

Hip Hop Narrators, Embeddable Metadata, and Nanopublications: New Ways of Thinking about Metadata and Annotated Publications for Web Discovery, Indexing, and Archiving

Jason A. Clark
Associate Professor
Head, Library Informatics & Computing
Montana State University (MSU) Library

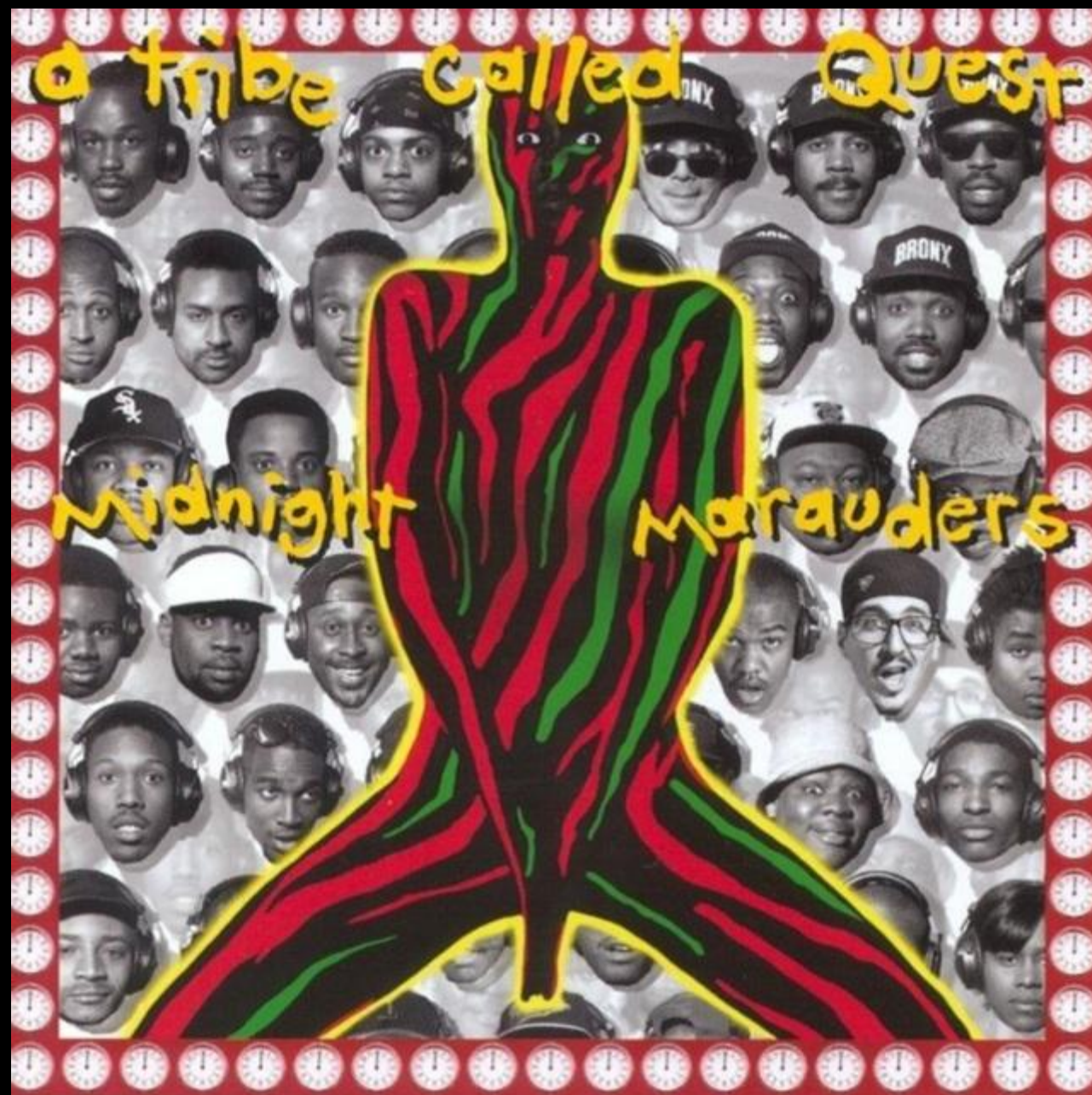
Digital Library Federation Forum
Vancouver, Canada
October 26, 2015

