Electronic Journal Publishing: Current Practices

Presented by

Bruce D. Rosenblum ブルース ローゼンブラム

CEO 最高経営責任者

Inera Incorporated

NII and SPARC/JAPAN, 2 November 2007

Introduction

- Part 1: Metadata publishing
- Part 2: Workflow transition
- Part 3: NLM DTD
- Appendix: Word 2007

Part 1: Metadata Publishing

- Why metadata is important
- Changing metadata sources
- Metadata quality issues
- Importance of metadata accuracy

Why Metadata Is Important

- Discovery in the print world (1660 to 1999)
  - Correspond with colleagues
  - Visit a research librarian
- Discovery in the electronic world (2000 to …)
  - emails/blogs with colleagues
  - PubMed
  - Web of Science or Scopus
  - Specialized databases
  - Google Scholar
- The “amateur” searcher vs. the trained librarian

Changing Metadata Sources

- ◆ Print world
  - Trained indexers manually rekeyed from print journal
  - Indexers understood requirements for *their* system
- ◆ Online world
  - Metadata originates with publishers
  - Many translations
    - Peer review → Production System → XML file → XML schema for indexer
    - Last translation done by publisher, not indexer

Metadata Quality Issues

- ◆ Librarians (indexers) understand their metadata requirements; publishers do not
- ◆ Example
  - PubMed
    - Indexed by librarians
    - Publisher data proofed by librarians
    - Not perfect, but often considered authoritative
  - CrossRef
    - Data supplied by publishers
    - Consistency and accuracy is a problem

Metadata Differences

- White, K. P., Speechley, M., Harth, M., and Ostbye, T. (2000). Co-existence of chronic fatigue syndrome with fibromyalgia syndrome in the general population--a controlled study. *Scandinavian Journal of Rheumatology* 29, 44–51.

PubMed Metadata

```
<PubmedArticle>
  <MedlineCitation Owner="NLM" Status="MEDLINE">
    <PMID>10722257</PMID>
    <Article PubModel="Print">
      <Journal>
        <ISSN>0300-9742</ISSN>
        <JournalIssue><Volume>29</Volume><Issue>1</Issue><PubDate><Year>2000</Year></PubDate></JournalIssue>
      </Journal>
      <ArticleTitle>Co-existence of chronic fatigue syndrome with fibromyalgia syndrome in the general population. A
    controlled study.</ArticleTitle>
      <Pagination><MedlinePgn>44-51</MedlinePgn></Pagination>
      <AuthorList CompleteYN="Y">
        <Author ValidYN="Y"><LastName>White</LastName><ForeName>K P</ForeName><Initials>KP</Initials></Author>
        <Author ValidYN="Y"><LastName>Speechley</LastName><ForeName>M</ForeName><Initials>M</Initials></Author>
        <Author ValidYN="Y"><LastName>Harth</LastName><ForeName>M</ForeName><Initials>M</Initials></Author>
        <Author ValidYN="Y"><LastName>Ostbye</LastName><ForeName>T</ForeName><Initials>T</Initials></Author>
      </AuthorList>
    </Article>
  </PubmedData>
</PubmedArticle>
```

CrossRef Metadata

```
<?xml version="1.0" encoding="UTF-8" ?>
<crossref_result>
  <body>
  <query key="" status="resolved">
  <doi>10.1080/030097400750001798</doi>
  <issn type="print">03009742</issn>
  <issn type="electronic">00000000</issn>
  <journal_title>Scandinavian Journal of Rheumatology</journal_title>
  <author>White, Mark Speechley, Manfred Hart</author>
  <volume>29</volume>
  <issue>1</issue>
  <first_page>44</first_page>
  <year>2000</year>
  <publication_type>full_text</publication_type>
  <article_title>Co-existence of chronic fatigue syndrome with fibromyalgia syndrome in the general population: A
    controlled study</article_title>
  </query>
  </body>
</crossref_result>
```

```
<xsd:element name="surname">
  <xsd:simpleType>
    <xsd:restriction base="xsd:string">
      <xsd:maxLength value="35"/>
      <xsd:minLength value="1"/>
    </xsd:restriction>
  </xsd:simpleType>
</xsd:element>
```

