 8859-1 encoding of	Text string				DC.Description		<figure> <p>	Will normally be generated by OCR software, and may

	text on Page.								or may not be corrected. No markup should be included in text.
Page-Location	System subpath (within Collection) to image file	Text string	0012/0003/	X		Component of DC.Identifier	856  d Component of 856  u	Component of SYSTEM specification for entity definition.	The path to the image should consist only of that part of the full path local to the collection; that is, the directories under /db/otmap/[Collection-ID]/htdocs/data/images/.
Page-Filename	Base filename for image of this Page	Text string	0120030029	X		Component of DC.Identifier	Component of 856  f Component of 856  u	Component of SYSTEM specification for entity definition.	The filename extension should not be included; the extension for image files will be derived from <a href="#">Page-Format</a> , while the extension for OCR text files will be assumed to be ".txt". This implies that if OCR text is stored in a file corresponding to a page image, the filenames must be identical except for the extension.
Page-Format	Type of file created	Valid <a href="#">MIME Media Type</a>	image/tiff	X		DC.Format	856  q	Used to construct component of SYSTEM specification for entity definition.	Used to determine filename extension for image files.
Page-Notes	Additional information about the source which might impact scanning quality, such as film type, print type, bound or unbound volumes, etc.	Text string	unbound cutup issue						Not currently used in retrieval or interface processing. Derived from former field Page-Source-Details.

**Object: Standard-Number**

Standard identifier according to some published scheme.

Data Element				Req/Rep		Crosswalk			Comments
Name	Definition	Format	Example	Req	Rep	DC	MARC	TEI	
Std-No-Type	Type of standard identifier	Text string	ISSN SICI			DC.Identifier [scheme]		type attribute (<idno>)	Must be supplied for each Std-No-Value.
Std-No-Value	A standard number or identifier, such as ISSN, ISBN, or URN.	Text string	0021-9584  0277- 786X( ) 364< 123:COIPDA >2.0.TX;2- S			DC.Identifier	020 022 024	<idno>	

† required when applicable

**Key to heading abbreviations**

<b>Req</b>	<b>Required</b>
<b>Rep</b>	<b>Repeatable</b>
<b>DC</b>	<b>Dublin Core</b>
<b>MARC</b>	<b>Machine-Readable Cataloging record</b>
<b>MODS</b>	
<b>TEI</b>	<b>Text Encoding Initiative</b>

**Contributors**

This document is the result of a number of discussions held over several years, most recently in February 2006. Participants include:

- Steven Dast
- Kirstin Dougan
- Mark Foster
- Peter Gorman

- Heather McCullough
- Amy Rudersdorf
- Rose Smith
- Jessica Williams

---

## References

[W3C-DTF] Wolf, Misha; Wicksteed, Charles (World Wide Web Consortium). *Date and Time Formats*<URL: <http://www.w3.org/TR/NOTE-datetime>>