

ISO (2010) *ISO 2146:2010 Information and documentation – Registry services for libraries and related organizations*. Geneva, Switzerland: ISO

McGuire, H. (2013) 'A publisher's job is to provide a good API for books'. Available at: [toc.oreilly.com/hughm](http://toc.oreilly.com/hughm) (accessed 15 October 2013).

O'Leary, B. (2013) 'Context, not container'. In H. McGuire and B. O'Leary (eds), *A Futurist's Manifesto*. Available at: [book.pressbooks.com](http://book.pressbooks.com) (accessed 15 October 2013).

UK Data Archive (2011) *Managing and Sharing Data*, 3rd edn. Available at: [www.data-archive.ac.uk/create-manage](http://www.data-archive.ac.uk/create-manage) (accessed 15 October 2013).

Wikipedia (2013) 'Database' entry. Available at: [en.wikipedia.org](http://en.wikipedia.org) (accessed 29 July 2013).

Wyman, P. (2013) 'Mashups, models, and monetization: making your indexes go all the way.' Presentation to O'Reilly Tools of Change for Publishing Conference, New York, 12–14 February.

*Sally Goodenough is an indexer and metadata specialist who has worked at defining the semantics of schemas, data elements, and data values in government agencies and academia for more than ten years.*  
Email: [sgoodenough@acm.org](mailto:sgoodenough@acm.org)

# The Matrix: creating an active index in all kinds of formats, from all kinds of tools

*Jan Wright, Glenda Browne, Michele Combs, David Ream and Pilar Wyman*

## Introduction

The tables in this paper are based on presentations given by the authors in the course of 2013, and are intended to help indexers choose tools and processes when clients need an active index (that is, an index where clicking on an entry takes the reader to the text location).

The Matrix covers print, HTML/XHTML, PDF, and the most popular and widely used eBook formats. The Matrix does not cover apps, as each app developer seems to have their own process for achieving their end product, and the process is likely to be dictated to the indexer by the client.

The Matrix also covers tools for indexing: standalone software packages, Microsoft Word, XML editors, Framemaker, InDesign, and text extraction tools.

'Single sourcing' is a term that has been used by technical writers for decades to describe the process of getting multiple outputs (differing file formats like print, HTML, and PDF) from one set of content files. Publishers also use 'digital first' or 'born digital', which essentially mean that the digital edition of a piece will dictate the work flow. When a client is 'single sourcing' or doing 'digital first', indexers can refer to the Matrix to help choose methods that will work for the desired output.

For an active index, the tools need to create an output which allows users to click on entries or locators and be taken to the target location. An active index with links from the index to the relevant text can be created by:

- embedding or encoding the index terms in the content files and relying on mechanisms or macros to generate the index and make the index links active, or
- creating locator hyperlinks to anchors (location indicators or unique IDs) that have been inserted (manually or automatically) into the text.

There are advantages and disadvantages to each method.

For index entries embedded in content files, the indexer creates the index entries directly in the file, at the point where the term or concept occurs. The index is then generated from these tags using either a built-in compiler (such as in MS Word or InDesign) or an external script or process (such as an XSL style sheet for XML). Depending on the software used, an additional step (such as a script, macro or plug-in) may be needed to create and insert active links for each entry or locator.

For indexes created as separate files with pointers to unique IDs (also known as anchors), the indexer can use traditional standalone indexing software. These IDs may be added by the content developer prior to the delivery of the files to the indexers, or they may be added by the indexer as part of the indexing process. Whoever adds them, the result is a set of targets that are unique and can be used to form the locator link string. The files and the IDs need to be in an agreed format so that book production macros or software mechanisms will understand the coded links and then match them to the IDs.

### *A moving target*

The Matrix represents a snapshot of a moving target in the software world. Techniques and technologies are constantly changing. To help keep it up to date, if you see a new tool or method, please email ASI's DTF team ([dttf@asindexing.org](mailto:dttf@asindexing.org)), or post news on the DTF LinkedIn site (<http://tinyurl.com/DTTF1inkedin>).

