2024 ASIS&T Webinar
June 13, 2024

“On the Duality of Metadata: Potential Uses and Abuses of Provenance Information (DCMI)”

Presenter: Amelia Acker

webinars@asist.org
Background

• Creating data together
• Mobile phones, apps, social media platforms
• Platformization of digital cultural memory
• Critical Data Studies perspectives on metadata
• Method for “reading” metadata

What is this talk about

- What is metadata?
- What is social media metadata? Metadata manipulation?
- What does it mean to study metadata?

- the generation, access and enclosure of platform metadata are epistemic marks of data cultures
- how does metadata connect to the “platformization” of digital cultural memory
- Some future directions in provenance information and AI
Some considerations, caveats

Surfacing priors from critical information studies, archival science, and scientific data management research.

Sharing techniques of media manipulation and disinformation campaigns has consequences.
Defining metadata

• “data about data”
• Data is potential information, prefix ‘meta’ means something beyond what is given
• Naming, ordering, identifying data with intention
• More complex systems, higher rates of abstraction and hidden-ness (“seamlessness”)
• Contemporary metadata applications typically concern reconciling multiple ontologies or migrating legacy systems
• The stakes of naming data are high
Metadata as **product**, metadata **as process**

• Metadata can be an outcome or a goal of systems

• Metadata can be fundamental to processes and arrangements

• Metadata coordinates stakeholders, ensures evidence and accountability

The **social life of metadata in platforms**

- Increasingly we build subjective identity through metadata encounters
- by advancing a particular worldview, all subject taxonomies (metadata) are inherently political
- Hidden histories of metadata are about control
- Most categories from everyday life are fuzzy and not objective

2015 FB changed gender selection from dropdown lists, tags, to open text fields
Metadata are Corporate Business Strategies

- Increasingly, metadata are essential features to corporate strategies and competitive advantage

- Typically involve collecting personal information for personalization and ad targeting

- Metadata development is increasingly enclosed and not open standards processes


Media manipulation occurs when metadata are gamed, exploited, break, or drift in platforms.
Why metadata for disinformation studies?

**Opportunities**
- A way to study virality, spread, audiences, engagement across platforms
- A way of reading cross-walking phenomena “hopscotching” or “platform filtering”
- A way of anticipating how (new) platform features will be gamed, capitalized upon, networked, possibly governed
- Increasing public data literacy

**Challenges**
- Increasingly seen as private data
- Opacity from state/corporate secrecy, technical challenges and resource constraints, or ML at scale
- Accessibility of Provenance information
- Platform lifecycle updates, less access for researchers
- Metadata is mundane
Data craft

practices that create, rely, or even play with the proliferation of data on social media by engaging with new computational and algorithmic mechanisms of organization and classification.

the **craftwork** of metadata manipulation
mimicking legitimacy to users
Reading Metadata

The chart captures a step-by-step process for reading metadata from social media content. The goal for each step is to evaluate different types of "red flags"—characteristics which can, when taken together, indicate likely manipulation and coordinated inauthentic behavior. More of these red flags can be interpreted as concrete evidence on their own. However, when taken together all of the following metadata categories—including interaction between other accounts—alerts readers, researchers, and users to see the traces of manipulative data craft. By examining the interaction between accounts and their followers, steps 4 and 5 allow readers to locate evidence of manipulation and disinformation resulting from coordinated engagement strategies that generate inauthentic behavior.

1. **Account Names**
   - Double or nonexistent accounts
   - Username different from user name
   - Name contains the words "Official" or "Real"

2. **Recycled Names, Bio, Profile Pics, or Context**
   - Available on Wikipedia, internet archive, or other public platforms
   - Related to other social media accounts for news articles, with "original" content
   - Lot of duplicate content over time

3. **Content**
   - Location check-ins or posts that contradict the content of posts
   - Total absence of comments or geotags
   - Absence of other accounts tagging themselves in content
   - Persistent use of link shorteners for authentic messaging and mass content posting

4. **Followers**
   - Recent account creation
   - Followed by large numbers of suspicious accounts
   - Nonsensical comments from followers
   - Sudden growth in followers or following

5. **Authentic Interaction**
   - Inaccurate descriptions of content suggesting misuse, dubious origins, or false context
   - Automated responses from other accounts (e.g., "Thanks for the follow! Check out my webpage!")
   - Low rates of being liked or shared by other accounts
   - Content promoted by ad networks

6. **To Verify**
   - Search for images with reverse-image search tools to discover copies or original image sources
   - Compare to profiles from other platforms with similar account names
   - Search for conversations, interaction, and activity between the account and followers
   - Assess whether followers interacting with the account are actually engaging in good faith, meaningful interaction, or unique responses
   - Confirm that responses are not simply automated messages, retweets, or responses with links
manipulating platforms with the weaponization of web archives

The Millennium report URL factcheck // mimicking legitimacy to platforms
• IA and Wayback now include yellow context banners for fact checked URLs
• Labeling effects of fact checkers
• ”safe” URLs for users
AI Disinformation campaigns

• Last week Microsoft Threat Analysis Center reported on Russian influence operations, the ramping up of disinformation campaigns focused on Paris Olympics
• Images, videos, and digitally generated content from bot accounts across platforms
• Assessing digitally generated and artificially spread
• Taken down, moderated, fact checked – {missing data}
futures: provenance research

- AI Elections Accord: mitigating risks of deceptive AI content with robust provenance methods
- Coalition for Content Provenance and Authenticity C2PA industry standard for certifying digital media
- US CHIPS Act Machine actionable data management
- NSF Cyberinfrastructure for Public Access and Open Science DCL to catalyze RDM for AI computing utilities
- Scientific Data Reuse (“from creators to reusers” by Borgman and Groth; Pasquetto et al.)
THANK YOU FOR ATTENDING

A copy of the recording and a follow-up survey will be emailed within 24 hours.

Submit a webinar proposal at https://www.asist.org/meetings-events/webinars/

webinars@asist.org