University of Kentucky
UKnowledge

Library Presentations | University of Kentucky Libraries

2009

Practical Metadata

Kathryn Lybarger
University of Kentucky, kathryn.lybarger@uky.edu

Right click to open a feedback form in a new tab to let us know how this document benefits you.

Follow this and additional works at: https://uknowledge.uky.edu/libraries_present

Part of the Library and Information Science Commons

Repository Citation
Lybarger, Kathryn, "Practical Metadata" (2009). *Library Presentations*. 22.
https://uknowledge.uky.edu/libraries_present/22

This Presentation is brought to you for free and open access by the University of Kentucky Libraries at UKnowledge. It has been accepted for inclusion in Library Presentations by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

Practical Metadata

Kathryn Lybarger

meta | morphosis

[the university of kentucky film-to-digital institute]

<METADATA>

meta | morphosis

[the university of kentucky film-to-digital institute]

What is metadata?

“data about data”

meta | morphosis

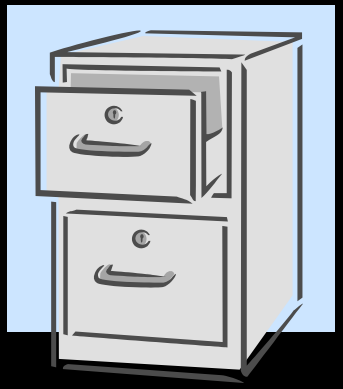
[the university of kentucky film-to-digital institute]

Types of Metadata

- Descriptive Metadata
- Structural Metadata
- Administrative Metadata
  - Preservation Metadata
  - Rights and Access Metadata
  - Technical Metadata

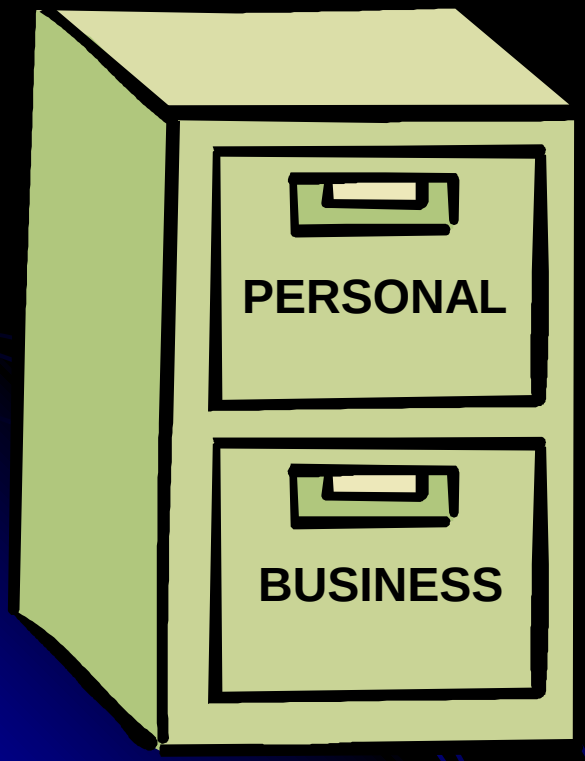
Examples

This space intentionally left blank.

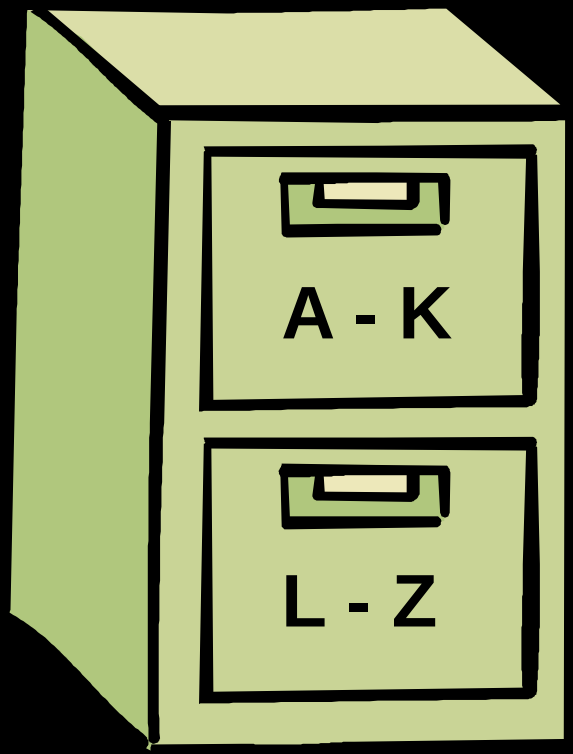


Examples

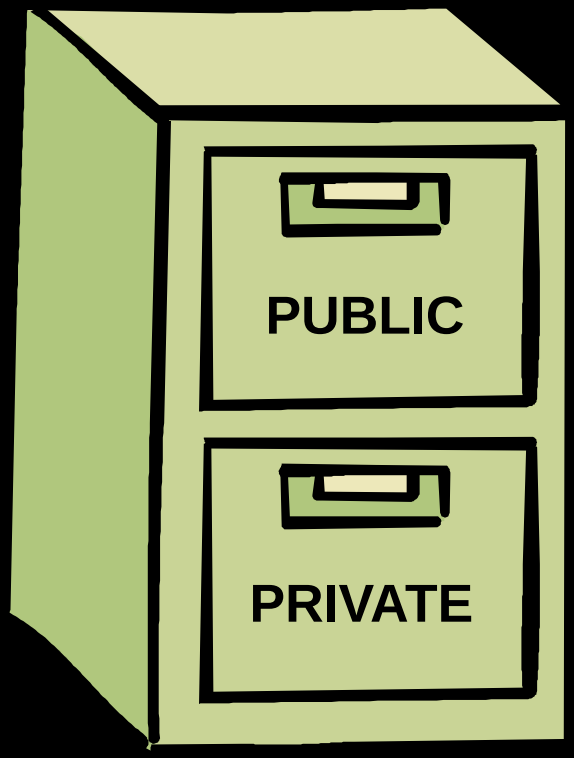
Descriptive



Structural



Administrative



What does metadata look like?

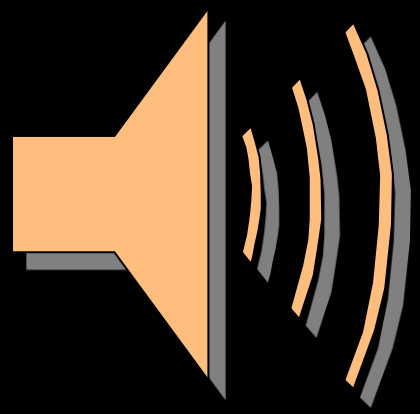
May be same format as data

Header added by Project Gutenberg

Ebook submitted

```
The Project Gutenberg EBook of Mavericks, by William MacLeod Raine

This eBook is for the use of anyone anywhere at no cost and with
almost no restrictions whatsoever.  You may copy it, give it away or
re-use it under the terms of the Project Gutenberg License included
with this eBook or online at www.gutenberg.net


Title: Mavericks

Author: William MacLeod Raine

Release Date: December 29, 2004 [EBook #14520]

Language: English

Character set encoding: ISO-8859-1

*** START OF THIS PROJECT GUTENBERG EBOOK MAVERICKS ***



Produced by Kathryn Lybarger and the PG Online Distributed
Proofreading Team




[Illustration: THE RIDER SLEWED IN THE SADDLE WITH HIS WHOLE ATTENTION
UPON POSSIBLE PURSUIT. _Frontispiece. Page 33_]

MAVERICKS
```

