

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

asis&t

Association for Information Science and Technology

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

facebook

The Obama White House
May 1, 2011 · ·

The United States Killed Osama bin Laden
May 1, 2011

President Obama addressed the nation to report to the American people and to the world that the United States conducted an operation that killed Osama bin Laden, the leader of al Qaeda and a terrorist responsible for the murder of thousands of innocent men, women, and children:
<http://go.wh.gov/oRscnZ>

127 Likes 43 Comments 3 Shares

Related Pages

- Public Figure
- Nonprofit Organization
- Democratic Party Political Party
- Nonprofit Organization
- Media/News Company
- Nonprofit Organization
- Public Figure
- Public Figure
- Politician
- Media/News Company

Acker, A. (2019). A death in the timeline: Memory and metadata in social platforms. *Journal of Critical Library and Information Studies*, 2(2).

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 hiddenness (“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

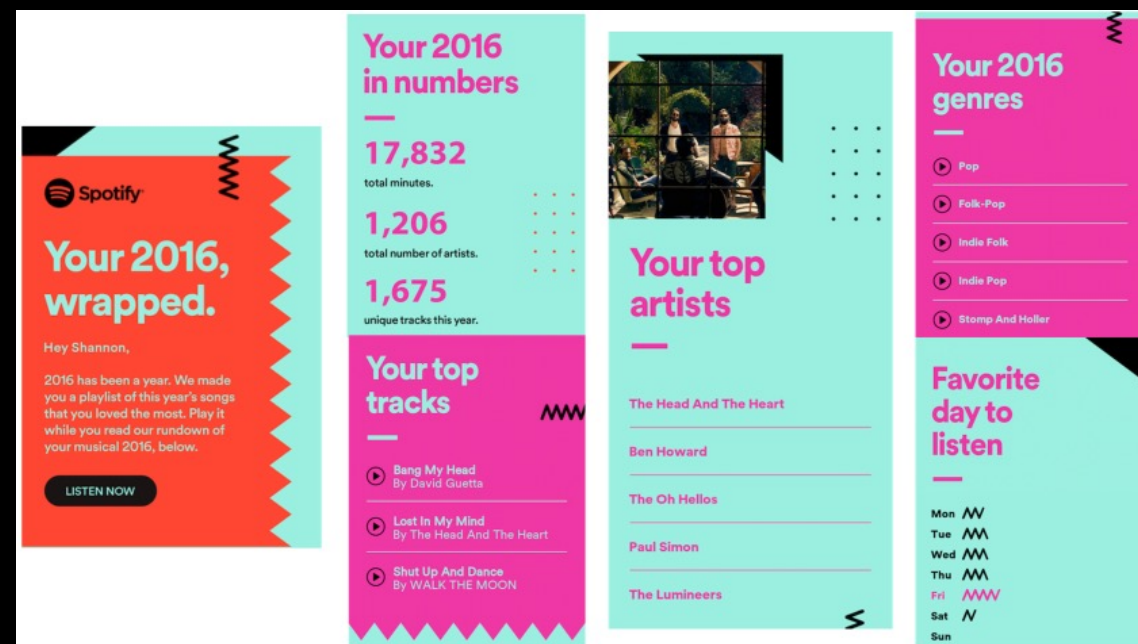
- Increasingly we build subjective identity through metadata encounters
- by advancing a particular worldview, all subject taxonomies (metadata) are inherently political
- Hidden histories of metadata are about control
- Most categories from everyday life are fuzzy and not objective

The screenshot shows the 'BASIC INFORMATION' section of a Facebook profile. The 'Birth Date' and 'Birth Year' fields are redacted with black boxes. The 'Gender' field is set to 'Custom' and is open, showing a text input field with 'Robot who doesn't know how to love' and a dropdown menu with the same text selected. Below it, a 'Preferred pronoun' dropdown is set to 'Male: "Wish him a happy birthday!"'. The page includes 'Save Changes' and 'Cancel' buttons.

