

Authority data

the good, the dirty and the semantic

E. Lynette Rayle

Cornell University

Dave Eichmann

The University of Iowa

Linked Data for Production Grant Funded by [Andrew W. Mellon Foundation](#)



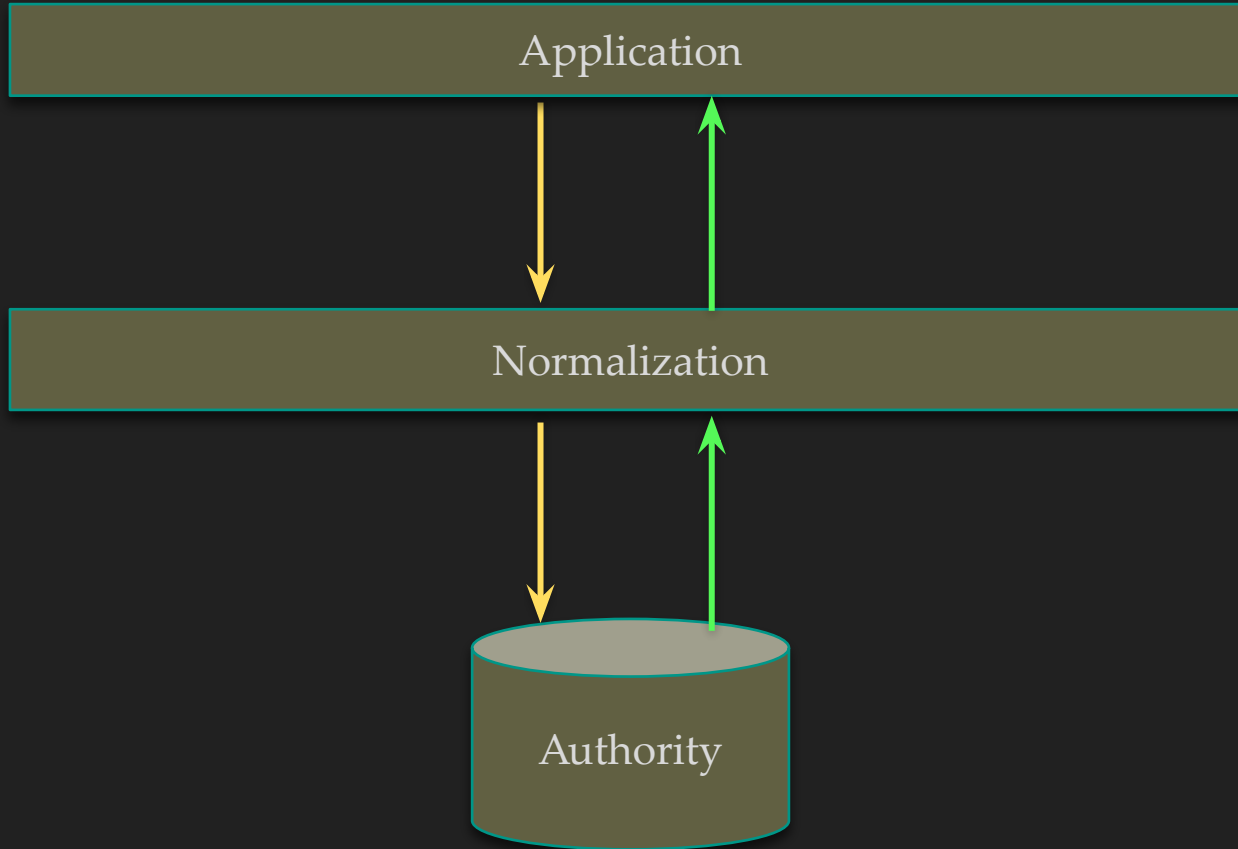
Who we are...

This work is part of the LD4P project and done in support of accessing controlled vocabularies from various authorities.

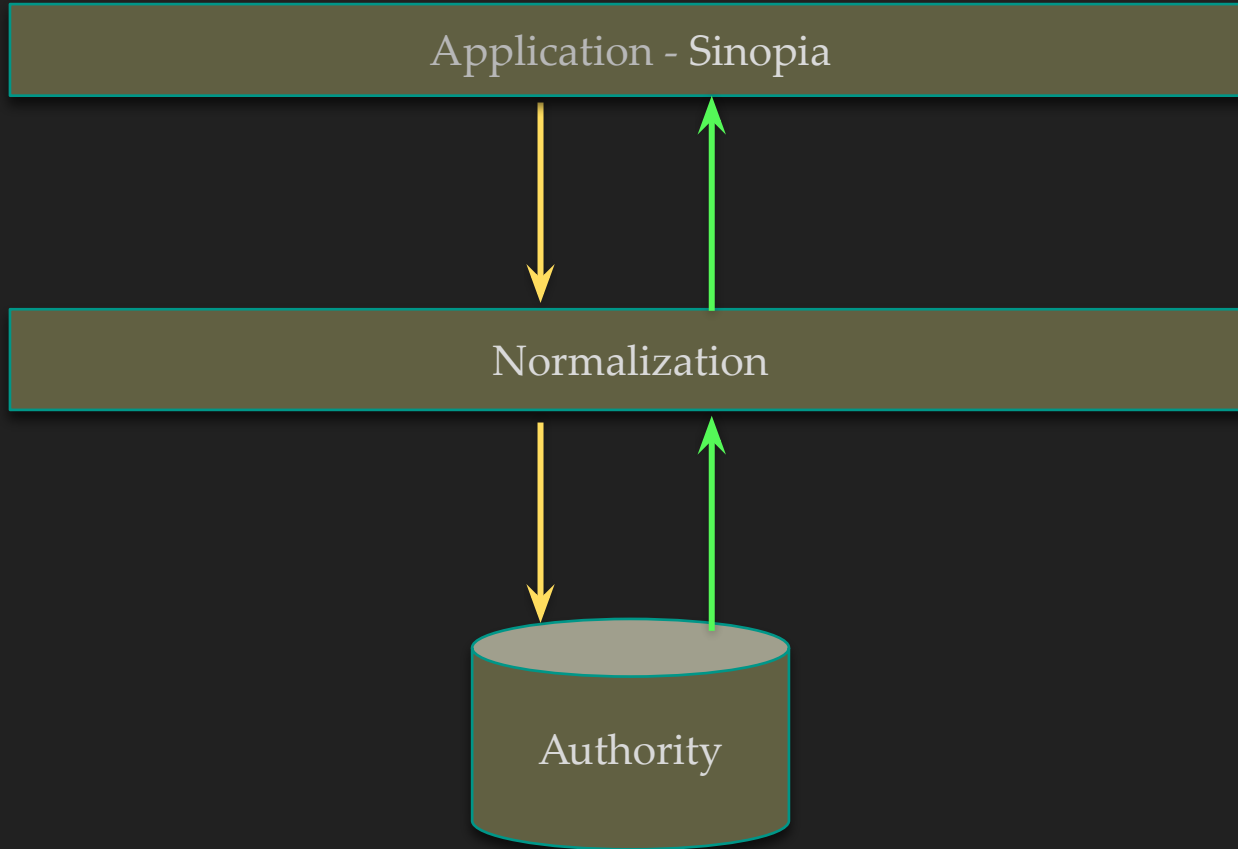
Lynette works on the normalization layer that provides a consistent layout of the data to allow the end applications to work with data in the same way regardless of where the data is coming from and what ontology is used to represent the data.