Metadata may have different format



WAV audio file

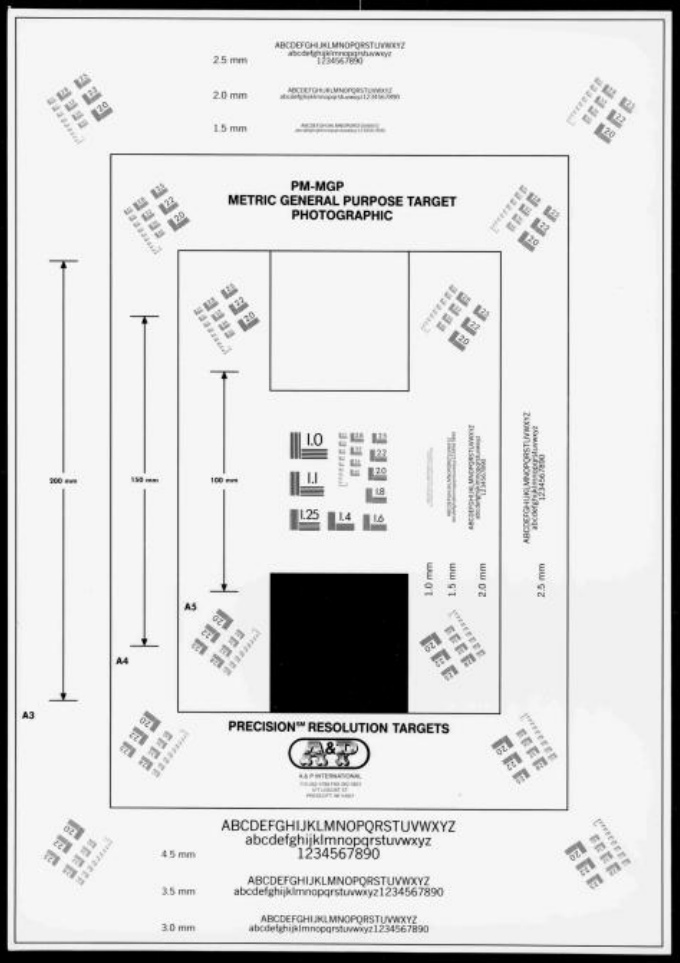
Text metadata

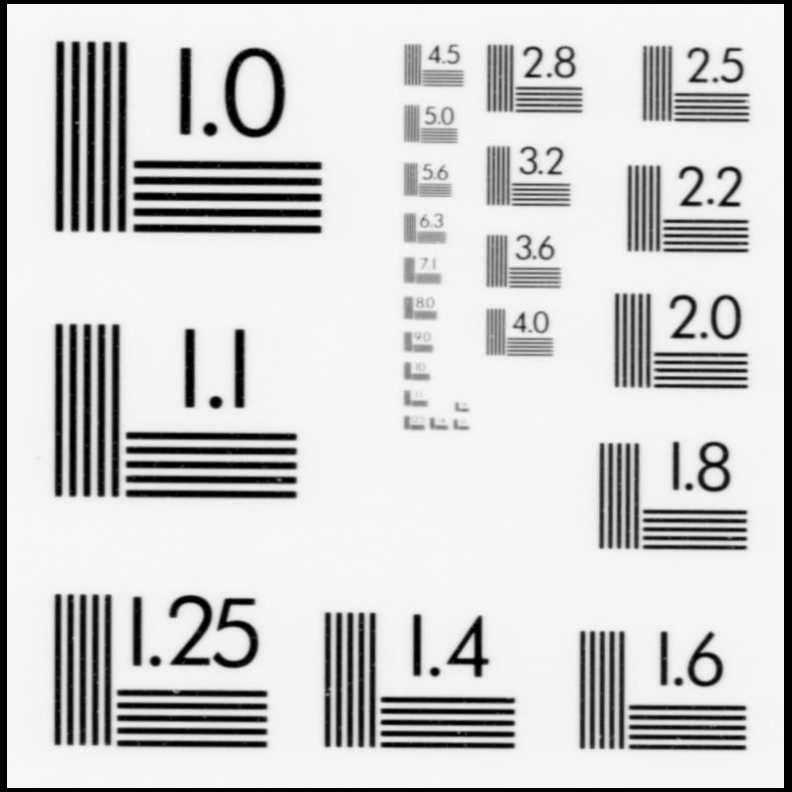
```
Master File Name: king_rpwcr03_t1_a
------------------------------------------------
Cassette Sleeve information:

"Maxell" Communicator C90 Series Audiocassette
                                  02OH108
                                   RPWCR 03
User Copy

An Unrehearsed interview with Martin Luther
King, Jr. for the University of Kentucky Libraries
Robert Penn Warren Civil Rights History Project.
The interview was conducted by Robert Penn Warren
on March 18, 1964
P.T.: 1 hr. 15 mins. Conditions: Fair
                                  Tape #1
------------------------------------------------
Hardware: Digidesign 002 mixing board
          Marantz SD4050 cassette deck, Dolby OFF

Software: Adobe Audition

Signal Processing: none

Settings: Uncompressed Windows PCM .wav file 1411K
          Sample Rate - 44.1 Hz
          Sample Size - 16 bit Stereo
          Run time - 32:45.220

------------------------------------------------
Derivative

Software: Adobe Audition

Processing:

File Type: 44100Hz, 16-bit, Mono
Uncompressed Size: 163 MB (171,814,012 bytes)
File Format: Windows PCM
             IEEE Float  (0.24 float type 3)
Last Written (local): 6/1/2006  11:30:46
Length: 32:28.004
```

Not all metadata is text

"Okay, it's September the 21, 1987, I'm in Frankfurt Kentucky in the home of Clarence Gunther, who was a World War II veteran, served in the Navy, entering service on March 6, 1940, and separated March 8, 1946 as a boats and mate first class. He was at Pearl Harbor…"

Not all metadata is verbal



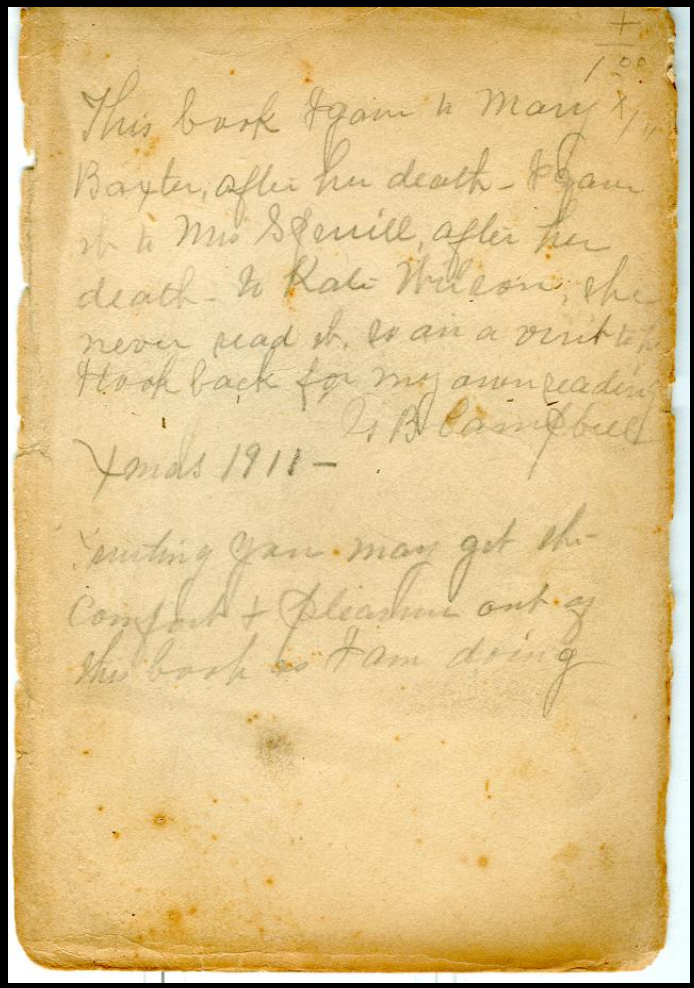


meta | morphosis

[the university of kentucky film-to-digital institute]

