Collections as/are Data[sets]: Connecting Research Data Management to Digital Collections

Jason A. Clark @jaclark

Montana State University
Research Data Access and Preservation Summit 2020

Outline

Datasets + Collection Worldviews

Collections are data

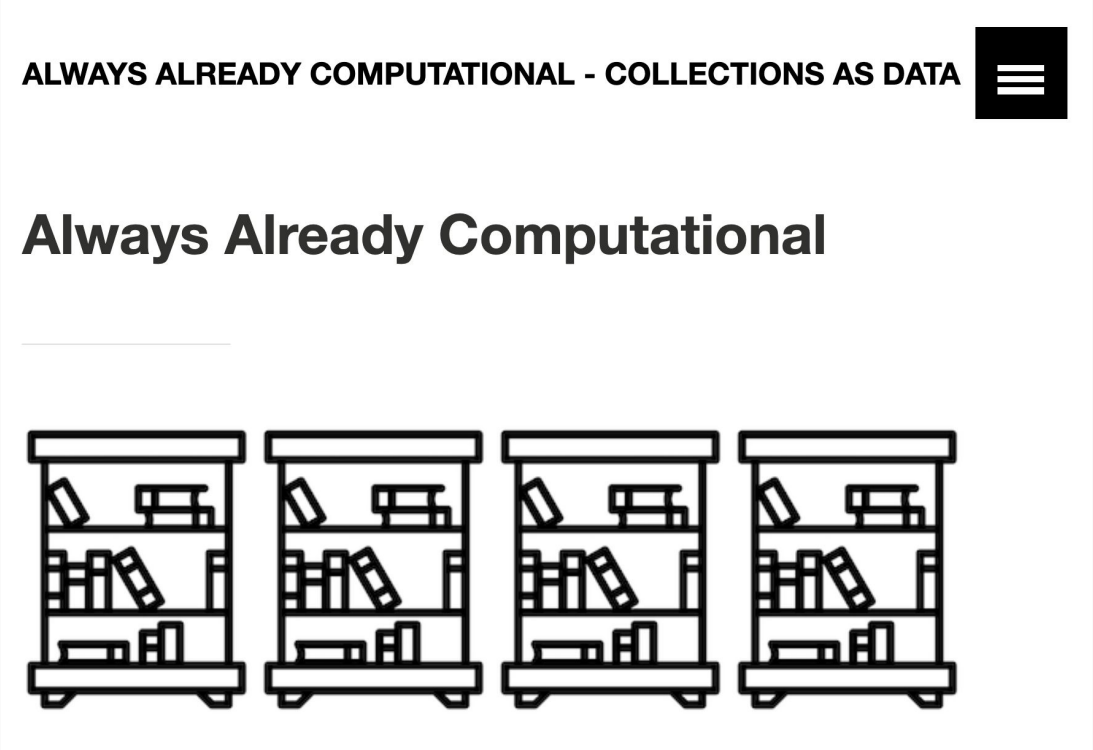
Benefits

Standards + Prototypes

Implications

Research Motivation

Research Motivation - Collections as/are Dataset[s]



Thomas Padilla

@thomasgpadilla

Research Motivation - Collections as/are Dataset[s]

A Dataset is a Worldview

On subjective data, why datasets should expire, & data sabotage.

Hannah Davis Follow

Mar 5 · 6 min read

This is a slightly expanded version of a talk given at the Library of Congress in September 2019.