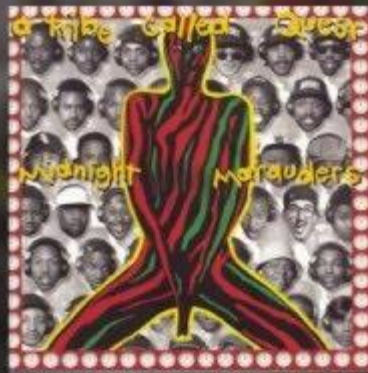
twitter as channel (#hashtag)

@jaclark #DLFforum

Overview

- WTH does Hip Hop have to do with metadata?
- Embedded metadata
- Embedded metadata in practice
- Questions

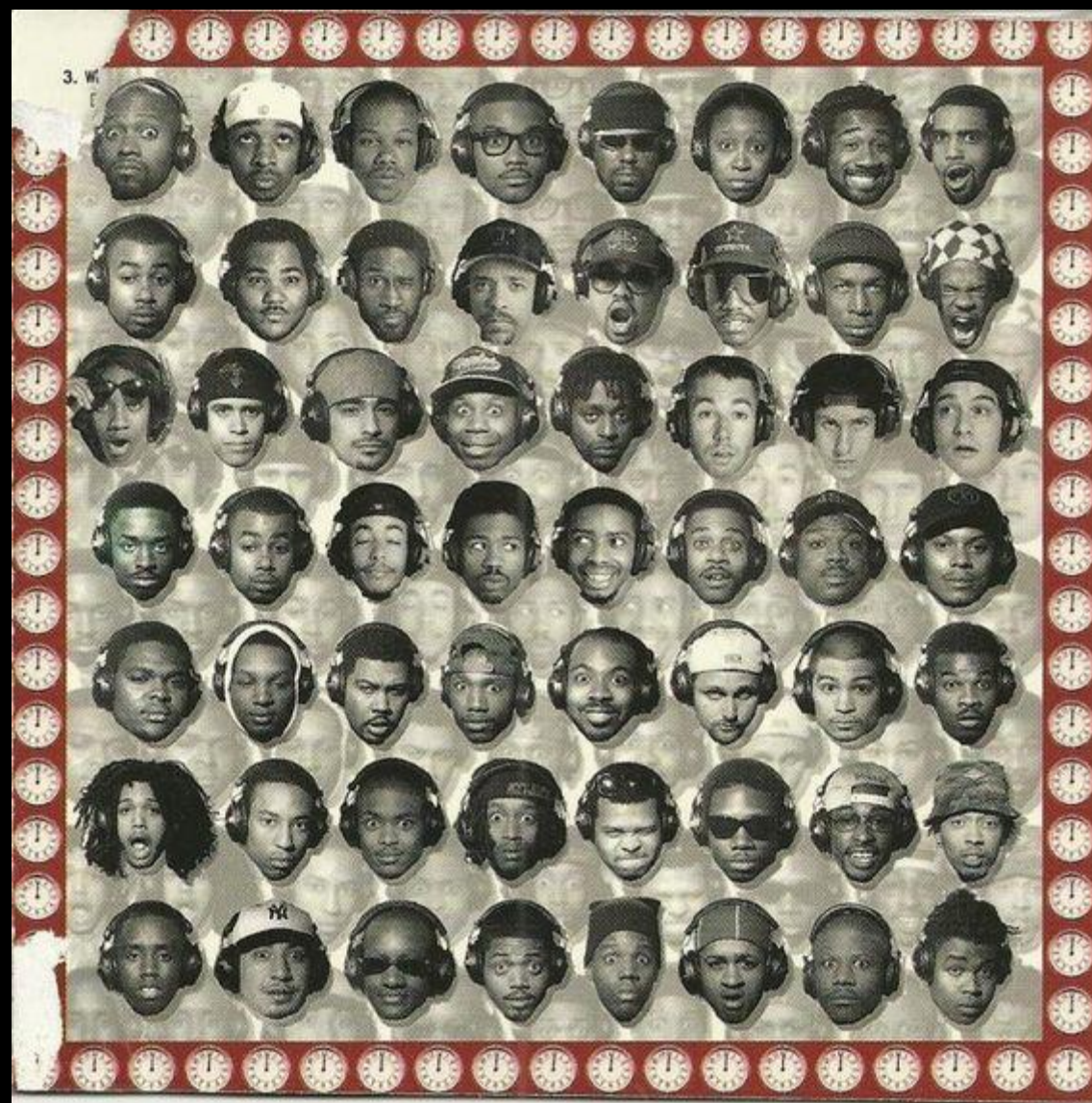




Midnight Marauders Tour Guide

A Tribe Called Quest

Midnight Marauders



"I dedicate this to all the MCs outta Queens
That goes for Onyx, LL, Run DMC
Akinyele, Nasty Nas and the Extra P
You need a chart, straight up and down
man, there ain't no other
Nuff respect to all my peeps that made the
album cover."