The Matrix, in its latest version, will be housed at <http://tinyurl.com/pqm6vmk> and <http://tinyurl.com/ngu5mae>

## Matrix I: Indexing tools and output options

Indexing tools	Supports print?	PDF?	
Stand-alone software	Yes.  NOTE: unique IDs must be available to use as locators.	Yes, using Sonar Activate.  NOTE: Locators in book must match Sonar's requirements. Volume numbers and using differing locator schemes in same piece confuse Sonar. Cross-references are not active.  NOTE: Acrobat 10 doesn't work with Sonar at this time. Use Sonar Activate 6 for Acrobat Pro 9 projects. Use Sonar Activate 5 for Acrobat 8 Pro projects.	
Microsoft Word	Yes. Supports alternate sort, italics/bold in entries, generic 'see' refs.  NOTE: Does not support n, nn, fig., t-style decorations or specialized locators, or cross-reference checking. Watch out for multiple targets from cross-references. Regenerate index before final printing.	Yes, using Sonar Activate.  NOTE: Cross-references are not active.	
XML editor (e.g. Oxygen, XMetaL, XMLSpy)	Yes. Requires a style sheet to output the desired format. Supports alternate sort, italics/bold in entries, generic 'see' refs.	Yes. Requires a style sheet to output the desired format.	
FrameMaker	Yes, using Frame's own module. Supports alternate sort, italics/bold in entries, generic cross-references.  NOTE: Regenerate before final printing. Does not support decorations or specialized locators. No cross-reference checking. Be careful with multiple targets from cross-references.	Yes.  NOTE: Cross-references are not active.	
InDesign	Yes, using InDesign's own indexing module (InDesign CC and all CS versions produce print indexes).  NOTE: Regenerate index before final printing. Don't count on having any italics or bold in entry text. The sort can be controlled. Be careful with multiple targets from cross-references. No cross-reference checking.  NOTE: To single-source both print and ebook editions from one set of files in older versions of InDesign CSx, use scripts (see Resources) and standalone indexing software. Full HTML anchors are placed in the text by the scripts and the page number is used as the HTML locator text in the hyperlinks. To get the print version, do a GREP search and replace to strip out the HTML codes in the locators.  There are plugins available to enhance the indexing module in InDesign (see Resources).	Yes.  NOTE: Can use Sonar Activate to activate the print index in a PDF that has an inactive index. Cross-references are not active.  NOTE: Rich Bines' indexing plug-ins can activate the print index for InDesign PDFs. See Resources.	