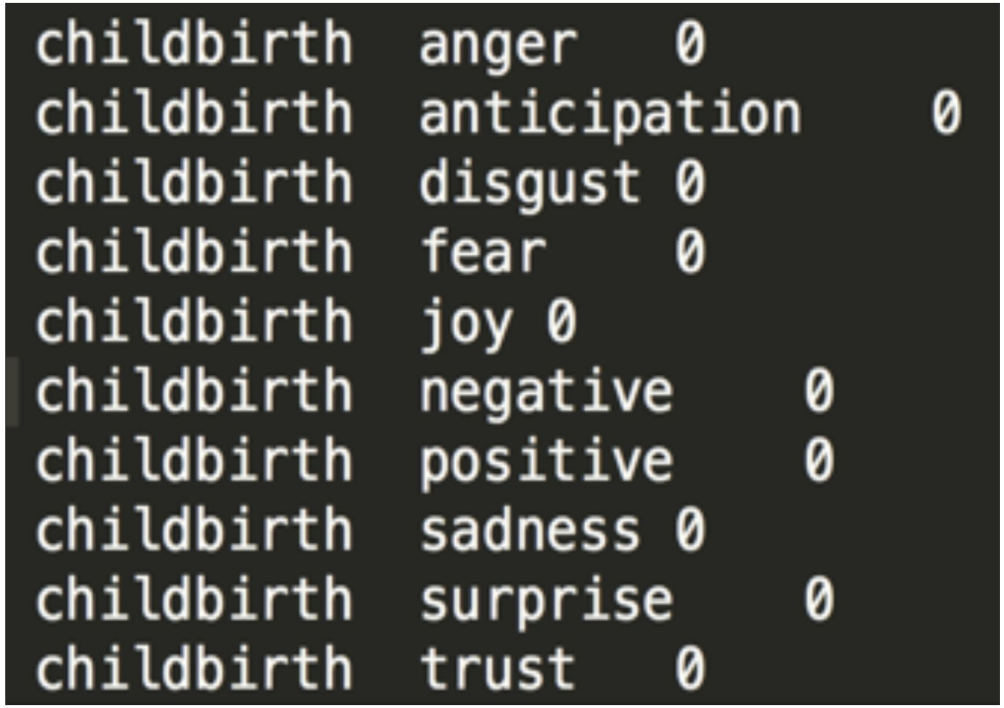
A Dataset is a Worldview

My name is Hannah Davis and I'm a research artist, generative musician, and data scientist working with machine learning. Most of my work deals with ideas around emotional data, algorithmic composition, and dataset creation.

Hannah Davis

@ahandvanish

Research Motivation - Collections as/are Dataset[s]



```
childbirth  anger    0
childbirth  anticipation    0
childbirth  disgust 0
childbirth  fear     0
childbirth  joy 0
childbirth  negative     0
childbirth  positive     0
childbirth  sadness 0
childbirth  surprise     0
childbirth  trust    0
```

The word "childbirth" had been tagged as being completely unemotional.

Research Motivation - Collections as/are Dataset[s]

"This led to a perspective that has informed all of my work since: a dataset is a worldview. It **encompasses the worldview of the people who scrape and collect the data, whether they're researchers, artists, or companies. It encompasses the worldview of the labelers**, whether they labeled the data manually, unknowingly, or through a third party service like Mechanical Turk, which comes with its own demographic biases."

Hannah Davis (2019)

https://towardsdatascience.com/a-dataset-is-a-worldview-5328216dd44d

Connecting Worldviews

MONTANA STATE UNIVERSITY

Jason A. Clark

RDAP 2020

Collections as/are Dataset[s]

Collections Worldview	Datasets Worldview
Findability	Reproducibility
Discovery	Sharing + Reuse
Inventory	Preservation
Linkable	Citeable

Connecting digital collection worldviews to dataset worldviews helps us realize that collections are datasets.

Benefits?