Dave works on the backend triple store caching system and indexing.

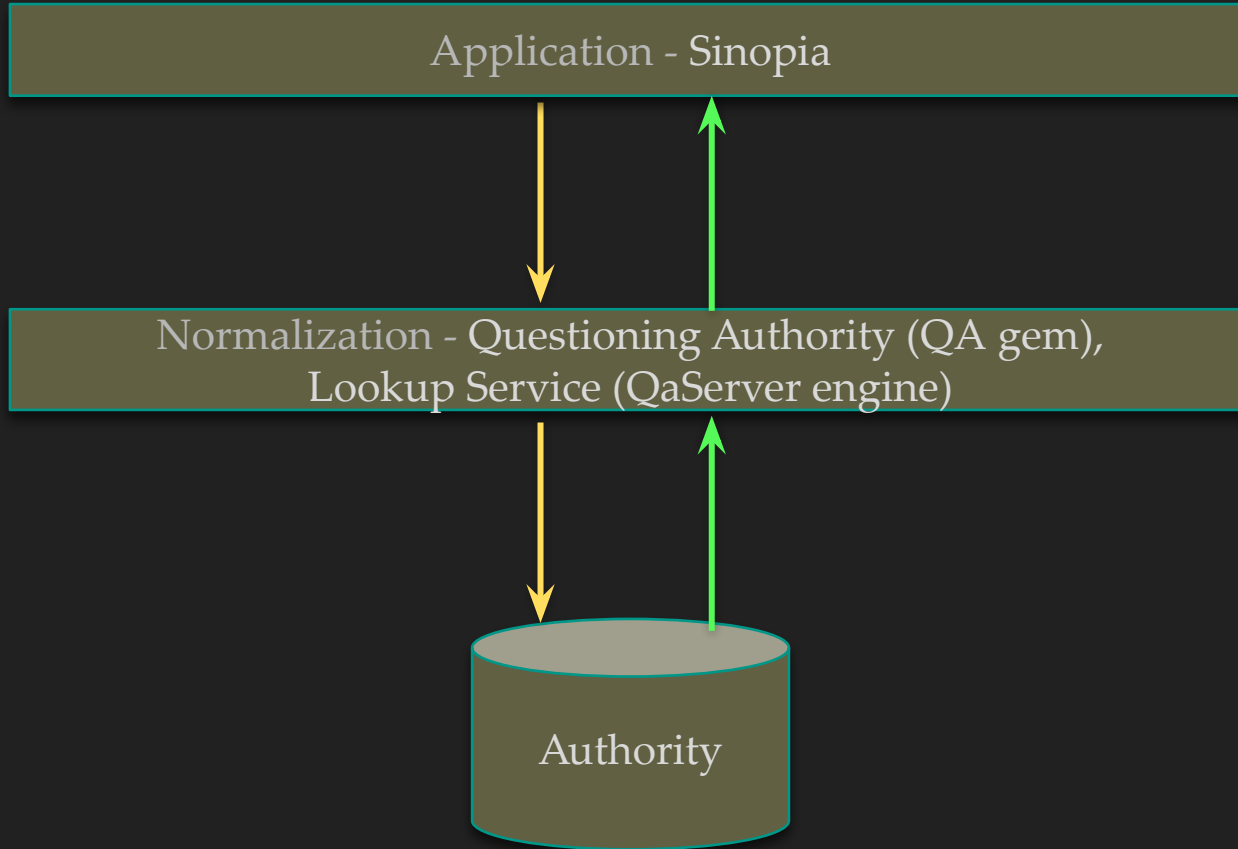
Architecture



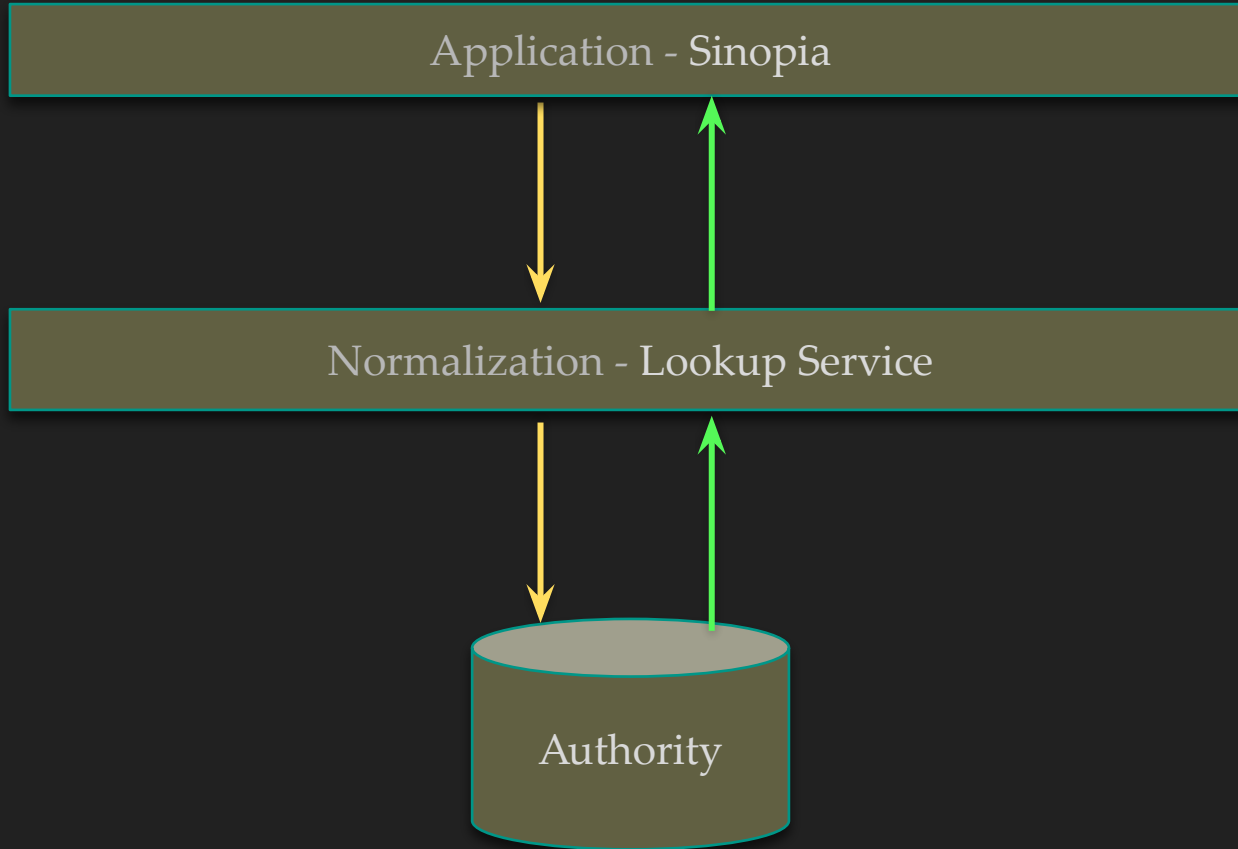
Architecture



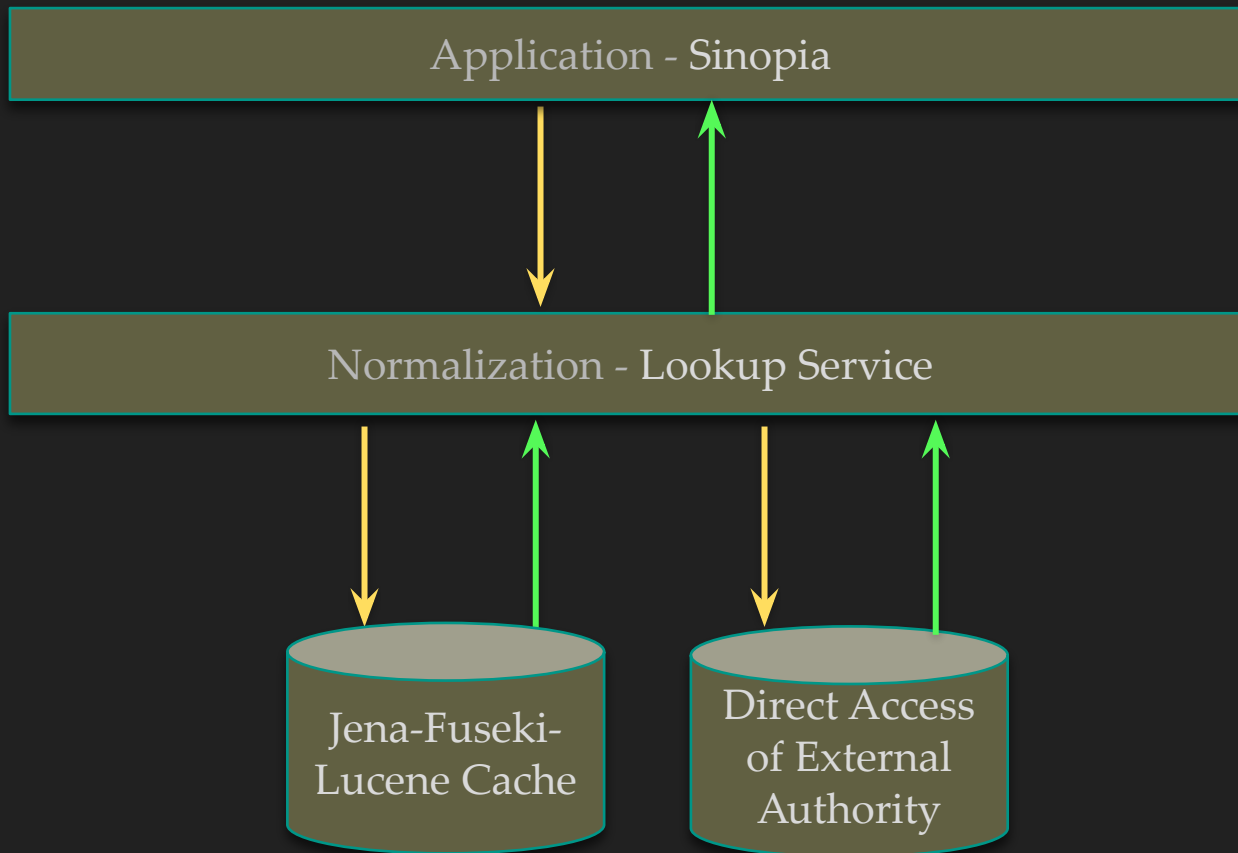
Architecture



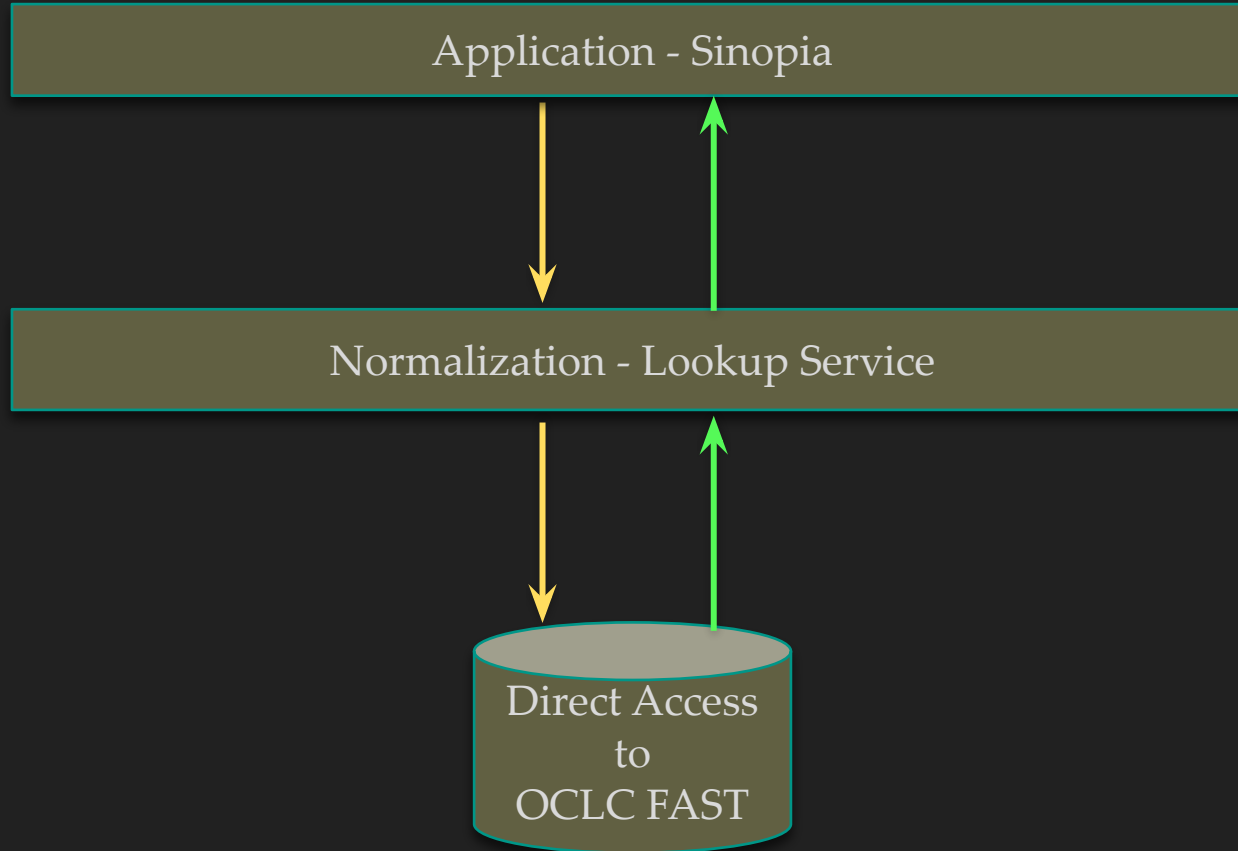
Architecture



Architecture



Lookup Search Request



Lookup Search Request

Application - Sinopia

https://lookup.ld4l.org/authorities/search/linked_data/oclc_fast/personal_name?q=twain&maximumRecords=2

Normalization - Lookup Service

Direct Access
to
OCLC FAST

Lookup Search Request

Application - Sinopia

http://lookup.ld4l.org/authorities/search/linked_data/oclc_fast/personal_name?q=twain&maximumRecords=2

Normalization - Lookup Service

<http://experimental.worldcat.org/fast/search?query=oclc.personalName+%22twain%22&sortKeys=usage&maximumRecords=2>

Direct Access
to
OCLC FAST