Importance of Metadata Accuracy

- Accurate metadata allows
  - Accurate indexing
  - Sophisticated relationship mappings
  - Readers to find *valuable* information
  - Traffic on your web site

For Better Metadata…

- Understand recipient requirements
- Build quality control systems
- Review process changes carefully
- Continually spot-check (beyond regular checks)

Part 2: Workflow Transition

- A little history
- The online imperative
- Paper workflows
- Electronic workflow 1.0
- Electronic workflow 2.0

A Little History

- ◆ Gutenberg
- ◆ Oldenburg
- ◆ Linotype
- ◆ Photon
- ◆ PostScript

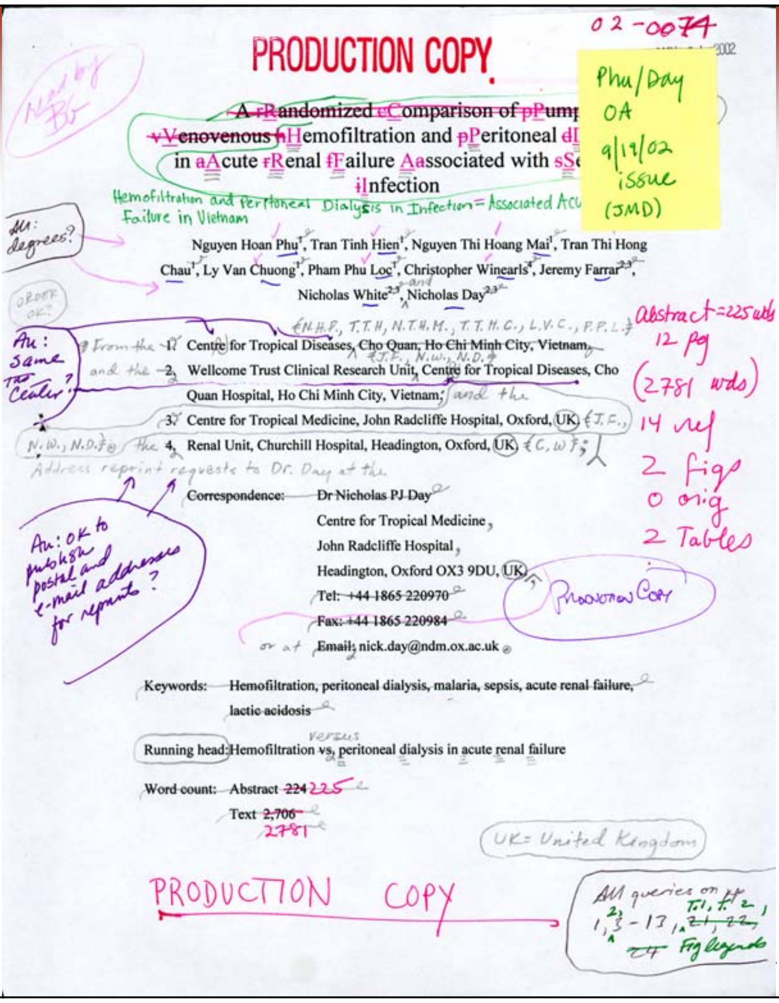
PHILOSOPHICAL
TRANSACTIONS:
GIVING SOME
ACCOMPT
OF THE PRESENT
Undertakings, Studies, and Labours
OF THE
INGENIOUS
IN MANY
CONSIDERABLE PARTS
OF THE
WORLD

Vol. I.
[illegible]

[illegible]

Last Fifteen Years…

- ◆ More change than the last 500
- ◆ Starting points are different
  - 1992: Paper
  - 2007: Electronic files
- ◆ Ending points are different
  - 1992: Print
  - 2007: Print, PDF, CD-ROM, XML, HTML



02-00~~7~~4

PRODUCTION COPY

Read by BB

Phu/Day
OA
9/19/02
issue
(JMD)

~~A rRandomized cComparison of pPump~~
~~vVenovenous h~~Hemofiltration and pPeritoneal dD
in aAcute rRenal fFailure Aassociated with sS
iInfection

Hemofiltration and Peritoneal Dialysis in Infection-Associated Acu
Failure in Vietnam

Au: degrees?

ORDER OK?

