Better Living through Linking

Daniel Chudnov - 2009-05-27 - dchud at umich edu

Texas Conference on Digital Libraries 2009

An extract of the full presentation
which is available here:

http://onebiglibrary.net/story/tcdl-2009-talk-better-living-through-linking

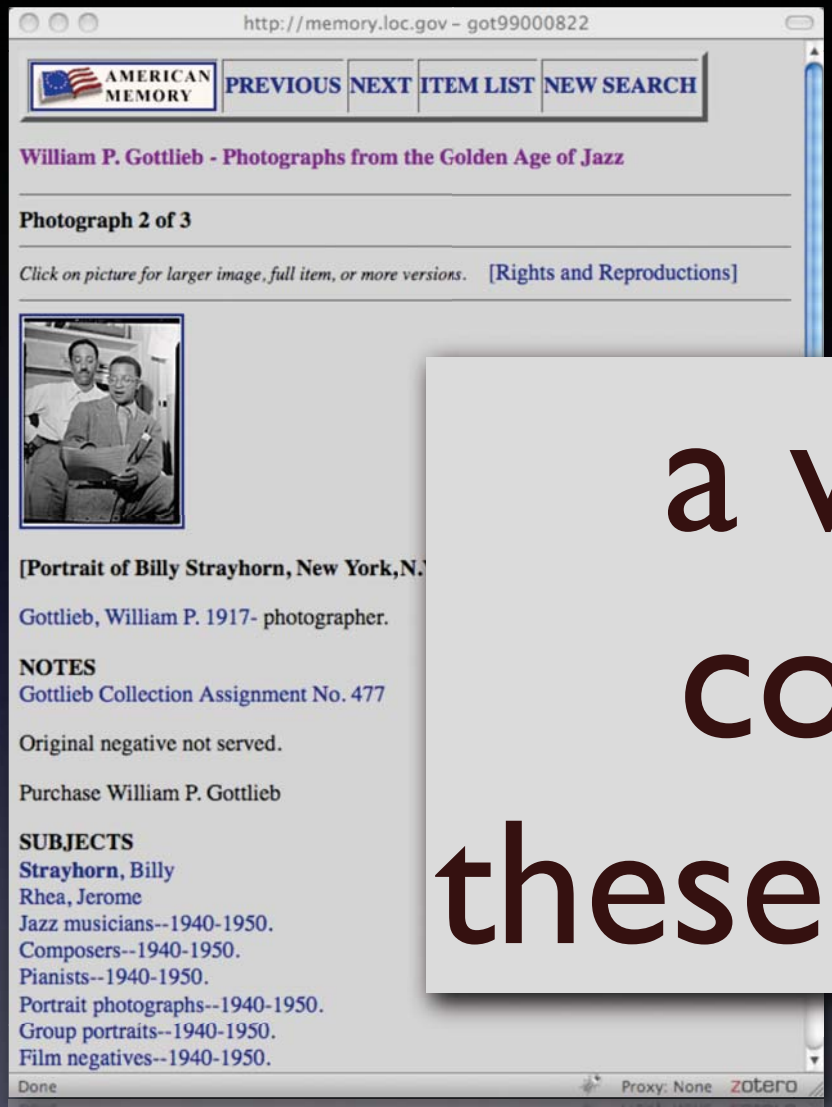
“Linked Data”
is
worth a close look

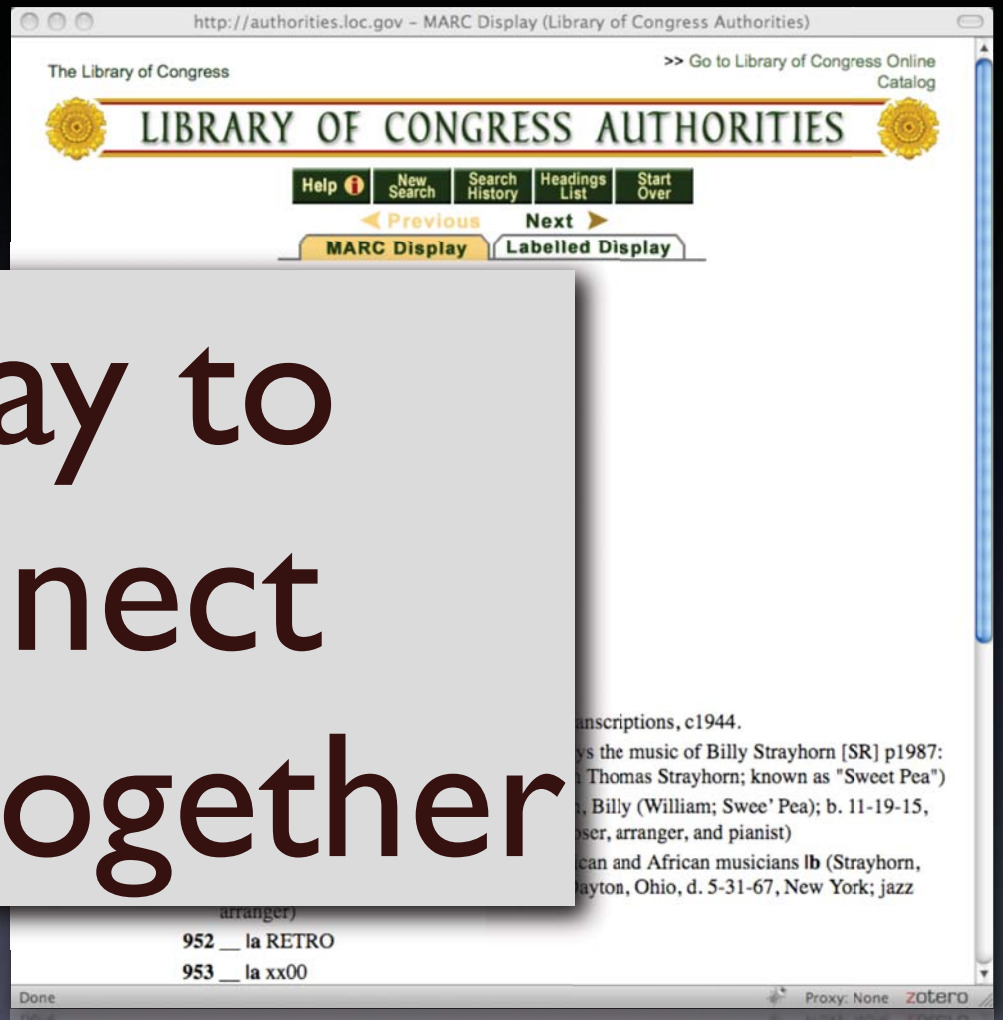
so what is it?

a way to
link
better

a way to
make
concepts explicit

a way to
connect
our stuff





a way to connect these together

item record

authority record

draw our stuff
deeper
into the web

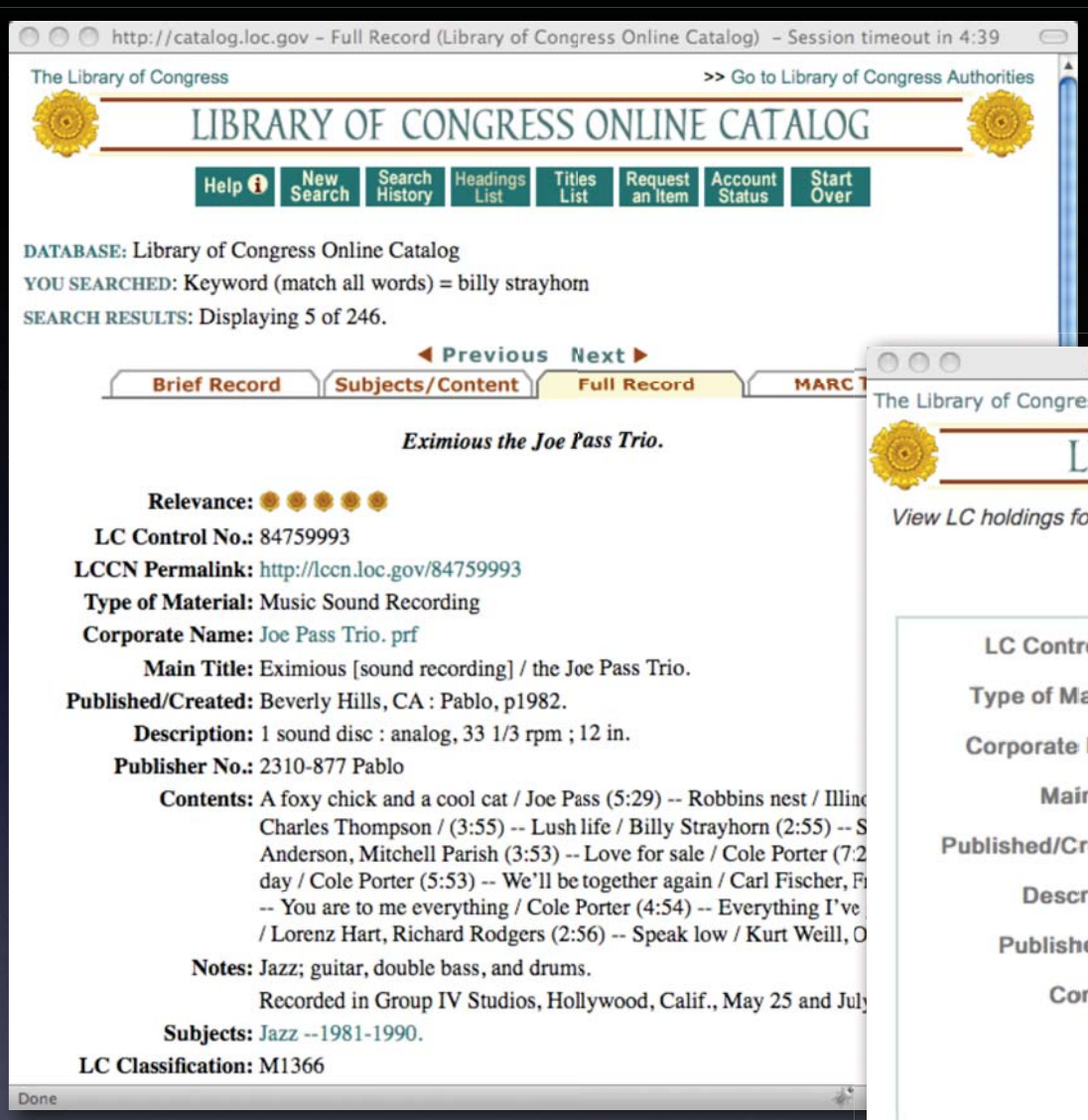
not just
files to download
but part of the web

how?

