

# **Quality Management for Electronic Journal Production and Archiving**

**Presented by  
Bruce D. Rosenblum  
CEO  
Inera Incorporated**

**Rudjer Boskovic Institute Library, 24 October 2005**



# DOI Stands For...



# DOI Stands For...

- ◆ **Digital Object Identifier**



# DOI Stands For...

- ◆ **Digital Object Identifier**
- ◆ **Dusty Old Issue**



# DOI Stands For...

- ◆ **Digital Object Identifier**
- ◆ **Dusty Old Issue**
- ◆ **Death Of Ink**



# Death Of Ink

**Gutenberg:** (Surveying the Frankfurt Book Fair)

**“This commodity must be as precious as gold!”**

**Gates:** **“Cheap as dirt, actually. And on its way out.**

**It’s called print. You invented it, or so history claims”**

*John Updike*

*Print: A Dialog (1995)*



# Dusty Old Issue

- Gates:** “The card catalogues of entire libraries—the bulging groaning repositories of the fading, crumbling fruit of your revolution—have been reduced to computer memory, exhaustively searched in a twinkling!”
- Gutenberg:** “...And—if I may point out a technical problem—its product never achieves autonomy from its means of delivery. A book can lie unread for a century, and all it needs to come to life is to be scanned by a literate brain. This CD-ROM of yours—what machine will be able to read it a hundred years from now?”

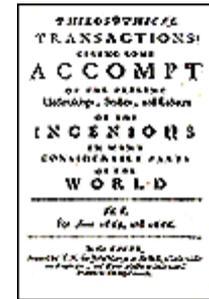
*John Updike*

*Print: A Dialog (1995)*



# A Little History

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



# Last Fifteen Years...

- ◆ **More change than the last 500**
- ◆ **Starting points are different**
  - 1990: Paper
  - 2005: Electronic files
- ◆ **Ending points are different**
  - 1990: Print
  - 2005: Print, PDF, CD-ROM, XML, HTML



# Agenda

- ◆ **New workflows, new quality issues**
- ◆ **Specific quality problems**
- ◆ **Establishing quality systems**
- ◆ **Library impact**
- ◆ **Conclusions**



# Journal Production, circa 1980

## ◆ Paper-based workflow

- Submit typewritten or longhand manuscript
- Submit paper artwork (line art, photo)
- Edit on paper
- Typeset
- Proof
- Typeset corrections
- Print



# PRODUCTION COPY

02-0074

2002

read by JB

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

A Randomized Comparison of Pump  
Venovenous Hemofiltration and Peritoneal Dialysis  
in Acute Renal Failure Associated with Sepsis and Infection

Hemofiltration and Peritoneal Dialysis in Infection - Associated Acute Renal Failure in Vietnam

MA: degrees?

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

ORDEF OK?

An: same as the Center?

1. Centre for Tropical Diseases, Cho Quan, Ho Chi Minh City, Vietnam  
2. Wellcome Trust Clinical Research Unit, Centre for Tropical Diseases, Cho Quan Hospital, Ho Chi Minh City, Vietnam  
3. Centre for Tropical Medicine, John Radcliffe Hospital, Oxford, UK  
4. Renal Unit, Churchill Hospital, Headington, Oxford, UK

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

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

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

PRODUCTION COPY

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

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

Word count: Abstract 224  
Text 2,706  
2781

UK = United Kingdom

# PRODUCTION COPY

All queries on pp  
1, 3 - 13, 21, 22  
Fig legends

# Quality Control, circa 1980

## ◆ Proofing paper

- Proofread Text
- Check special characters
- Check table data
- Check image quality



# Production Quality Issues, c. 1980

- ◆ Unclear original copy
- ◆ Poor editorial markup
- ◆ Error-prone typesetters
- ◆ Images
  - Poor original quality
  - Mishandled original
  - Halftone clarity



# New Electronic Workflow

- ◆ **Submit electronic text file**
- ◆ **Submit electronic artwork**
- ◆ **Edit electronic text**
- ◆ **Format conversion 1?**
- ◆ **Typeset**
- ◆ **Proof**
- ◆ **Typeset corrections**
- ◆ **Create Print and PDF**
- ◆ **Format conversion 2?**



# Different Workflow

- ◆ **Different starting point (data inputs)**
  - **Electronic text file vs. paper manuscript**
  - **Electronic art files vs. paper artwork**
- ◆ **Different production tactics**
  - **Reuse submitted electronic data vs. re-key**
  - **Electronic editing**
- ◆ **Different products (data outputs)**
  - **Print, PDF, XML, HTML vs. print only**



