
Hearst Newspapers

300 W. 57th Street
New York, NY 10019

Request for Proposals

Digital news taxonomy audit & realignment

February 1, 2022

OVERVIEW

Hearst Newspapers seeks a consultant to audit and analyze the taxonomic infrastructure of our content management systems. The end product should be a set of actionable recommendations for overhauling our data structure, system integrations and work processes, and a roadmap for addressing years' worth of inconsistently categorized content.

Deadline to respond

February 28, 2022

OVERVIEW	1
Deadline to respond	1
GENERAL INFORMATION	3
About Hearst Newspapers	3
Purpose of RFP	3
DESIRED OUTCOMES	4
Goals	4
Deliverables	4
Work Timeline	5
RESPONSE & SELECTION PROCESS	6
Response Requirements Checklist	6
Evaluation criteria	6
Finalist presentations	7
QUESTIONNAIRE	8
A.1 PROBLEMS OBSERVED	9
Inconsistency	9
Lack of categorical frameworks	9
Manual entry	9
Legacy clutter	10
A.2 SCOPE OF WORK	11
People	11
Content Types	11
Products	12
Metadata	12
Systems	12

GENERAL INFORMATION

About Hearst Newspapers

Hearst Newspapers publishes 42 dailies and weeklies that serve a wide range of communities across the country. They include a mix print and digital, subscriber-only sites (sfchronicle.com, timesunion.com), free sites (sfgate.com, Chron.com), metered sites (timesunion.com, nhregister.com, beaumontenterprise.com), and some smaller community sites that are digital-only (bigrapidsnews.com).

One in 4 Americans lives in a community served by Hearst Newspapers, and we reach approximately 89 million readers per month.

Purpose of RFP

Hearst Newspapers is seeking competitive proposals for an audit and analysis of our data infrastructure. The primary focus will be on our digital editorial content, streamlining the metadata used across markets, deprecating old data structures, and implementing a new schema that can be maintained into the future. In addition to traditional editorial content, the scope of this project will also touch puzzles, commerce, and other non-editorial content that lives on our websites and is delivered through our CMSes.

This taxonomy overhaul is a core part of a larger replatforming initiative. We have begun work on a unified design system and universal page templates, and over the next year will implement a centralized API and new tech stack. A strong taxonomical foundation is key to the success of those endeavors, and we will devote appropriate resources to its maintenance and evolution.

DESIRED OUTCOMES

Through a comprehensive audit & analysis of our taxonomy, we hope to improve our back-end systems, boost efficiency across several teams and departments, and improve experiences for our readers.

Goals

1. Increase content viewability
2. Increase engagement, reader satisfaction and recirculation
3. Streamline work processes
4. Smoother integration between systems
5. More reliable and accessible data
6. Support replatforming
7. Make us more nimble in responding to user and business needs

Deliverables

1. Current-state audit & mapping

- Systems & processes
- Content & media types
- Data input
- Data output
- Third-party partners & feeds
- Gaps & risks

2. Analyses

- Competitive analysis
- Identify gaps & risks
- Incorporate user personas, journey mapping or other frameworks for conceptualizing a new model
- Identification and comparison of any existing industry standard taxonomies

3. Recommendations and proposed roadmap

- System updates, additions or integrations
- A new data management schema
- Specific attention to photo and rights management
- Specific attention to SEO (possibly in coordination with some of our other consultants)

- An actionable roadmap for data cleanup, migration and archival with clear processes for accomplishing this
- A forward-looking user testing and validation process
- Documentation and governance

4. Implementation support (post-delivery)

- Assist with prioritization
- Review implementation strategies
- Post-hoc review and follow-up

All deliverables must be provided electronically for sharing and archival purposes. At minimum, a live or virtual presentation will also be required for the recommendation & roadmap.