Metadata may have no structure

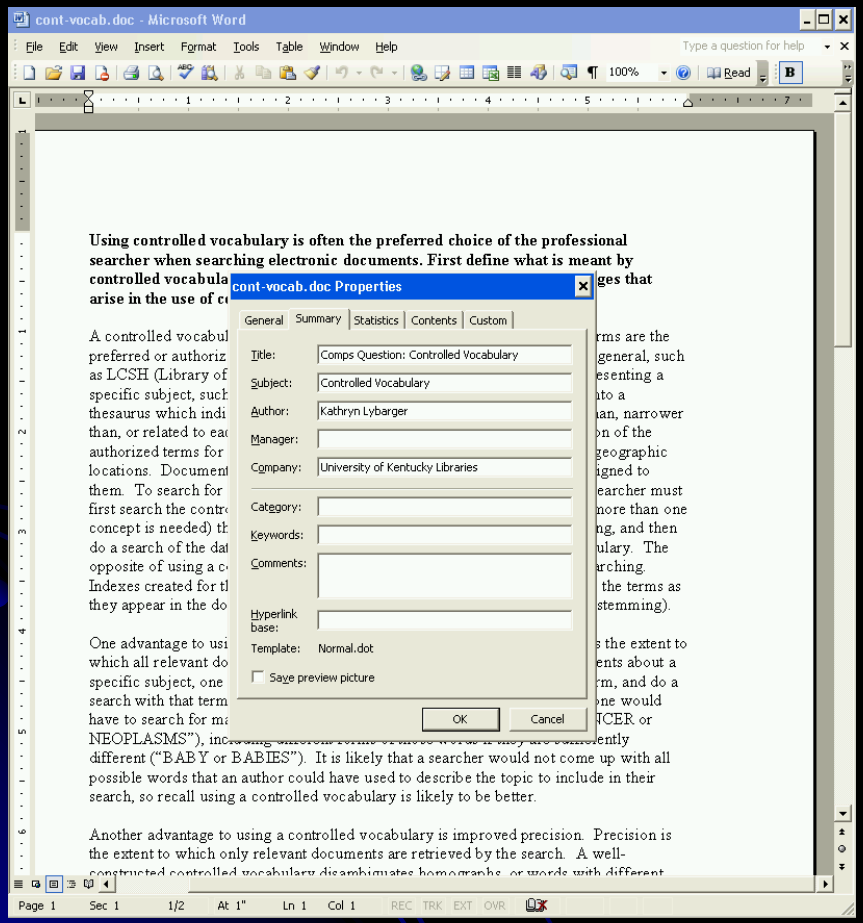
"This book I gave to Mary Baxter. After her death, I gave it to Mrs. Spruill. After her death, to Kate Wilson. She never read it, so on a visit to her, I took back for my own reading."



meta | morphosis

[the university of kentucky film-to-digital institute]

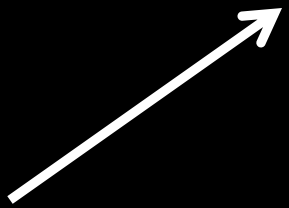
Metadata may have some structure

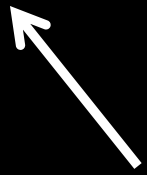


- Word processors allow "document properties"
- Anything can go in these fields

File names are metadata

cont-vocab.doc





Some indication of content

File type

Metadata may have rich structure

- Example: MARC record
- Requires expertise to read and create
- Allows very detailed searching

020 __ **|a** 9780756608347 (pbk.)
020 __ **|a** 0756608333 (plc)
020 __ **|a** 9780756608330 (plc)
029 1_ **|a** NZ1 **|b** 8900160
035 __ **|a** (OCoLC)56755783
037 __ **|a** 1236030 **|b** QBI
042 __ **|a** pcc
043 __ **|a** n-us---
050 00 **|a** E457.905 **|b** .S76 2005
082 00 **|a** 973.7/092 **|a** B **|2** 22
049 __ **|a** KUKK
100 1_ **|a** Stone, Tanya Lee.
245 10 **|a** Abraham Lincoln / **|c** Tanya Lee Stone.
250 __ **|a** 1st American ed.
260 __ **|a** New York : **|b** DK Pub., **|c** c2005.
300 __ **|a** 128 p. : **|b** ill. (some col.) ; **|c** 20 cm.
440 _0 **|a** DK biography

XML: eXtensible Markup Language

- Many rich metadata formats are encoded as XML
- A schema or DTD specifies rules which a document must follow
- Examples: XHTML, EAD, TEI, NDNP