## Matrix I continued

<b>Supports active indexes in:</b>	
<b>HTML / Web (HTML, XHTML)?</b>	<b>eBook (XHTML)?</b>
<p>Yes, as long as locators are or include anchor IDs (&lt;a id='xx'&gt; tags) in content code.</p> <p>NOTE: HTML/Prep™ in conjunction with CINDEXTM or Sky Index™ can be used to generate the appropriate files with the appropriate links. Cross-references are active. Indexes can be split by letter.</p>	<p>Yes, as long as locators are or include anchor IDs (&lt;a id='xx'&gt; tags) in content code or utilize Canonical Fragment IDs (CFIs).</p> <p>NOTE: HTML/Prep™ in conjunction with CINDEXTM or SKY Index™ can be used to generate the appropriate files with the appropriate links. Cross-references are active. Indexes can be split by letter.</p> <p>NOTE: TExtract™, which allows an indexer to add, edit, and control extracted index entries, now supports anchors for eBook index output. This is new functionality that has not been tested yet by the authors.</p>
<p>Yes, with third-party macros written by client</p> <p>NOTE: Requires specialized macros to insert anchor tags (&lt;a id='xx'&gt;) and get the index to link to them. Word is programmed to look only for the page the text is on at the time of generation.</p>	<p>Yes, with third-party macros written by client</p> <p>NOTE: Requires specialized macros to insert anchor tags (&lt;a id='xx'&gt;) and get the index to link to them. Word is programmed to just look for the page the text is on at the time of generation. TExtract may be of use (see Resources).</p> <p>NOTE: Although Amazon Create Space and Kindle Direct will accept a Word document as input, the resulting index will NOT be active.</p>
<p>Yes. Can link to individual word level. Requires a style sheet to output the desired format and to insert appropriate anchors and links in the output. Oxygen ships with a DocBook-to-HTML stylesheet that can be applied to generate a readable/proofable index when you compile the book. Supports italics and bold in entries, active and generic cross-references.</p>	<p>Yes. Can link to individual word level. Requires a style sheet to output the desired format and to insert appropriate anchors and links in the output. Oxygen ships with a DocBook stylesheet that can be applied to generate a readable/proofable index when you compile the book. Supports italics and bold in entries, active and generic cross-references.</p>
<p>May be possible, but would likely take a lot of work and specialized macros.</p>	<p>Adobe reports that Frame does convert indexes to EPUB. It is possible to convert a document created in Frame to EPUB via RoboHelp, but the resulting index is not particularly good.</p>
<p>Yes, with InDesign CC.</p> <p>NOTE: Older versions of InDesign strip out index entries for HTML and XHTML exports. HTML anchors need to be inserted into the content, and used as locators.</p> <p>NOTE: InDesign CC leaves index anchors in the HTML, but the indexing works only if you create the book as one long file. Traditional Book File and Chapter File structuring doesn't work.</p>	<p>Yes, with InDesign CC's EPUB export.</p> <p>NOTE: Indenting is not perfect, but the CSS style sheet can be edited (see Resources). Cross-references are not active.</p> <p>NOTE: Kindle formats are not supported.</p> <p>NOTE: Prior versions of InDesign strip out index entries. For prior versions, use scripts or plug-ins, or a standalone index using HTML anchors inserted in InDesign content (see Resources).</p>

## Matrix 2: Linking to entries versus embedding entries in content files

Considerations	Entries link to unique IDs (i.e. creating separate index file)	Entries embedded in content (i.e. some form of embedded indexing)
What's inserted into the content files (either by the publisher or the indexer)?	HTML anchors or unique IDs. IDs must be able to be converted by compiler into HTML anchors or other anchor format required for links to work	Index entries are inserted directly into content as fields, XML elements, or using the program's own unique marking system.
Single sourcing capabilities: Can you use one set of files to get indexes in all kinds of output (PDF, HTML, XML, eBook?)	<p>Yes.</p> <p>NOTE: If content is later edited, unique IDs may be lost.</p> <p>NOTE: Most workflows will rely on an outside compiler to generate different outputs, and to generate the index and create links to unique IDs. If the publishing software doesn't natively provide good output, publisher may have to tweak manually.</p>	<p>Yes.</p> <p>NOTE: InDesign Creative Cloud (CC) outputs EPUB with active indexes; it also outputs active indexes in HTML, but only if the book is one long file. Older versions do not output indexes for HTML files or eBooks. Indesign and Frame can output active PDF indexes.</p> <p>NOTE: Word can output a print index. Word does not export active indexes in PDF. If the publishing process requires importing Word into InDesign, note that Word's index markers don't always import correctly.</p> <p>NOTE: XML can support any type of output but requires an XSL style sheet or other automated process to do so.</p>
Granularity: When you click on a link in the index, how close do you get to the referenced content in the text?	<p>Links can only go to where unique IDs are located (e.g. if they are at the paragraph level, links go to paragraph level).</p> <p>NOTE: Display of content will vary depending on device size and settings. The user may have to scroll a bit to get to the referenced content (e.g., a 'page' may consist of several 'screens').</p> <p>NOTE: Indexers should consider which ID to use (e.g. choose the ID closest to the term or choose the ID for the nearest section title to provide context).</p>	<p>Links can go to any level.</p> <p>NOTE: Different workflows will dictate what is displayed as locators. Indexes must be recompiled following changes to content.</p> <p>NOTE: Indexers should consider where to embed index markers (e.g. choose a location closest to the term or choose a location in the nearest section title to provide context). The publisher may have guidelines on where they want index markers placed.</p>
File management: Who has the files?	<p>Indexer has index files. Publisher has content files (indexer has copy). Publisher needs to notify indexer when/if content changes.</p> <p>NOTE: Publisher can either run scripts to insert anchors and hand off copies of files to indexer, or indexer can run scripts and return a set of files to publisher. Publisher can continue tweaking files once anchors/unique IDs are in, as long as they don't delete anchors/unique IDs.</p>	<p>Indexer has copy of the content files.</p> <p>NOTE: May be done in three ways: (1) Indexer works directly in content files; (2) Indexer works outside the files using unique IDs as interim locators, client or indexer later runs macro to merge index entries into the files; (3) Indexer works in copy of content files which will later be merged via software workflow with the 'master copy' held by the publisher.</p> <p>NOTE: Unless publisher has an accurate process for merging changed files, they should not make changes to content files once the indexer begins work. If publisher does make changes to content files, the workflow should include error checking.</p>