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



Spotify Wrapped in 2016

*{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

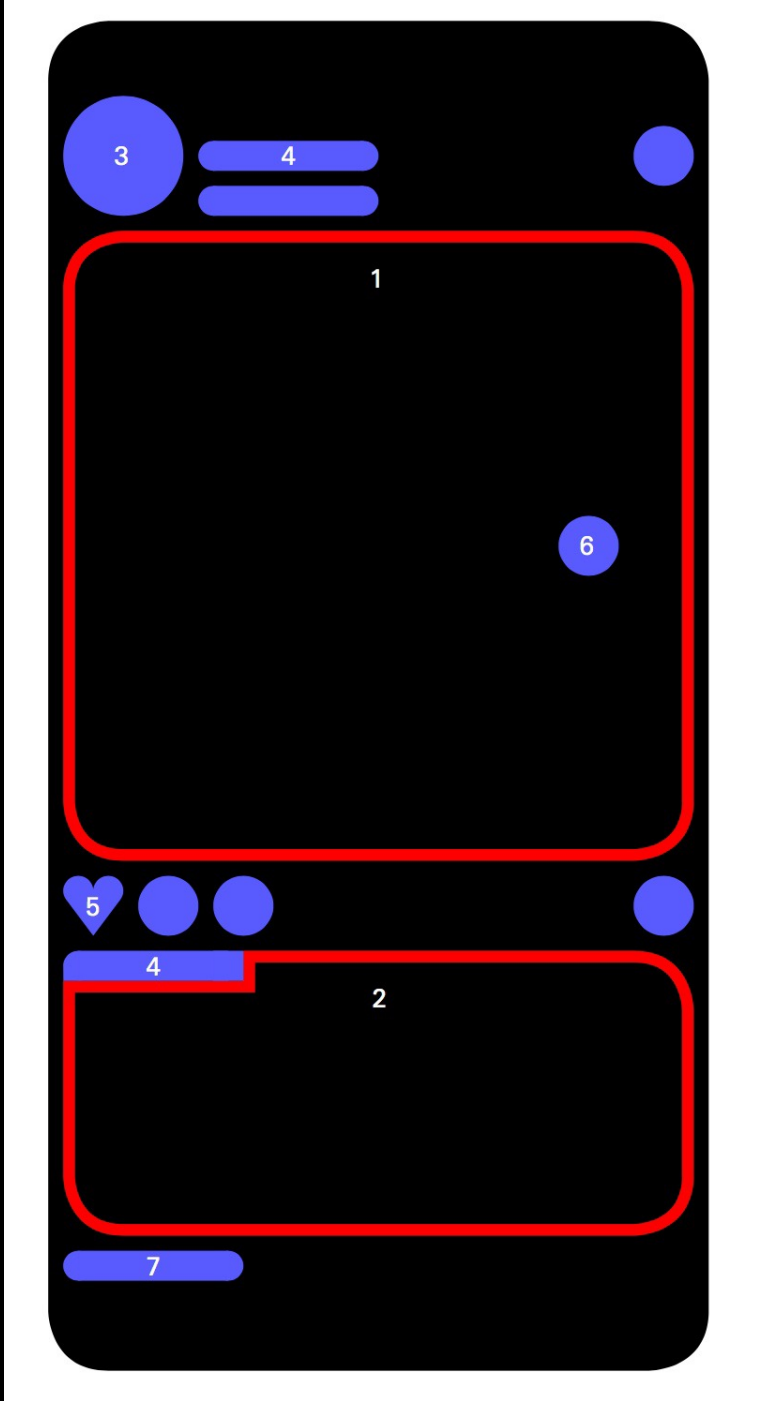
Data craft

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.