* search of cache performed via Lucene index

Lookup Search Request

Application - Sinopia

http://lookup.ld4l.org/authorities/search/linked_data/oclc_fast/personal_name?q=twain&maximumRecords=2

Normalization - Lookup Service

<http://experimental.worldcat.org/fast/search?query=oclc.personalName+%22twain%22&sortKeys=usage&maximumRecords=2>

Direct Access
to
OCLC FAST

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:coord="http://whatever/java/ORG.oclc.util.CoordToDecimal"
  xmlns:naco="http://whatever/java/ORG.oclc.util.NacoNormalize"
  xmlns:norm="http://whatever/java/ORG.oclc.util.NormalFormC" xmlns:schema="http://schema.org/"
  xmlns:madsrdf="http://www.loc.gov/mads/rdf/v1#"
  xmlns:skosxl="http://www.w3.org/2008/05/skos-xl#" xmlns:dct="http://purl.org/dc/terms/"
  xmlns:foaf="http://xmlns.com/foaf/0.1/" xmlns:void="http://rdfs.org/ns/void#"
  xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:mx="http://www.loc.gov/MARC21/slim"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xml:base="http://id.worldcat.org/fast"/>
  <rdf:Description rdf:about="31622">
    <dct:identifier>31622</dct:identifier>
    <dct:replaces rdf:resource="1986616"/>
    <skos:inScheme rdf:resource="ontology/1.0/#fast"/>
    <rdf:type rdf:resource="http://schema.org/Person"/>
    <skos:inScheme rdf:resource="ontology/1.0/#facet-Personal"/>
    <skos:prefLabel>Twain, Mark, 1835-1910</skos:prefLabel>
    <schema:name>Twain, Mark, 1835-1910</schema:name>
    <skos:altLabel>Clemens, Samuel Langhorne, 1835-1910</skos:altLabel>
    <schema:name>Clemens, Samuel Langhorne, 1835-1910</schema:name>
    <skos:altLabel>Make Teviin, 1835-1910</skos:altLabel>
```

Lookup Search Request

Application - Sinopia

http://lookup.ld4l.org/authorities/search/linked_data/oclc_fast/personal_name?q=twain&maximumRecords=2