Nguyen Hoan Phu[1], Tran Tinh Hien[1], Nguyen Thi Hoang Mai[1], Tran Thi Hong Chau[1], Ly Van Chuong[1], Pham Phu Loc[1], Christopher Winearls[4], Jeremy Farrar[2,3], and Nicholas White[2,3], Nicholas Day[2,3]

{N.H.P., T.T.H, N.T.H.M., T.T.H.C., L.V.C., P.P.L.}

abstract=225 wds
12 pg
(2781 wds)
14 ref
2 figs
0 orig
2 Tables

Au: same the Center?

From the ~~1.~~ Centre for Tropical Diseases, Cho Quan, Ho Chi Minh City, Vietnam, {J.F., N.W., N.D.}
and the ~~2.~~ Wellcome Trust Clinical Research Unit, Centre for Tropical Diseases, Cho Quan Hospital, Ho Chi Minh City, Vietnam; and the
~~3.~~ Centre for Tropical Medicine, John Radcliffe Hospital, Oxford, UK {J.F., N.W., N.D.}; the ~~4.~~ Renal Unit, Churchill Hospital, Headington, Oxford, UK {C.W.};

Address reprint requests to Dr. Day at the

~~Correspondence:~~ ~~Dr Nicholas PJ Day~~
Centre for Tropical Medicine,
John Radcliffe Hospital,
Headington, Oxford OX3 9DU, UK;
~~Tel: +44 1865 220970~~
~~Fax: +44 1865 220984~~
or at ~~Email:~~ nick.day@ndm.ox.ac.uk

Au: OK to publish postal and e-mail addresses for reprints?

PRODUCTION COPY

~~Keywords: Hemofiltration, peritoneal dialysis, malaria, sepsis, acute renal failure, lactic acidosis~~

Running head: Hemofiltration ~~vs.~~ versus peritoneal dialysis in acute renal failure

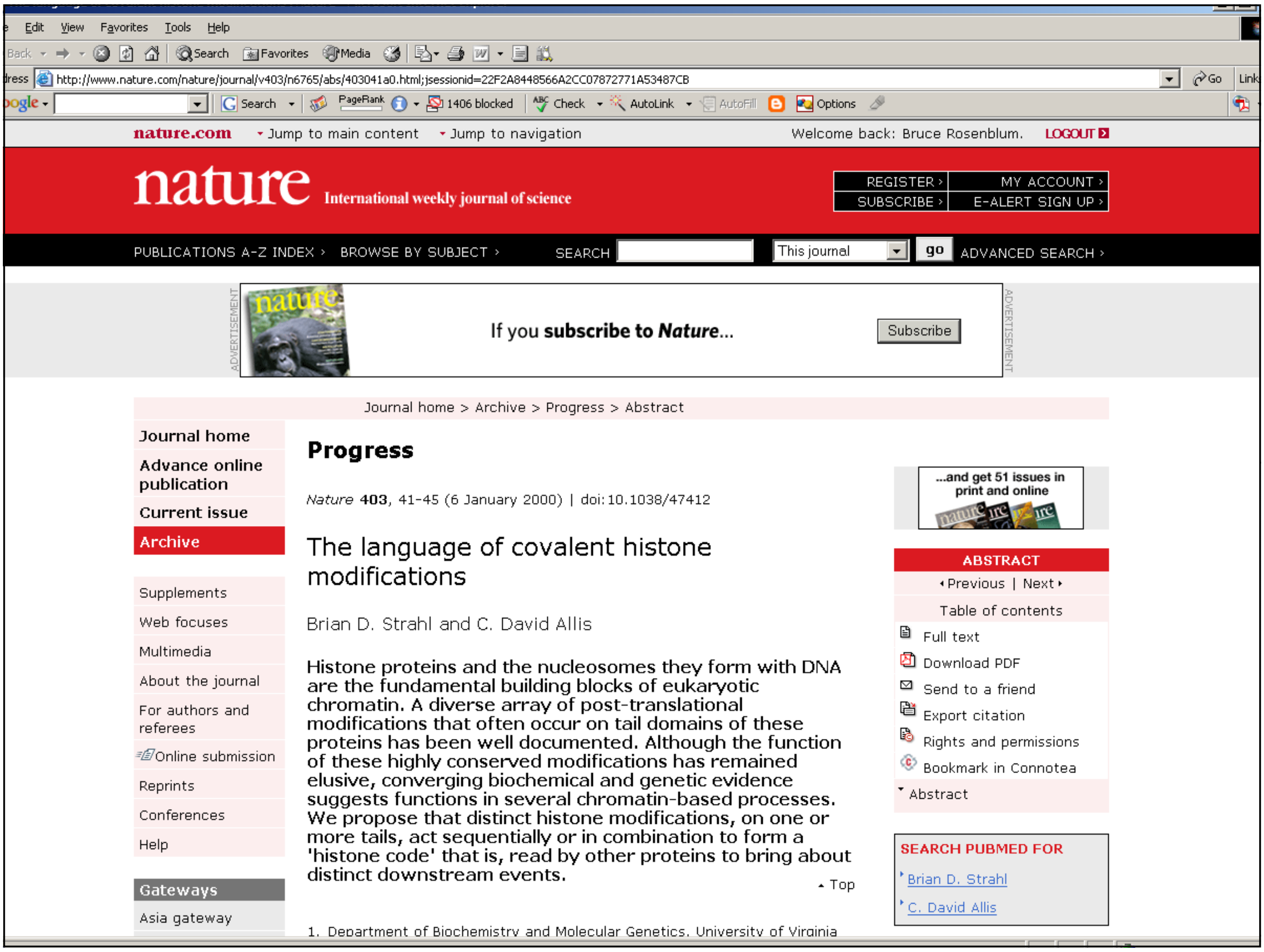
~~Word count: Abstract 224~~ 225
~~Text 2,706~~ 2781