Acker, A. (2018). Data craft: The manipulation of social media metadata. Data & Society Research Institute <https://datasociety.net/library/data-craft/>

the **craftwork** of metadata manipulation

content vs metadata





babin.official

Follow

23 posts 119 followers 223 following

Rep. Brian Babin

Honored to serve citizens of #TX36 in U.S. Congress. Chairman of @HouseScience Space Subcommittee and Member of House @Transport Committee. babin.house.gov

POSTS

TAGGED



repbrianbabin

Follow

13 posts 201 followers 287 following

Brian Babin

Honored to serve citizens of #TX36 in U.S. Congress. Chairman of @HouseScience Space Subcommittee and Member of House @Transport Committee. babin.house.gov

POSTS

TAGGED



2018 Congressional Art Competition

Awards Reception & Gallery Showing
Saturday April 21, 2018 from 11:00 am - 1:00 pm
Dayton Community Center
801 South Cleveland Street - Dayton, TX - 77535

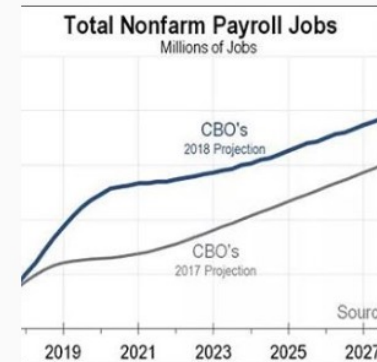
2018 THEME:
Hurricane Harvey Through Your Eyes

First Place
Award will be displayed for one year in a national exhibit, representing the people of the 36th District, in the Gallery Tunnel of the United States Capitol building in Washington, D.C. Additionally, the winner will participate in the national awards reception in Washington, D.C. Award tray pictures in Washington will be provided.

Second & Third Place
Award displayed for one year in US Representative Brian Babin's district office.

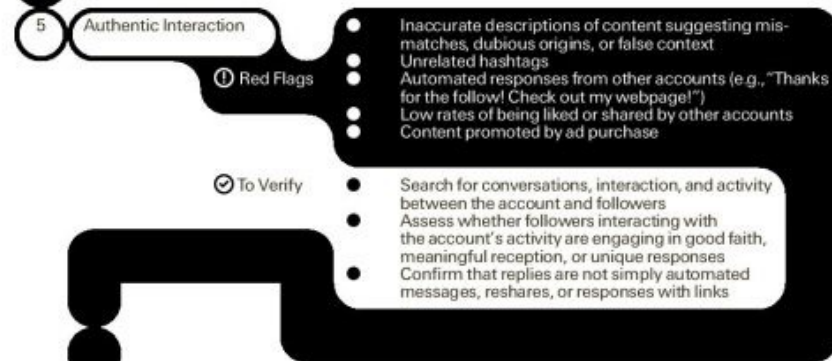
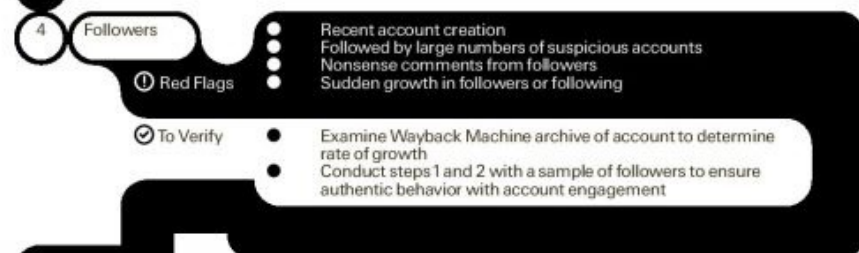
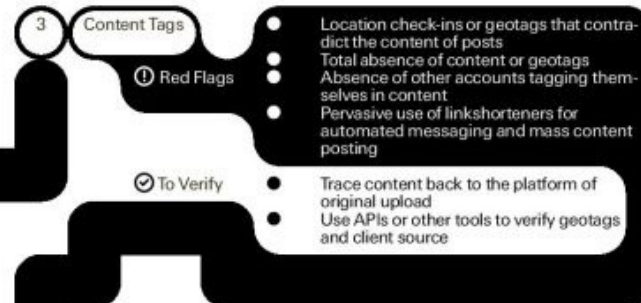
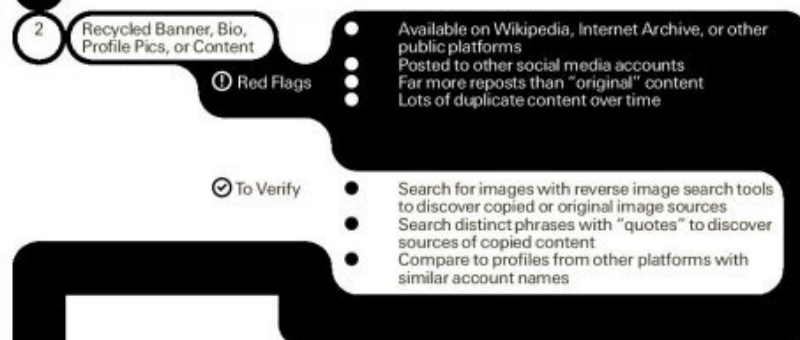
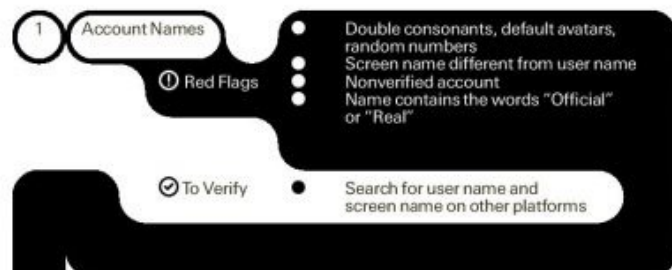
Depictions may include:
- Weather events
- First Responders
- Neighbors Helping Neighbors
- Disaster Relief Efforts