# Different Performance

## ◆ Production time

- **10 years ago: Paper**
  - Print in 6 to 12 months, maybe more
- **Today: Electronic**
  - Online in less than one day



# The Problem Summary

- ◆ **Different inputs**
  - Require new initial quality checks
- ◆ **Different production processes**
  - Require different production quality checks
- ◆ **More outputs**
  - Require new proofing systems
- ◆ **Shorter production time**



# Attacking The Problem

## ◆ New workflow

- Most publishers in transition or completed

## ◆ New quality processes

- Too busy establishing new workflow
- It's enough to get through the day
- Result: quality is reactive, not proactive



# Quality in Electronic Workflow

## ◆ Where to start?

- **Review *entire* workflow**

- Look at your desired result first
- Then look at your raw material (e.g. author manuscripts)
- Then determine optimal path from start to finish
- Build in quality checks at every step

- **Don't be afraid to scrap earlier investments**

- Some organizations are already on workflow version 3 or 4
- Software lifecycles are short



# Quality Truisms

- ◆ **Errors are more expensive to fix at each successive stage**
- ◆ **More warm bodies worsen the problem**
- ◆ **Consistent procedures yield consistent quality**
- ◆ **Continued quality requires constant vigilance**



# XML Solution Concepts

- ◆ **Quality requires documented standards**
  - Content standards
  - PDF and XML standards
  - Process standards
- ◆ **Standards require enforcement**
- ◆ **Selective automation reduces errors**
- ◆ **Training reduces errors**
- ◆ **Quality improvement requires self-assessment**



# Electronic Text Problems

## ◆ Open the file accurately

- Viruses
- Format and program version conversions (Word, Word Perfect, Other)
- Special characters & missing fonts

## ◆ The author as amateur “typesetter”

- File clean up
- Special character cleanup
  - degrees: ° ° ° °
  - Plus/minus: ± ±
  - Beta and s-set: ß ß
- Tables, equations



# Electronic Citation Problems

- ◆ **Electronic reference management tools**
  - **Incorrect templates**
  - **Incorrect data**
    - Incorrect author entry
    - Online service inaccuracies
- ◆ **Author error**
  - **See Lukić, et al. Clin Anat. 2004;17:534-539**



# Electronic Text Solutions

- ◆ **Editorial preflight group**
  - Virus Scan, File conversion
- ◆ **Editorial training**
  - Word for editors vs. Word for authors
- ◆ **Editorial tools**
  - Automate tedious tasks
  - Don't turn editors into "taggers"
  - Allow editors to concentrate on text readability
- ◆ **Editorial is the least expensive place to fix author errors!**



# Electronic Graphic Problems

## ◆ Formats

- EPS, GIF, JPEG, TIFF

and an alphabet soup of others

## ◆ Resolution

- 72 DPI
- Screen vs. printer resolution

## ◆ Color space

- Color depth
- RGB vs. CMYK



# Electronic Graphic Solutions

## ◆ Set standards

- Formats, color space, resolution, etc.

## ◆ Supplier preflight applications

- Supplier tools
- Integrated submission and peer review tools



# PDF Problems

- ◆ **Distiller version**
- ◆ **File size**
- ◆ **Graphic resolution**
  - **Print production vs. Online rendering**
- ◆ **Metadata**
  - **Title, author, keywords, etc.**



# PDF Solutions

- ◆ **Set standards**
  - **Set distiller standards**
  - **Review supplier's settings**
- ◆ **Check metadata**
  - **Visually**
  - **Electronically**
- ◆ **PDF is easy...**
  - **It just has to look good**



# **It parsed. It's done. Not!**

**“In 1999, we were grateful just for  
receiving files that parsed”**

**– Kathryn Henniss, Highwire Press**



# XML Syntax and Semantics

- ◆ XML is not just about tags...
  - Syntax is simple
  - Semantics are challenging



# XML: What's Between The Tags

- ◆ **The XML you can see**
  - Text
  - Tables
- ◆ **The XML you can't see**
  - Metadata (~25% of tags)
  - Inconsistent formatting
  - Hyperlinks
  - Reference granularity
  - Reference links
- ◆ **The “utterly useless” file**



# Metadata: Journal ID

- ◆ **What's in a name?**
  - **Genes & Genetic Systems**
  - **Genes and Genetic Systems**
- ◆ **Treated as different journals**
- ◆ **No ISSN cross-correlation**



# Metadata: Article ID

## ◆ Set standard

- Avoid special characters
- Avoid XML meta-characters (< > ")  
`<doi>10.1002/(SICI)1097-010X(19960501)275:1<20::AID-JEZ4>3.0.CO&semi;2-C</doi>`
- Avoid page number dependency