Benefits

Manifest for search engines

Metadata

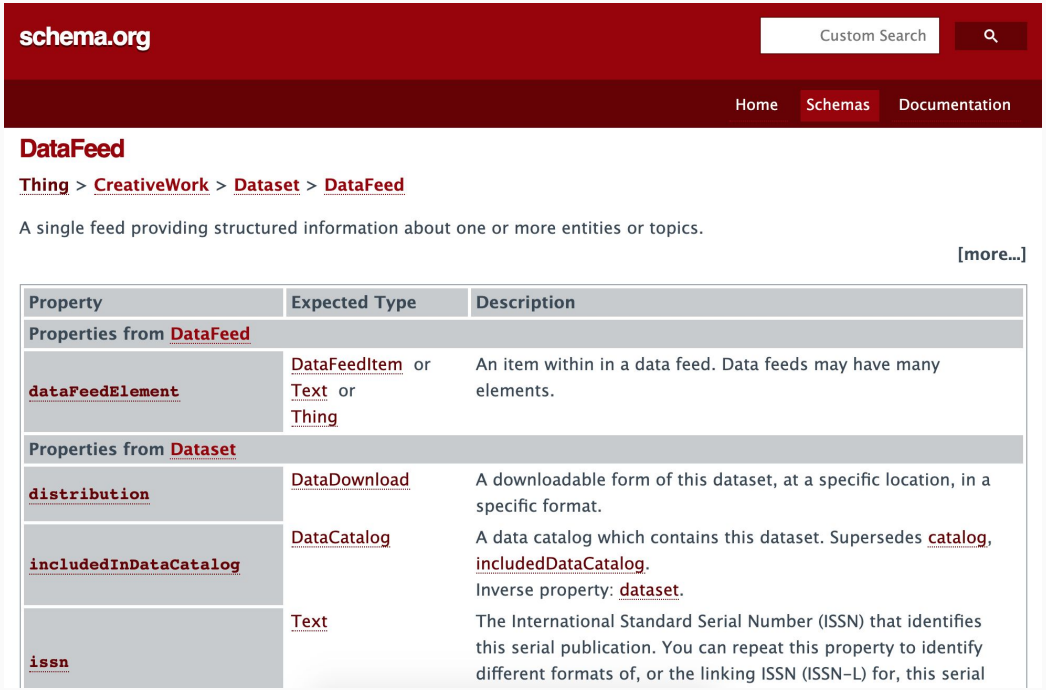
Reuse

Preservation

Citeable

Standards + Prototypes

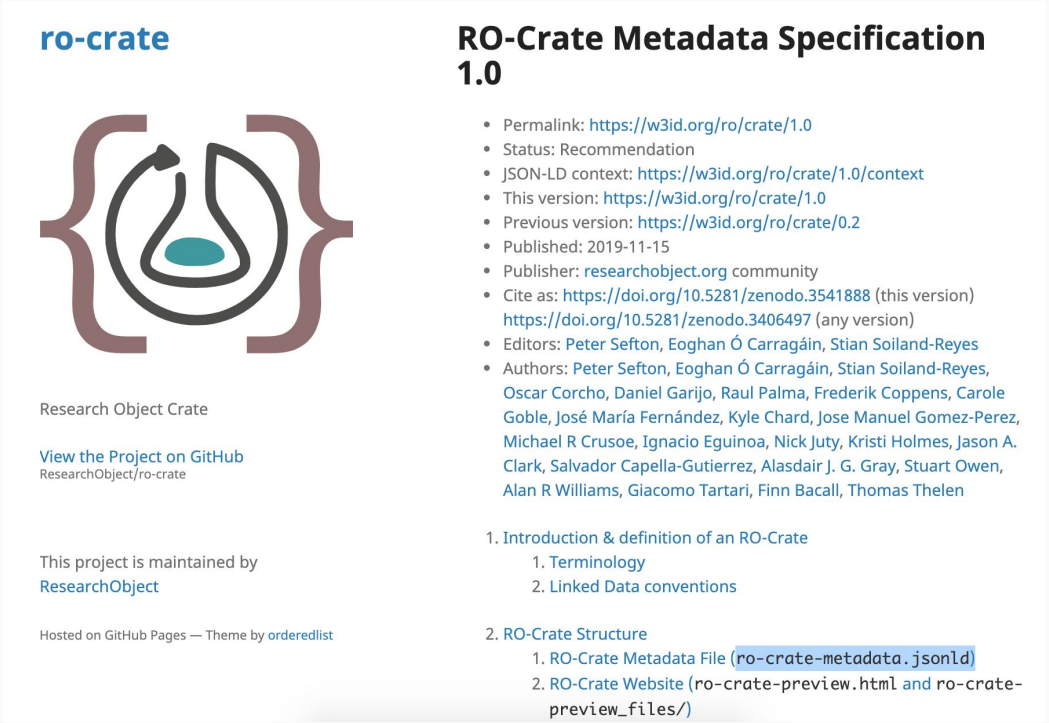
Standards - Schema.org



https://schema.org/Dataset

https://schema.org/DataFeed

Standards - Research Object Crate (RO-Crate)