use URIs as names for things
use HTTP URIs
provide useful information
include links to other URIs

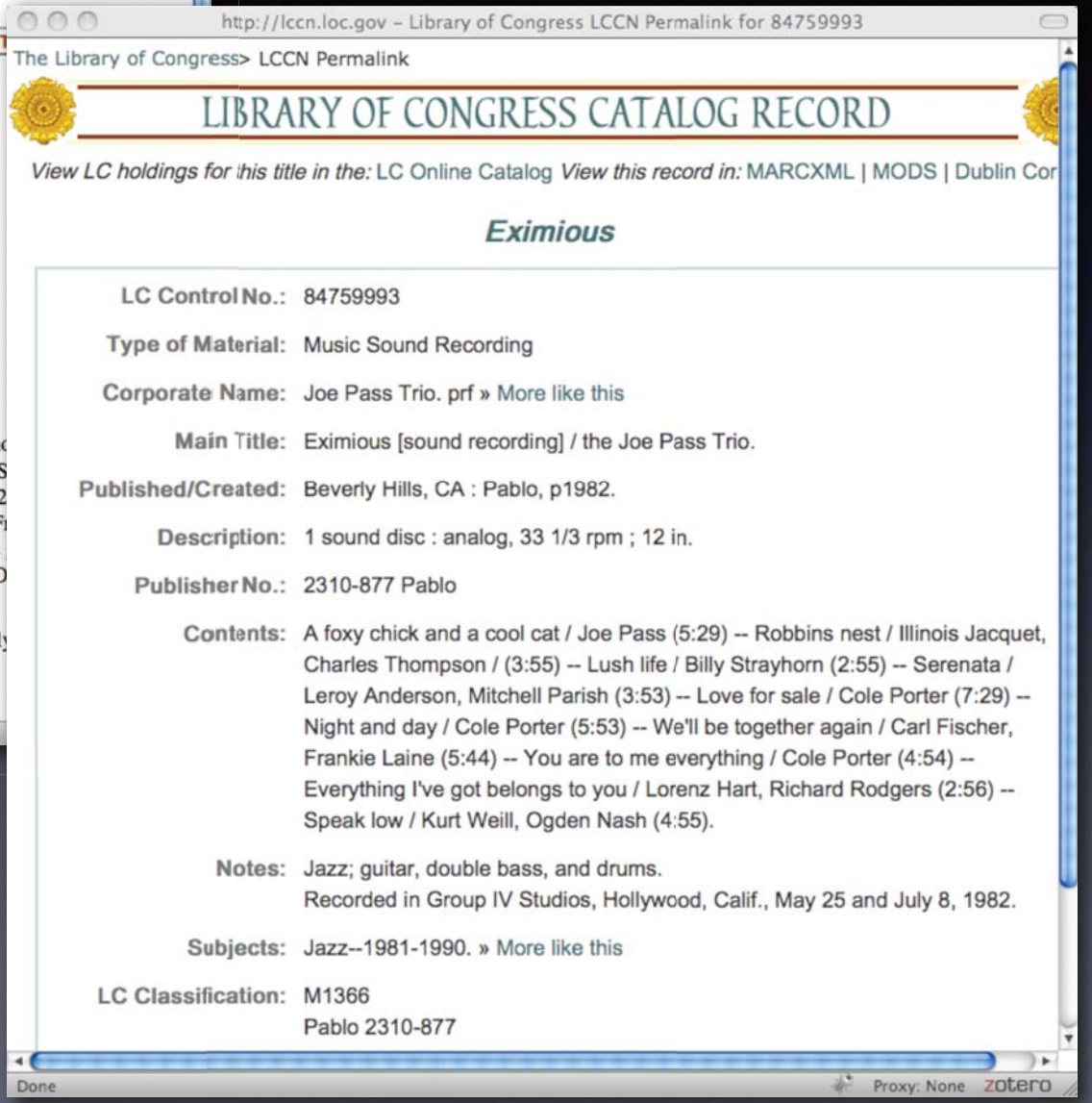
http://www.w3.org/DesignIssues/LinkedData.html

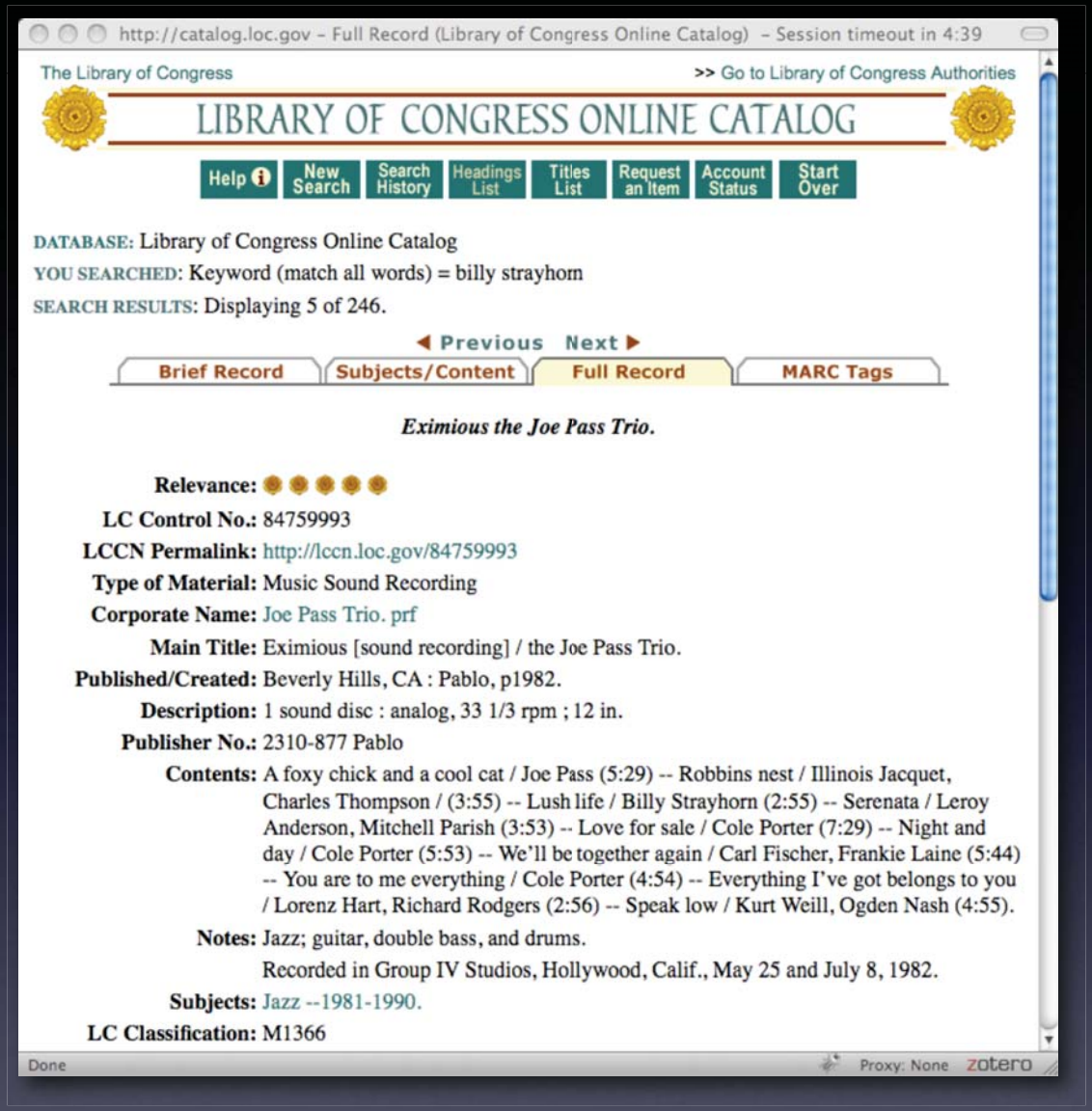
break down concepts
make them “pages”