## ◆ Use and validate check digits

- PII: S0092-8674(00)00189-6

## ◆ *Never* re-key article IDs

- From editorial → composition → PDF → XML



# Metadata: Article Type

- ◆ **Accuracy (“Review” instead of “Obituary”)**
- ◆ **Case (UPPER, Title Case, Sentence case)**
- ◆ **Singular vs. Plural**
- ◆ **Why is it important?**
  - **Drives online template**
  - **Drives online table of contents**



# Inconsistent Formatting

## ◆ Degree of XML format retention

**Figure 1.** This is a caption

<b>Publisher</b>	<b>Corresponding XML</b>
Blackwell	<code>&lt;no&gt;Figure 1. &lt;/no&gt;&lt;caption&gt;&lt;p&gt;This is a caption&lt;/p&gt;&lt;/caption&gt;</code>
Elsevier	<code>&lt;no&gt;Figure 1&lt;/no&gt;&lt;caption&gt;&lt;p&gt;This is a caption&lt;/p&gt;&lt;/caption&gt;</code>
Highwire	<code>&lt;no&gt;&lt;b&gt;Figure 1.&lt;b&gt; &lt;/no&gt;&lt;caption&gt;&lt;p&gt;This is a caption&lt;/p&gt;&lt;/caption&gt;</code>

- ◆ Inconsistent formatting yields inconsistent look
- ◆ Affects citations, references, figures, tables



# Missing links: Internal

## ◆ **Uncited references**

- **Author error**
- **Name/date mismatches**
  - Accents (e.g. “Ruiz-Garcia et al. 1997” vs. “Ruiz-García et al. 1997”)
  - “l” vs. “1” (e.g. “200l” and “2001”)

## ◆ **Uncited figures, tables**

- **Cause rendering and XML failures**



# Missing Links: URLs

## ◆ Special characters

- **Word's AutoCorrect inserts incorrect special characters in URLs (hyphen vs. en-dash)**
- **Coding of “&” in URLs**
  - & vs. &amp; vs. %26
- **Residual spaces from typesetting**
  - <http://www.inera.com/resources/xml/information/DTDs/more/text/very/long/url/page.shtml>



# References: Granularity

## ◆ Print Version

Neutra, R., Shusterman, D. (1991) Hypotheses to explain the higher symptom rates observed around hazardous waste sites. *Environmental Health Perspectives* **94**, 31–38.

## ◆ Unstructured XML

```
<bib id="b7"><other-ref>Neutra, R., Shusterman, D. (1991)
  Hypotheses to explain the higher symptom rates observed around hazardous waste sites.
  <i>Environmental Health Perspectives</i> <b>94</b>,
  31&ndash;38.</other-ref></bib>
```

## ◆ Structured XML

```
<bb id="b7"><jnlref>
  <au><snm>Neutra</snm>, <fnms>R.</fnms></au>,
  <au><snm>Shusterman</snm>, <fnms>D.</fnms></au> (
  <cd year="1991">1991</cd>) <tl>Hypotheses to explain the higher
  symptom rates observed around hazardous waste sites.</tl>
  <pubtl>Environmental Health Perspectives</pubtl> <vid>94</vid>,
  <ppf>31</ppf>&ndash;<ppl>38</ppl>.</jnlref></bb>
```

## ◆ No reference linking without structure



# References: Online Links

## ◆ Active links require

- Accurate journal name, year, volume, first page
- OpenURL links fail otherwise

[http://www.crossref.org/openurl?issn=03770273&aulast=Walker  
&volume=54&spage=117&date=1983](http://www.crossref.org/openurl?issn=03770273&aulast=Walker&volume=54&spage=117&date=1983)

## ◆ Common errors

- Incorrect journal name
  - Ann. Int. Med. vs. Ann. Intern. Med.
- Incorrect page number
- Volume and Issue number confused by author



# Online Links: “Exceptional” Cases

## ◆ Non-standard information presents electronic system problems

- **Clinical Orthopaedics and Related Research**

- Issue number without volume

- **Supplements with page 1**

- Ambiguous linking; Use “S1”

- **MMWR Citations**

Centers for Disease Control and Prevention. General recommendations on immunization: recommendations of the Advisory Committee on Immunization Practices and the American Academy of Family Physicians. *MMWR* 2002;51(No. RR-2):1-35.