a lightweight approach to packaging research data with their metadata

https://w3id.org/ro/crate/1.0

#digital-library-and-repository-content

Collections API as DataFeed

Application Programming Interface

API Payload as a DataFeed (.jsonld)


```
{
"@context": "http://schema.org/",
"@type": "DataFeed",
"name": "James Willard Schultz Collection",
"description": "API result for James Willard Schultz Collection",
"license": "https://creativecommons.org/licenses/by/4.0/",
"identifier": "https://arc.lib.montana.edu/schultz-0010/",
"dateModified": "2015-01-02",
"dataFeedElement": [
  {
    "@type\": \"DataFeedItem": [
      {
        "item": {
          "recordInfo_recordIdentifier": "64",
          "identifier_ark": "ark:/75788/m4sw9r",
          "identifier": "064.jpg",
          "proxyIdentifier": "",
          "titleInfo_title": "Photo of a young man in ceremonial dress playing a drum, Browning, Montana",
          "abstract": "Boy in full Indian dress standing on a rug outsideof a building and playing an Indian drum. Nothing written on verso.",
          "extension": "clothing and dress, drums, blackfeet indians, indians of north america, montana",
          "name_namePart": "James Willard Schultz",
          "originalMetadata_name": "Unknown",
          "originInfo_dateIssued": "1859-1947",
          "originalMetadata_dateIssued": "Unknown",
          "originInfo_dateCreated": "2005-09-09",
          "location_physicalLocation": "Montana State University Library Collection 10 - James Willard Schultz Papers, 1867-1969",
          "typeOfResource": "1 Photographic Print, B&W, 15 x 9 cm ",
          "originInfo_publisher": "Montana State University--Bozeman",
          "accessCondition": "https://creativecommons.org/licenses/by-nc-sa/3.0/us/",
          "subject_topic1": "Montana--History",
          "subject_topic2": "Indians of North America--Montana",
          "subject_topic3": "Schultz, James Willard, 1859-1947",
          "subject_topic4": "Native American Music",
          "note": "None",
          "genre": "Still Image",
          "physicalDescription_internetMediaType": "image/jpeg",
```


https://arc.lib.montana.edu/schultz-0010/api.php?v=1&type=search&q=boy&limit=10&format=json

James Willard Schultz

Progressive Web App (PWA) built using DataFeed and RO-Crate standards as datastore (.jsonld)

Prototype

James Willard Schultz Collection

Search Browse About

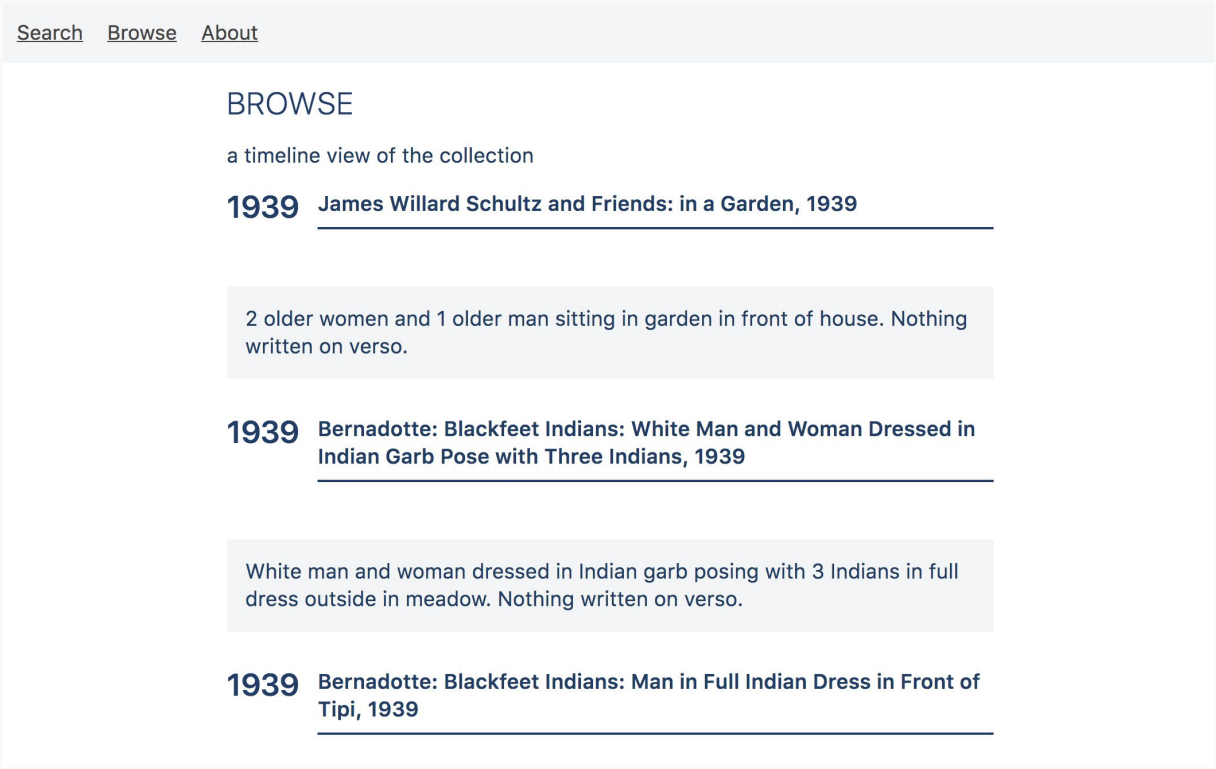
James Willard Schultz Photos & Personal Papers Collection includes photos & documents of Blackfeet, Blood, Kootenai, Shoshone and Arapaho Native Americans, Glacier and Waterton Lakes National Parks, as well as his professional writing career and personal papers.