the old

the new





the old

bad URL

screenscraping required

few links out

no links in

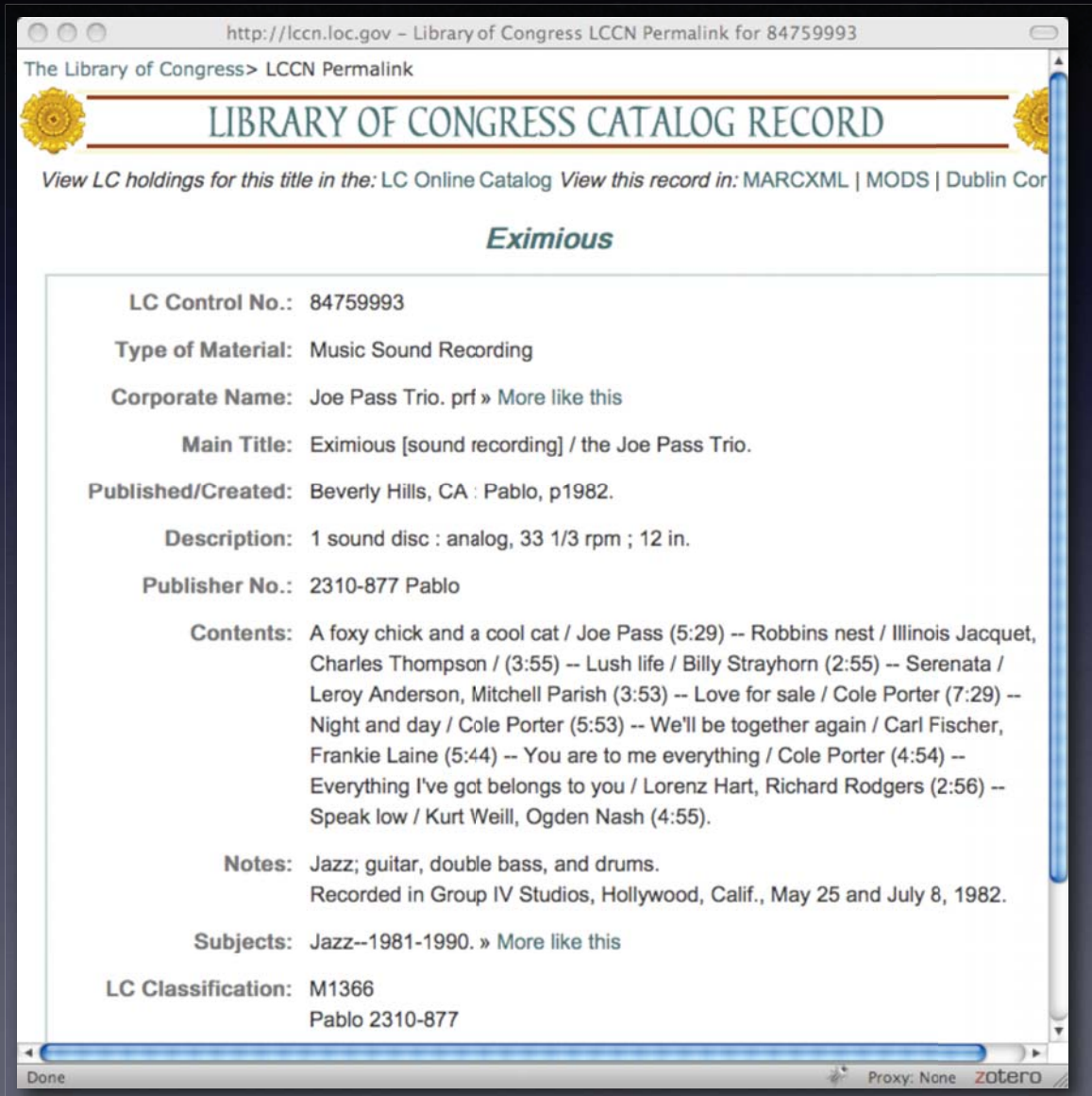
the new

good URL

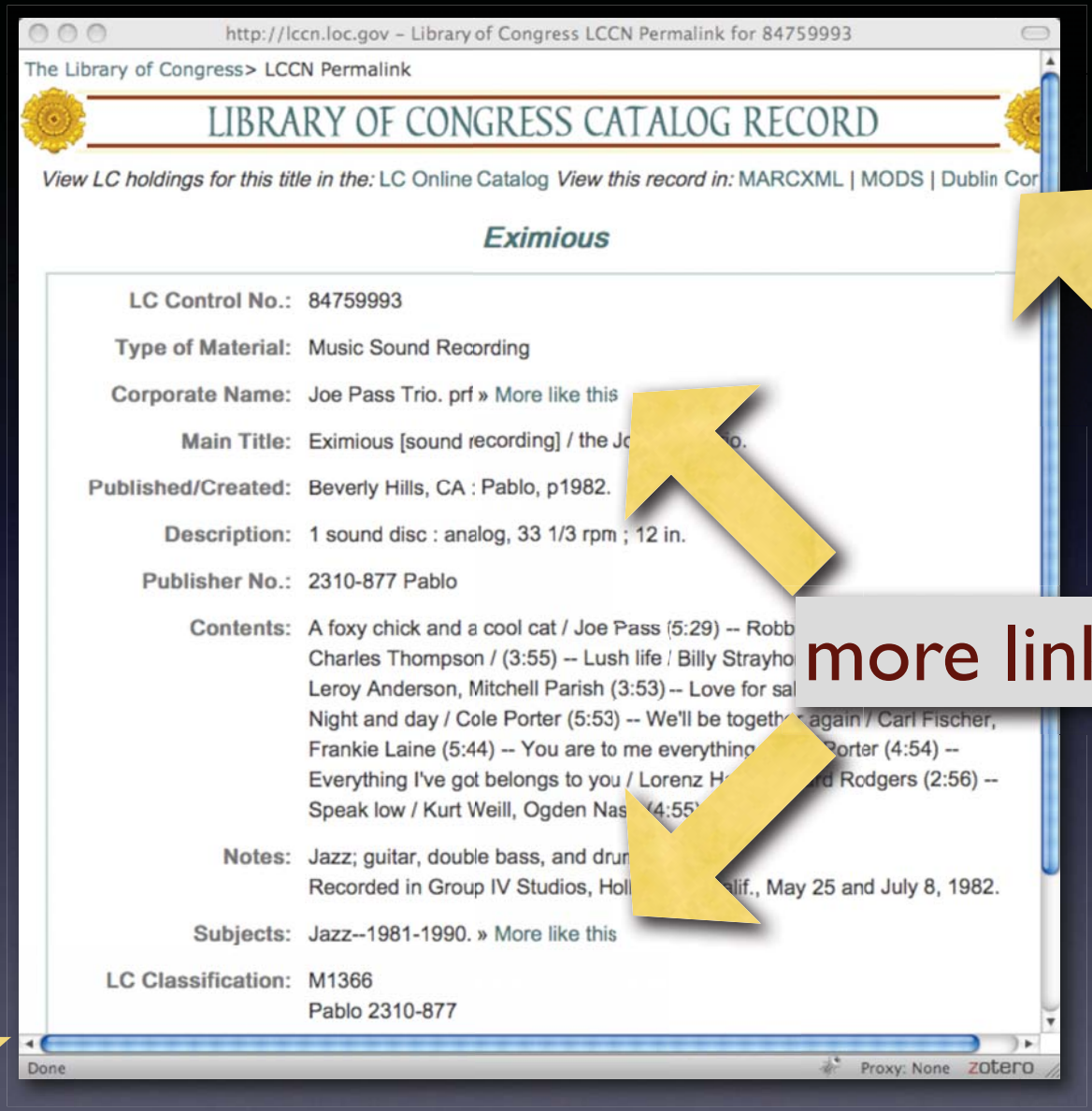
alternate formats

some links out

more links in



LCCN Permalink



alternate formats

more links out

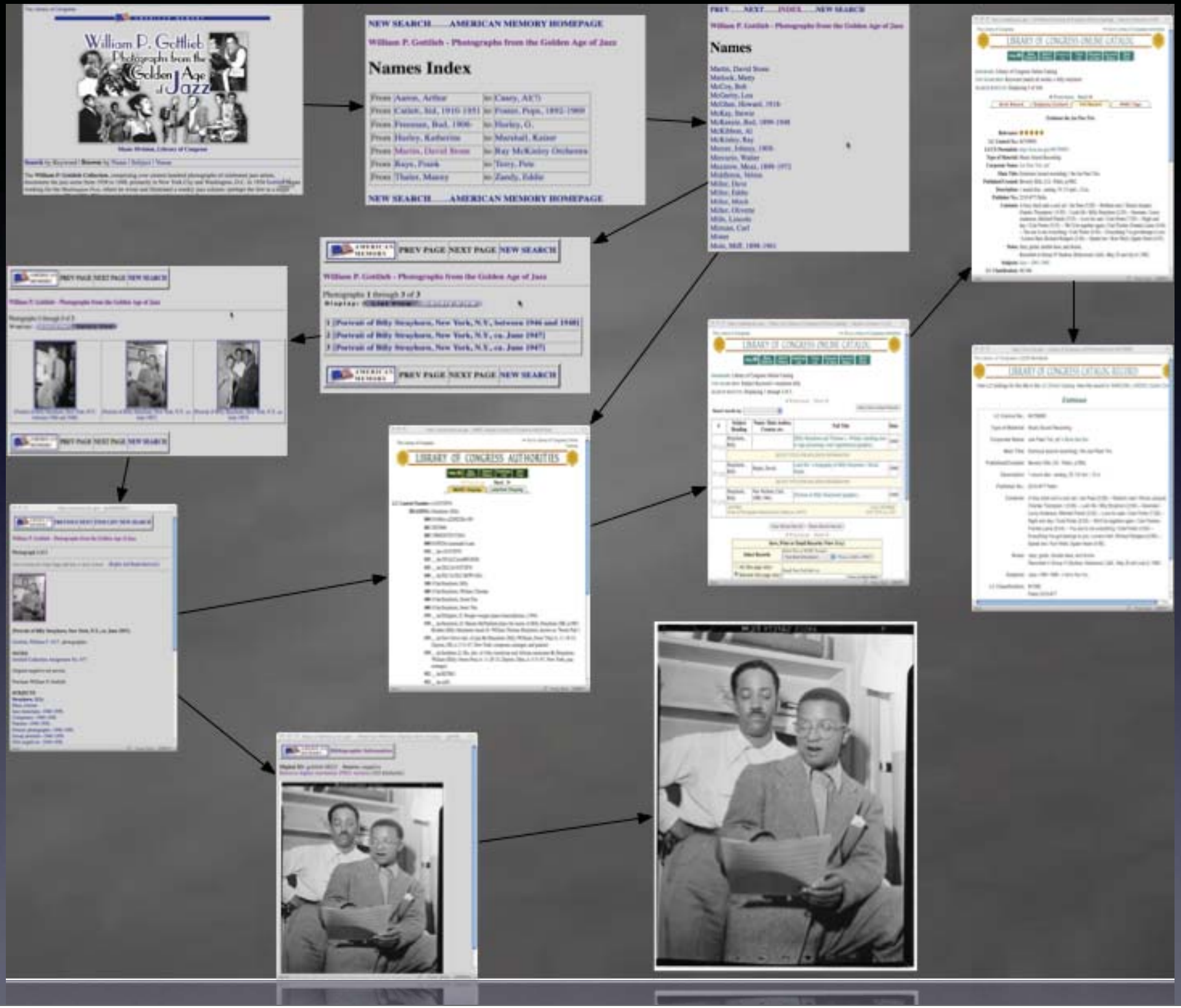
clean URI

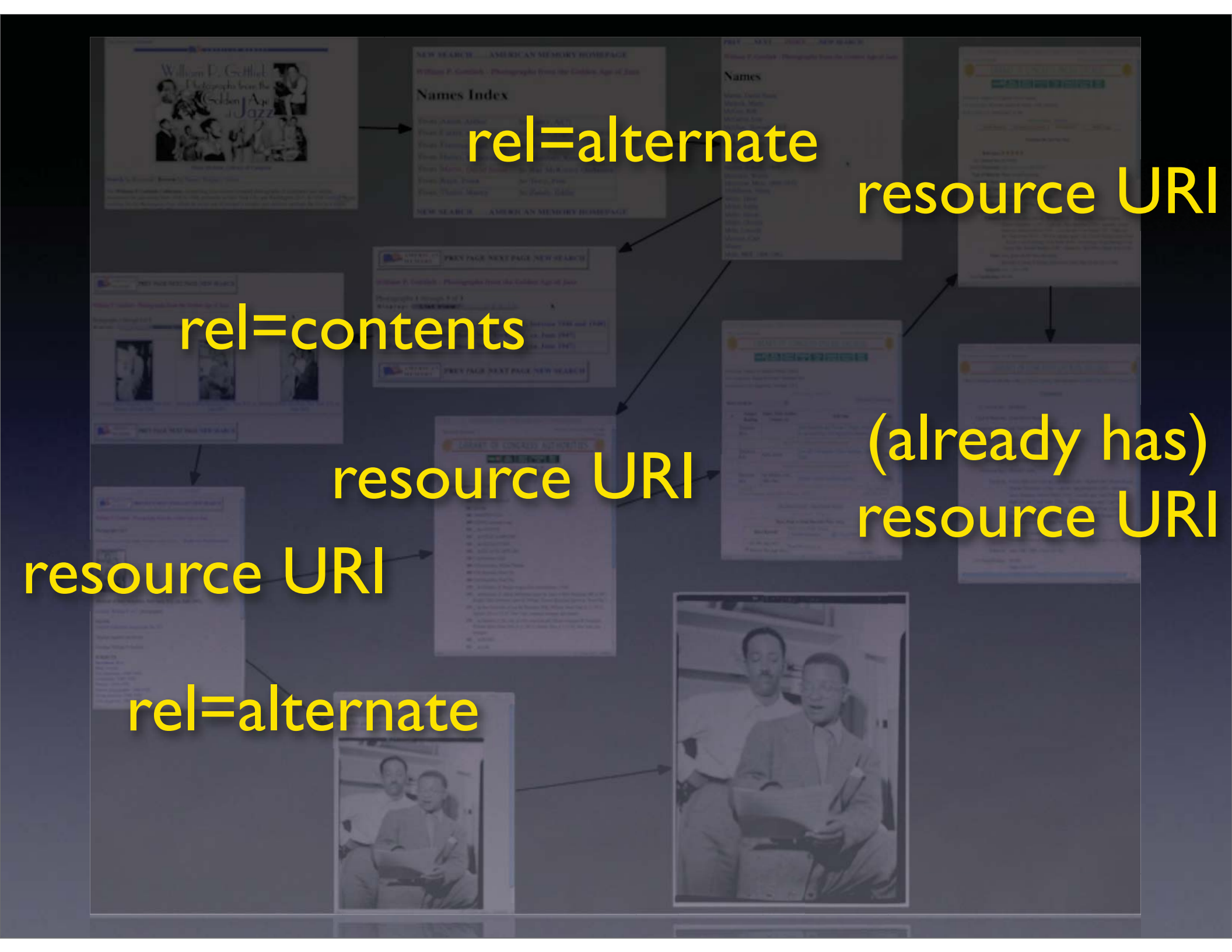
http://lccn.loc.gov/84759993

```
<link rel="alternate" type="application/mods+xml" href="http://lccn.loc.gov/84759993/mods">
<link rel="alternate" type="application/marc+xml" href="http://lccn.loc.gov/84759993/marcxml">
<link rel="alternate" type="application/dc+xml" href="http://lccn.loc.gov/84759993/dc">
<link rel="alternate" type="application/oai_dc+xml" href="http://lccn.loc.gov/84759993/oai_dc">
```

discoverable, parseable alternate formats

apply these rules
everywhere


William P. Gottlieb
Photographs from the Golden Age of Jazz
NEW SEARCH.....AMERICAN MEMORY HOMEPAGE
William P. Gottlieb - Photographs from the Golden Age of Jazz