*MMWR vs. MMWR Morb Mortal Wkly Rep vs. MMWR Recomm. Rep vs. MMWR Surveill Summ*



NEJM -- Prophylaxis against Rabies - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Copy Paste

Address <http://content.nejm.org/cgi/content/full/351/25/2626?hits=20&where=fulltext&andorexactfulltext=and&sea> Go Links >> Google Heart Disease >>

---



# The NEW ENGLAND JOURNAL of MEDICINE

---

HOME | SEARCH | CURRENT ISSUE | PAST ISSUES | COLLECTIONS | HELP

---

You are signed in as [bruce.rosenblum](#) at [Subscriber level](#) | [Sign Out](#) | [Edit Your Information](#) | [CiteTrack Personal Alerts](#) | [Personal Archive](#)

**CLINICAL PRACTICE**

◀ [Previous](#) **Volume 351:2626-2635**    December 16, 2004    **Number 25** [Next](#) ▶

## Prophylaxis against Rabies

*Charles E. Rupprecht, V.M.D., M.S., Ph.D., and Robert V. Gibbons, M.D., M.P.H.*

This *Journal* feature begins with a case vignette highlighting a common clinical problem. Evidence supporting various strategies is then presented, followed by a review of formal guidelines, when they exist. The article ends with the authors' clinical recommendations.

**THIS ARTICLE**

- ▶ [Return to Search Result](#)
- ▶ [PDF](#)
- ▶ [PDA Full Text](#)
- ▶ [CME Exam](#)

---

**COMMENTARY**

- ▶ [Letters](#)

**A six-month-old girl presents for a "well-baby" appointment in New Jersey. The**

Internet



# MMWR Recomm Rep Citation...

NEJM -- Prophylaxis against Rabies - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address <http://content.nejm.org/cgi/content/full/351/25/2626?hits=20&where=fulltext&andorexactfulltext=and&sea...> Go Links >> Google Heart Disease >>

39. Chakravarty A. Neurologic illness following post-exposure prophylaxis with purified chick embryo cell antirabies vaccine. *J Assoc Physicians India* 2001;49:927-928. [\[Medline\]](#)
40. Briggs DJ, Dreesen DW, Nicolay U, et al. Purified Chick Embryo Cell Culture Rabies Vaccine: interchangeability with Human Diploid Cell Culture Rabies Vaccine and comparison of one versus two-dose post-exposure booster regimen for previously immunized persons. *Vaccine* 2000;19:1055-1060. [\[CrossRef\]](#) [\[ISI\]](#) [\[Medline\]](#)
41. **General recommendations on immunization:** recommendations of the Advisory Committee on Immunization Practices (ACIP) and the American Academy of Family Physicians (AAFP). *MMWR Recomm Rep* 2002;51:135. [\[Medline\]](#)
42. Siber GR, Werner BG, Halsey NA, et al. Interference of immune globulin with measles and rubella immunization. *J Pediatr* 1993;122:204-211. [\[ISI\]](#) [\[Medline\]](#)
43. Messenger SL, Smith JS, Rupprecht CE. Emerging epidemiology of bat-associated cryptic cases of rabies in humans in the United States. *Clin Infect Dis* 2002;35:738-747. [\[CrossRef\]](#) [\[ISI\]](#) [\[Medline\]](#)
44. Gibbons RV, Holman RC, Mosberg SR, Rupprecht CE. Knowledge of bat rabies and human exposure among United States cavers. *Emerg Infect Dis* 2002;8:532-534. [\[ISI\]](#) [\[Medline\]](#)
45. Human death associated with bat rabies -- California, 2003. *MMWR Morb Mortal Wkly Rep* 2004;53:33-35. [\[Medline\]](#)
46. Gibbons RV. Cryptogenic rabies, bats, and the question of aerosol transmission. *Ann Emerg Med* 2002;39:528-536. [\[CrossRef\]](#) [\[ISI\]](#) [\[Medline\]](#)
47. Update: investigation of rabies infections in organ donor and transplant recipients — Alabama, Arkansas, Oklahoma, and Texas, 2004. Atlanta: Centers for Disease Control and Prevention, 2004. (Accessed November 19, 2004, at <http://www.cdc.gov/mmwr/preview/mmwrhtml/mm53d709.htm>.)
48. Thraenhart O, Marcus I, Kreuzfelder E. Current and future immunoprophylaxis against human rabies: reduction of treatment failures and errors. *Curr Top Microbiol Immunol* 1994;187:173-194. [\[Medline\]](#)

Internet

# ...Leads to the Wrong Abstract

Entrez PubMed - Microsoft Internet Explorer

Address [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=retrieve&db=pubmed&list\\_uids=12002171&dopt=Abstract](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=retrieve&db=pubmed&list_uids=12002171&dopt=Abstract)