XML: Example

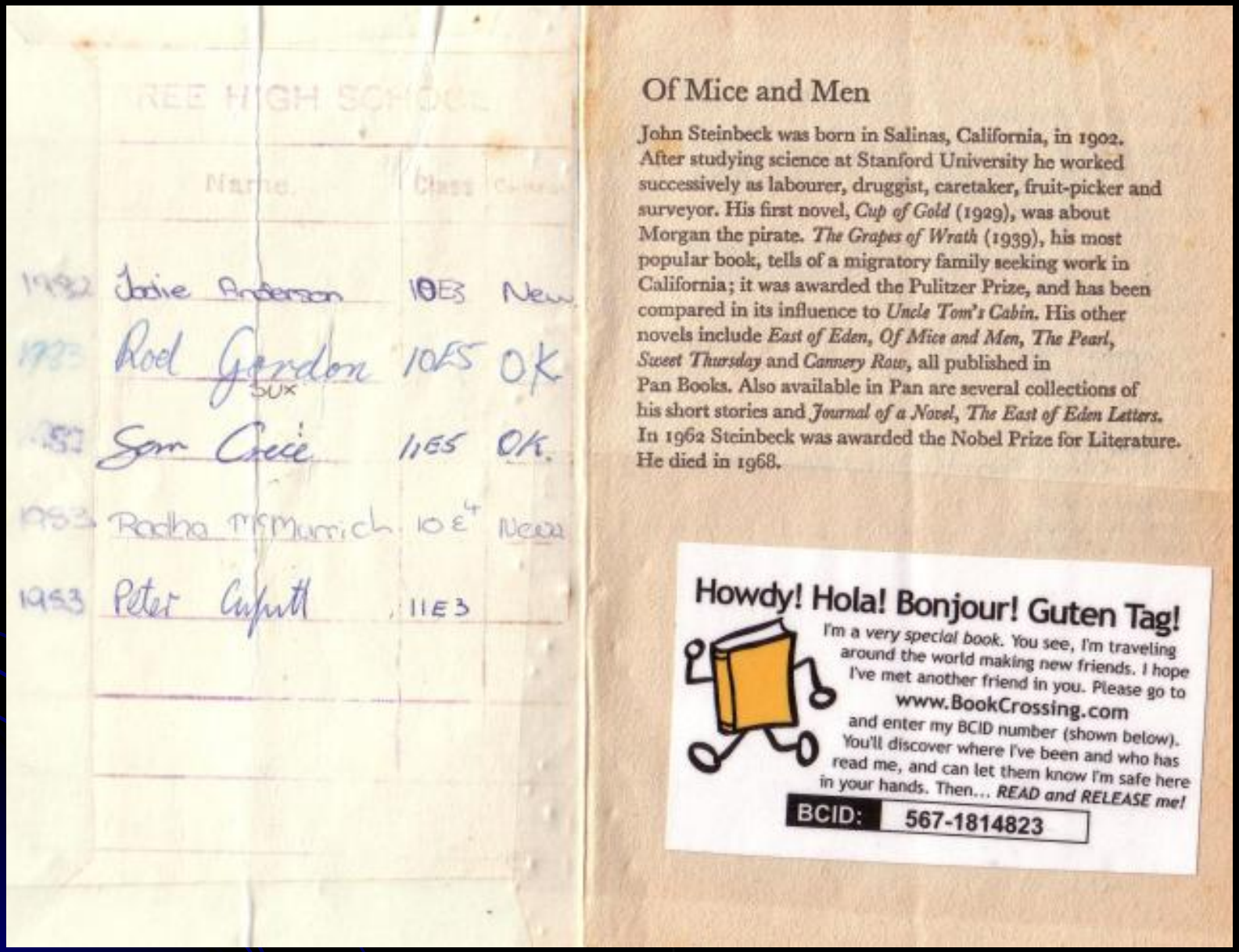
```
<book isbn="978-0898713619">
  <title>Numerical Linear
  Algebra</title>
  <author>Lloyd N. Trefethen</author>
  <author>David Bau, III</author>
  <publisher>SIAM</publisher>
</book>
```

XML: Advantages

- “self-describing”
- Validation catches many errors
- XML tools may be used for any XML language
  - searching
  - transformation
  - communication

When is metadata created?

Who creates metadata?



REE HIGH SCHOOL

	Name	Class	Comm
1982	Jodie Anderson	10E3	New
1983	Rod Gordon SUX	10E5	OK
1983	Sam Crice	11E5	OK.
1983	Radha McMurrich	10E4	New
1983	Peter Cuputt	11E3	

Of Mice and Men

John Steinbeck was born in Salinas, California, in 1902. After studying science at Stanford University he worked successively as labourer, druggist, caretaker, fruit-picker and surveyor. His first novel, *Cup of Gold* (1929), was about Morgan the pirate. *The Grapes of Wrath* (1939), his most popular book, tells of a migratory family seeking work in California; it was awarded the Pulitzer Prize, and has been compared in its influence to *Uncle Tom's Cabin*. His other novels include *East of Eden*, *Of Mice and Men*, *The Pearl*, *Sweet Thursday* and *Cannery Row*, all published in Pan Books. Also available in Pan are several collections of his short stories and *Journal of a Novel*, *The East of Eden Letters*. In 1962 Steinbeck was awarded the Nobel Prize for Literature. He died in 1968.

Howdy! Hola! Bonjour! Guten Tag!

I'm a *very special book*. You see, I'm traveling around the world making new friends. I hope I've met another friend in you. Please go to

www.BookCrossing.com

and enter my BCID number (shown below). You'll discover where I've been and who has read me, and can let them know I'm safe here in your hands. Then... ***READ and RELEASE me!***

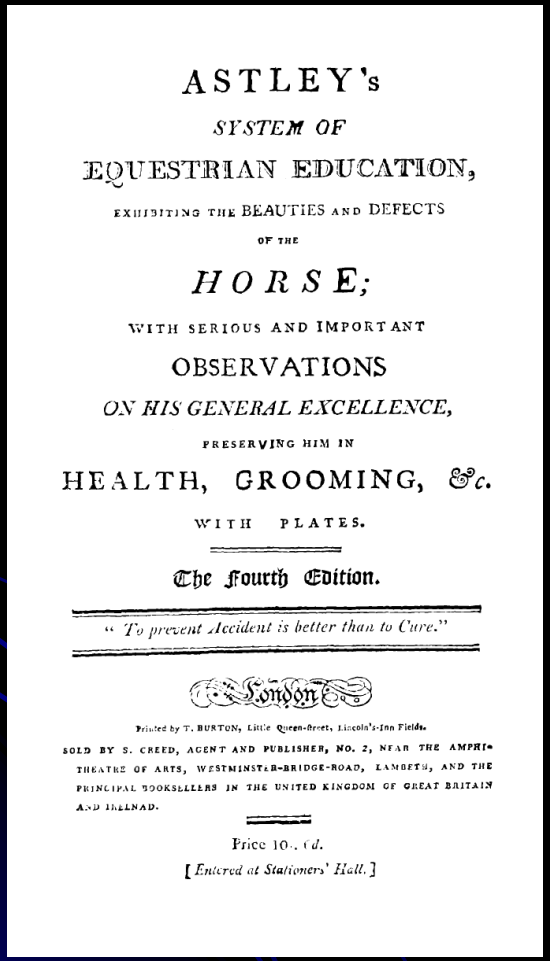
BCID: 567-1814823

meta | morphosis

[the university of kentucky film-to-digital institute]

Where is metadata?

Metadata may be inside the data



ASTLEY's

SYSTEM OF

EQUESTRIAN EDUCATION,

EXHIBITING THE BEAUTIES AND DEFECTS

OF THE

HORSE;

WITH SERIOUS AND IMPORTANT

OBSERVATIONS

ON HIS GENERAL EXCELLENCE,

PRESERVING HIM IN

HEALTH, GROOMING, *&c.*

WITH PLATES.

The Fourth Edition.

"*To prevent Accident is better than to Cure.*"

London

Printed by T. BURTON, Little Queen-Street, Lincoln's-Inn Fields.

SOLD BY S. CREED, AGENT AND PUBLISHER, NO. 2, NEAR THE AMPHITHEATRE OF ARTS, WESTMINSTER-BRIDGE-ROAD, LAMBETH, AND THE PRINCIPAL BOOKSELLERS IN THE UNITED KINGDOM OF GREAT BRITAIN AND IRELNAD.

Price 10*s*. 6*d*.