```
[{"uri": "http://id.worldcat.org/fast/31622", "id": "31622",  
  "label": "Twain, Mark, 1835-1910"},  
 {"uri": "http://id.worldcat.org/fast/365563", "id": "365563",  
  "label": "Twain, Shania"}]
```

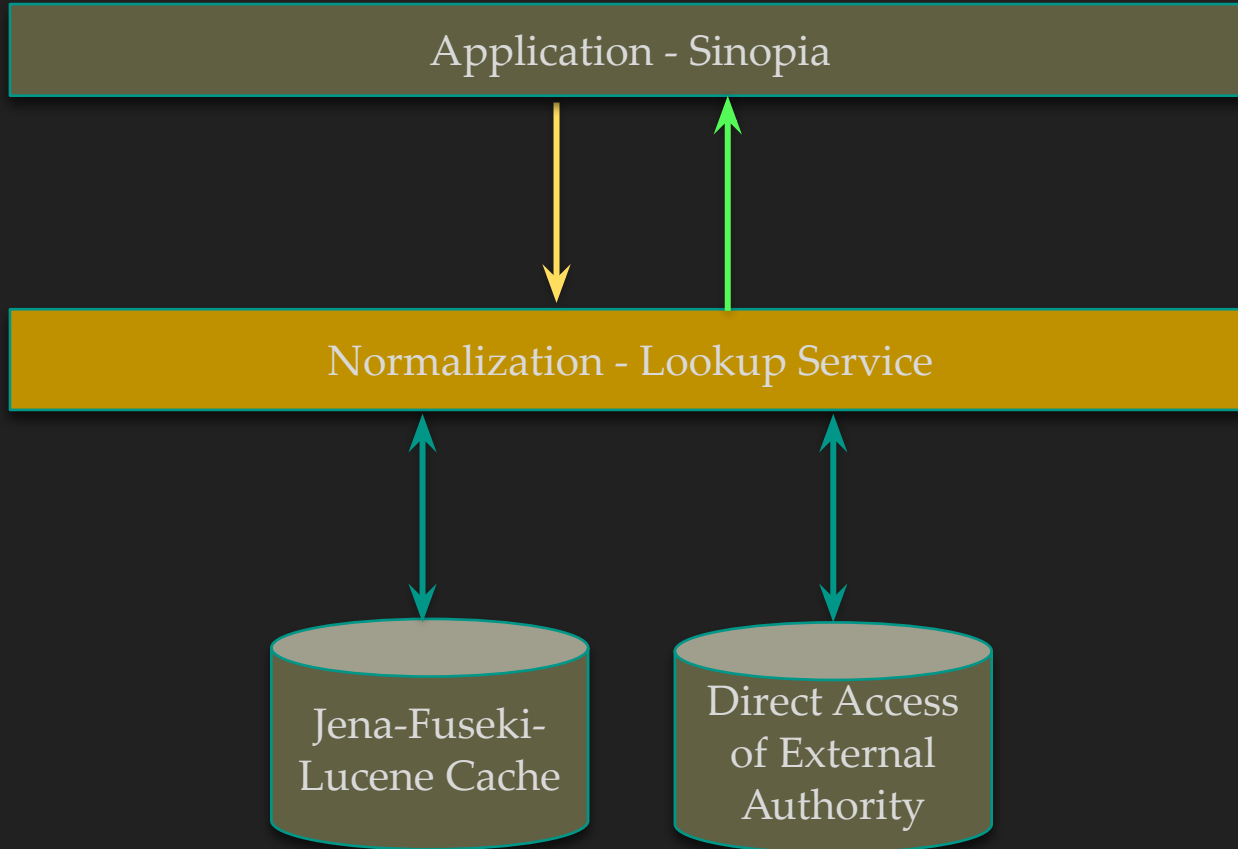
Normalization - Lookup Service

<http://experimental.worldcat.org/fast/search?query=oclc.personalName+%22twain%22&sortKeys=usage&maximumRecords=2>

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"  
  xmlns:dc="http://purl.org/dc/elements/1.1/"  
  xmlns:coord="http://whatever/java/ORG.oclc.util.CoordToDecimal"  
  xmlns:naco="http://whatever/java/ORG.oclc.util.NacoNormalize"  
  xmlns:norm="http://whatever/java/ORG.oclc.util.NormalFormC" xmlns:schema="http://schema.org/"  
  xmlns:madsrdf="http://www.loc.gov/mads/rdf/v1#"  
  xmlns:skosxl="http://www.w3.org/2008/05/skos-xl#" xmlns:dct="http://purl.org/dc/terms/"  
  xmlns:foaf="http://xmlns.com/foaf/0.1/" xmlns:void="http://rdfs.org/ns/void#"  
  xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:mx="http://www.loc.gov/MARC21/slim"  
  xmlns:skos="http://www.w3.org/2004/02/skos/core#"  
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xml:base="http://id.worldcat.org/fast/">  
  <rdf:Description rdf:about="31622">  
    <dct:identifier>31622</dct:identifier>  
    <dct:replaces rdf:resource="1986616"/>  
    <skos:inScheme rdf:resource="ontology/1.0/#fast"/>  
    <rdf:type rdf:resource="http://schema.org/Person"/>  
    <skos:inScheme rdf:resource="ontology/1.0/#facet-Personal"/>  
    <skos:prefLabel>Twain, Mark, 1835-1910</skos:prefLabel>  
    <schema:name>Twain, Mark, 1835-1910</schema:name>  
    <skos:altLabel>Clemens, Samuel Langhorne, 1835-1910</skos:altLabel>  
    <schema:name>Clemens, Samuel Langhorne, 1835-1910</schema:name>  
    <skos:altLabel>Make Teviin, 1835-1910</skos:altLabel>
```

Direct Access
to
OCLC FAST

Architecture



Initial Basic Use Case

Simple auto-complete for a field -- type search query

Creator required

The person or group responsible for the work. Usually this is the author of the content. Personal names should be entered with the last name first, e.g. "Smith, John."

Requirements

Describe your work

Visibility

- Open Access** Everyone. Check out [SHERPA/RoMEO](#) for specific publishers' copyright policies if you plan to patent and/or publish your Demo Work in a journal.
- Institution** Restrict access to only users and/or groups from Institution
- Embargo**
- Lease**
- Private** Only users and/or

Normalized JSON from the Lookup Service

```
[
  {
    "uri":"http://id.loc.gov/authorities/names/n85389596",
    "id":"n 85389596",
    "label":"Smith, John W. (John Williamson)"
  },
  {
    "uri":"http://id.loc.gov/authorities/names/n86851637",
    "id":"n 86851637",
    "label":"Smith, John W. (John Williamson), 1897-1986"
  },
  {
    "uri":"http://id.loc.gov/authorities/names/n88005259",
    "id":"n 88005259",
    "label":"Simon, John S. (John Smith), 1843-1933"
  },
  ...
]
```


Simple auto-complete for a field -- make a selection

Creator required

The person or group responsible for the work. Usually this is the author of the content. Personal names should be entered with the last name first, e.g. "Smith, John."

- Smith, John W. (John Williamson)
- Smith, John W. (John Williamson), 1897-1986
- Simon, John S. (John Smith), 1843-1933
- St. John-Smith, C. (Christopher), 1956-
- Smith, John Austin, 1954-
- Walker-Smith, John A.
- Smith, John, 1921-

Requirements

- Describe your work

Visibility

- Open Access** Everyone. Check out [SHERPA/RoMEO](#) for specific publishers' copyright policies if you plan to patent and/or publish your Demo Work in a journal.
- Institution** Restrict access to only users and/or groups from Institution
- Embargo**
- Lease**
- Private** Only users and/or

Simple auto-complete for a field -- set value in field

Creator required

The person or group responsible for the work. Usually this is the author of the content. Personal names should be entered with the last name first, e.g. "Smith, John."