NCBI PubMed National Library of Medicine NLM

My NCBI [Sign In] [Register]

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search PubMed for [ ] Go Clear

Limits Preview/Index History Clipboard Details

Display Abstract Show 20 Sort by Send to

All: 1 Review: 0

1: MMWR Recomm Rep. 2002 Apr 12;51(RR-3):1-31. [Related Articles, Links](#)

[LinkOut to related resource](#)

**Prevention and control of influenza. Recommendations of the Advisory Committee on Immunization Practices (ACIP).**

[Bridges CB, Fukuda K, Uyeki TM, Cox NJ, Singleton JA; Centers for Disease Control and Prevention, Advisory Committee on Immunization Practices.](#)

Division of Viral and Rickettsial Diseases, National Center for Infectious Diseases, USA.

This report updates the 2001 recommendations by the Advisory Committee on Immunization

# XML Solutions: Editorial

## ◆ Copy editing accuracy

- More important than print-only era
- XML tagging is based on patterns

Neutra, R., Shusterman, D. (1991) Hypotheses to explain the higher symptom rates observed around hazardous waste sites. *Environmental Health Perspectives* **94**, 31–38.

## ◆ Copy editor training

- Understand the process
- Learn the tools

## ◆ Automate error-prone editorial functions

## ◆ Validate references in editorial (PubMed, CrossRef)



# XML Solutions: Selection

## ◆ DTD selection

- Meets production requirements
- Meets archive requirements

## ◆ Whose DTD?

- Publisher's
- Supplier's (Keton, Pubitem)
- Aggregator's (Highwire, Ovid)

## ◆ Neutral DTD

- NLM Journal Publishing and Archiving DTD Suite



# XML Solutions: Documentation

- ◆ **Read my mind or read my book**
- ◆ **DTD Documentation**
  - High-level structures
  - Tag usage
  - Samples
  - Publisher interpretation
- ◆ **Examples**
  - **Elsevier's Tag-by-Tag**  
[http://support.sciencedirect.com/tectext\\_sgml.shtml](http://support.sciencedirect.com/tectext_sgml.shtml)
  - **CrossRef Deposit Schema**  
<http://doi.crossref.org/doc/tech/CrossRefSchemaDocumentation.pdf>



# XML Solutions: Validation Tools

- ◆ Parsing is not enough
- ◆ W3C XML Schemas are not enough
- ◆ Build validation tools based on *your* requirements
- ◆ Deploy tools to suppliers
- ◆ Demand (don't ask) suppliers to integrate tools
- ◆ Recheck XML on receipt
  - “Trust but verify”



# Validation Tools: What To Check

- ◆ **Check what's important to you**
  - **Missing metadata fields (e.g. journal title)**
  - **Valid metadata content (ISSN is correct)**
  - **Errata cross-references original item**
  - **Uncited figures, tables, references**
  - **Too many unstructured references**
- ◆ **Check *everything* that can be checked!**
  - **Ask the engineer about the number of error messages**
- ◆ **Validate your validation tools**



# Validation Tools: Examples

- ◆ Elsevier VTool
- ◆ NLM quality check
- ◆ XML 2003
  - Piez: XSLT for Quality Checking in a Publication Workflow

[http://www.idealliance.org/papers/dx\\_xml03/papers/04-04-02/04-04-02.pdf](http://www.idealliance.org/papers/dx_xml03/papers/04-04-02/04-04-02.pdf)



# Visual Proofing Tools

- ◆ **Integrated visual proofing tools**
  - **Strategic line breaks**
  - **Color coding**
  - **Tagged and tagless proofing**



# Supplier Quality Audit

- ◆ **Walk every step of the workflow**
- ◆ **Examine quality processes**
- ◆ **Review suppliers process documentation**
- ◆ **Make them demonstrate with *your* file**
- ◆ **Verify quality tools in action**
  - **Can the operator bypass the tool?**
- ◆ **Meet the people doing the work**
  - **Quiz them on your DTD**
  - **Check to see who's doing it 6 months later**



# Offshore Suppliers

- ◆ **All process issues are the same**
  
- ◆ **What's different?**
  - **Language skills**
  - **Culture of quality**



# Publisher Challenges

- ◆ **Understand new quality problems**
- ◆ **Educate staff about quality impact**
- ◆ **Document standard processes**
- ◆ **Build tools to aid in electronic quality management**
- ◆ **Enforce usage of quality tools**



# Library Impact

- ◆ **New content sources**
- ◆ **Finding the “right” content**
- ◆ **Metadata quality**
- ◆ **Archiving**