Names Index
NEW SEARCH.....AMERICAN MEMORY HOMEPAGE
PREV.....NEXT.....INDEX.....NEW SEARCH
Names
PREV PAGE NEXT PAGE NEW SEARCH
Photographs 1 through 3 of 3
1 [Portrait of Billy Strayhorn, New York, N.Y., between 1946 and 1948]
2 [Portrait of Billy Strayhorn, New York, N.Y., ca. June 1947]
3 [Portrait of Billy Strayhorn, New York, N.Y., ca. June 1947]
LIBRARY OF CONGRESS AUTHORITIES


rel=alternate
resource URI
rel=contents
resource URI
(already has)
resource URI
resource URI
rel=alternate

arrows
become
cross-references

why?

it becomes
crawl-able
mine-able

integrated access
through
smart crawl / index

precision improves
with
rich metadata
explicitly linked

recall improves
with
explicit links
through authority recs

bring
precision / recall
of clunky OPACs
to open web

this is
doing web stuff
better

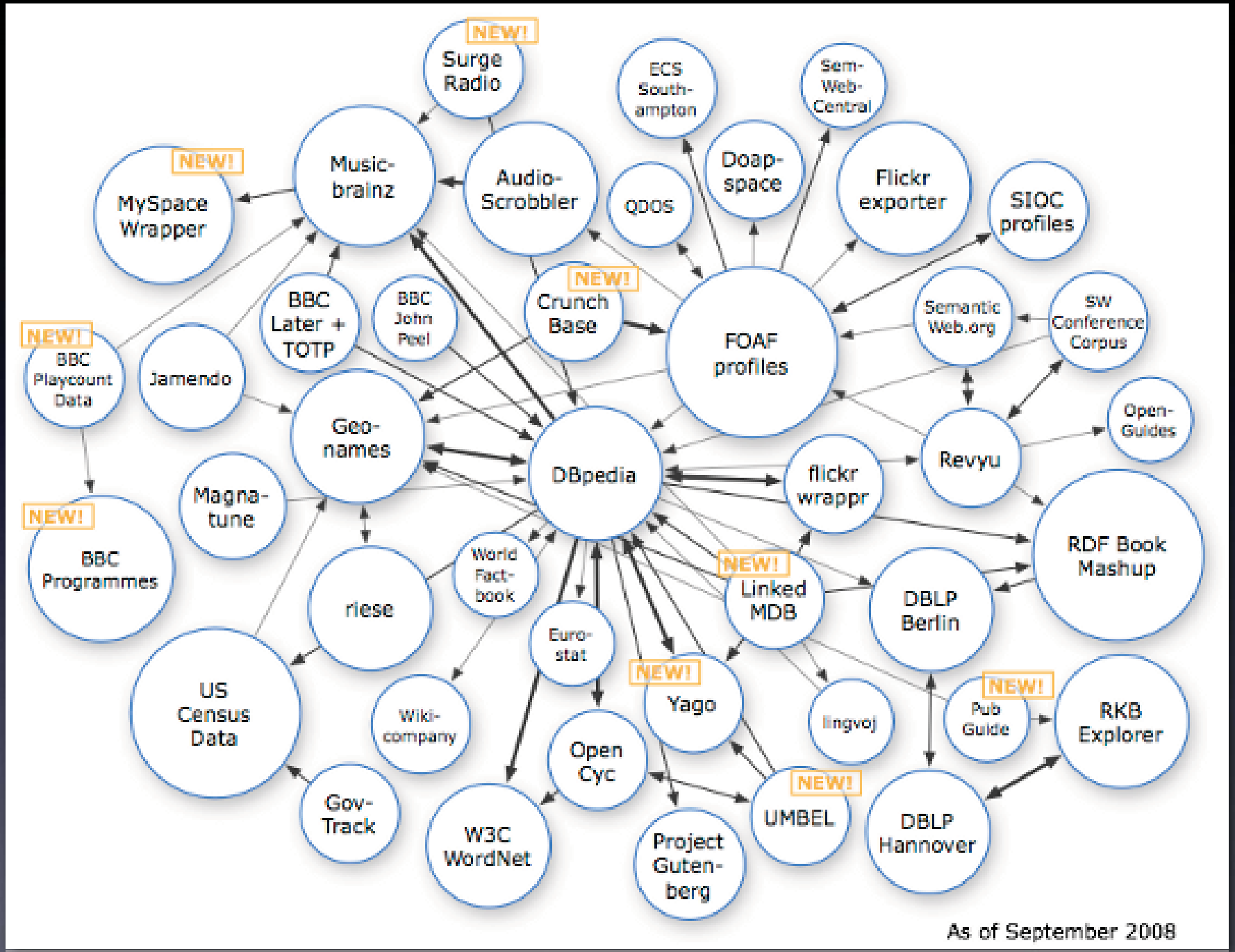
just by doing
HTML / HTTP

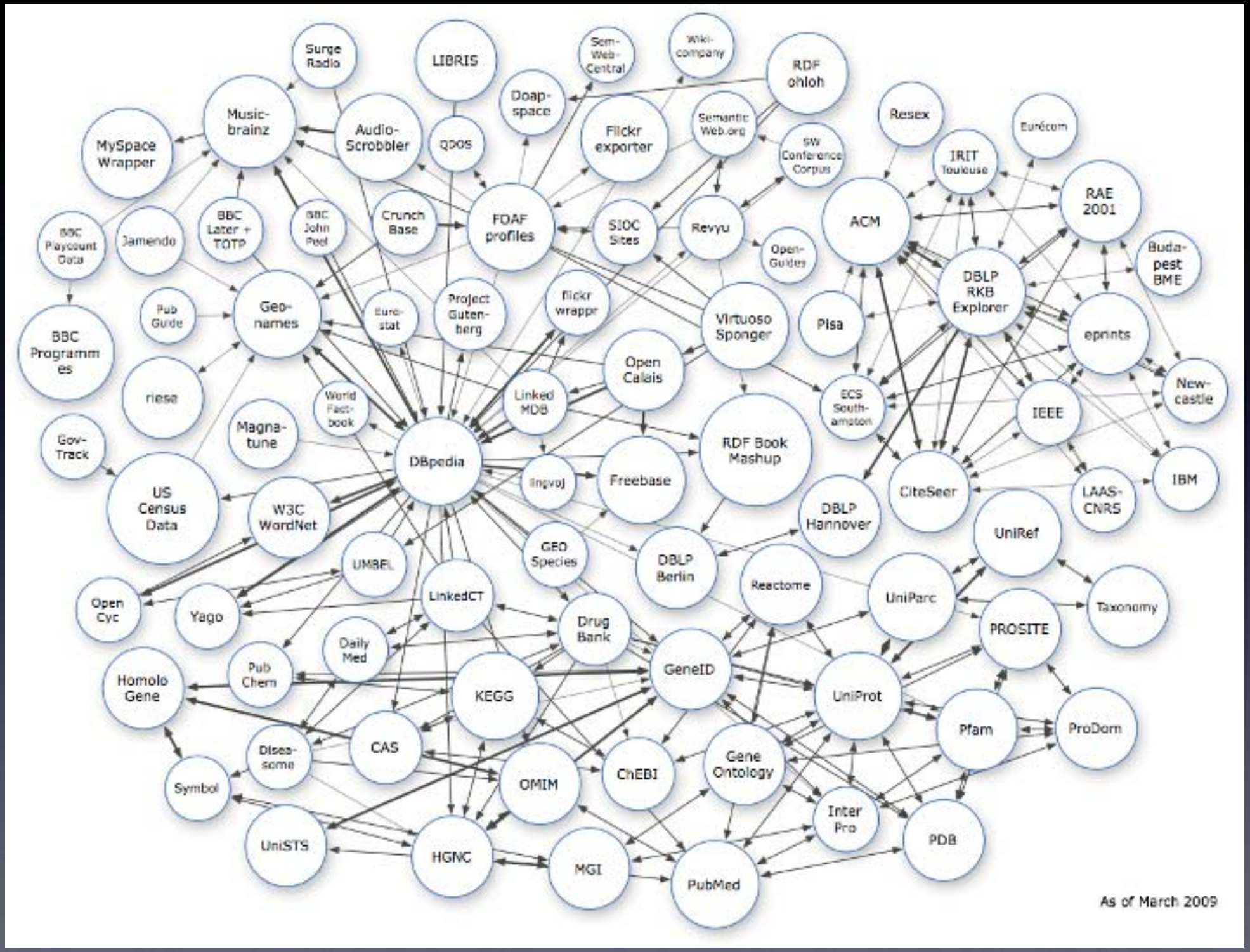
use URIs as names for things
use HTTP URIs
provide useful information
include links to other URIs