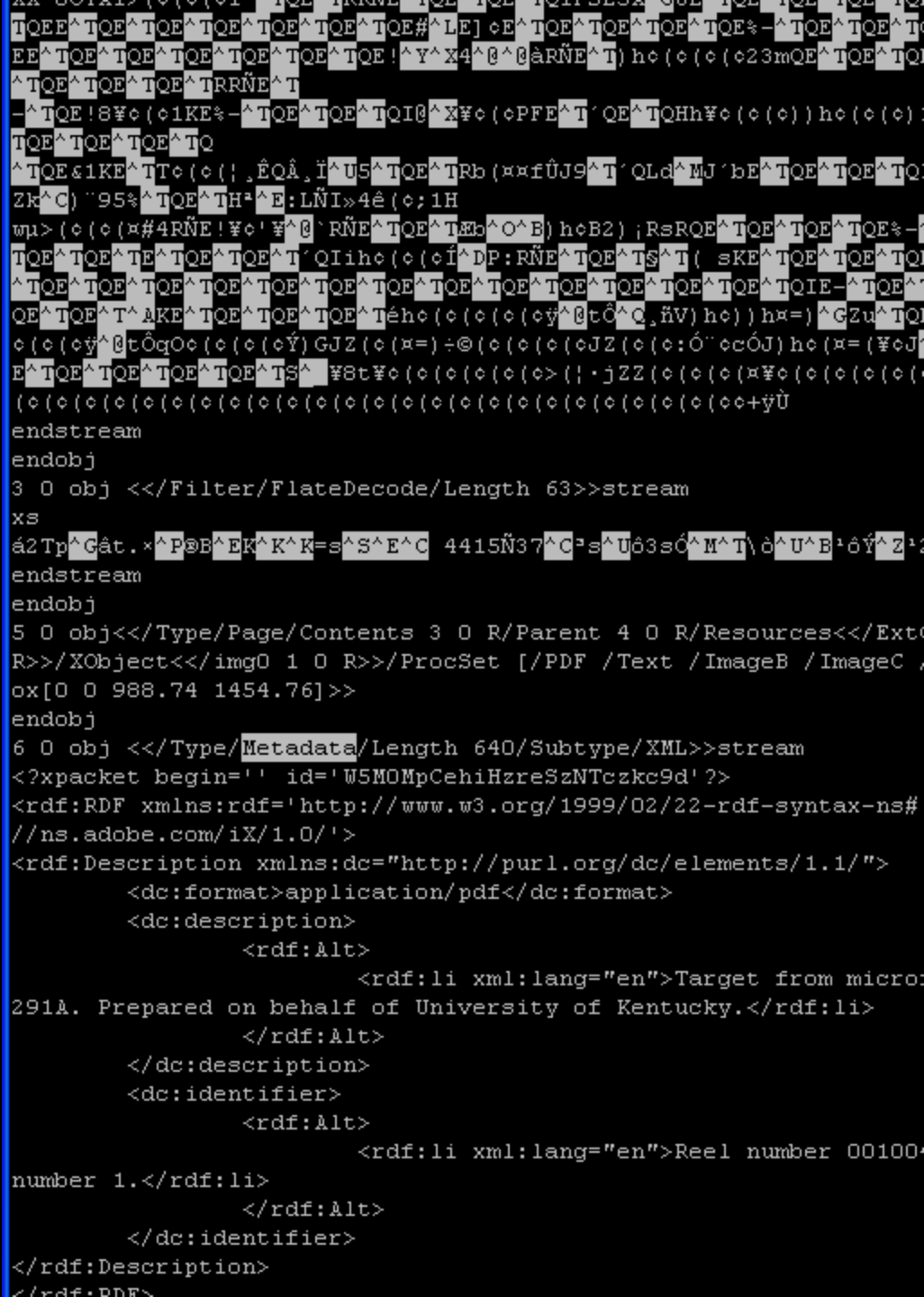
[*Entered at Stationers' Hall.*]

- Physical:
  - Title page
  - Table of contents
  - Index
- Digital:
  - Header information

Binary data →

Header information in an image file

XML metadata →



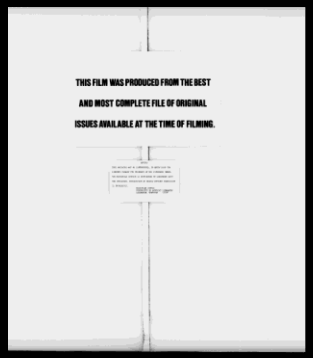
Metadata can be near the data

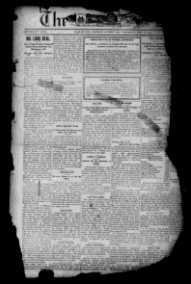
- Title and author on the spine of a book
- Associated .txt file with a .wav file
- Alternate data streams (Windows)

Metadata can be gathered elsewhere

- Card catalog
- Index
- Search engine

Metadata can be multiple places





microfilm

catalog

box lid

Bee **S-50**
Earlington, KY
98'
1892 **negative**

Brief View | Full View | MARC Format

The bee.

Relevance: ■■■■■
Title: Bee (Earlington, Ky.)
The bee.
Publisher: Earlington, Ky. : Bee Publishing Co.,
Description: v. : ill.
Frequency: Semiweekly

Location: Young Library - Periodicals Desk Microfilm
Call Number: S-50
Status: No information available
Library has: 1892:Apr.28-1906:Dec.20,

How is metadata different from normal data?

- No clear distinction!
- Metadata is also data
- Metadata can have metadata

Meta-metadata?

meta | morphosis

[the university of kentucky film-to-digital institute]

How much metadata?

- Too little metadata?
  - Different objects may have the same metadata
- Too much metadata?
  - You may never get started
  - Collection may take too long
  - Collection may be inconsistent / incomplete

What is good metadata?

Metadata should be accessible

- Easy to find
- Readable
  - Physical: legible, permanent
  - Digital: standard, non-proprietary format

Metadata should be meaningful

- Relationship to data should be clear

- Digital
  - Encoded content should be parse-able
  - XML should be well-formed, valid

Metadata should be accurate

- Adds to recall and precision in searching
- Not all metadata is apparent from looking at the data itself
- False metadata may lead to false conclusions about the data!

False metadata: Example

AMERICA'S TURF AUTHORITY

Daily Racing Form

VOL. LXX. No. 112 2 1 CHICAGO, ILL., SATURDAY, MAY 9, 1964 PRICE 50 CENTS (PAY NO MORE)

14 Eye $40,450 Illinois Derby; Mr. Brick Faces 15 in Withers

Copy Chief Tops Louisville 'Cap

Eight Compete in Feature Attraction; Lemon Twist Is Rated as Close Contender

By BUD LYON

Staff Correspondent

CHURCHILL DOWNS, Louisville, Ky., May 8.—Fourth Estate Stable's Copy Chief, fresh from victory in Keeneland's Ben Ali Handicap, heads a field of eight for Saturday's 27th running of the $56,450 Louisville Handicap while under top weight of 118 pounds and with the leading rider at the meeting aboard, Don Brumfield.

Slated closest to the lightweight in this 1 1-16 miles test is Theodore D. Bell's Lemon Twist, third in last week's Churchill Downs Cup, who will be under 116 pounds with Kelly Broussard. Hugh A. Grant's Argentine-bred Piave at 115 and R. J. Campbell, Wilson and McDermott's Piper's Son, 114 and William M. Cook, Mrs. O. A. Bentzinger's Poet of Stress at 113 with R. J. Gonzalez, Mrs. J. M. Brandon's Blood Lick, at 112 and Jimmy Nichols, Forrest H. Lindsey's Below Deck with Jacinto Vasquez and 111; and C. A. Dubois' Pottery Man at 110 pounds with no rider named complete the field.