OPEN TO STUDENTS
GRADES 9 - 12
Participating students must reside within the 36th Congressional District

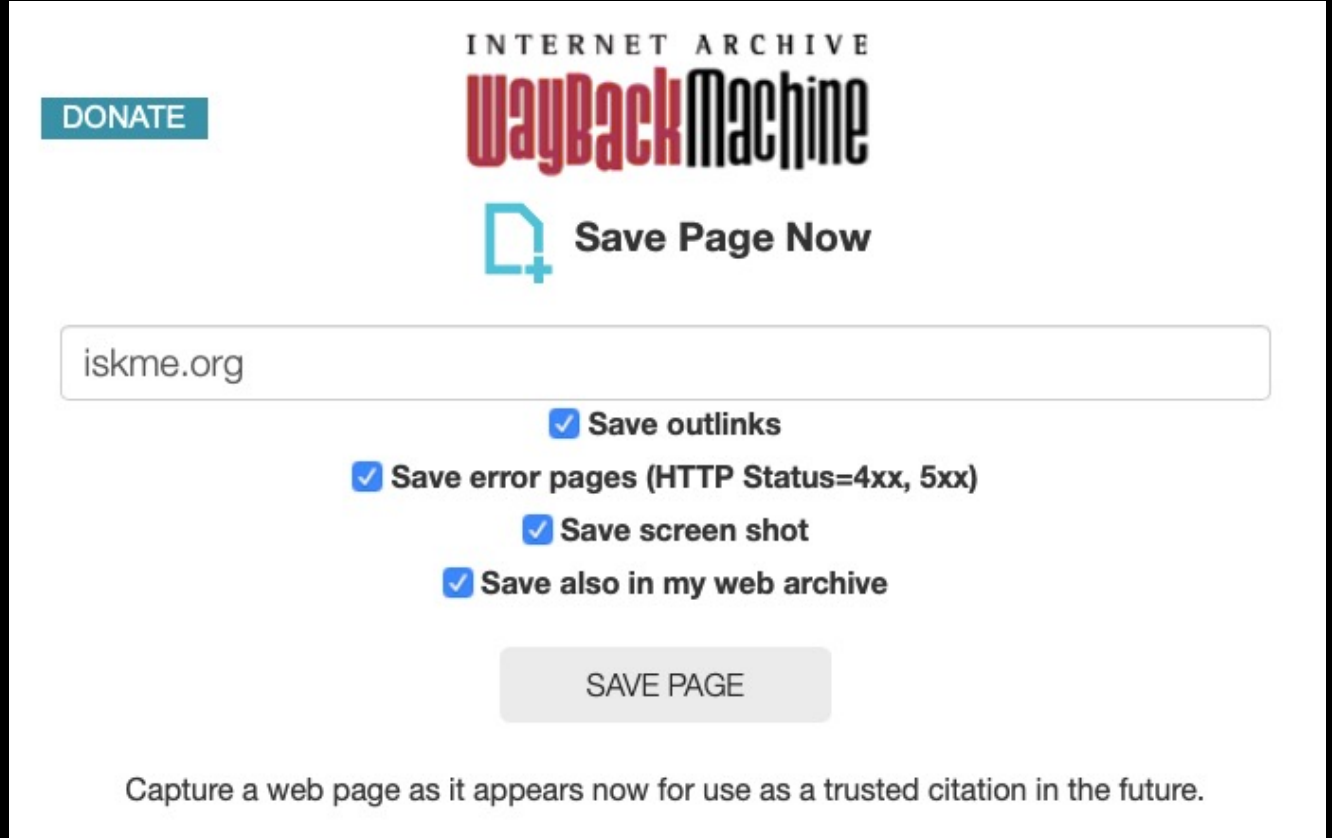



Reading Metadata

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. None 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—allows 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.



manipulating
platforms
with the
weaponization
of web archives



The screenshot shows the 'Save Page Now' interface on the Wayback Machine website. At the top left is a 'DONATE' button. The 'INTERNET ARCHIVE' logo is at the top center, with the 'WayBack Machine' logo below it. A teal icon of a document with a plus sign is to the left of the 'Save Page Now' text. Below this is a search input field containing 'iskme.org'. Underneath the input field are four checked options: 'Save outlinks', 'Save error pages (HTTP Status=4xx, 5xx)', 'Save screen shot', and 'Save also in my web archive'. A 'SAVE PAGE' button is centered below the options. At the bottom, a line of text reads: 'Capture a web page as it appears now for use as a trusted citation in the future.'

Acker, A., & Chaiet, M. (2020). The weaponization of web archives: Data craft and COVID-19 publics. *Harvard Kennedy School Misinformation Review*, 1(3).

Create Post

Mitch Chalet

Friends

Additional Reporting On This

Before you share this content, you might want to know there's additional reporting from PolitiFact and Science Feedback.

Pages and websites that repeatedly publish or share false news will see their overall distribution reduced and be restricted in other ways. Learn More