UK= United Kingdom

PRODUCTION COPY

All queries on pp 1, 2, 3-13, T.1, T.2, ~~21, 22,~~ ~~24~~ Fig legends

All Rights Reserved



Edit View Favorites Tools Help
Back Search Favorites Media
http://www.nature.com/nature/journal/v403/n6765/abs/403041a0.html;jsessionid=22F2A8448566A2CC07872771A53487CB
Go Links
Search PageRank 1406 blocked Check AutoLink AutoFill Options

nature.com · Jump to main content · Jump to navigation · Welcome back: Bruce Rosenblum. LOGOUT

nature International weekly journal of science

REGISTER › MY ACCOUNT › SUBSCRIBE › E-ALERT SIGN UP ›

PUBLICATIONS A-Z INDEX › BROWSE BY SUBJECT › SEARCH This journal go ADVANCED SEARCH ›

ADVERTISEMENT If you **subscribe to *Nature***... Subscribe

Journal home > Archive > Progress > Abstract

Journal home
Advance online publication
Current issue
Archive
Supplements
Web focuses
Multimedia
About the journal
For authors and referees
Online submission
Reprints
Conferences
Help
Gateways
Asia gateway

Progress

Nature **403**, 41-45 (6 January 2000) | doi:10.1038/47412

The language of covalent histone modifications

Brian D. Strahl and C. David Allis

Histone proteins and the nucleosomes they form with DNA are the fundamental building blocks of eukaryotic chromatin. A diverse array of post-translational modifications that often occur on tail domains of these proteins has been well documented. Although the function of these highly conserved modifications has remained elusive, converging biochemical and genetic evidence suggests functions in several chromatin-based processes. We propose that distinct histone modifications, on one or more tails, act sequentially or in combination to form a 'histone code' that is, read by other proteins to bring about distinct downstream events.

Top

1. Department of Biochemistry and Molecular Genetics, University of Virginia

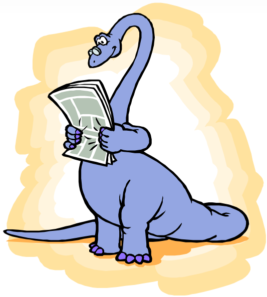
...and get 51 issues in print and online

ABSTRACT
‹ Previous | Next ›
Table of contents
Full text
Download PDF
Send to a friend
Export citation
Rights and permissions
Bookmark in Connotea
Abstract