Spinters in Roamer Handicap

Racing secretary Alex Lewis in addition to the normal fixture has carded an excellent supporting program including the $7,500 Roamer Handicap as the seventh race which attracted a well-matched field of nine top (late) sprinters going 6 furlongs. Heading this group is Reinmar at

Continued on Page Nine

Card Preakness Prep As Exhibition Event

Knightly Manner, Thanks Dec, Big Pete Vie at 1 1-16 Miles

By JOE HIRSCH

PIMLICO, Baltimore, Md., May 8.—Headed by William Haggin Perry's Knightly Manner, winner of Laurel's Chesapeake Stakes, and Mrs. Harriet N. Ball's Big Pete, fourth in the recent Delaware Valley Stakes at Garden State Park, a trio of three-year-olds drawn today for the one-hundredth Belmont's $4,000 Preakness Prep at a 1 1-16 miles, a stepping stone to next week end's 89th running of the $150,000 Preakness, middle jewel of the Triple Crown. The Prep will be presented on a better exhibition, between the seventh and eighth races on tomorrow's card.

Originally scheduled as a $7,500 feature, the Preakness Prep loomed a match race this morning, when the names of Kentucky Manner and Big Pete were the only ones dropped in the entry box. Entries closed late, however, and that afternoon track officials announced that E. J. Cheery's Thanks Dee will also participate in the Prep, with the value of the race cut back

Continued on Page Nine

Bold Destroyer, Geology Vie For Favoritism at Sportsman's

Floral Shop, Van Berg Duo, Canal, Shield of Valor Rate Highly in 1⅛ Miles Feature

By OSCAR OTIS

SPORTSMAN'S PARK, Cicero, Ill., May 8.—Headed by stakes winners Bold Destroyer and Geology, a full field of 14 three-year-olds has been nominated for the ninth running of the $44,650 Illinois Derby at 1 1-8 miles here today. If all 14 start, the winner will collect earnings of $28,292.50.

Bold Destroyer, winner of the James H. Bright Memorial at Tropical Park and also the Illinois Derby Prep here earlier in the season, will vie for public favoritism with Clifford Lussky's Geology, who was entered up from Churchill Downs only Thursday to take a shot at this richest and most prestige-laden event of Sportsman's spring stakes schedule.

Pierce Sr. to Saddle Pair

Both Bold Destroyer, owned by Marguerite Daniel, and Geology drew favorable post positions. Bold Destroyer is break from number two spot, Geology from number seven. As for those two stakes winners, who must share top weight of 121 pounds, a Florida-Kentucky rivalry has developed. Bold Destroyer being Florida-owned and bred, Geology, Kentucky-owned and -bred. Under the allowance conditions of the Illinois Derby, those two must concede four pounds to all other rivals. Geology's stakes victory came during the winter at the City of Miami at Tropical.

While most of the starters have established form at Sportsman's Park, one "last minute" surprise entrant was H. Amiro's Nenkia, who was rushed over from Arlington Park by trainer Joe H. Pierce Jr., who blows out for the race Thursday morning.

Continued on Page Six

Castle Forbes Two, Half Lengths Winner at Big A

AQUEDUCT, N. Y., May 8.—Today's featured $14,000 Sparkle Purse for three-year-old fillies best south of Jamaica when Jacques D. Wimpfheimer's stakes winning Petite Rouge, who would probably have been second choice to Mrs. H. C. Phipps' successful Castle Forbes, was found to be running a fever and had to be scratched. Of the six who started, all but Lewsdeck Stable's Su Vel are eligible to next Saturday's $18,000 added Acorn Stakes at one mile.

Castle Forbes rewarded the favorite players in the crowd of 38,461 that turned out in mid-summer weather by scooting with authority under John L. Rotz. At the end of 7 furlongs in 1:20⅘, the favorite led Jerry Reske's Nashaikh by two and a half lengths and paid $5.16. Mrs. Mendy Adler's Trebuy was a half length farther back and a head before Cal. Isidor Bieber's Petitesal.

Continued on Page 10

CLIFFORD LUSSKY—Owner-breeder of the invading Geology, one of the favorites in the Illinois Derby.

Eight Go in Feature Event at Hazel Park

Rob Roy III. Heads Line-up; Topweight at 118 in Mile Test

By JACK WILSON

Staff Correspondent

HAZEL PARK, Mich., May 8.—Dolce Carter Jr.'s Rob Roy III. heads a strong field of eight middle distance performers in the $10,000 Inaugural Handicap, top offering on the first week end program at Hazel Park. The Argentine import, hero of Oaklawn Park's Southland Handicap, must shoulder top impost of 118 pounds in Saturday's headliner at one mile.

Trotting the Carver runner will be Marion H. Vee Berg's Red Cargo, easy winner of the Cheyenne Handicap in his last start at Sportsman's Park. The Plaudette Fleet, six-year-old gets in with 116 pounds and will be handled by Frank Perkins. Elsie Coffman has been named on Rob Roy III.

Completing the inaugural line-up will be E. O. Guthrie's Devil's Advocate, 115 and R. J. Bright; Mrs. B. B. Bowden's Cup Terry, 113 and John Bey; B. V. Hockett's Sailing Chance, 108, no boy; Rosco and Adelbert's Jart Titan, 108 and Dick Barrett; T. A. Grissom's Mr. Buddy, 107 and Elwood Marks, and the Elias Bros.' I Recall, 107 with no rider named as yet.

Supporting feature on the interesting 10-race card which gets underway at 2:00 p.m. will be the $5,000 Sophomore Handicap which will bring out several Michigan Derby hopefuls. Scheduled to go the 6 furlongs test for three-year-olds will be C. O.

Continued on Page Eight

Sturgis Speedster Surprise Starter

Webster, Calumet Farm Send Couplings; Aqueduct Mile Feature Event Worth $63,200

By BOB HOKWOOD

Staff Correspondent

AQUEDUCT, N. Y., May 8.—A field of 10 three-year-olds was entered this morning for Saturday's 38th running of the $63,200 Withers Mile. There is one surprise candidate in the person of Hoy Sturgis' Mr. Brick, who could be favored though he finished a tiring sixth in last Saturday's Kentucky Derby. Early in the week, trainer Jimmy Pyros said that he would consider starting Mr. Brick if the Withers should attract a small field, which has never seemed likely. Whether or not the sub of Johns Joy will actually start in this lucky field is conjectural, but by the early speed he avoid trouble and Bob Thorpe has been engaged to ride.

If all 10 start in this edition of the Withers, which is run likely, the winner will collect a tidy $41,080. All starters are to carry 126 pounds.

Couplings Present Challenge

The chief rivals to Mr. Brick would appear to be the Reginald N. Webster entry of Traffic and National, Calumet Farm's Kentucky Jug and Ky. Pioneer, Joseph M. Roebling's Black Mountain and W. L. McKnight's A. Dack who will be saddled by Johnny Nerud; A. B. Hancock's Alphabet and Charles W. Boudreau's Bon Beaut, who will be coupled because of a commonality of ownership; George D.

Continued on Page 10

Prospect of 'Off' Track For $15,350 Ascot Cup

Filatee II., Shop, Causeway Top Invaders in Fine Field of 14

By BOB HURLEY

Staff Correspondent

ASCOT PARK, Cuyahoga Falls, Ohio, May 8.—Summer thunderstorms which made the prospect of an "off" track likely cost a couple switches in rider assignments were the only late changes for the 54th running of the Ascot Gold Cup at 2 miles on Saturday.