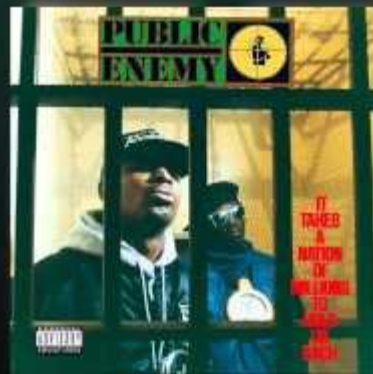
from "God Lives Through" on Midnight Marauders

<http://genius.com/A-tribe-called-quest-god-lives-through-lyrics>

A culture of attribution

embedded attribution

“citations” inline at the album
and track level



Terminator X To The Edge Of Panic

Public Enemy

It Takes A Nation Of Millions To Hold
Us Back

Embedded metadata

Metadata that resides in the
object itself

Precedence

Adobe's XMP for .pdfs EXIF for media

<https://en.wikipedia.org/wiki/Exif>

https://en.wikipedia.org/wiki/Extensible_Metadata_Platform

All rights reserved. No part of this book may be reproduced in any form or by any electronic or mechanical means, including information storage and retrieval systems, without written permission from the publisher, except by a reviewer who may quote passages in a review.

British Library Cataloguing in Publication Information Available

Library of Congress Cataloging-in-Publication Data

Clark, Jason A.

Responsive web design in practice / Jason A. Clark.

pages cm.

Includes bibliographical references and index.

ISBN 978-1-4422-4368-2 (cloth : alk. paper) — ISBN 978-1-4422-4369-9 (pbk. : alk. paper) —

ISBN 978-1-4422-4370-5 (ebook)

1. Web site development. 2. Web sites—Design. 3. Library Web sites. I. Title.

TK5105.888.C5425 2015

006.7—dc23

2015013754

 TM The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences Permanence of Paper for Printed Library Materials, ANSI/NISO Z39.48-1992.

Printed in the United States of America

Cataloging in Publication

www.loc.gov/publish/cip/techinfo/cipdata.html

Embedded Metadata in Practice

New affordances in the web
environment favor embedded
metadata



RANGE SCIENCE INFORMATION SYSTEM (RSIS)

BROWSE

[Article Titles](#)
[Journal Titles](#)
[MLRA Ecoregions](#)
[Vegetation Types](#)
[AGROVOC Subjects](#)
[Keywords](#)

ABOUT

Article citations on: riparian ecosystems, rangeland management, range environments.

- [+ About RSIS Article Types](#)
- [+ About RSIS Citation Database and Sponsors](#)
- [+ More About Searching](#)



sagegrouse chicks

SAMPLE CITATIONS

Predicting plant species' responses to grazing

P. A. Vesk, and M. Westoby in *Journal of Applied Ecology* (2001)
 Keyword(s): Australia, consistency, ecological synthesis, grazing intensity, increaser,...

Fish and grazing relationships: The facts and some pleas

J. N. Rinne in *Fisheries* (1999)
 Keyword(s): grazing, fish, fish-grazing management, southwest, riparian ecosystems, wat...

Restoration of a native shrubland impacted by exotic grasses, frequent fire, and nitrogen deposition in southern California

N. K. Cione, P. E. Padgett, and E. B. Allen in *Restoration Ecology* (2002)
 Keyword(s): coastal sage scrub, historical grazing, herbicide selectivity, mulch, nitro...

CREATED AND MAINTAINED BY



A digital collection built for:

discovery

archiveability

agency

schema.org web page types

+

manifest.json

+

item level nanopub

A prescribed architecture for
discovery

What is Schema.org?

This site provides a collection of schemas that webmasters can use to markup HTML pages in ways recognized by major search providers, and that can also be used for structured data interoperability (e.g. in JSON). Search engines including Bing, Google, Yahoo! and Yandex rely on this markup to improve the display of search results, making it easier for people to find the right Web pages.