http://www.w3.org/DesignIssues/LinkedData.html

doesn't *fit* the web,
it *is* the web

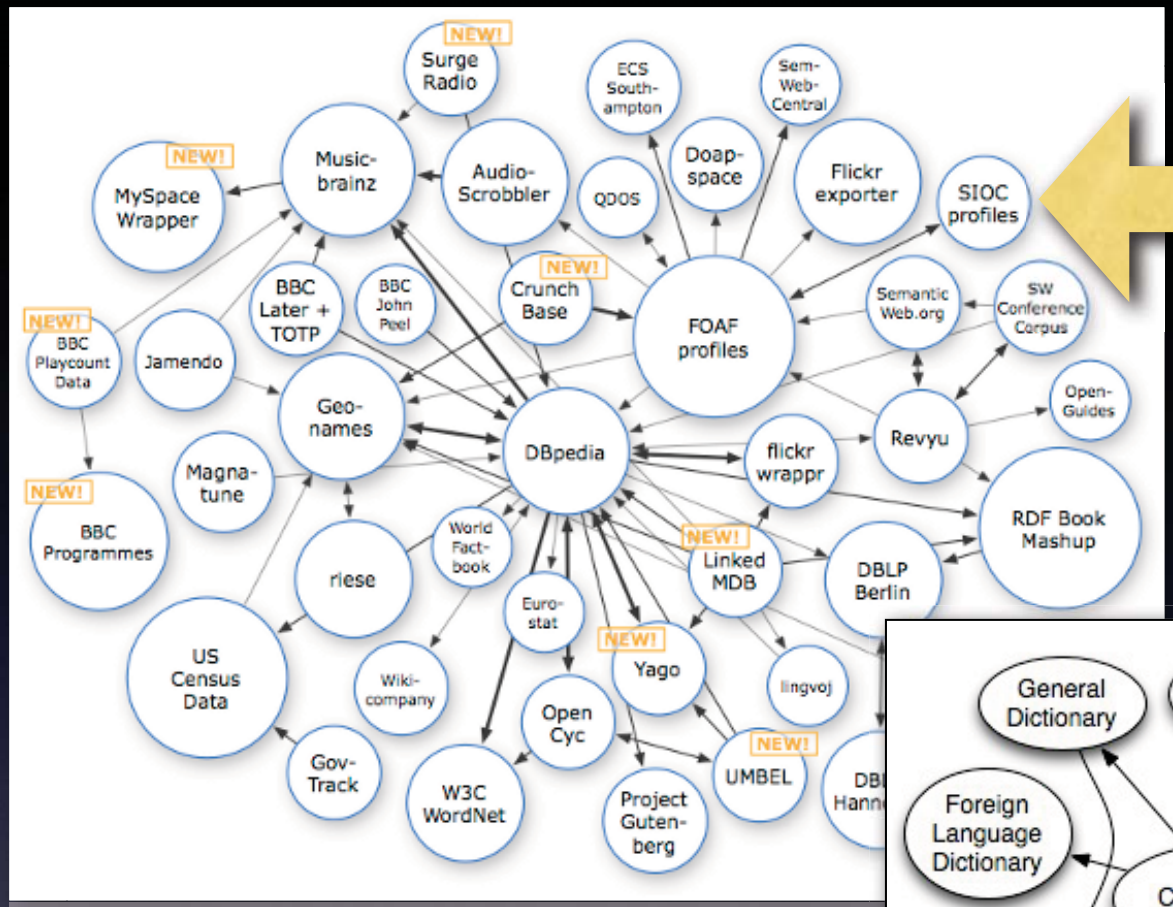
many examples already


NEW!
Surge Radio
ECS South-ampton
Sem-Web-Central
NEW!
MySpace Wrapper
Music-brainz
Audio-Scrobbler
QDOS
Doap-space
Flickr exporter
SIOC profiles
BBC Later + TOTP
BBC John Peel
NEW!
Crunch Base
FOAF profiles
Semantic Web.org
SW Conference Corpus
NEW!
BBC Playcount Data
Jamendo
Geo-names
DBpedia
Revyu
Open-Guides
flickr wrappr
Magna-tune
NEW!
BBC Programmes
RDF Book Mashup
World Fact-book
NEW!
Linked MDB
DBLP Berlin
riese
Euro-stat
US Census Data
NEW!
Yago
NEW!
Pub Guide
RKB Explorer
Wiki-company
Open Cyc
lingvoj
Gov-Track
NEW!
UMBEL
W3C WordNet
Project Guten-berg
DBLP Hannover
As of September 2008


Surge Radio
LIBRIS
Sem-Web-Central
Wiki-company
RDF ohloh
Doap-space
Music-brainz
Audio-Scrobbler
Flickr exporter
Semantic Web.org
SW Conference Corpus
Resex
Eurécom
IRIT Toulouse
MySpace Wrapper
QDOS
RAE 2001
BBC Playcount Data
Jamendo
BBC Later + TOTP
BBC John Peel
Crunch Base
FOAF profiles
SIOC Sites
Revyu
ACM
Open-Guides
Buda-pest BME
Pub Guide
Geo-names
Euro-stat
Project Guten-berg
flickr wrappr
Virtuoso Sponger
Pisa
DBLP RKB Explorer
eprints
BBC Programmes
riese
Open Calais
ECS South-ampton
New-castle
World Fact-book
Linked MDB
IEEE
Magna-tune
Gov-Track
DBpedia
RDF Book Mashup
Freebase
Ingvoj
CiteSeer
LAAS-CNRS
IBM
US Census Data
W3C WordNet
DBLP Hannover
UniRef
UMBEL
GEO Species
DBLP Berlin
Reactome
UniParc
Taxonomy
Open Cyc
Yago
LinkedCT
Drug Bank
PROSITE
Daily Med
GeneID
Homolo Gene
Pub Chem
KEGG
UniProt
Pfam
ProDom
Disea-some
CAS
Gene Ontology
Symbol
OMIM
ChEBI
Inter Pro
UniSTS
HGNC
MGI
PubMed
PDB
As of March 2009

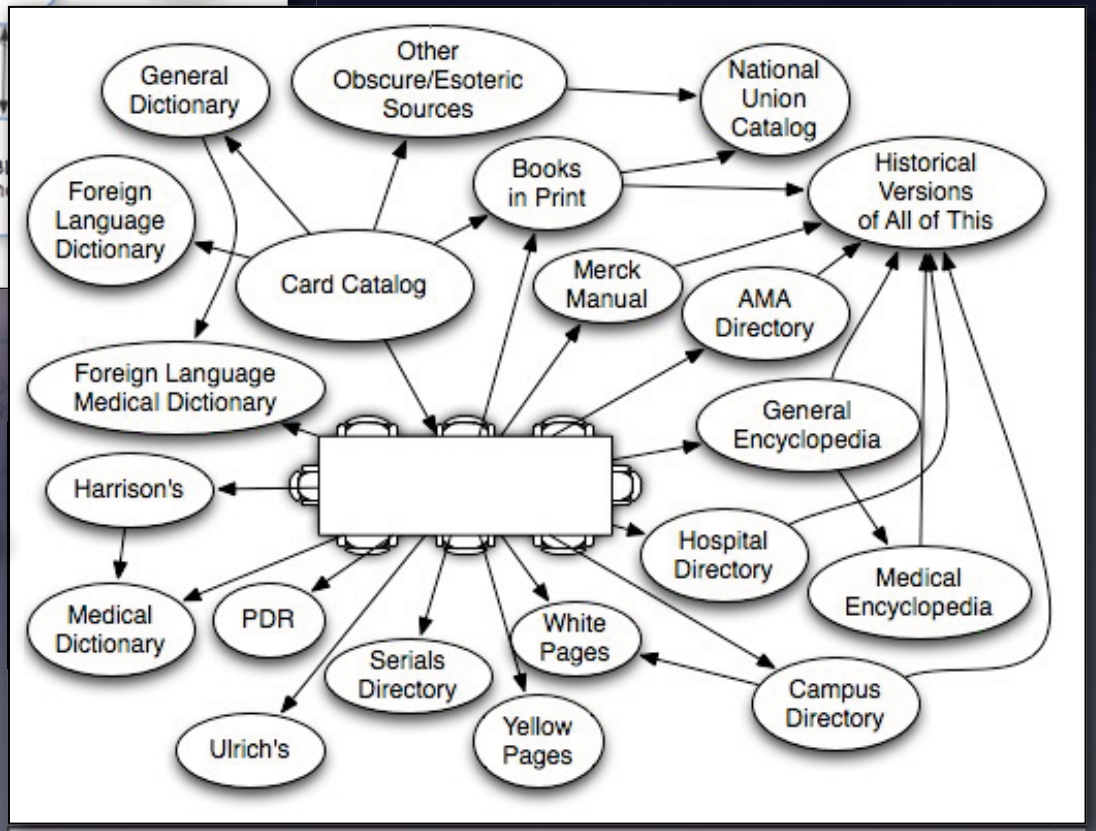
LIBRIS
id.loc.gov
chronicling america

if i squint just right...