SEARCH PUBMED FOR
Brian D. Strahl
C. David Allis

Original Paper Workflow

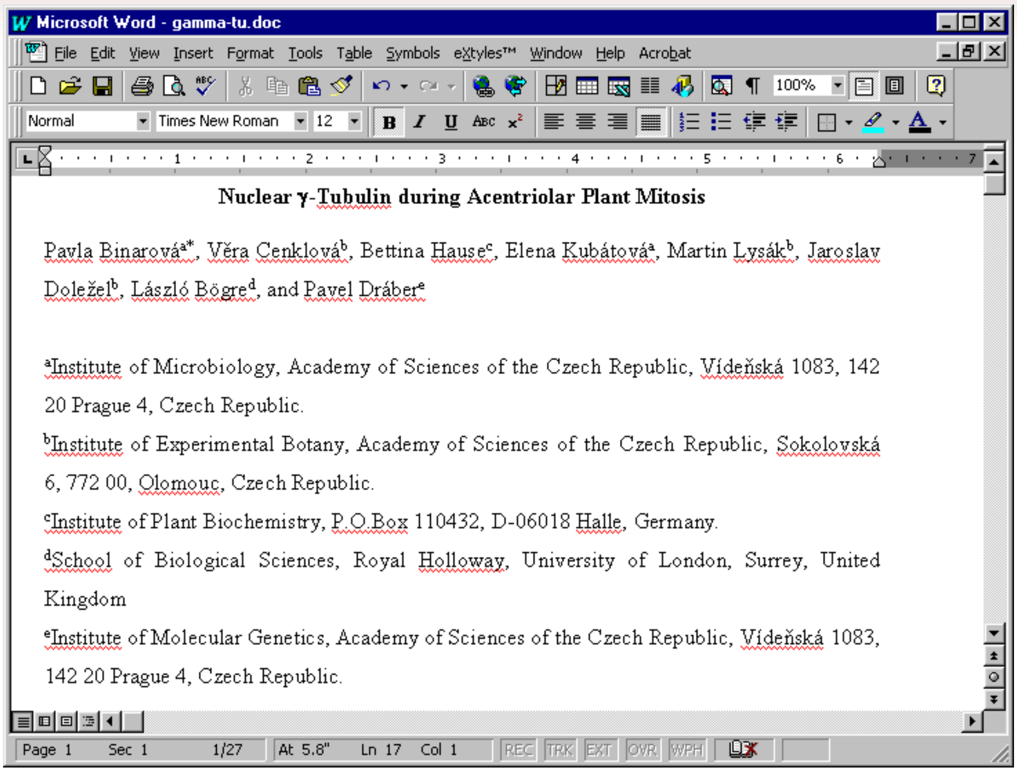
- Submit and edit on paper
- Keyboard for typesetting
- Proof
- Typeset corrections
- Print



Electronic Workflow 1.0

- Author submit electronic or paper
- Convert to "coded" file
- Edit coded file
- Typeset from coded file
- Re-key tables and math
- Proof and typeset corrections
- Print and create PDF
- Create SGML

Author’s File in Word



Microsoft Word - gamma-tu.doc
File Edit View Insert Format Tools Table Symbols eXtyles™ Window Help Acrobat
100%
Normal
Times New Roman
12
Page 1 Sec 1 1/27 At 5.8" Ln 17 Col 1 REC TRK EXT OVR WPH

Nuclear γ-Tubulin during Acentriolar Plant Mitosis

Pavla Binarová[a*], Věra Cenklová[b], Bettina Hause[c], Elena Kubátová[a], Martin Lysák[b], Jaroslav Doležel[b], László Bögre[d], and Pavel Dráber[e]

[a]Institute of Microbiology, Academy of Sciences of the Czech Republic, Vídeňská 1083, 142 20 Prague 4, Czech Republic.

[b]Institute of Experimental Botany, Academy of Sciences of the Czech Republic, Sokolovská 6, 772 00, Olomouc, Czech Republic.

[c]Institute of Plant Biochemistry, P.O.Box 110432, D-06018 Halle, Germany.

[d]School of Biological Sciences, Royal Holloway, University of London, Surrey, United Kingdom

[e]Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Vídeňská 1083, 142 20 Prague 4, Czech Republic.