Smith, John W. (John Williamson), 1897-1986

Requirements

Describe your work

Visibility

- Open Access** Everyone. Check out [SHERPA/RoMEO](#) for specific publishers' copyright policies if you plan to patent and/or publish your Demo Work in a journal.
- Institution** Restrict access to only users and/or groups from Institution
- Embargo**
- Lease**
- Private** Only users and/or

Simple auto-complete for a field -- also store URI

Creator required

The person or group responsible for the work. Usually this is the author of the content. Personal names should be entered with the last name first, e.g. "Smith, John."

Smith, John W. (John Williamson), 1897-1986

Uri: <http://id.loc.gov/authorities/names/n86851637>

Requirements

Describe your work

Visibility

- Open Access** Everyone. Check out [SHERPA/RoMEO](#) for specific publishers' copyright policies if you plan to patent and/or publish your Demo Work in a journal.
- Institution** Restrict access to only users and/or groups from Institution
- Embargo**
- Lease**
- Private** Only users and/or

Extended Context

Lookup with Context

LOC Names: locnames_id4l_cache/person*

LITERAL WITH DEFAULT

DLC Language: English

- Barcode*
- Enumeration and chronology

Frequency (RDA 2.14)

Note on Frequency (RDA 2.17.12)

Note

Note

Lookup with Context

LOC Names: locnames_id4l_cache/person*

- + Barcode*
- + Enumeration and chronology

Frequency (RDA 2.14)

Note on Frequency (RDA 2.17.12)

Note

- + Note

LITERAL WITH DEFAULT

DLC

Lookup with Context -- type search query

LOC Names: [locnames_id4l_cache/person*](#)

Barcode*

Enumeration and chronology

[Frequency \(RDA 2.14\)](#)

[Note on Frequency \(RDA 2.17.12\)](#)

Note

Note

LITERAL WITH DEFAULT

DLC X

Normalized JSON + context from the Lookup Service

```
[
  {
    "uri": "http://id.loc.gov/authorities/names/n87118067",
    "id": "n 87118067",
    "label": "Taylor, Sam",
    "context": [
      ...
      {
        "property": "Variant Label",
        "values": ["Taylor, Samuel L., 1916-1990"],
        "Selectable": false, "drillable": false },
      {
        "group": "Dates", "property": "Birth: ",
        "values": ["(edtf) 1916-07-12"],
        "Selectable": false, "drillable": false },
      {
        "group": "Dates", "property": "Death: ",
        "values": ["(edtf) 1990-10-05"],
        "Selectable": false, "drillable": false },
      {
        "property": "Occupation",
        "values": ["Clarinetists, Jazz musicians, Saxophonists"],
        "Selectable": false, "drillable": false },
      ...
    ]
  }
]
```


Lookup with Context -- results + context in UI

Statement of Responsibility Relating to Title Proper (RD)
Statement of Responsibility Relating to Title Proper (R

Notes about the instance

Note

+ Note

LOC Names: locnames_id4l_cache/person

sam taylor

Search

Search

Search

LOC person [names] (QA)

- Taylor, Sam**
Variant Label: Taylor, Samuel L., 1916-1990
Birth: : (edtf) 1916-07-12
Death: : (edtf) 1990-10-05
Occupation: Clarinetists, Jazz musicians, Saxophonists
- Taylor, Sam, 1970-**
Birth: : (edtf) 1970
Occupation: Translators
- Taylor, Sam, 1979-**
- Taylor, Sam, 1968-**
Birth: : 19680319
Occupation: College teachers
Field of Activity: Creative writing
- Taylor, Sam S.**

LOC person [subjects] (QA)

- Taylor's checkerspot**
- Shorthand--Taylor**
- Taylor comet**

Lookup with Context -- selected entity

LOC Names: [locnames_id4l_cache/person*](#)

Taylor, Sam

Search

- + Barcode*
- + Enumeration and chronology

[Frequency \(RDA 2.14\)](#)

Frequency (RDA 2.14)

[Note on Frequency \(RDA 2.17.12\)](#)

Note

Mint URI Add

- + Note

LITERAL WITH DEFAULT

LITERAL WITH DEFAULT

DLC X Language: English

Lookup with Context -- application stores URI

LOC Names: [locnames_id4l_cache/person*](#)

Taylor, Sam

Search

Uri: <http://id.loc.gov/authorities/names/n87118067>

+ Barcode*

+ Enumeration and chronology

Frequency (RDA 2.14)

Frequency (RDA 2.14)

Note on Frequency (RDA 2.17.12)

Note Mint URI Add

+ Note

LITERAL WITH DEFAULT

LITERAL WITH DEFAULT

DLC Language: English

Bringing in data to multiple fields

(On the near horizon for Sinopia)

LINKED DATA EDITOR

[Search](#)[Resource Templates](#)[Load RDF](#)**Under Development - Not Exact Design**Search 

For searching Sinopia, use a * to wildcard y



Sinopia is a project c
available on Sinopia
[Dedication.](#)



Frank Sinatra

Format:**Label:** **Type:** ArtistFrank Sinatra - Frank Sinatra's Greatest Hits
(1968)**Format:** Vinyl,LP,Compilation**Label:** Reprise Records **Type:** MasterFrank Sinatra - Frank Sinatra Sings For Only The
Lonely (1958)**Format:** Vinyl,LP,Album,Mono**Label:** Capitol Records **Type:** MasterFrank Sinatra - A Jolly Christmas From Frank
Sinatra (1957)

Mellon Foundation. All metadata
[CC0 1.0 Universal Public Domain](#)

Discogs converted to Bibframe - Full record as json-ld

```
{
  "@context": {
    "bf2": "http://id.loc.gov/ontologies/bibframe/",
    "rdfs": "http://www.w3.org/2000/01/rdf-schema#",
    "xsd": "http://www.w3.org/2001/XMLSchema#",
    "rdf": "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  },
  "@graph": [
    {
      "@id": "Work1",
      "@type": [
        "bf2:Work",
        "bf2:Audio"
      ],
      "bf2:title": {
        "@id": "Work1Title"
      },
      "bf2:genreForm": [
        {
          "@id": "http://id.loc.gov/authorities/genreForms/gf2014026879"
        }
      ],
      "bf2:contribution": [
        {
          "@id": "Work1PrimaryContribution1"
        },
        {
          "@id": "Work1PrimaryContribution2"
        }
      ],
      {
        "@id": "Work1Title",
        "bf2:mainTitle": "Sorry / Why Remind Me"
      },
      {
        "@id": "http://id.loc.gov/authorities/genreForms/gf2014026879",
        "@type": "bf2:GenreForm",
        "rdfs:label": "Jazz"
      }
    },
    {
      "@id": "http://vocab.getty.edu/aat/300014502",
      "@type": "bf2:BaseMaterial",
      "rdfs:label": "Vinyl"
    },
    {
      "@id": "http://id.loc.gov/vocabulary/carriers/sd",
      "@type": "bf2:Carrier",
      "rdfs:label": "audio disc"
    },
    {
      "@id": "IdentifierPrimary",
      "@type": "bf2:Identifier",
      "rdf:value": {
        "@value": "11620608",
        "@type": "xsd:integer"
      }
    },
    {
      "@id": "ProvisionActivityDate",
      "@type": "bf2:ProvisionActivity",
      "bf2:date": {
        "@value": "1950",
        "@type": "xsd:date"
      }
    },
    {
      "@id": "ProvisionActivity1",
      "@type": "bf2:ProvisionActivity",
      "bf2:agent": {
        "@id": "Columbia"
      }
    }
  ]
}
```

