Investing in the Persistence of News

eDesiderata Forum 2017
Forum Agenda

Session I: The Digital News Environment and Marketplace
12:00-2:00 p.m. CT

• Conversation 1: Inside the Online News Machine
• Conversation 2: Digital News and its Scholarly Uses
• Conversation 3: The Library Supply Chain for News

Session II: Strategies & Models for Library Investment in News Access
2:00-3:00 p.m. CT

• Conversation 1: Site Licensing
• Conversation 2: Direct Library Investment in News Access and Preservation
Session I: The Digital News Environment and Marketplace
Conversation 1: Inside the Online News Machine

Bernard Reilly, Moderator
President
Center for Research Libraries
Conversation 1: Inside the Online News Machine

Evan Sandhaus
Executive Director, Knowledge & Metadata Management
The New York Times
Administering a 166-year-old Database

Tagging, Taxonomy and Better Living Through Metadata

Evan Sandhaus
Director Search, Archives, Semantics
The New York Times
@kansandhaus
On The Menu

- The New York Times
- (me)
- The Archive
- A Brief History Of Tagging
- Tagging In 2015
  - Tagging Process
  - Tagging Example
  - Tagging Applications
  - The Future of Tagging
- One More Thing
The Problem By The Numbers

Almost

60,000

Issues Published Since September 18, 1851
The Problem By The Numbers

3,500,000+

Unique Pages Printed Since September 18, 1851
The Problem By The Numbers

15,000,000+

Articles Published
September 18, 1851
A Brief History of Tagging*

* At New York Times
Meet The Tag

• A consistent string used to indicate that an authority or a subject is a principal topic of a news article.
  • Entity:
    • Person
    • Place
    • Organization
    • Title of Creative Work (Book Title, Movie Title, etc)
  • Subject
    • Everything else:
      • Eggs
      • Iraq War (2003 - 2011)
  • We’ve been tagging for a LONG TIME
INDEX TO THE NEW-YORK TIMES.
From September 1851 to September 1858.

The figures represent the number of the Times.

A

Adams (John Q.)

Abel (John) shoots John Emerson at Memphis 1804
Abbeville appointed French minister to India 184
Aberdeen (Earl of) former ministry 310. Personal affairs 1049
Abraham, George, re-arrested 2006, 2085
Abbeville appointed French minister to India 184
Academy of Music opened 900
Academy of Medicine 351
Academy of Music opened 900
Academy of Medicine 351
Academy of Medicine (N.Y.) report on Public Health 235
Academy of Medicine (N.Y.) report on Public Health 235
Academy of Sciences, London 330
Academy of Sciences, London 330
Adams (John Q.) on the Fisheries 319
Adams (John Q) on the Fisheries 319
Adams (John Q.) the death and burial 266
Adams (John Q.) the death and burial 266
March 28, 1969

The Tag Goes Digital

NY Times to Offer Information Service

NEW YORK — The New York Times has announced plans for a computerized service that will for the first time permit outsiders, as well as company employees, to have access to materials the newspaper’s information files.

The system, called the New York Times Information bank, is scheduled to go into operation early in 1971. It will utilize information gathered from the Times and some 50 other publications.

John Rothman, director of information services for the Times, said the bank would be the nation’s first fully automated general information system.

Ivan Veit, a vice president of the Times, said: “The New York Times intends to enhance its reputation, through its information retrieval system, as one of the world’s most reliable and authoritative sources of information.”

He added: “We feel the potential market for the services that the system will be capable of producing extends into many areas.

“The services could be put to invaluable use by government agencies engaged in social research, scholars preparing such major documents as doctoral dissertations, general business services conducting research in specific areas for various clients and journalists marshaling material for books and articles.