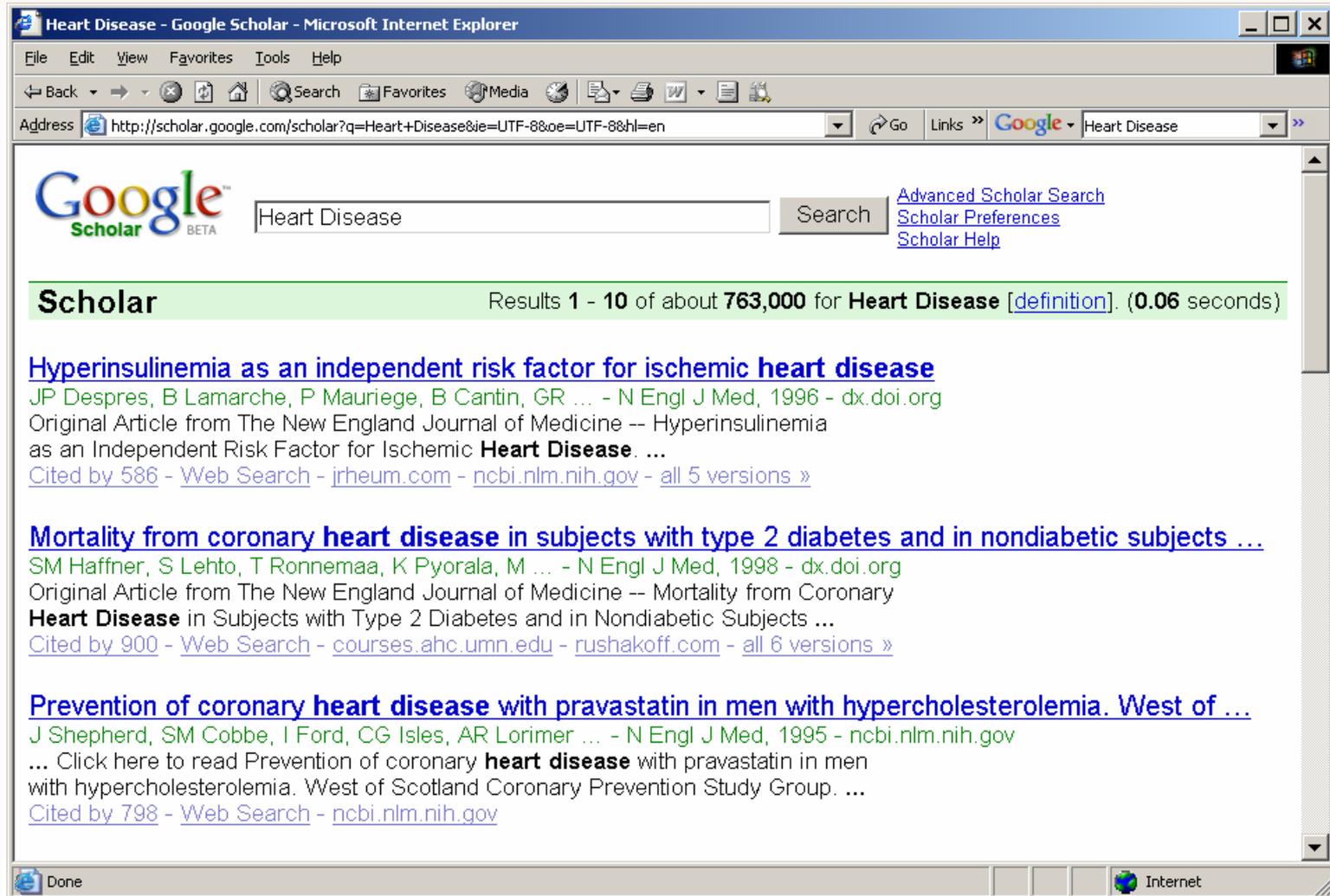


# New Content Sources

- ◆ **No longer printed books and journals**
  - Grey literature
  - Students and Google
- ◆ **Google Scholar**
  - Automatic indexing
  - Sort order
- ◆ **The dilemma: The “amateur” searcher in contrast to the trained librarian**



# Google Scholar Results



The screenshot shows a Microsoft Internet Explorer browser window with the title "Heart Disease - Google Scholar - Microsoft Internet Explorer". The address bar contains the URL "http://scholar.google.com/scholar?q=Heart+Disease&ie=UTF-8&oe=UTF-8&hl=en". The search bar contains the text "Heart Disease" and a "Search" button. To the right of the search bar are links for "Advanced Scholar Search", "Scholar Preferences", and "Scholar Help".