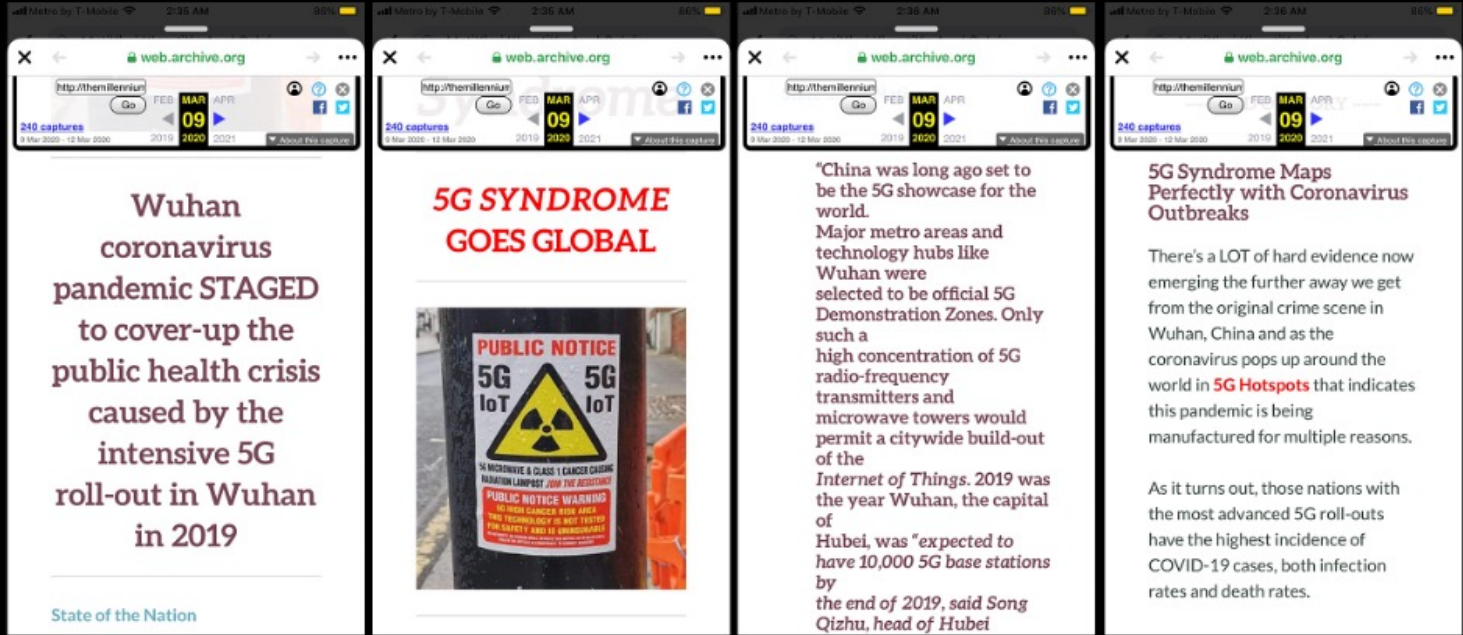
FALSE PolitiFact **Fact-Check**
 False: No evidence that coronavirus was faked to cover up '5G Syndrome'
 PolitiFact is a fact-checking website that rates the accuracy of clai...

Science Feedback **Fact-Check**
 False: The cause of COVID-19 has been definitively identified in scientific studies as a novel coronavirus, not 5G
 There is no causal link between COVID-19 and 5G technology. Th...

Share Anyway Cancel

Add to Your Post

Post



	Original URL	Archived URL
Facebook Reactions (Likes+)	7,606	16,583
Facebook Comments	6,050	14,318
Facebook Shares	5,770	10,113
Facebook Interactions (Total)	19,426	41,014

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

Internet Archive Blogs

Blog Announcements archive.org About Events Developers Donate

Fact Checks and Context for Wayback Machine Pages

Posted on [October 30, 2020](#) by [Mark Graham](#)

Fact checking organizations and origin websites sometimes have information about pages archived in the Wayback Machine. The Internet Archive has started to surface some of these annotations for Wayback Machine users. We are attempting to preserve our digital history but recognize the issues around providing access to false and misleading information coming from different sources. By providing convenient links to contextual information we hope that our patrons will better understand what they are reading in the Wayback Machine.

As an example, Politifact has investigated a claim included in a webpage that we [archived](#). [Our.news](#) has matched this URL to [the Politifact review](#) which allowed us to provide a yellow context banner for Wayback Machine patrons.



In a different case, we surfaced the discovery that [a webpage](#) is part of a disinformation campaign according to the researchers at [Graphika](#) and link to their research [report](#).