Coded File Example

```
<ATL>Nuclear &gamma;-Tubulin during Acentriolar Plant Mitosis</ATL>
<AUG>Pavla Binarov&aacute;<SUP>a</SUP>*, V&ecaron;ra Cenklov&aacute;<SUP>b</SUP>, Bettina
    Hause<SUP>c</SUP>, Elena Kub&aacute;tov&aacute;<SUP>a</SUP>, Martin
    Lys&aacute;k<SUP>b</SUP>, Jaroslav Dole&zcaron;el<SUP>b</SUP>, L&aacute;szl&oacute;
    B&ouml;gre<SUP>d</SUP>, and Pavel Dr&aacute;ber<SUP>e</SUP></AUG>
<AFF><SUP>a</SUP>Institute of Microbiology, Academy of Sciences of the Czech Republic,
    V&iacute;de&ncaron;sk&aacute; 1083, 142 20 Prague 4, Czech Republic.</AFF>
<AFF><SUP>b</SUP>Institute of Experimental Botany, Academy of Sciences of the Czech
    Republic, Sokolovsk&aacute; 6, 772 00, Olomouc, Czech Republic.</AFF>
<AFF><SUP>c</SUP>Institute of Plant Biochemistry, P.O.Box 110432, D-06018 Halle,
    Germany.</AFF>
<AFF><SUP>d</SUP>School of Biological Sciences, Royal Holloway, University of London,
    Surrey, United Kingdom</AFF>
<AFF><SUP>e</SUP>Institute of Molecular Genetics, Academy of Sciences of the Czech
    Republic, V&iacute;de&ncaron;sk&aacute; 1083, 142 20 Prague 4, Czech Republic.</AFF>
<COR>*To whom correspondence should be addressed. E-mail
    <UNL>binarova@biomed.cas.cz</UNL>; fax 420-2-4752384.</COR>
<RRH>Running title: &gamma;-Tubulin in Plant Mitosis</RRH>
```

Electronic Workflow 1.0 Summary

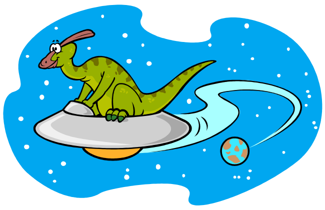
- Advantages
  - Better than paper
  - Avoided SGML tool limitations
  - Minimized Training costs
- Disadvantages
  - Three file conversions
  - Error-prone editorial workflow
  - Errors discovered in SGML creation

The Online Imperative

- Researchers today
  - Work online
  - Expect fast publication
  - Expect electronic publication before print
- Publishers must
  - Meet researcher expectations
  - Or see the best manuscripts sent to other journals

Electronic Workflow 2.0

- Submit electronic file
- Edit in Microsoft Word
  - TeX converted to Word
- Convert Word to XML
  - math and tables, too
- Typeset from XML
- Proof and typeset corrections
- Post article PDF and XML online when ready
- Print and mail issue



Electronic Workflow 2.0 Summary

- Advantages
  - Minimizes format conversions
  - Minimizes Training costs
  - Allows online publication before print

Part 3: NLM DTD

- NLM DTD background
- NLM DTD today
- Integrated XML publishing

NLM DTD background

- 2001: Many proprietary DTDs
  - Version proliferation was uncontrolled
  - The DTD "version of the week"
- Publishing partner interchange
  - Mildly frustrating
  - Numerous bi-lateral conversions
- Libraries were concerned with eJournal archives

PDF Archive Issues

- ◆ PDF Advantages
  - Preserves exact look
  - Near-universal acceptance
- ◆ PDF Disadvantages
  - Proprietary format
  - Lacked semantic markup
- ◆ Not a useful archive format

XML Archive Issues

- ◆ XML Advantages
  - Human readable
  - Standards-based
  - Non-proprietary
    - – Well… sort of
- ◆ XML Disadvantages
  - SGML/XML was a Tower of Babel
  - Publisher nuisance was a library headache

The Study

- E-Journal Archival DTD Feasibility Study
  - Harvard University E-Journal Archiving Project
  - September to December 2001
- Cross-sectional analysis of 10 scholarly DTDs
  - Analysis of conversion problems
  - Analysis of quality problems
- http://www.diglib.org/preserve/hadtdfs.pdf

Study DTD Recommendations

- Standards-based (XML, Unicode, CALS, MathML)
- Modular and extensible
- Permit multiple markup models
  - e.g. MathML/TeX math or XHTML/CALS tables
- Optional preservation of generated text
- Well documented