[Browse](#)[Editor](#)[Import Resource Template](#)

Under Development - Not Exact Design

[Preview RDF](#)

Resource Template Label

[Clone|Edit] *Name of Resource*

Instance of*

BIBFRAME Work

[Mint URI](#) [Add](#)

- ✚ Title Information
- ✚ (Time) Coverage of the Content (RDA 7.3)
- ✚ Notes about the Work
- ▣ Content Type (RDA 6.9)

Content Type (RDA 6.9)

- ✚ Illustrative Content (RDA 7.15)
- ✚ Color Content (RDA 7.17)
- ✚ Has BIBFRAME Instance
- ✚ Authorized Access Point Representing the Work (RDA 6.27.1)

Title Information

Instance Title

[Mint URI](#) [Add](#)

▣ Title Proper (RDA 2.3.2) (BIBFRAME: Main title)

Title Proper (RDA 2.3.2)
(BIBFRAME: Main title)

- ✚ Other Title Information (RDA 2.3.4) (BIBFRAME: Subtitle)
- ✚ Part number
- ✚ Part name
- ✚ Note on title

Discogs

▣ Genre Form (BIBFRAME: Genre Form)

Genre Form

(BIBFRAME: Genre Form)

▣ Base Material (BIBFRAME: Base Material)

Base Material

(BIBFRAME: Base Material)

▣ Carrier (BIBFRAME: Carrier)

Carrier

(BIBFRAME: Carrier)

[Browse](#)[Editor](#)[Import Resource Template](#)

Under Development - Not Exact Design

[Preview RDF](#)

Resource Template Label

[Clone|Edit] *Name of Resource*

Instance of*

BIBFRAME Work

[Mint URI](#) [Add](#)

- + Title Information
- + (Time) Coverage of the Content (RDA 7.3)
- + Notes about the Work
- Content Type (RDA 6.9)

Content Type (RDA 6.9)

- + Illustrative Content (RDA 7.15)
- + Color Content (RDA 7.17)
- + Has BIBFRAME Instance
- + Authorized Access Point Representing the Work (RDA 6.27.1)

Title Information

Instance Title

[Mint URI](#) [Add](#)

- Title Proper (RDA 2.3.2) (BIBFRAME: Main title)

Title Proper (RDA 2.3.2)
(BIBFRAME: Main title)

- + Other Title Information (RDA 2.3.4) (BIBFRAME: Subtitle)
- + Part number
- + Part name
- + Note on title

Discogs

- Genre Form (BIBFRAME: Genre Form)

Genre Form

(BIBFRAME: Genre Form)

- Base Material (BIBFRAME: Base Material)

Base Material

(BIBFRAME: Base Material)

- Carrier (BIBFRAME: Carrier)

Carrier

(BIBFRAME: Carrier)

[Browse](#)[Editor](#)[Import Resource Template](#)

Under Development - Not Exact Design

[Preview RDF](#)

Resource Template Label

[Clone|Edit] *Name of Resource*

Instance of*

BIBFRAME Work

[Mint URI](#) [Add](#)

- ✚ Title Information
- ✚ (Time) Coverage of the Content (RDA 7.3)
- ✚ Notes about the Work
- ☑ Content Type (RDA 6.9)

Content Type (RDA 6.9)

- ✚ Illustrative Content (RDA 7.15)
- ✚ Color Content (RDA 7.17)
- ✚ Has BIBFRAME Instance
- ✚ Authorized Access Point Representing the Work (RDA 6.27.1)

Title Information

Instance Title

[Mint URI](#) [Add](#)

- ☑ Title Proper (RDA 2.3.2) (BIBFRAME: Main title)

Title Proper (RDA 2.3.2)
(BIBFRAME: Main title)

- ✚ Other Title Information (RDA 2.3.4) (BIBFRAME: Subtitle)
- ✚ Part number
- ✚ Part name
- ✚ Note on title

Discogs

- ☑ Genre Form (BIBFRAME: Genre Form)

Genre Form

(BIBFRAME: Genre Form)

- ☑ Base Material (BIBFRAME: Base Material)

Base Material

(BIBFRAME: Base Material)

- ☑ Carrier (BIBFRAME: Carrier)

Carrier

(BIBFRAME: Carrier)

[Browse](#)[Editor](#)[Import Resource Template](#)

Under Development - Not Exact Design

[Preview RDF](#)

Resource Template Label

[Clone|Edit] *Name of Resource*

Instance of*

BIBFRAME Work

Mint URI

Add

- ✚ Title Information
- ✚ (Time) Coverage of the Content (RDA 7.3)
- ✚ Notes about the Work
- ☑ Content Type (RDA 6.9)

Content Type (RDA 6.9)

Lookup

- ✚ Illustrative Content (RDA 7.15)
- ✚ Color Content (RDA 7.17)
- ✚ Has BIBFRAME Instance
- ✚ Authorized Access Point Representing the Work (RDA 6.27.1)

Title Information

Instance Title

Mint URI

Add

- ☑ Title Proper (RDA 2.3.2) (BIBFRAME: Main title)

Title Proper (RDA 2.3.2)
(BIBFRAME: Main title)

Sorry / Why Remind Me

- ✚ Other Title Information (RDA 2.3.4) (BIBFRAME: Subtitle)
- ✚ Part number
- ✚ Part name
- ✚ Note on title

Discogs

- ☑ Genre Form (BIBFRAME: Genre Form)

Genre Form

(BIBFRAME: Genre Form)

Jazz

- ☑ Base Material (BIBFRAME: Base Material)