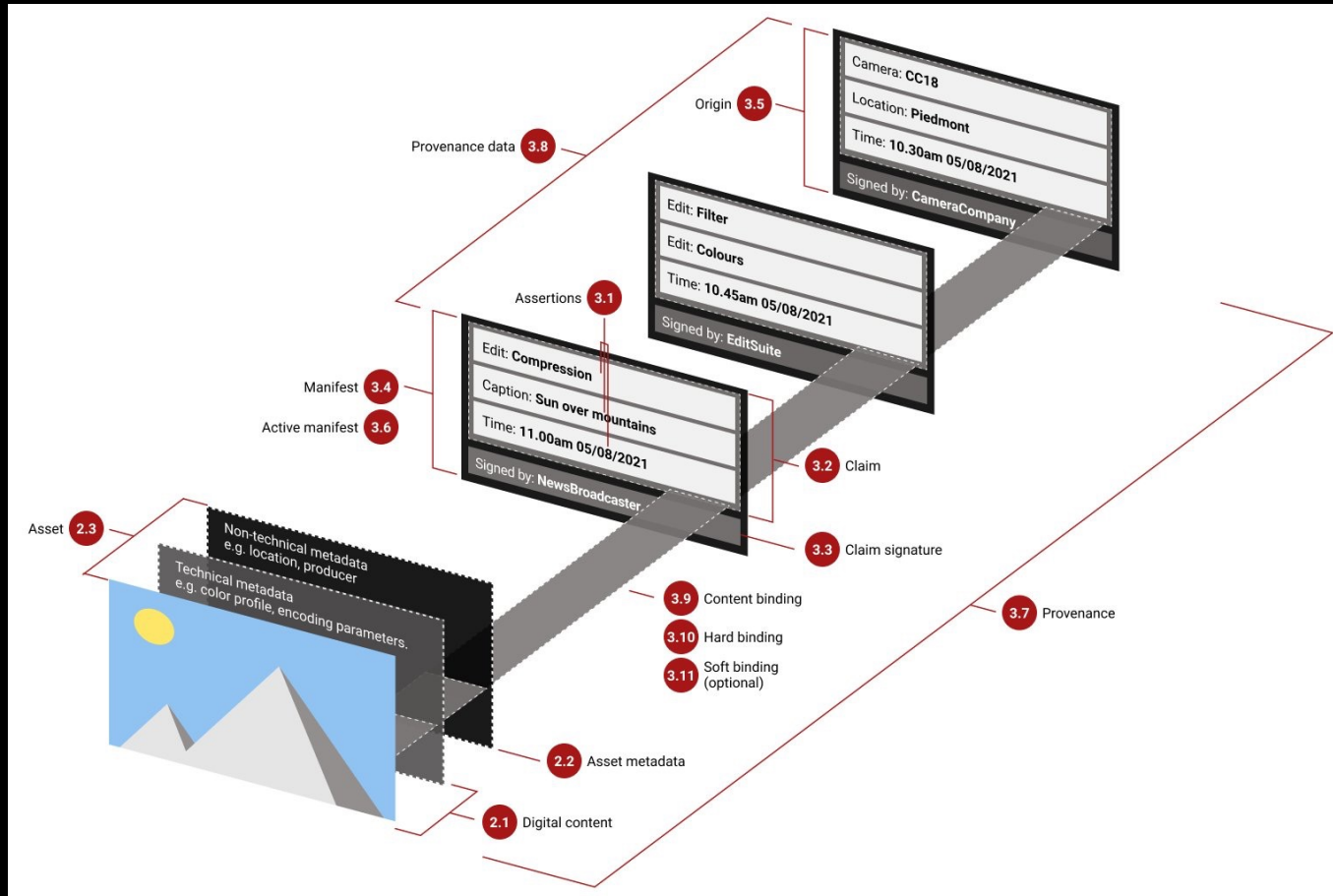
AI Disinformation campaigns

- 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.



C2PA Standard

THANK YOU FOR ATTENDING



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



webinars@asist.org

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

asis&t

Association for Information Science and Technology