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
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
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.
TK5105.888.C5425 2015
006.7—dc23
2015013754


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
https://arc.lib.montana.edu/range-science/
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.
Index page

```html
<link rel="canonical" href="http://arc.lib.montana.edu/schultz-0010/"/>
...
<div id="main" vocab="http://schema.org/" typeof="CollectionPage">

Topics page

```html
<div id="main" vocab="http://schema.org/" typeof="AboutPage">
<meta property="isPartOf" content="http://arc.lib.montana.edu/schultz-0010/"/>

Item page

```html
<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/
"default_locale": "en",
"lang": "en-US",
"dir": "ltr",
"background_color": "#f8f8f6",
"name": "Range Science Information System (RSIS)",
"short_name": "RSIS",
"description": "The Range Science Research Information System (RSIS) has over 1,400 bibliographic citations to journal articles",
"related_applications": [
    {
        "platform": "web",
        "url": "https://github.com/jasonclark/clark-website"
    }
],
"topic": [
    {
        "name": "Semantic Web",
        "uri": "http://dbpedia.org/resource/Semantic_Web"
    },
    {
        "name": "Open Web Platform",
        "uri": "http://dbpedia.org/resource/Open_Web_Platform"
    },
    {
        "name": "User interface design",
        "uri": "http://dbpedia.org/resource/User_interface_design"
    },
    {
        "name": "Javascript",
        "uri": "http://dbpedia.org/resource/Javascript"
    }
]
Activity stream metadata for agency
Nanopub - Ecological approaches to riparian restoration in northeast Oregon


http://dx.doi.org/10.2307/4002860

Article Summary / Main Points

Senock, et. al. look at how...

http://arc.lib.montana.edu/range-science/item/619#summary

Summary of Methods

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

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