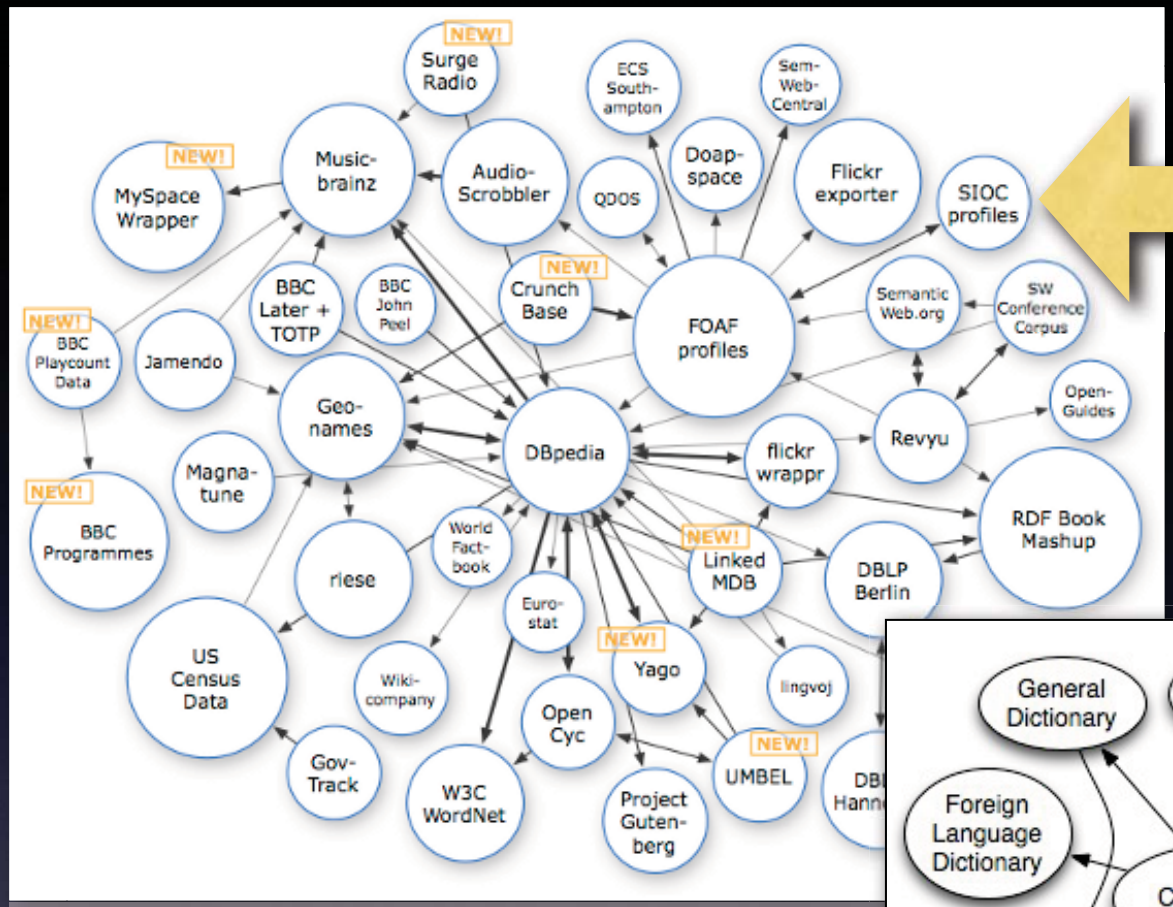


the "LOD cloud"

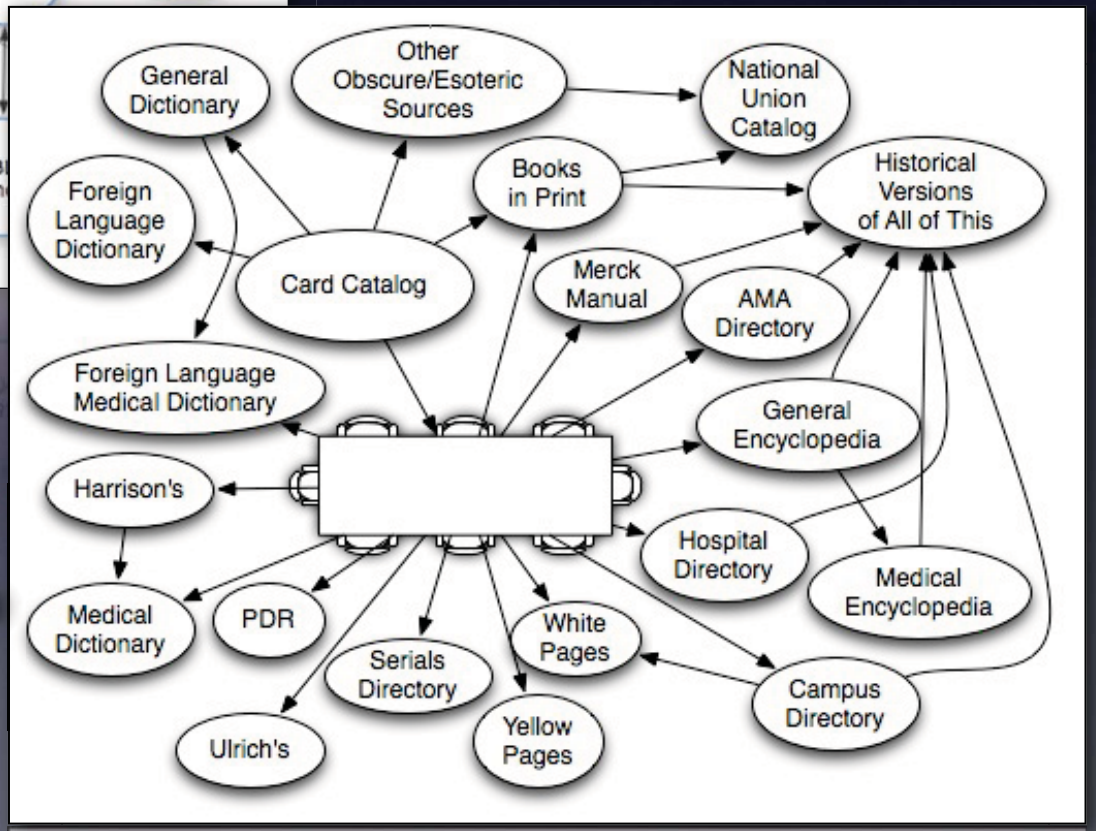
a "ready reference" collection



make this useful



turn this into linked data



seems obvious,
right?

3.

steps you can take today

how to
link your data

(or, About All That Other Stuff You Have)

use URIs as names for things
use HTTP URIs
provide useful information
include links to other URIs

http://www.w3.org/DesignIssues/LinkedData.html

first:
clean, cool URIs
for data concepts

instead of this:
/search?subject=Strayhorn,+Billy

do this:
/subject/lccn/n81072976

easy to implement
today

easy to redirect
tomorrow

next:
offer data views

use

<a rel=____ href=____>

<link rel=___>

offer up
MARC,
or JSON,
or RDF,
or ...

next:
relate your data
to established
authorities

your search results
backed by shared
authority records

put it all together

visit /subject/lccn/n81072976

- get your content results
- get alternate data views (concept record)
- get links to other useful data
- get links to authorized form