Work Timeline

	Work Phase	Deliverable
March 2022	Vendor Selection	Contract signed
March	Discovery	1. Current-state audit 2. Analysis
April	Recommendations	3. Recommendations & Proposed Roadmap
May-June	Implementation support	4. Review & feedback
Through EOY 2022	Post-hoc review and follow-up	Responses upon request

RESPONSE & SELECTION PROCESS

Response Requirements Checklist

Proposals must include	
Cover sheet	<ul style="list-style-type: none">• Company name & contact information• General information about your company - size, location, age• Brief statement of company's financial health• Primary personnel on this project• Overview of your proposal
Relevant experience	<ul style="list-style-type: none">• CV or Resume(s)• Summary of prior work• References (optional)• Work samples (optional)
Proposal	<ul style="list-style-type: none">• Detail your proposal for completing this work• Roadmap for deliverables 1-4 listed above
Questionnaire	<ul style="list-style-type: none">• Response to the 7 questions below
Financial proposal	<ul style="list-style-type: none">• Cost breakdown, including any license, storage, subcontractor or implementation costs that may arise.

Responses should arrive no later than **February 28, 2022** to jessica.parks@hearst.com.

Questions can be directed to:

Jessica Parks
Director of Product, Newsroom Tools
(267) 838-2279
jessica.parks@hearst.com

Freyja Balmer
Vice President, Product
freyja.balmer@hearst.com

Evaluation criteria

- Relevant experience
- Familiarity with news publishing
- Alignment with our user and business needs

-
- Ability to tie recommended actions to expected outcomes
 - Availability and adherence to timeline

Finalist presentations

Proposals will be evaluated on the criteria above, and finalists will be invited to deliver a presentation or demo to the HNP Product leadership team (via Zoom).

This timeline represents our best estimate, but may be subject to change. All applicants would be notified in the event of such a change.

Vendor Selection Timeline	
RFP issued	Feb. 1, 2022
Proposals due	Feb. 28, 2022
Notification of finalists	March 4, 2022
Vendor presentations	March 7-11, 2022
Vendor selected	March 18, 2022

QUESTIONNAIRE

Please answer the following questions in moderate detail. Links and examples are welcome.

1. Please describe your discovery process.
2. Please summarize the tools and formats you will use for this work.
3. Please propose a roadmap for the deliverables listed above.
4. What intermediate information could we expect prior to the final report?
5. How much time do you anticipate spending per week on this project? How available will you be for questions and feedback from our team?
6. How might human behavior and change management be incorporated into your report?
7. Do you need any additional information to compile a comprehensive proposal?

A.1 PROBLEMS OBSERVED

Hearst has a complicated patchwork taxonomy that includes sometimes overlapping and conflicting business rules. New products and systems have been added without fully deprecating the old ones, leaving behind years of dead ends, duplication and defunct data. We have observed the following problems stemming from messy taxonomy, but this is by no means a comprehensive list.

Inconsistency

Markets have historically been free to set their own standards and work practices. There has never been a comprehensive effort to standardize taxonomy across the network.

Example: Our premium sites cover largely the same traditional news categories — sports, news, business, entertainment. San Francisco classifies these into 422 different channels, while Houston has 902. In their main navigational menu, San Francisco lists 89 channels while San Antonio highlights 35. Each newspaper needs flexibility to customize taxonomy to its market, but they sometimes diverge in ineffective ways — “Cars” vs “Autos,” “Education” vs. “Schools,” etc.

Lack of categorical frameworks

Section and keyword taxonomies are powered largely by tags. There is little hierarchy, and no intersectional framework on which to build. Different metadata types need to be defined and their intersections mapped to facilitate better decision-making.

Example: The Commerce team wants to add a new content type for Product Reviews, and make it distinct from Shopping Deals, Sponsored Content, and editorial content like Food Critic reviews. Should this new content be implemented as a new subsection of shopping? Or using a tag like ‘Reviews’? A new template or revenue profile? Any of the above could deliver the reader experience desired by the Commerce team, but they would all have different implications for other ways our data is used.

Manual entry

In both the digital CMS and the print CMS, categorization is almost entirely manual and relies on front-line journalists to both know and follow the taxonomical guidelines they are given. Those guidelines may vary from market to market, even manager to manager.