**Scholar** Results 1 - 10 of about 763,000 for **Heart Disease** [definition]. (0.06 seconds)

[Hyperinsulinemia as an independent risk factor for ischemic heart disease](#)  
JP Despres, B Lamarche, P Mauriege, B Cantin, GR ... - N Engl J Med, 1996 - dx.doi.org  
Original Article from The New England Journal of Medicine -- Hyperinsulinemia as an Independent Risk Factor for Ischemic **Heart Disease**. ...  
[Cited by 586](#) - [Web Search](#) - [jrheum.com](#) - [ncbi.nlm.nih.gov](#) - [all 5 versions](#) »

[Mortality from coronary heart disease in subjects with type 2 diabetes and in nondiabetic subjects ...](#)  
SM Haffner, S Lehto, T Ronnemaa, K Pyorala, M ... - N Engl J Med, 1998 - dx.doi.org  
Original Article from The New England Journal of Medicine -- Mortality from Coronary **Heart Disease** in Subjects with Type 2 Diabetes and in Nondiabetic Subjects ...  
[Cited by 900](#) - [Web Search](#) - [courses.ahc.umn.edu](#) - [rushakoff.com](#) - [all 6 versions](#) »

[Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. West of ...](#)  
J Shepherd, SM Cobbe, I Ford, CG Isles, AR Lorimer ... - N Engl J Med, 1995 - ncbi.nlm.nih.gov  
... Click here to read Prevention of coronary **heart disease** with pravastatin in men with hypercholesterolemia. West of Scotland Coronary Prevention Study Group. ...  
[Cited by 798](#) - [Web Search](#) - [ncbi.nlm.nih.gov](#)



# Finding the “Right” Content

## ◆ **Appropriate copy**

- **OpenURL and link resolvers**
- **Institutional registry**

## ◆ **Correct item**

- **Does the link resolve**
  - **Is the information correct?**
- **Does the link go to the correct item**



# Metadata Quality

- ◆ **Librarians understand their metadata requirements; publishers do not**
- ◆ **Example**
  - **PubMed**
    - Indexed by librarians
    - Publisher data proofed by librarians
    - Not perfect (e.g. accents, Greek), but often considered authoritative
  - **CrossRef**
    - Data supplied by publishers
    - Accuracy is a significant 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>`



# A JBJS Am Citation on PubMed

The screenshot shows the PubMed website interface in a Microsoft Internet Explorer browser window. The address bar displays the URL: [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list\\_uids=12377907&query\\_hl=1](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=12377907&query_hl=1). The page features the NCBI logo, the PubMed logo, and the National Library of Medicine (NLM) logo. A navigation bar includes links for All Databases, PubMed, Nucleotide, Protein, Genome, Structure, OMIM, PMC, Journals, and Books. A search bar contains the text "PubMed" and "for". Below the search bar are buttons for Limits, Preview/Index, History, Clipboard, and Details. The display settings are set to "Abstract", "Show 20", and "Sort by". The search results show one entry: "1: J Bone Joint Surg Am. 2002 Oct;84-A(10):1775-81." with a red button for "Full text article at www.ejbs.org". The article title is "Shoulder arthroplasty for arthritis after instability surgery." and the authors are "Sperling JW, Antuna SA, Sanchez-Sotelo J, Schleck C, Cofield RH." The page also includes a sidebar with links for About Entrez, Text Version, and Entrez PubMed Overview, Help | FAQ, Tutorial, New/Noteworthy, and E-Utilities.

# A JBJS Am Citation on Highwire

The screenshot shows a Microsoft Internet Explorer browser window displaying the article page for "Shoulder Arthroplasty for Arthritis After Instability Surgery" by Sperling et al. in JBJS Am, Volume 84, Issue 10, 1999. The page includes a navigation menu, a search box, and a list of links for the article.

**Shoulder Arthroplasty for Arthritis After Instability Surgery -- Sperling et al. 84 (10): 1775 - Microsoft Internet Explorer**

File Edit View Favorites Tools Help

Address <http://www.ejbs.org/cgi/content/abstract/84/10/1775> Go Links >>

**JBJS** When Surgery Depends on Accuracy. AXIEM™ Electromagnetic Navigation. **Medtronic**

**QUICK SEARCH:** [advanced]  
Author: Keyword(s):  
Go    
Year:  Vol:  Page:

**JBJS CME HELP FEEDBACK SUBSCRIPTIONS ARCHIVE SEARCH TABLE OF CONTENTS**

*The Journal of Bone and Joint Surgery (American)* 84:1775-1781 (2002)  
© 2002 [The Journal of Bone and Joint Surgery, Inc.](#)

