Assessing the markup layer in U.S.-based digital news publishing

Bernat Ivancsics[1], Eve Washington[1], Helen Yang[1], Emily Sidnam-Mauch[2], Ayana Monroe[4], Errol Francis II[2], Joseph Bonneau[3], Kelly Caine[2], Susan E. McGregor[1]

Columbia U[1], Clemson U[2], NYU[3], UNC Chapel Hill[4]

Metadata schemas in the markup of digital news

This study analyzes the markup layer containing Schema.org, openGraph and NewsML tags in a corpus constituting 2,226 digital news stories gathered from the main pages of 742 national, local, Black, and other identity-based news organizations (3 stories each). Through this study we identify existing publishing practices and map various markup strategies that allow news stories to be "read," ranked, and distributed by curation algorithms that ultimately shape news organizations' participation in the wider platform economy of digital content distribution.

```
    </ul>                HDN.dataLayer.sharing.openGraphUrl = "https://www.sfchronicle.com/
    </section>           HDN.dataLayer.sharing.openGraphType = "article";
</div>
<script type="application/ld+json">
{ "@context" : "https://schema.org",
  "@type" : "Organization",
  "url" : "http://www.bakersfield.com",
  "sameAs" : ["https://www.facebook.com/
}
</script>
<div class="clearfix"></div></div></div>
```

Figure 1. Sample of embedded Schema.org & openGraph type markup language in local news organizations' HTML layer

Analysis 1: Does a CMS affect markup styles?

Seeking to identify the extent to which a publisher's choice of CMS appeared to influence the various metadata tags that appear on its published articles, we designed an R script that compared the tags present on web pages across all publishers using a given CMS. We then examined what percentage of publishers using a given CMS had the exact same Schema tags on their pages. Our analysis revealed that news outlets publishing with the most popularly used CMSs used the same Schema tags 80% of the time.

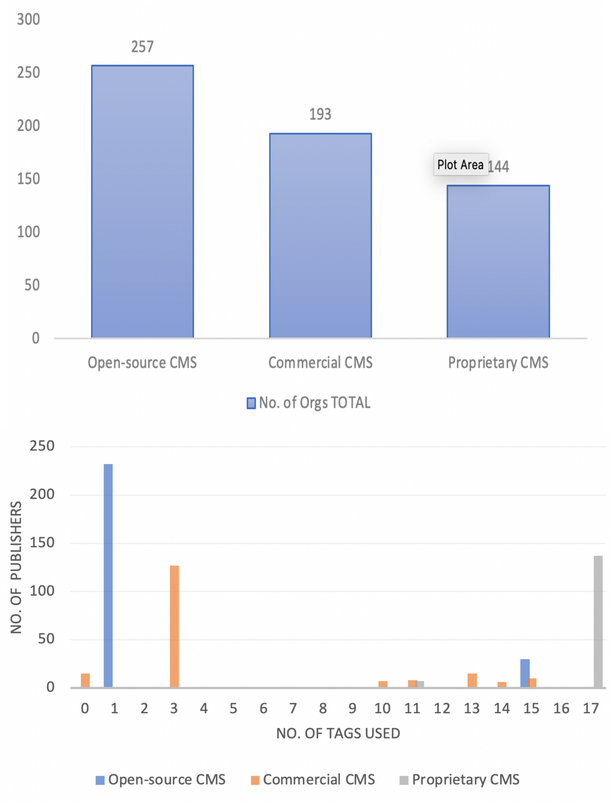

Figure 2. Top chart: No. of top CMS types by category. Bottom chart: Number of uniformly used tags by CMS type and no. of publishers

Analysis 2: How do publishers mark up their digital content?

Markup and metadata constitute the machine-readable interface that determines a digital news content's rank and classification in a list of search results or cascading content feeds, and thus play an invisible but significant role in determining the discoverability and representation of particular news stories. Since search engines and social media providers rely on the metadata standards provided by Schema.org and other content-markup systems derived from it, publishers would benefit from providing rich and consistent metadata in marking up their content. In reality, however, we have found that **smaller publishers** relying on free and open-source publication systems **rarely mark up their online news stories**, while more well-resourced publishers using unique, proprietary CMSs tend to provide richer metadata.

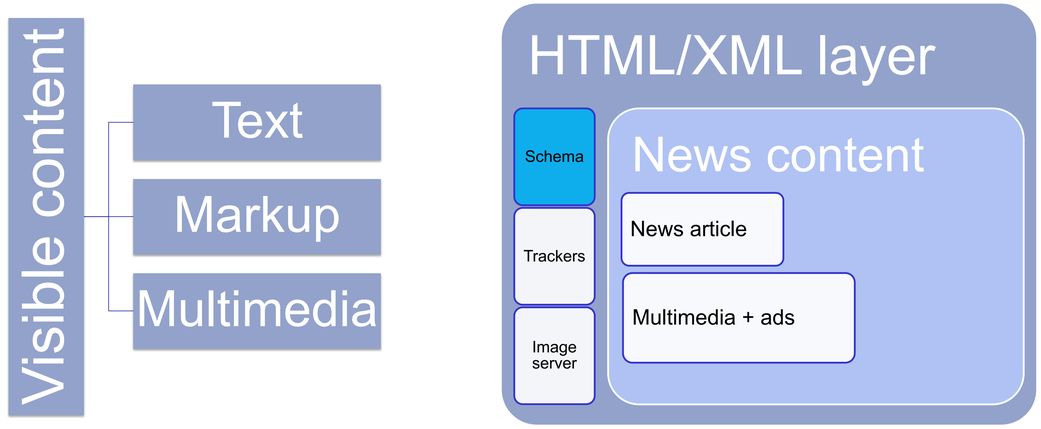

Figure 3. Models for layering technical components of digital news content

Discussion: Ways of enhancing discoverability in digital news

Could enhanced technical training in web development and digital publishing tools improve digital publishers' markup practices, with a special focus on the technical-administrative work of editors and developers who work for Black or ethnic media organizations? These research questions may be proposed within the context of the broader institutional analysis of the costs, affordances, and interoperability characteristics of the various CMSs.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 1940670, 1940713 and 1940679.

References

Ananny, M. and Crawford, K. (2015). A Liminal Press: Situating news app designers within a field of networked news production. Digital Journalism 3.2, pp. 192–208.

Guha, R. V., Brickley, D., and Macbeth, S. (2016). Schema. org: evolution of structured data on the web. Communications of the ACM 59.2, pp. 44–51.

Sunne, S. (2021). An Introduction to Schemas for Journalists. Nieman Foundation.