all that's easy

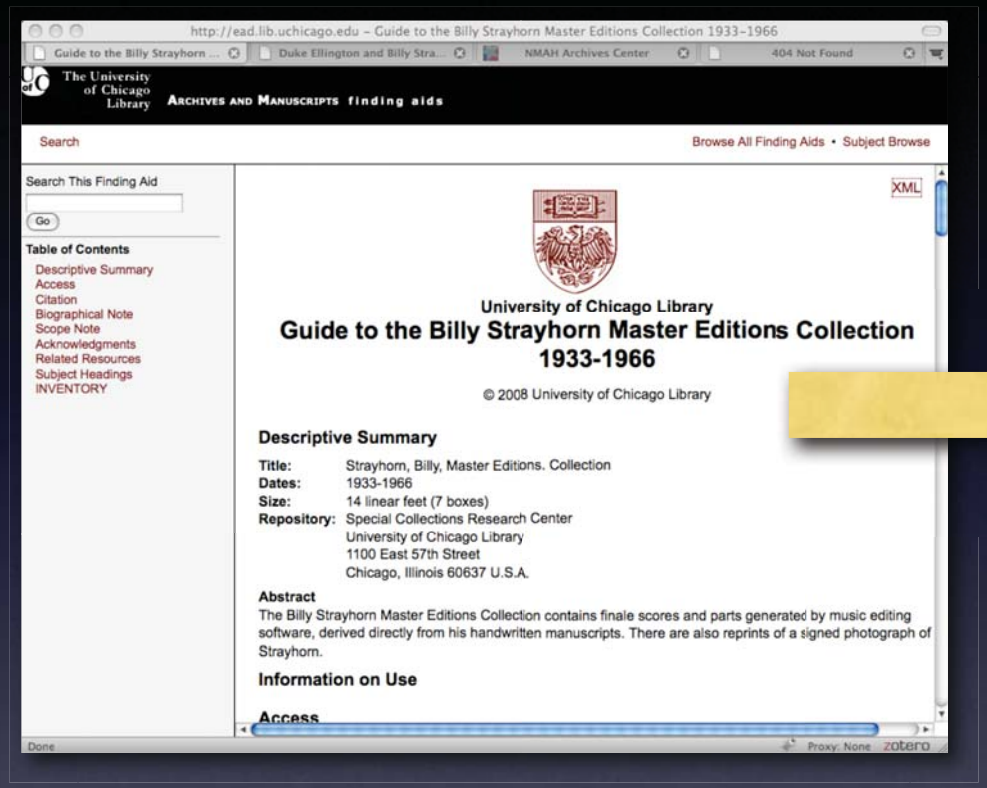
i promise

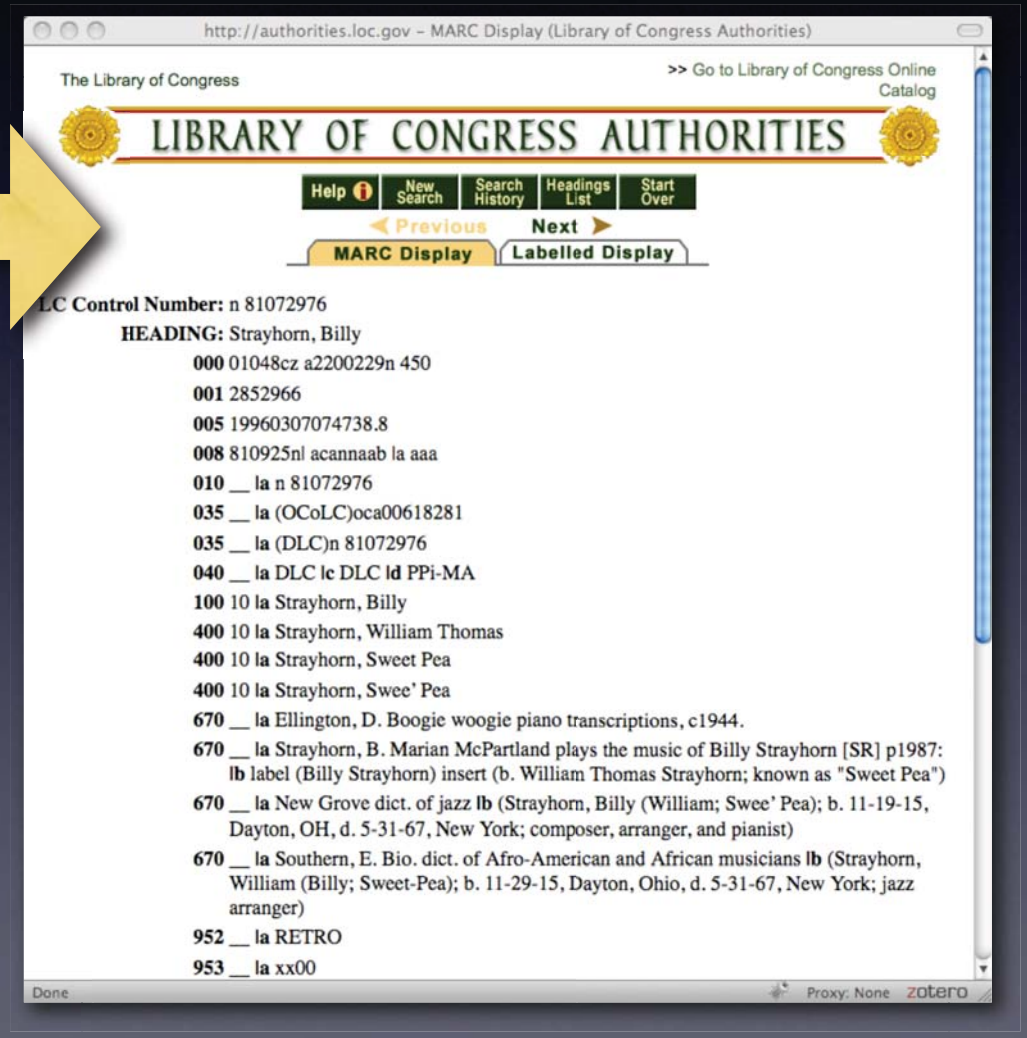
the next part's hard

how to make it last

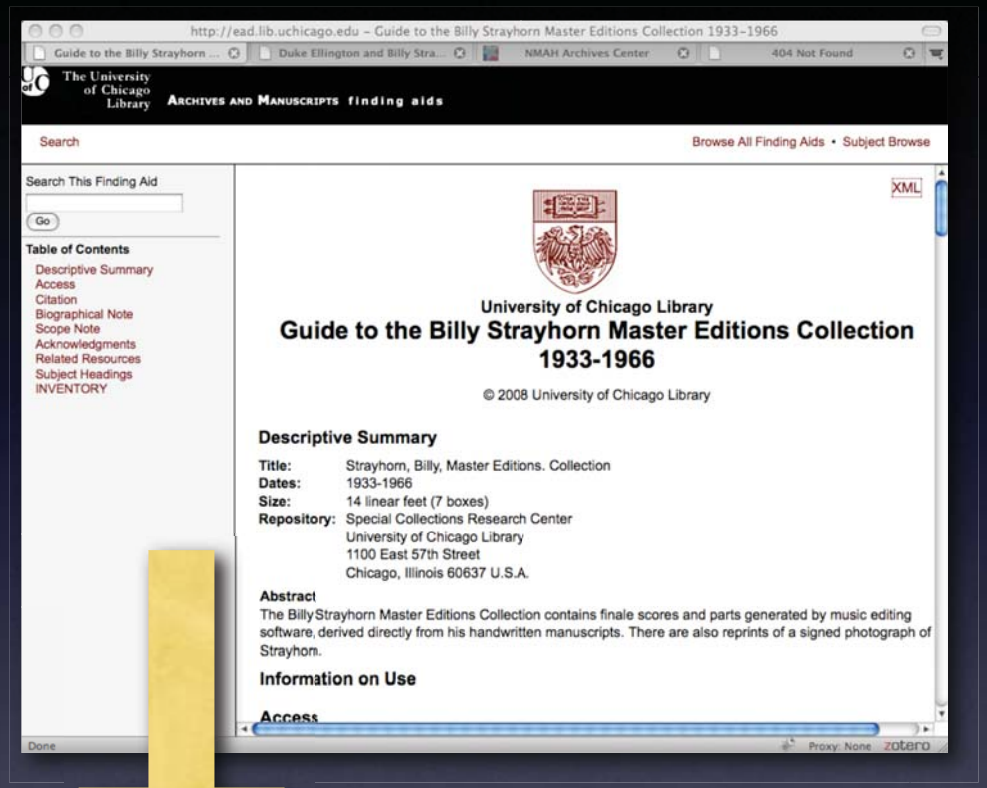
if your site breaks
when links break
cache and link yourself