Three separate downpours drenched the area this morning making the track muddy underfoot and the report from the Cuyahoga Akron weather station was for more of the same in the forenoon tomorrow.

The field for the $15,350 Gold Cup remains at 14. There is a chance that both Kilyhoon and Honey Rush, who are on the also eligible list, will get a chance to draw in for trainer Hal Russell, for he is reluctant to start Blazer Boxer under 118 pounds if the footing is not fast. Only 12 horses will be permitted to start and the winner will earn $8,050.

Seven invaders, headed by Mrs. Adele

Seven Invaders

Apparent from file:

4800 x 6800 pixels

Metadata: 400dpi

Conclusion:

Image: 12in x 17in

Paper: 11in x 16in

meta | morphosis

[the university of kentucky film-to-digital institute]

False metadata: Example

AMERICA'S TURF AUTHORITY

Daily Racing Form

VOL. LXX. No. 112 2 1 CHICAGO, ILL., SATURDAY, MAY 9, 1964 PRICE 50 CENTS (PAY NO MORE)

14 Eye $40,450 Illinois Derby; Mr. Brick Faces 15 in Withers

Copy Chief Tops Louisville 'Cap

Eight Compete in Feature Attraction; Lemon Twist Is Rated as Close Contender

By BUD LYON

Bold Destroyer, Geology Vie For Favoritism at Sportsman's

Florsl Shop, Van Berg Duo, Canal, Shield of Valor Rate Highly in 1⅛ Miles Feature

By OSCAR OTIS

Sturgis Speedster Surprise Starter

Webster, Calumet Farm Send Couplings; Aqueduct Mile Feature Event Worth $63,200

By BOB HOKHOOD

Eight Go in Feature Event at Hazel Park

Rob Roy III. Heads Line-up; Topweight at 118 in Mile Test

By JACK WILSON

Card Preakness Prep As Exhibition Event

Knightly Manner, Thanks Doc, Big Pete Vie at 1 1-16 Miles

By JOE HIRSCH

Castle Forbes Two, Half Lengths Winner at Big A

Prospect of 'Off' Track For $15,350 Ascot Cup

Filatee II., Shop, Causeway Top Invaders in Fine Field of 14

By BOB HURLEY

Metadata: 200dpi

Conclusion:

Image: 24in x 34in

Paper: 22in x 32in

meta | morphosis

[the university of kentucky film-to-digital institute]

OCR: Optical character recognition

- An automated process of turning images of letters into (searchable) text
- Very common metadata for images of books/newspapers
- Often uncorrected, somewhat inaccurate

Uncorrected OCR: Example

THREE DS TRIUMPH

Zacaweista Easily Qutsprints T S

Jordan and Morsun

Stadium Purse Feature of Wash ¬

ington Park ProgramWeather

and Track Conditions Ideal

HOMEWOOD Ill June 2 Zacaweista the

good son of High Time in the Three Ds

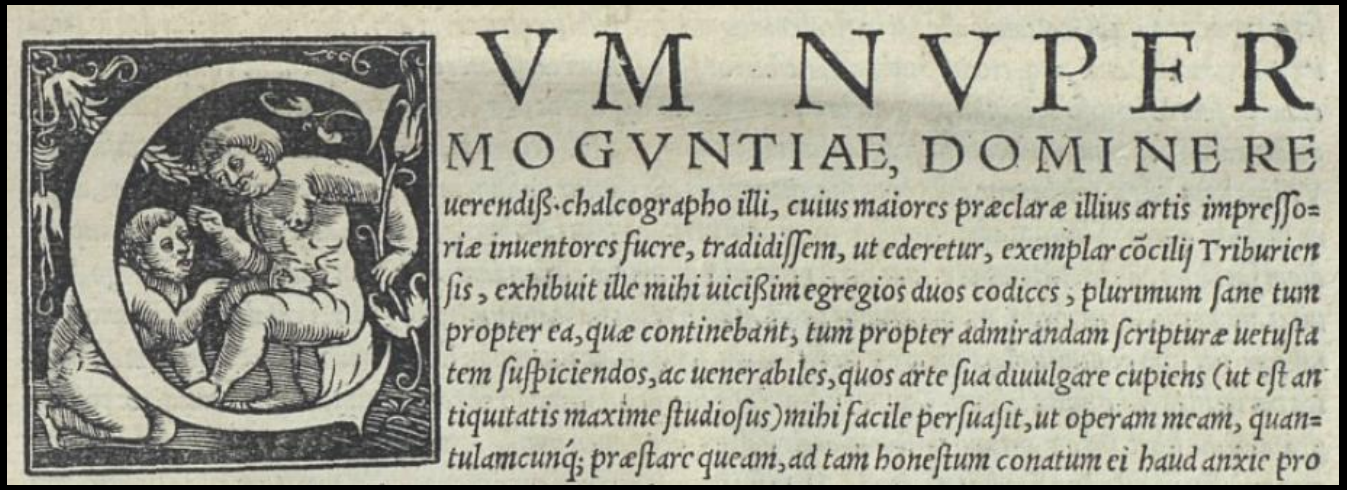
THREE D'S TRIUMPH

Zacaweista Easily Outsprints T. S. Jordan and Morsun.

Stadium Purse Feature of Washington Park Program—Weather and Track Conditions Ideal.

HOMEWOOD, Ill., June 2.—Zacaweista, the good son of High Time, in the Three D's Stock Farm Stable, figured in his first winning effort since the winter season when he easily accounted for the Stadium Purse,

Uncorrected OCR: Example



VM NVPER
MOGVNTIAE, DOMINE RE
uerendiß·chalcographo illi, cuius maiores præclaræ illius artis impreſſo=
riæ inuentores fuere, tradidiſſem, ut ederetur, exemplar cõcilij Triburien
ſis, exhibuit ille mihi uiciſſim egregios duos codices, plurimum ſane tum
propter ea, quæ continebant, tum propter admirandam ſcripturæ uetuſta
tem ſuſpiciendos, ac uenerabiles, quos arte ſua diuulgare cupiens (ut eſt an
tiquitatis maxime ſtudioſus) mihi facile perſuaſit, ut operam meam, quan=
tulamcunq; præſtare queam, ad tam honeſtum conatum ei haud anxie pro

ucrendifi-chdlcogrdpboilli, cuiusmahm pr£chr& iUiusdrtis imprcffos

rie inuentorcsfucre, trddidifftm, utedcretur, exempkrcocilijTriburicrt

Why metadata?