## Matrix 2 continued

Considerations	Entries link to unique IDs (i.e. creating separate index file)	Entries embedded in content (i.e. some form of embedded indexing)
What if pagination changes and content moves around?	This is only a problem if using page numbers as locators.	Index must be recompiled after each content change and the index reviewed for accuracy or errors.
What about translating content and index into other languages?	<p>Translation of entries: Index is translated as a complete document OR publisher may choose to request a complete re-index of the new translation.</p> <p>Locators: If using page numbers as locators, pagination may change because some languages are 'wordier' than others and/or different editions are paginated differently.</p>	<p>Translation of entries: The embedded entries must be translated along with the text.</p> <p>Locators: Not an issue, since locators will be generated when the index is regenerated</p>
Chunking content: Can you create smaller books from selected chapters and still have a workable index?	Yes. Potential issues include orphaned sub-heads, missing cross-references, and broken links.	Yes. Potential issues include orphaned sub-heads, missing cross-references, and broken links.
Updates and revisions: What about next year/month when we revise it?	<p>If original entries can be sorted consecutively (that is, in order of appearance in the content), indexer can review the index for changed paragraphs and page numbers.</p> <p>NOTE: If later editions are anticipated, consider planning for this capability.</p>	Depends on software used. The ability to view index entries in context is helpful, as is a tool that highlights differences in versions of files
Can legacy books with an index be published as ebooks with an active index?	<p>Yes. (1) Obtain content files (e.g., PDFs, eBook), including index; (2) If necessary, insert unique anchors into the content files using a tool like LevTech's EPUB Index Manager™, TExtract, or the InDesign scripts from Ole Kvern; (3) sort existing index into locator order; (4) enter new locators that point to the anchors rather than the page numbers; (5) output the index in the appropriate format to include in the eBook.</p> <p>NOTE: Consider where unique anchors need to be/should be inserted: page breaks, section level, paragraph level. The presence/absence of unique IDs dictates where entries will be able to link to.</p>	If legacy book is in a layout/desktop publishing software, and if the software has a later version that supports active ebook export, convert files to the later version.

## Resources for active indexing in ebooks and other formats

### News and information

- ASI Digital Trends Task Force (DTTF): <http://bit.ly/rFp0Re>
- SI: [www.ptg-indexers.org.uk](http://www.ptg-indexers.org.uk)
- ANZSI ebooks page: [www.anzsi.org/site/ebooks.asp](http://www.anzsi.org/site/ebooks.asp)
- LinkedIn News on Indexing and eBooks: <http://tinyurl.com/DTTFLinkedIn>
- Twitter:
  - @ASIndexing
  - @indexers
  - @isc/csi

- @The\_Indexer
- #dailyentry
- #ePrdctn hour (follow the discussion on Wednesdays: 11 am EST / 10 am CST / 9 am MST / 8 am PST)
- #indexing

### Tools and technology

- InDesign tools
  - Plugins to enhance the indexing module of InDesign, including importing, exporting, and activating indexes, by Rich Bines, Kerntiff Publishing Systems Ltd, [www.kerntiff.co.uk](http://www.kerntiff.co.uk)
  - Scripts for indexing in older versions of InDesign for ebook export: <http://tinyurl.com/WrightInDesignscripts>