Many sites are generated from structured data, which is often stored in databases. When this data is formatted into HTML, it becomes very difficult to recover the original structured data. Many applications, especially search engines, can benefit greatly from direct access to this structured data. On-page markup enables search engines to understand the information on web pages and provide richer search results in order to make it easier for users to find relevant information on the web. Markup can also enable new tools and applications that make use of the structure.

A shared markup vocabulary makes it easier for webmasters to decide on a markup schema and get the maximum benefit for their efforts. So, in the spirit of sitemaps.org, search engines have come together to provide a shared collection of schemas that webmasters can use.

We invite you to [get started!](#)

View our blog at blog.schema.org.

CollectionPage, AboutPage, ItemPage

Index page

```
<link rel="canonical" href="http://arc.lib.montana.edu/schultz-0010/" />
```

...

```
<div id="main" vocab="http://schema.org/" typeof="CollectionPage">
```

Topics page

```
<div id="main" vocab="http://schema.org/" typeof="AboutPage">
```

```
<meta property="isPartOf" content="http://arc.lib.montana.edu/schultz-0010/" />
```

Item page

```
<div id="main" vocab="http://schema.org/" typeof="ItemPage">
```

```
<meta property="isPartOf" content="http://arc.lib.montana.edu/schultz-0010/" />
```

A web app manifest for archiving

A manifest.json file

<https://www.w3.org/TR/appmanifest/>


```
1  {
2    "default_locale": "en",
3    "lang": "en-US",
4    "dir": "ltr",
5    "background_color": "#f8f8f6",
6    "name": "Range Science Information System (RSIS)",
7    "short_name": "RSIS",
8    "description": "The Range Science Research Information System (RSIS) has over 1,400 bibliographic citations to journal articles",
9    "related_applications": [
10     {
11       "platform": "web",
12       "url": "https://github.com/jasonclark/clark-website"
13     }
14   ],
15   "topic": [
16     {
17       "name": "Semantic Web",
18       "uri": "http://dbpedia.org/resource/Semantic_Web"
19     },
20     {
21       "name": "Open Web Platform",
22       "uri": "http://dbpedia.org/resource/Open_Web_Platform"
23     },
24     {
25       "name": "User interface design",
26       "uri": "http://dbpedia.org/resource/User_interface_design"
27     },
28     {
29       "name": "Javascript",
```

Activity stream metadata for
agency

```
1  {
2    "@context": "http://www.w3.org/ns/activitystreams",
3    "name": "Nanopub – Ecological approaches to riparian restoration in northeast Oregon",
4    "id": "http://arc.lib.montana.edu/range-science/item/619",
5    "published": "2014-12-12T12:12:12Z",
6    "type": "Collection",
7    "totalItems": 4,
8    "partOf": "Restoration and Management Notes",
9    "items": [
10     {
11       "type": "Article",
12       "name": "Ecological approaches to riparian restoration in northeast Oregon",
13       "actor": "S. R. Senock, D. M. Anderson, L. W. Murray, and G. B. Donart",
14       "id": "http://dx.doi.org/10.2307/4002860",
15       "object": "http://dx.doi.org/10.2307/4002860"
16     },
17     {
18       "type": "Note",
19       "name": "Article Summary / Main Points",
20       "actor": "RSIS editors",
21       "content": "Senock, et. al. look at how...",
22       "id": "http://arc.lib.montana.edu/range-science/item/619#summary",
23       "object": "http://arc.lib.montana.edu/range-science/item/619#summary"
24     },
25     {
26       "type": "Note",
27       "name": "Summary of Methods",
28       "actor": "RSIS editors",
29       "content": "The methods used in ...",
```

a nanopub expression for each
item

<https://www.w3.org/TR/2016/CR-activitystreams-core-20160906/>

Next Steps

- Continued testing with web crawler and web archiving tools
- Continue to refine the activity stream metadata
- Test utility of activity stream metadata as a browse point

Flow

<http://genius.com/discussions/7182-What-exactly-does-flow-and-delivery-mean>

Working with the “Flow” of the Web

There is a beat on the web.

How do we build a metadata
flow that works with it?

Acknowledgements

I would like to thank Montana State University and my excellent colleagues at MSU Library for supporting my research.

I am distributing this work under the Attribution-Noncommercial-Share Alike 3.0 Unported license.

Code

.html and .json examples
web crawler .py

<https://github.com/jasonclark/metadata-nanopub>

Questions?

@jaclark

www.jasonclark.info