**Scientific Article**

**Shoulder Arthroplasty for Arthritis After Instability Surgery**

John W. Sperling, MD, Samuel A. Antuna, MD,  
Joaquin Sanchez-Sotelo, MD, Cathy Schleck, BS and  
Robert H. Cofield, MD

**This Article**

- [Full Text](#)
- [Full Text \(PDF\)](#)
- [\[Supplementary Material\]](#)
- [Letters to the Editor: Submit a response](#)
- [Alert me when this article is cited](#)
- [Alert me when Letters to the Editor are posted](#)
- [Alert me if a correction is posted](#)

**Services**

- [Email this article to a friend](#)
- [Similar articles in this journal](#)
- [Similar articles in PubMed](#)
- [Alert me to new issues of the journal](#)
- [Download to citation manager](#)

Internet

## ◆ Highwire DTD definition for Volume, circa 1999

vol      NUMBER      #REQUIRED



# Highwire Citation Manager Download

JBJS -- Download Directly to Citation Manager - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address <http://www.ejbs.org/cgi/citmgr?gca=jobojos;84/10/1775> Go Links >>

The Journal of Bone & Joint Surgery  
**eJBJS**

JBJS CME HELP FEEDBACK SUBSCRIPTIONS ARCHIVE SEARCH TABLE OF CONTENTS

**QUICK SEARCH:** [advanced]  
Author: Keyword(s):  
Go    
Year:  Vol:  Page:

## Download Directly to Citation Manager

Download the citation below by selecting one of the formats on the left ([HELP](#) is below):

- [EndNote](#)  
- EndNote format  
(Mac & Win)
- [Reference Manager](#)  
\*  
- Ris format  
(Win only)

Sperling, John W., Antuna, Samuel A., Sanchez-Sotelo, Joaquin, Schleck, Cathy, Cofield, Robert H.  
**Shoulder Arthroplasty for Arthritis After Instability Surgery**  
J Bone Joint Surg Am 2002  
84: 1775-1781

Internet

# EndNote Download Record

**%A Sperling, John W.**

**%A Antuna, Samuel A.**

**%A Sanchez-Sotelo, Joaquin**

**%A Schleck, Cathy**

**%A Cofield, Robert H.**

**%T Shoulder Arthroplasty for Arthritis After Instability Surgery**

**%0 Journal Article**

**%D 2002**

**%J J Bone Joint Surg Am**

**%P 1775-1781**

**%V 84**

**%N 10**

**%U <http://www.ejbs.org/cgi/content/abstract/84/10/1775>**

**%8 October 1, 2002**



# The JBJS “Problem”

## ◆ Consider these steps

- Author finds Sperling, et al on Highwire
- Downloads to citation manager as “84”, not 84-A”
- Inserts reference into new article
- Editor fails to catch missing “-A”
- Article is published online, and fails to link because PubMed requires “84-A”
- Reader is frustrated; JBJS loses content views



# The JBJS Citation Status

## ◆ The current situation

- Highwire fixed their DTD

vol CDATA #REQUIRED

- PubMed matches on “84-A” *and* “84”, now

## ◆ However

- JBJS supplier still submits to Highwire without “-A”
- PubMed does not match “84A”
- Researchers have incorrect data in citation manager databases



# Archiving (1)

- ◆ **Who is responsible for content archive?**
  - Publishers
  - Librarians
  - Aggregators
- ◆ **Key archive issues**
  - Format (PDF, XML)
  - Access (Open, Delayed, or Dark access)
  - Redundancy



# Archiving (2)

## ◆ Who is archiving

- JSTOR/Portico
- British Library, Dutch National Library, Library of Congress
- Highwire/LOCKSS

## ◆ Archive content quality

- Dependent on quality of publisher-supplied data
- Portico's challenges



# Summary of Issues

## ◆ Key Quality Ingredients

- Standards
- Documentation
- Training
- Automation
- Validation
- Vigilance
- Self-assessment
- Manage for quality
- Culture of quality



# Conclusions

- ◆ **Electronic journal quality problems are significant**
  - Publishers need a better understanding of problems
  - Aggregators contribute to problems
  - Librarians suffer from problems
- ◆ **Greater awareness of problems by all parties is the first step to solutions**



# Questions?



**Bruce Rosenblum**  
**Inera Incorporated**  
**+1 (617) 969 - 3053**  
**[brosenblum@inera.com](mailto:brosenblum@inera.com)**  
**[www.inera.com](http://www.inera.com)**