Metadata can be used to identify data

- Title / author on the spine of a book
- File names
- Labels

Metadata can be used to interpret data

- Instructions for tax forms
- Language / character encoding
- XML DTD or schema

Metadata can be used to search data

- Card catalog
- Index
- Search engine

Metadata can be used to manage data

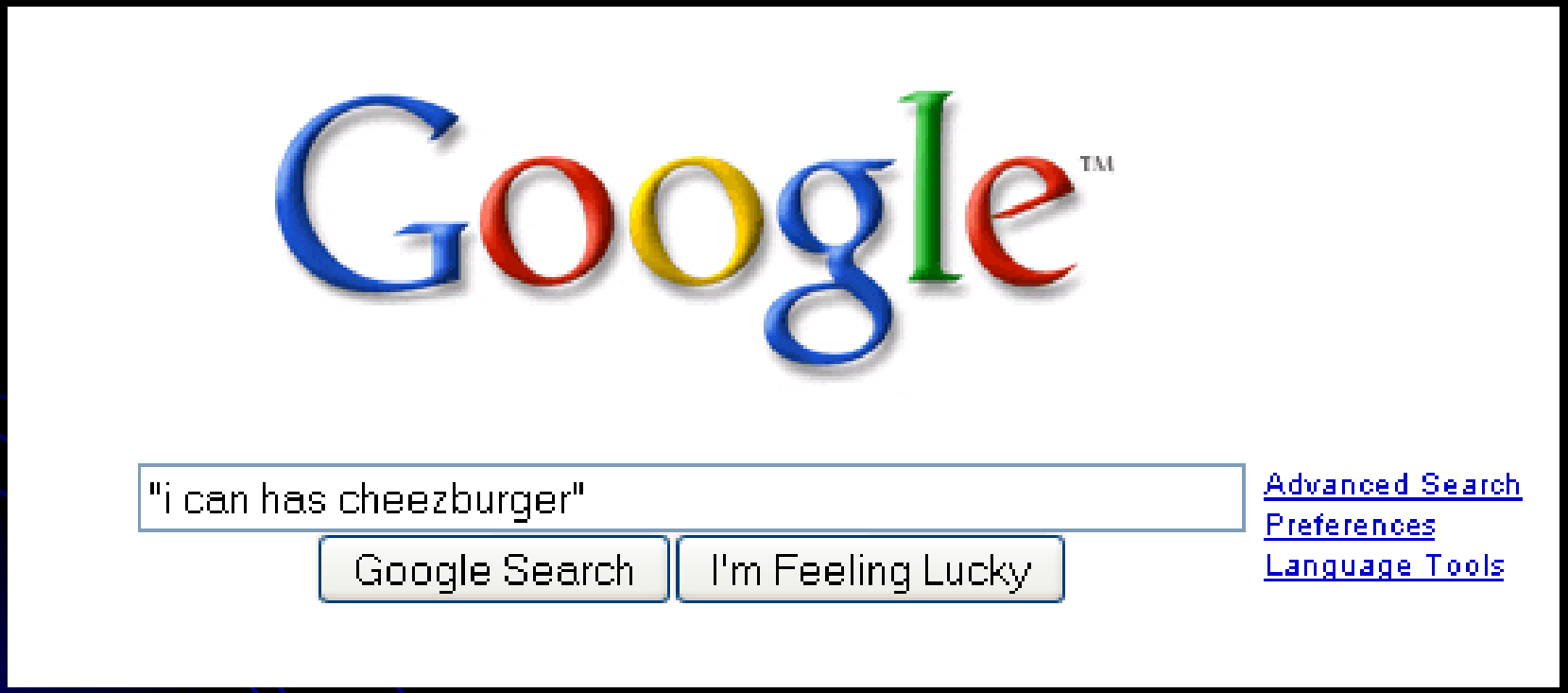
- Call numbers
- “Burn by” dates on file boxes
- Rights and access

Metadata can be used to communicate about data

- Finding aids
- Abstracts
- OAI

Who uses metadata?

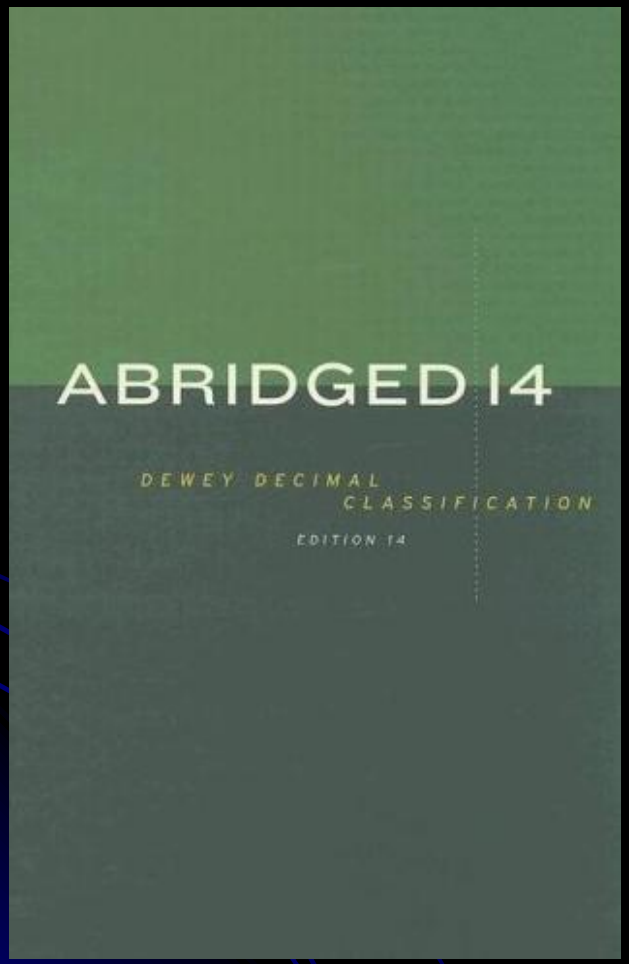
Internet users use metadata

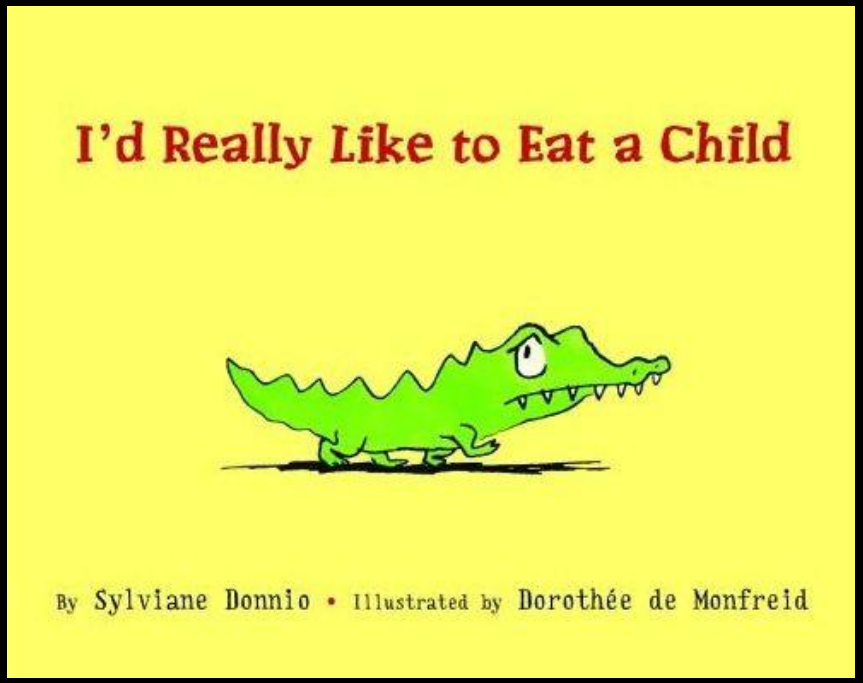


Librarians use metadata

- Reference
- Cataloging
- Collection Development

Children use metadata





What is metadata?

meta | morphosis

[the university of kentucky film-to-digital institute]

You already understand metadata!

meta | morphosis

[the university of kentucky film-to-digital institute]

</METADATA>

meta | morphosis

[the university of kentucky film-to-digital institute]

Any questions about metadata?