DTD Development Group

- Funded by
  - National Library of Medicine
  - Mellon Foundation
- Developed by
  - Jeff Beck (NCBI)
  - Deborah Lapeyre (Mulberry Technologies, Inc.)
  - Bruce Rosenblum (Inera Inc.)
- Started: April 2002; First release: April 2003

Project Methodology

- Document Analysis
  - Reviewed hundreds of journals
  - Special focus on non-life sciences titles
- DTD Analysis
  - Reviewed more than 35 journal publishing DTDs
- Feasibility Study
  - DTD Comparison
  - Conversion Issues

The NLM DTD

- ◆ Use for all scholarly disciplines
  - Document analysis focused on non-life sciences
  - e.g. Archeology, Physics, Economics, History…
- ◆ Flexible markup of special semantic cases
  - Open ended elements, e.g. <named-content>
  - Numerous attributes, e.g. list-type
- ◆ Freely customizable
  - Freely available
  - No copyright

Modular Design

- ◆ Shared core modules
  - display elements
  - list elements
  - reference elements
  - etc. (~25 modules, plus MathML and ISO entities)
- ◆ Different top level modules to define
  - Top element hierarchy
  - Parameter entities for element models
  - Degree of strictness in models

NLM DTD Today

- Used by
  - Publishers (a small sampling)
    - BioMed Central, CFA Institute, Croatian Medical Journal, CSIRO, Haworth Press, Lippincott Williams & Wilkins, Science, Society for General Microbiology, World Health Organization…
  - Aggregators
    - Atypon, Highwire, Ingenta
  - Archives
    - PubMed Central, British Library, Library of Congress, Portico (6060+ journals participating from 45 publishers)

Review Board

- Review board initiated for DTD maintenance
  - Jeff Beck (Moderator), National Library of Medicine
  - Alex Brown, Griffin Brown
  - Mark Doyle, American Physical Society
  - Beth Friedman, Data Conversion Laboratory
  - Linda Good, Cadmus Communications
  - Kathryn Henniss, HighWire Press
  - Laura Kelly, National Library of Medicine
  - Debbie Lapeyre (Tag Set Secretariat), Mulberry Technologies, Inc.
  - Nikos Markantonatos, Atypon Systems, Inc.
  - John Meyer, Portico
  - Jules Milner-Brage, HighWire Press
  - Tom Mowlam, BioMed Central
  - Evan Owens, Portico
  - Bruce Rosenblum, Inera, Inc.
  - B. Tommie Usdin, Mulberry Technologies, Inc.
- Meets ~4 times/year by conference call

NLM DTD Versions

- Version 1.0: April 2003
- Version 1.1: November 2003
- Version 2.0: August 2004
- Version 2.1: June 2005
- Version 2.2: May 2006
- Version 2.3: May 2007
- Version 3.0: Expected early 2008
  - NISO registration with 3.0

NLM DTD Suite Family

- Original
  - Green (archiving), e.g. ISSN optional
  - Blue (publishing), e.g. ISSN required
- Today
  - Green (archiving)
  - Blue (publishing)
  - Pumpkin, for NLM authoring
  - Book, for different metadata and wrappers
  - Historical Book, for digitizing older books

Integrated XML Publishing

- XML source drives
  - Composition and PDF
  - HTML web pages
  - Metadata deposits
    - PubMed Central → PubMed
    - CrossRef via CrossRef Schema or NLM DTD
  - Publisher interchange
  - Archive

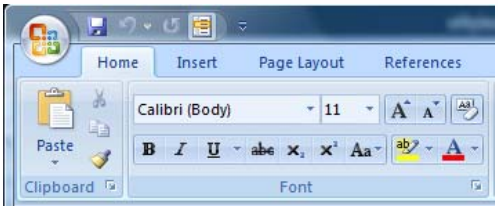
Why Adopt The NLM DTD?