- How to edit CSS style sheet to fix indenting in InDesign exports:  
<http://tinyurl.com/InDesignindenting>
- TExtract: [www.texyz.com/textract](http://www.texyz.com/textract)
- LevTech's HTML/Prep for use with CINDEXT and SKY Index (web, EPUB2/3, XHTML output):  
<http://tinyurl.com/levtechhtml-prep>
- XML editors (e.g. oXygen [www.oxygenxml.com](http://www.oxygenxml.com), XMetaL, <http://xmetal.com>, XMLSpy, [www.altova.com/xml-editor/](http://www.altova.com/xml-editor/))
- Sonar Activate:  
<http://tinyurl.com/sonaractivateproducts>
- IDPF EPUB 3.01 IWG Specification:  
[www.idpf.org/epub/idx/](http://www.idpf.org/epub/idx/)

#### Legacy index conversions

- 'Converting a legacy print book to an EPUB with pinpoint index kinking' – David K. Ream,  
<http://tinyurl.com/levtecrepurposingindex>
- LevTech's EPUB Index Manager and legacy indexes workflow processes: [www.levtechinc.com](http://www.levtechinc.com) or [DaveReam@levtechinc.com](mailto:DaveReam@levtechinc.com)

Jan Wright, [jancw@wrightinformation.com](mailto:jancw@wrightinformation.com), @windexing  
Glenda Browne, [glendabrowne@gmail.com](mailto:glendabrowne@gmail.com)  
Michele Combs, [mrothen2@twcny.rr.com](mailto:mrothen2@twcny.rr.com)  
David K. Ream, [DaveReam@levtechinc.com](mailto:DaveReam@levtechinc.com)  
Pilar Wyman, [pilarw@wymanindexing.com](mailto:pilarw@wymanindexing.com)

# InDesign

Indexing tools designed for indexers who use InDesign.

Import from DIS, RTF,

Embedding,

Hyperlinking,

Clickable,

ePub ready,

Export to DIS.

Services designed for indexers who do not use InDesign.

WE DO IT FOR YOU.

[email.editmarks@btinternet.com](mailto:email.editmarks@btinternet.com)  
[www.kerntiff.co.uk](http://www.kerntiff.co.uk)

## Translating Zembla; or, how to finish *Pale fire*

In a conversation with René Alladaye about his brilliant new book (*The darker side of Pale fire* – the best introduction to *Pale fire* currently available, although it's only available through Amazon.fr currently) at the recent Nabokov and France conference, the question of translating the index came up. In most languages, this is not a problem, because the final entry, 'Zembla, a distant northern land' (*PF*, 315), which works as a fitting conclusion to the narrative, will naturally come last as 'Z' is the last letter of the alphabet.

In non-Latin scripts, this is more problematic, most prominently in Nabokov's native tongue, Russian, where 'З' or 'Z' is ninth of 33 characters. Véra Nabokov's translation of *Pale fire* for Ardis Press works round this by rephrasing the entry:

ЯЧЕЙКА яшмы, Зембля, далекая северная страна.

[Orbicle of jasp, Zembla, a far northern country]

Я [ya] is the last letter of the Russian alphabet. This raises a further question of what 'orbicle of jasp' – a quotation

from line 558, 'Terra the Fair, an orbicle of jasp' (*PF*, 54) – is doing in front of Zembla to retain its position? I don't have any immediate answers, but a deeper analysis of the differences between the English and Russian index will surely help. Ultimately, the flow of the narrative is more important than the index's order, indicating the importance in the Nabokovs' collective mind of having the Zembla entry of the index close the text.

Thanks to Marina Savina for helping translating the Russian and finding the reference to 'orbicle of jasp' in the poem.

Simon Rowberry

### References

- Alladaye, R. (2013) *The darker side of Pale fire*. Paris: Michel Houdiard.  
Nabokov, V. (1962) *Pale fire*. New York: GP Putnam's & Son.  
Nabokov, V. (1983) *Бледный огонь [Pale fire]*. trans. V. Nabokov. Ann Arbor, Mich.: Ardis Press.