instead of just this

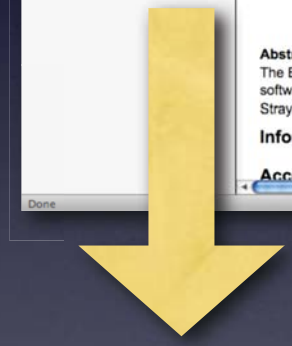




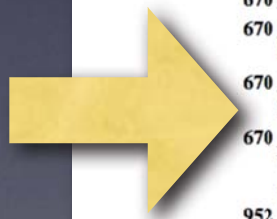
do this

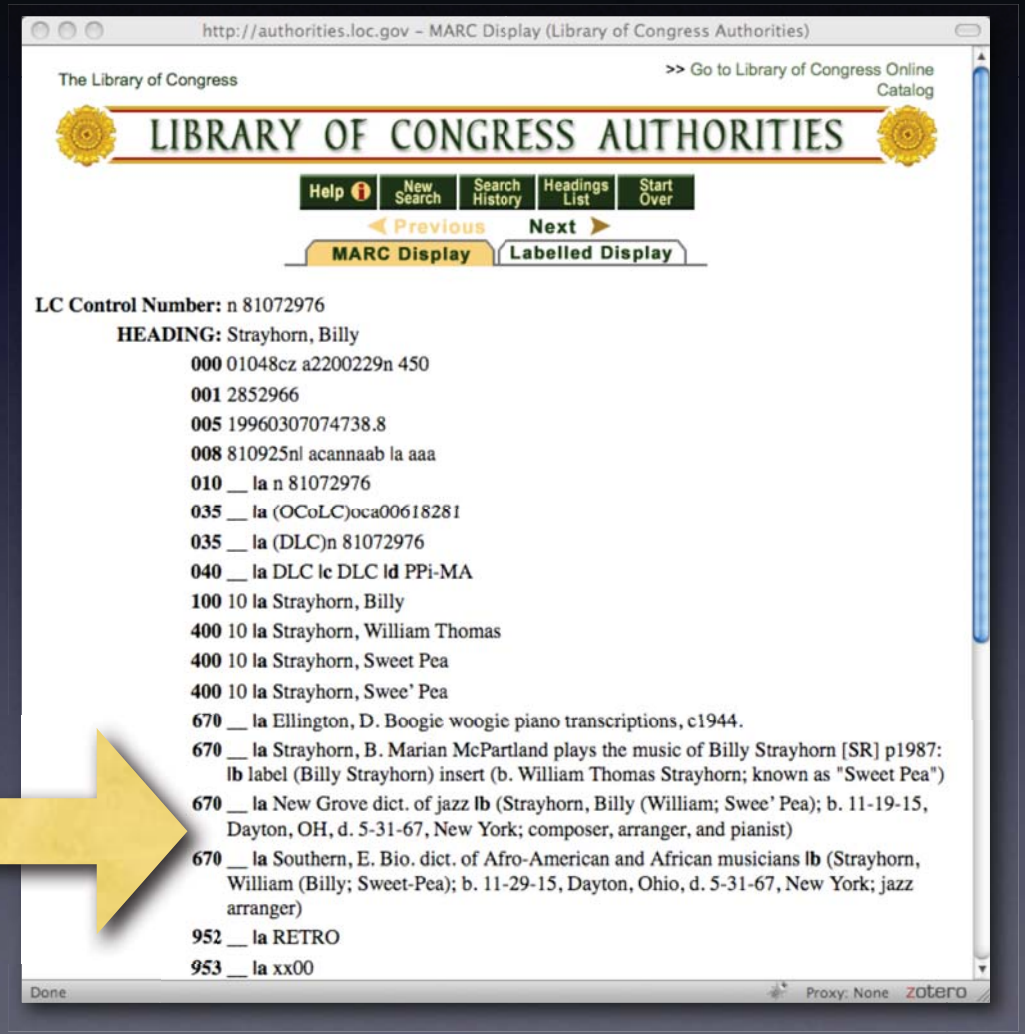


UChicago cache



same as





just like
we do
in ILS/OPACs

but, yknow,
the web way

so if a remote
link breaks

your local links
still work

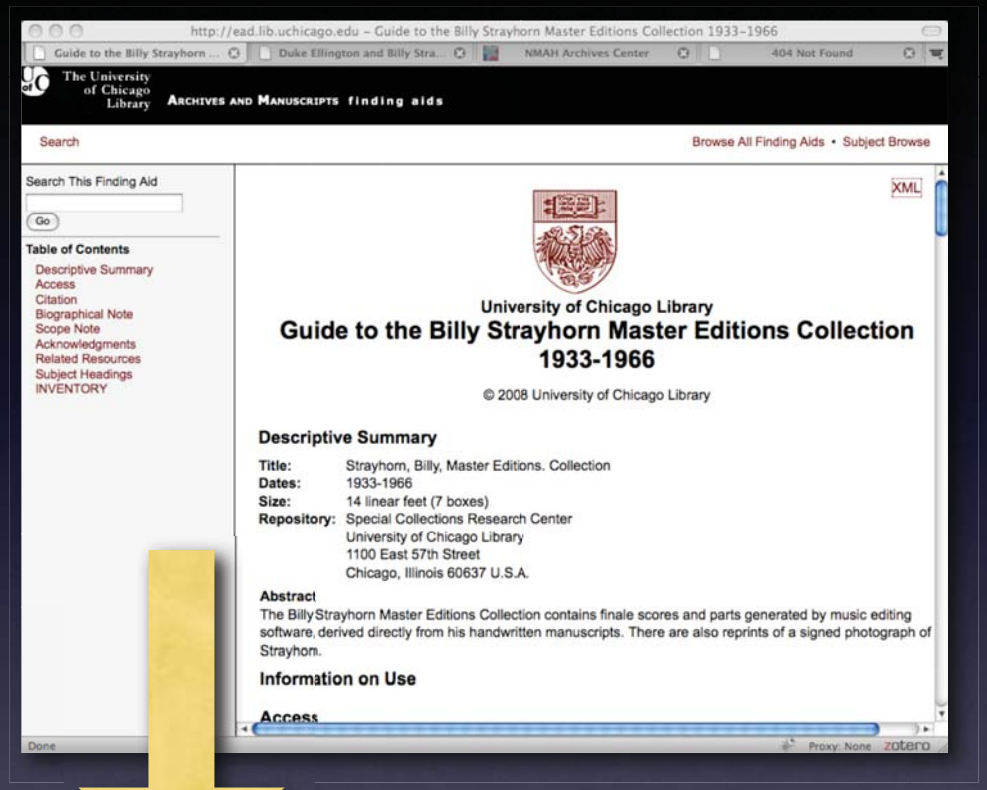
also,
proxy concepts

your
/subject/lccn/n81072976
==
my
/subject/lccn/n81072976

but in *your* app

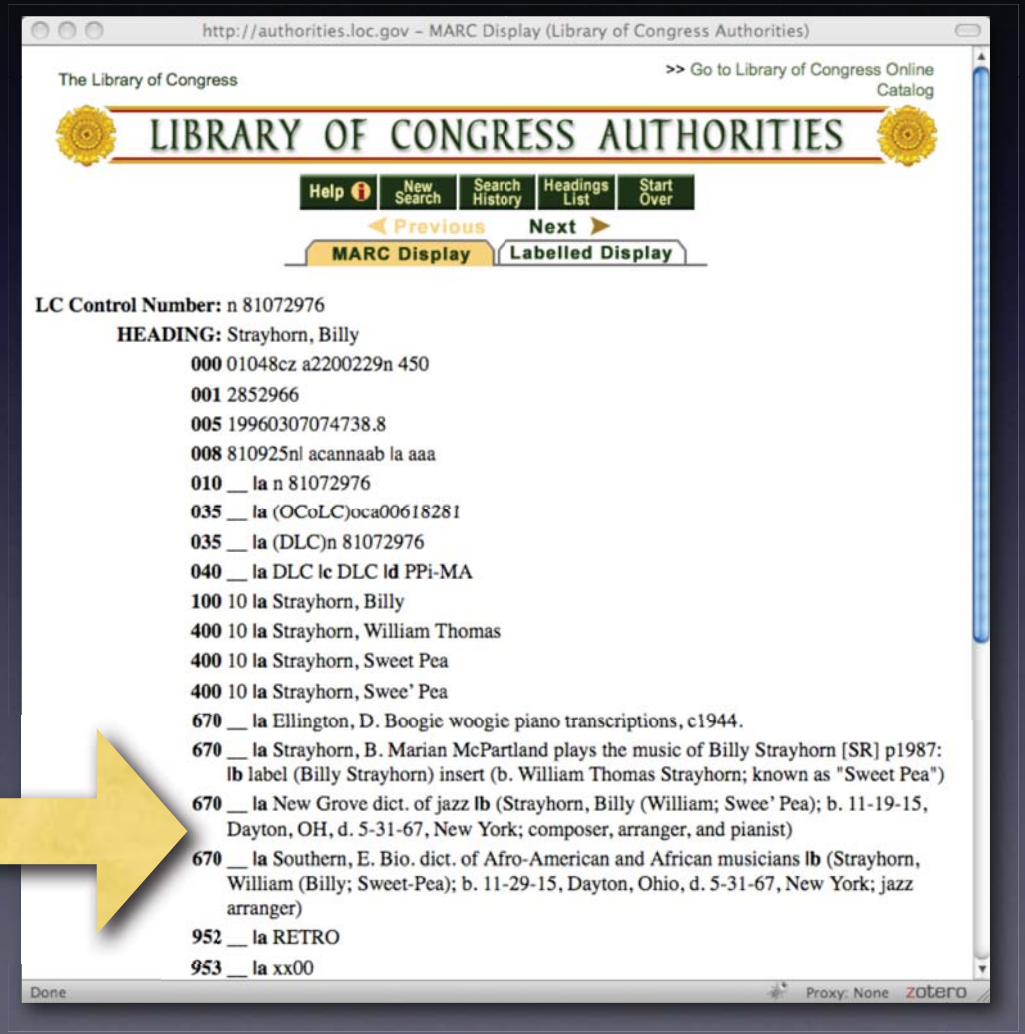
make every cache
its own
linked data source

instead of just this

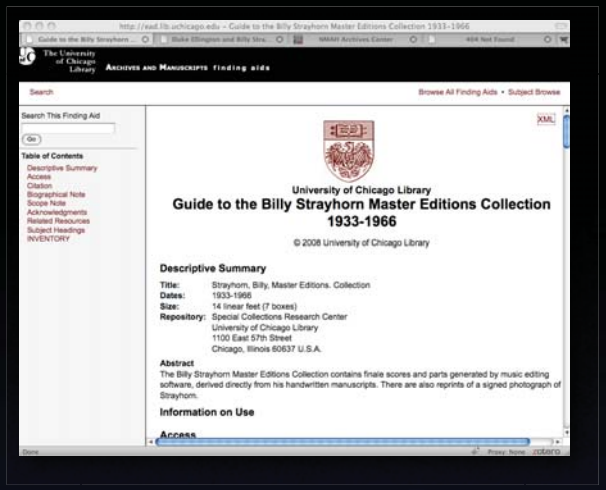


UChicago cache

same as

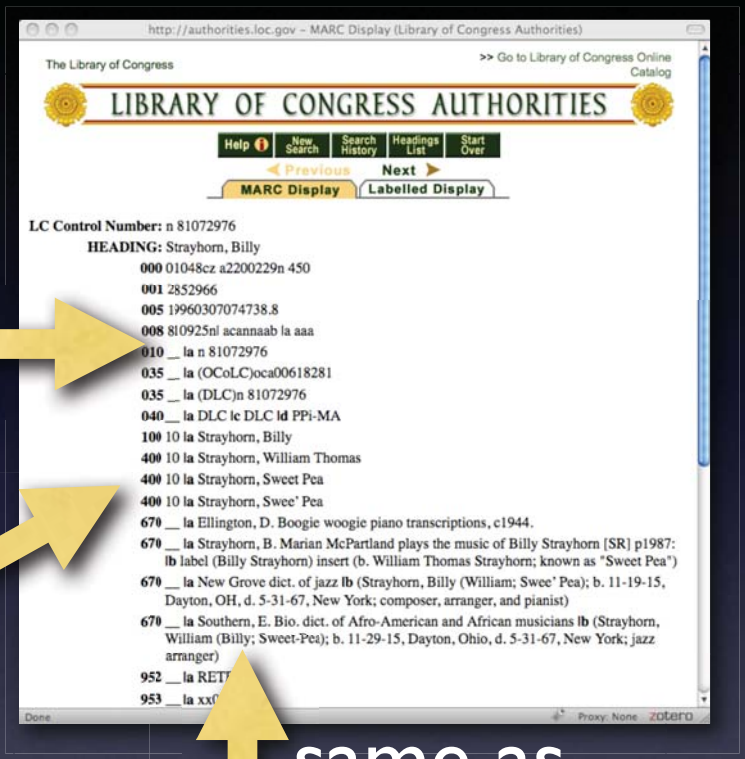


enable this



UChicago cache

same as

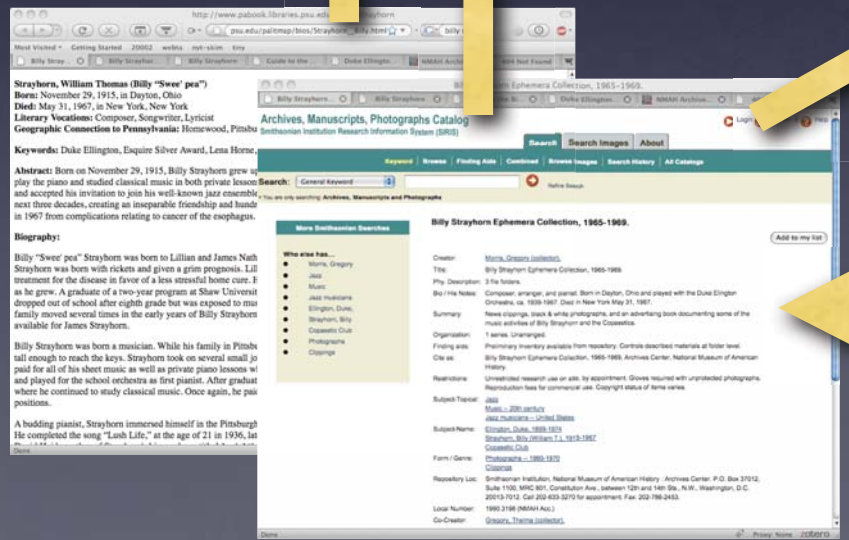


same as

same as

same as

same as



if one goes down
the others live on

if others live on
it's all still linked

if it's all still linked
we can keep adding

linked data
that lasts

i think

thanks!

dchud at umich edu
onebiglibrary.net
twitter.com/dchud