- Domain neutral
- Independently developed
- Common structural framework
- Well-documented
- Continually maintained
- Widely adopted
- Standard

XML Drives Workflow 2.0

- XML is a meta language
- More important:
  - XML drives integrated workflow
  - XML drives the business processes
  - XML drives new products
  - XML drives new knowledge
  - XML drives knowledge preservation
- XML can transform scholarly publication

Appendix: Word 2007



- User interface
  - Out: Menus, Toolbars
  - In: Ribbons, Task Panes
- New features for scholars
  - Citation Manager
  - Equation Editor
- XML File Format
- New add-in model

Equation Builder

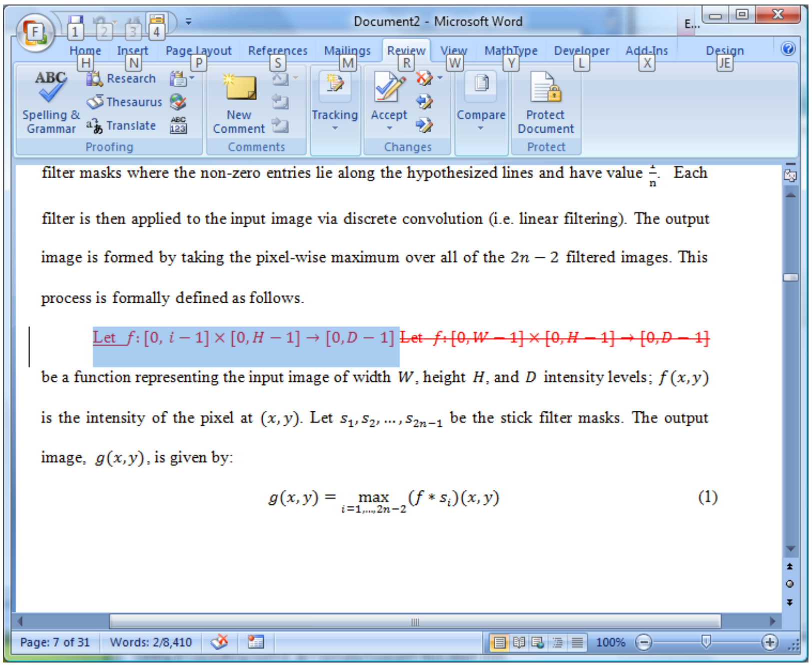
- MathType available for years
  - Equation Editor is slimmed version of MathType
- New Equation Builder based on Microsoft paper:
  - http://www.unicode.org/notes/tn28/UTN28-PlainTextMath.pdf
  - "Linear format" vs. "Built-up format"
  - Designed for ease of equation entry

Equation Builder Export Support

- MathML supported
  - Corrects our statement at XUG 2006
- No TeX support
- No graphic support (EPS or GIF)
  - May impact InDesign or Quark workflow

Equation Builder Problems

- ◆ Unreliable Save As RTF



filter masks where the non-zero entries lie along the hypothesized lines and have value $\frac{1}{n}$. Each filter is then applied to the input image via discrete convolution (i.e. linear filtering). The output image is formed by taking the pixel-wise maximum over all of the $2n-2$ filtered images. This process is formally defined as follows.

Let $f: [0, i-1] \times [0, H-1] \rightarrow [0, D-1]$ ~~Let $f: [0, W-1] \times [0, H-1] \rightarrow [0, D-1]$~~

be a function representing the input image of width W, height H, and D intensity levels; $f(x,y)$ is the intensity of the pixel at (x,y). Let $s_1, s_2, \ldots, s_{2n-1}$ be the stick filter masks. The output image, $g(x,y)$, is given by:

$$g(x,y) = \max_{i=1,\ldots,2n-2} (f * s_i)(x,y) \qquad (1)$$

Incorrect MathML Transformation

- ◆ Two equation editors produce different MathML

$$\sum_{i,j}$$

MathType	Equation Builder
<munder>	<munder>
<mo>&#x2211;</mo>	<mo>&#x2211;</mo>
<mrow>	<mi>i</mi>
<mi>i</mi>	<mo>,</mo>
<mo>,</mo>	<mi>j</mi>
<mi>j</mi>	</munder>
</mrow>	<mrow>
</munder>	</mrow>

- ◆ Test to determine if differences are significant!

Word 2007 Preparation Guide

- Short term
  - Prepare systems to receive DOCX files
  - Update author instructions
- Medium Term
  - Communicate strategy to management and staff
  - Check with suppliers and test external solutions
- Long term
  - Test and redevelop in-house macros
  - Test full workflow with Word 2007 *before* live deployment
  - Train staff *before* deployment

Questions?

Bruce Rosenblum
Inera Incorporated
+1 (617) 969 - 3053
brosenblum@inera.com
www.inera.com