“The list could be extended to include the news and public affairs departments of radio and television networks, advertising agencies and the research arms of many philanthropic foundations.”
Tagging In 2017
Tagging Today

- The Index was retired in 2016
- We also tag articles prior to publication using supervised automation.
- Tagging powers many parts of our operation
Tagging On NYTimes.com

- Untagged News Item
- Auto Tagging System
- Taxonomists (This is also important)
  - Suggested Tags
  - Human Being (This is important)
    - Tag Suggestions

- Tagged News Item!!!!
Tagging: An Example
Chosen Totally At Random
Just for Chiefs Fans, a Welcome Taste of Kansas City in Queens

By REEVE WIEDEMANN  NOV. 17, 2013

Kansas City fans gathered at John Brown Smokehouse in Long Island City, Queens, to watch their unbeaten team play Denver. Victor J. Blue for The New York Times

Few bar-stood debates are more tiresome than arguments over which team’s fans have suffered the most. But it’s hard to imagine a season much worse than the one Kansas City Chiefs fans endured last year. On the field, the team had the N.F.L.’s worst record; off it, a starting linebacker killed his girlfriend before shooting himself in front of the coach and general manager.

By season’s end, Chiefs fans were wearing black to Arrowhead Stadium.
- **Descriptors**
  - Barbecue
  - Football

- **Locations**
  - Queens (NYC)

- **People**
  - Sandhaus, Evan (Not Really)

- **Organizations**
  - Kansas City Chiefs
  - John Brown Smokehouse (Queens, NY, Restaurant)

- **Creative Works**
  - The Birth Of Tragedy (Book) (not really)

---

Victor J. Blue for The New York Times
Tagging: Why?
And More!

- Advertising
- Analytics
- Search
- Archives
The Future

• News Events
• New Tagging System
• Linked Data
…One More Thing
Given a word or tag, return the number of times that word appeared for every decade, year, month, and day in the New York Times Archive.
IRAQ (Place)
BOBSLEDDING (Subject)

1851-01-01

1851-12-31

2012-01-01

2012-12-31

The New York Times
BUSH, GEORGE (Person)
<table>
<thead>
<tr>
<th>Date</th>
<th>NYT Archive Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1851-01-01</td>
<td>1851-12-31</td>
</tr>
<tr>
<td>2012-01-01</td>
<td>2012-12-31</td>
</tr>
</tbody>
</table>

The New York Times
COLUMBUS DAY (Subject)
Conversation 1: Inside the Online News Machine

**Philip Spiegel**  
Senior Director for Content Management Operations  
LAC Group
The Digital News Environment and Marketplace

Philip Spiegel
Senior Director, Content Management Operations
LAC Group
pspiegel@lac-group.com
Overview

LAC Group is embedded within major news broadcaster supporting a number of critical services:

- Ingest of new content into production environment
- Manage media server space and traffic to/from digital archive
- Primary administrators of MAM system
- Create new records / edit – update existing records
- Manage digital and physical order / request traffic
- Pro-actively curate content / daily content highlights into easily accessible packets for all platforms to quickly access
- Long term preservation initiatives
- Dedicated team of information specialists support daily research / information needs using a range of resources.
Content

• Primarily video but also includes still images, audio, documents, graphics, logs, transcripts and other associated assets

• Content is born digital

• We also support legacy videotape and film assets in a range of formats
Accessibility

• News media archive resources are primarily provided to support on-going, daily internal editorial/production/business needs.

• Networks want to protect their content and IP and recognize and appreciate the value and responsibility of their media archives.

• Direct public access not always available but limited content is accessible via networks stock footage divisions/licensing partner(s)… may involve research and / or usage fees.

• Some news content available via academic resources.
Approach

• Workflows have evolved with technology
• User community expects/demands fingertip - desktop access to news archive content
• Generally, over-arching approach to media library management is similar across different environments but specifics vary by organization’s operation and cultural needs / priorities
• DAM/MAM applications usually sit within or just outside (with API’s) news digital production workflow, accessed via desktop IT network.
Librarians are MAM Gods!

- MAM and Digital Archives are built on complex technologies and tools but the success of these projects require library skills and best practice.

- Regardless of asset type, assets need to be well organized, with robust, concise and standardized metadata.

- Good practice brings order to the chaos and provides comprehensive “find-ability” and transparency into the depths of the collections assets.
What About the Long Term?

• Digital Archives require on-going management and maintenance

• The responsibility of digital archive storage is (theoretically)...forever
Conversation 1: Inside the Online News Machine

Discussion and Questions