Example: The Managing Editor in Albany is trying to make data-informed decisions about coverage and directs her newsroom to tag 2-3 sections and 4-8 keywords per article. The

Managing Editor in San Antonio has noticed that publishing volume correlates to higher conversions; his directives focus on article count & overall traffic, so his newsroom starts spending more time finding stories and less time producing them correctly.

Legacy clutter

Old sections and systems have not been properly deprecated, resulting in a lot of clutter both in the systems and the metadata. This leads to human error and duplication, inefficiency as staff navigate around defunct fields and tags, and downstream traffic implications.

Example: When a producer adds a new package to the homepage, they must define a “collection type.” The dropdown list has 37 choices, but only 9 are currently supported.

Example: An SEO optimization service (Botify) is identifying hundreds of thousands of batch redirects (301s) that are slowing down search engines’ ability to map our pages. Many of these are being traced to defunct site sections or blogs.

A.2 SCOPE OF WORK

The work product should encompass all of the following user stories, content types and systems.

People

Content Creators

- Reporters, editors, visual journalists, and web producers in free and premium newsrooms
- Shopping, Sponsored Content, and Performance Content (evergreen) teams

Business Operations

- Product Management & Operations
- Engineering
- Design
- Business Intelligence
- IT

Clients

- Readers of our content, ranging from new and flyby users to loyal subscribers
- Advertisers and sponsors

Content Types

Articles - any individual piece of journalism or advertising

- Traditional articles, longer narratives, briefs
- Special projects and custom interactives
- Photo essays and galleries
- Sponsored content and Shopping articles
- Wire and third-party content

Sections - pages that group together similar content

- Topic pages and section fronts
- Landing pages for puzzles and other revenue or engagement content
- Author pages and other ancillary

Homepages

- Front pages of each website

Products

The primary focus of this project is digital website taxonomy. However, the solutions explored there may be integrated to adjacent products as well:

- Print editions
- E-editions (aka “replica”)
- Mobile apps
- Newsletters

Metadata

Metadata involved in this project will include but are not limited to:

- Sites
- Sections
- Channels
- Tags
- Topics
- Keywords
- Authors
- Images
- Videos
- Galleries
- Captions
- Credits
- Design template
- Affiliate content
- Sponsored content
- Syndicated (came from a third-party)
- Entitlement (to another Hearst site)
- Feeds (to RSS, newsletters, aggregators)
- Index (for SEO)
- Timestamps
- Source
- Source system
- Semantic (AI-generated tags of people, places, orgs)
- Related Content
- Title(s)
- Description(s)
- Rights management
- Revenue Profile
- Free/premium
- Collection
- Slug
- User
- User group

Many tags are being used for more than one purpose: categorization, navigation, front-end display, distribution, rights management, analytics, syndication, and more. The framework proposed should group these into a more structured architecture.

Systems

WCM

Our proprietary digital CMS includes authoring and editing capabilities, photo and video asset management, packaging and front-end distribution. WCM is used by 650 active users managing 43 websites in 7 states. The number of active users will expand substantially as we continue the

digital-first transition in 2022. WCM intakes feeds from our print CMS as well as wire services and third-party partners.

Cue

Our print CMS was built by StiboDX (formerly CCI/Newsgate). Many of our newsrooms still do the bulk of their writing and editing in Cue, packaging pages for print and then sending feeds to WCM. This workflow will be reversed over the next year, but Cue will remain our CMS for packaging print pages.

Wordpress

A small number of our sites (mostly blogs and small community weeklies) are delivered via Wordpress instead of WCM. We have been migrating off this platform, but for ease of use and other reasons, it may remain in our CMS portfolio.

DAM

Not really a digital asset management system, ours is essentially just an ingestion service. Assets are not stored in our DAM, but are processed there before making their way into WCM and Wordpress. A re-evaluation of our DAM needs and capabilities is in scope for this taxonomy project and the broader replatforming.