FEATURED

1939 1939 1939 1939

1939 1939 1939 1932

Prototype


Search
Browse
About
BROWSE
a timeline view of the collection
1939
James Willard Schultz and Friends: in a Garden, 1939
2 older women and 1 older man sitting in garden in front of house. Nothing written on verso.
1939
Bernadotte: Blackfeet Indians: White Man and Woman Dressed in Indian Garb Pose with Three Indians, 1939
White man and woman dressed in Indian garb posing with 3 Indians in full dress outside in meadow. Nothing written on verso.
1939
Bernadotte: Blackfeet Indians: Man in Full Indian Dress in Front of Tipi, 1939

Manifest (.json)

```
{
  "name": "James Willard Schultz Collection",
  "short_name": "SchultzPWA",
  "start_url": "./?utm_source=homescreen",
  "display": "standalone",
  "background_color": "#213c69",
  "description": "A digital library template app",
  "theme_color": "#213c69",
  "prefer_related_applications": false,
  "categories": [
    "books",
    "education"
  ],
  "icons": [
    {
      "src": "./img/icons/icon-72x72.png",
      "sizes": "72x72",
      "type": "image/png"
    },
    {
      "src": "./img/icons/icon-96x96.png",
      "sizes": "96x96",
      "type": "image/png"
    },
    {
      "src": "./img/icons/icon-128x128.png",
      "sizes": "128x128",
      "type": "image/png"
    },
    {
      "src": "./img/icons/icon-144x144.png",
      "sizes": "144x144",
      "type": "image/png"
    },
    {
      "src": "./img/icons/icon-152x152.png",
      "sizes": "152x152",
```

Items (.jsonld)


```
{
"@context": "http://schema.org/",
"@type": "DataFeed",
"name": "James Willard Schultz Collection",
"description": "API result for James Willard Schultz Collection",
"license": "https://creativecommons.org/licenses/by/4.0/",
"identifier": "https://arc.lib.montana.edu/schultz-0010/",
"dateModified": "2015-01-02",
"dataFeedElement": [
  {
    "@type\": \"DataFeedItem": [
      {
        "item": {
          "recordInfo_recordIdentifier": "64",
          "identifier_ark": "ark:/75788/m4sw9r",
          "identifier": "064.jpg",
          "proxyIdentifier": "",
          "titleInfo_title": "Photo of a young man in ceremonial dress playing a drum, Browning, Montana",
          "abstract": "Boy in full Indian dress standing on a rug outsideof a building and playing an Indian drum. Nothing written on verso.",
          "extension": "clothing and dress, drums, blackfeet indians, indians of north america, montana",
          "name_namePart": "James Willard Schultz",
          "originalMetadata_name": "Unknown",
          "originInfo_dateIssued": "1859-1947",
          "originalMetadata_dateIssued": "Unknown",
          "originInfo_dateCreated": "2005-09-09",
          "location_physicalLocation": "Montana State University Library Collection 10 - James Willard Schultz Papers, 1867-1969",
          "typeOfResource": "1 Photographic Print, B&W, 15 x 9 cm ",
          "originInfo_publisher": "Montana State University--Bozeman",
          "accessCondition": "https://creativecommons.org/licenses/by-nc-sa/3.0/us/",
          "subject_topic1": "Montana--History",
          "subject_topic2": "Indians of North America--Montana",
          "subject_topic3": "Schultz, James Willard, 1859-1947",
          "subject_topic4": "Native American Music",
          "note": "None",
          "genre": "Still Image",
          "physicalDescription_internetMediaType": "image/jpeg",
```

Implications for Research Data Practice

Implications

Identifying collections as datasets

Encoding collections as datasets

Prioritizing reuse

Prioritizing reproducibility

Connecting research data metadata practices to collections metadata

Collections are datasets.

- emulatable and preservable

- set up for indexing and metadata (manifest file)

- static “databases” that can be reused/redeployed

How can we continue to connect our worldviews?

- Cross-departmental appointments
- Recognition of emerging dataset standards
- Release our collections as datasets

Thank you

Jason A. Clark
@jaclark