Base Material

(BIBFRAME: Base Material)

Vinyl

- ☑ Carrier (BIBFRAME: Carrier)

Carrier

(BIBFRAME: Carrier)

audio disc

[Browse](#)[Editor](#)[Import Resource Template](#)

Under Development - Not Exact Design

[Preview RDF](#)

Resource Template Label

[Clone|Edit] *Name of Resource*

Instance of*

BIBFRAME Work

Mint URI

Add

- ✚ Title Information
- ✚ (Time) Coverage of the Content (RDA 7.3)
- ✚ Notes about the Work
- ☑ Content Type (RDA 6.9)

Content Type (RDA 6.9)

Lookup

- ✚ Illustrative Content (RDA 7.15)
- ✚ Color Content (RDA 7.17)
- ✚ Has BIBFRAME Instance
- ✚ Authorized Access Point Representing the Work (RDA 6.27.1)

Title Information

Instance Title

Mint URI

Add

- ☑ Title Proper (RDA 2.3.2) (BIBFRAME: Main title)

Title Proper (RDA 2.3.2)
(BIBFRAME: Main title)

Sorry / Why Remind Me

- ✚ Other Title Information (RDA 2.3.4) (BIBFRAME: Subtitle)
- ✚ Part number
- ✚ Part name
- ✚ Note on title

Discogs

- ☑ Genre Form (BIBFRAME: Genre Form)

Genre Form

(BIBFRAME: Genre Form)

Jazz

- ☑ Base Material (BIBFRAME: Base Material)

Base Material

(BIBFRAME: Base Material)

Vinyl

- ☑ Carrier (BIBFRAME: Carrier)

Carrier

(BIBFRAME: Carrier)

audio disc

Lots of Linked Data Authorities

Direct Access

- AGROVOC
- DBPEDIA_DIRECT *(term fetch only)*
- GEONAMES_DIRECT
- LOC_DIRECT *(term fetch only)*
- NALT_DIRECT *(term fetch only)*
- OCLCFAST_DIRECT

Exploring

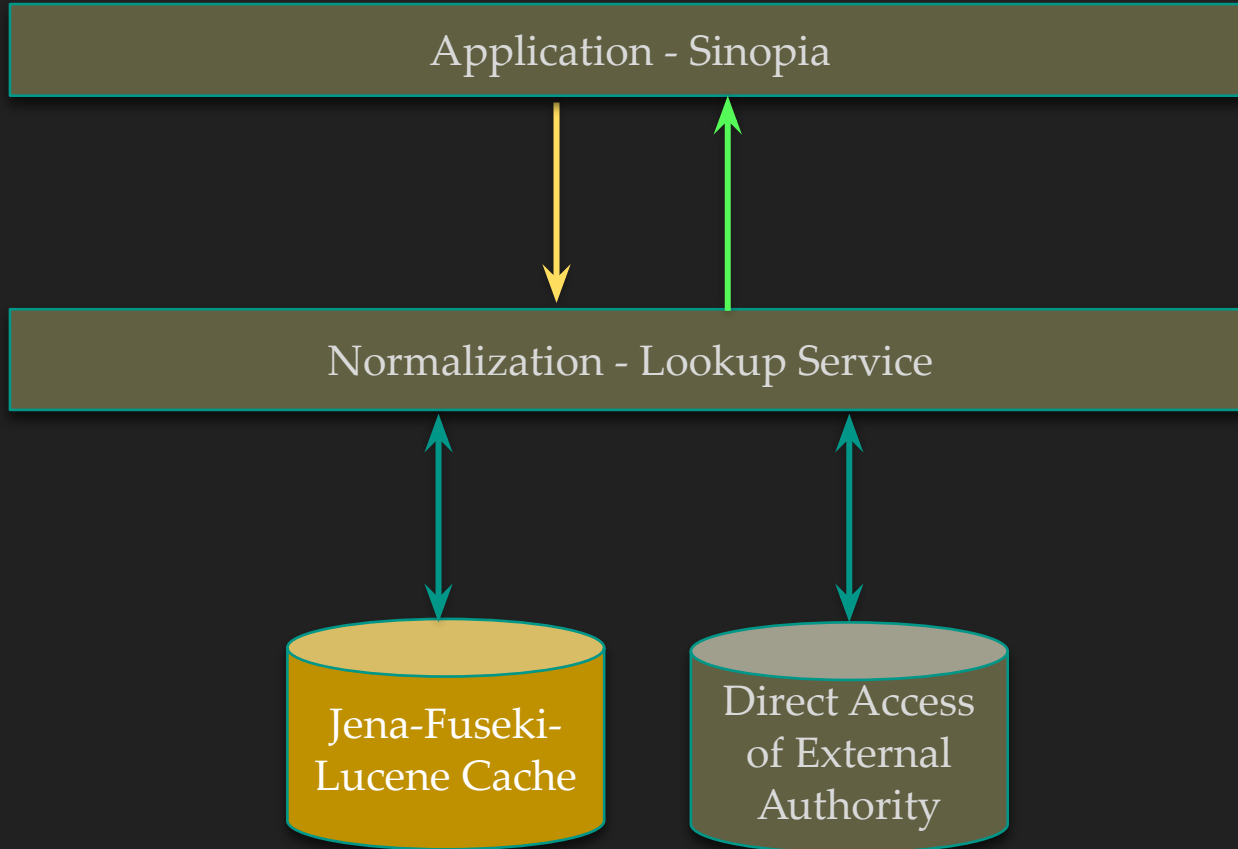
- Wikidata *(APIs are not linked data)*
- ISNI *(not available as linked data... yet)*

LD4P Cache

- AGROVOC_LD4L_CACHE
- DBPEDIA_LD4L_CACHE
- GEONAMES_LD4L_CACHE
- Getty
 - GETTY_AAT_LD4L_CACHE
 - GETTY_TGN_LD4L_CACHE
 - GETTY_ULAN_LD4L_CACHE
- Library of Congress
 - LOCDEMOGRAPHICS_LD4L_CACHE
 - LOGENRES_LD4L_CACHE
 - LOCNAMES_LD4L_CACHE
 - LOCPERFORMANCE_LD4L_CACHE
 - LOCSUBJECTS_LD4L_CACHE
- MESH_LD4L_CACHE
- NALT_LD4L_CACHE
- OCLCFAST_LD4L_CACHE
- RDA Registry
- Share-VDE data for 20 institutions

A few more authorities available through QA that access authority APIs that are not supporting linked data.

Architecture



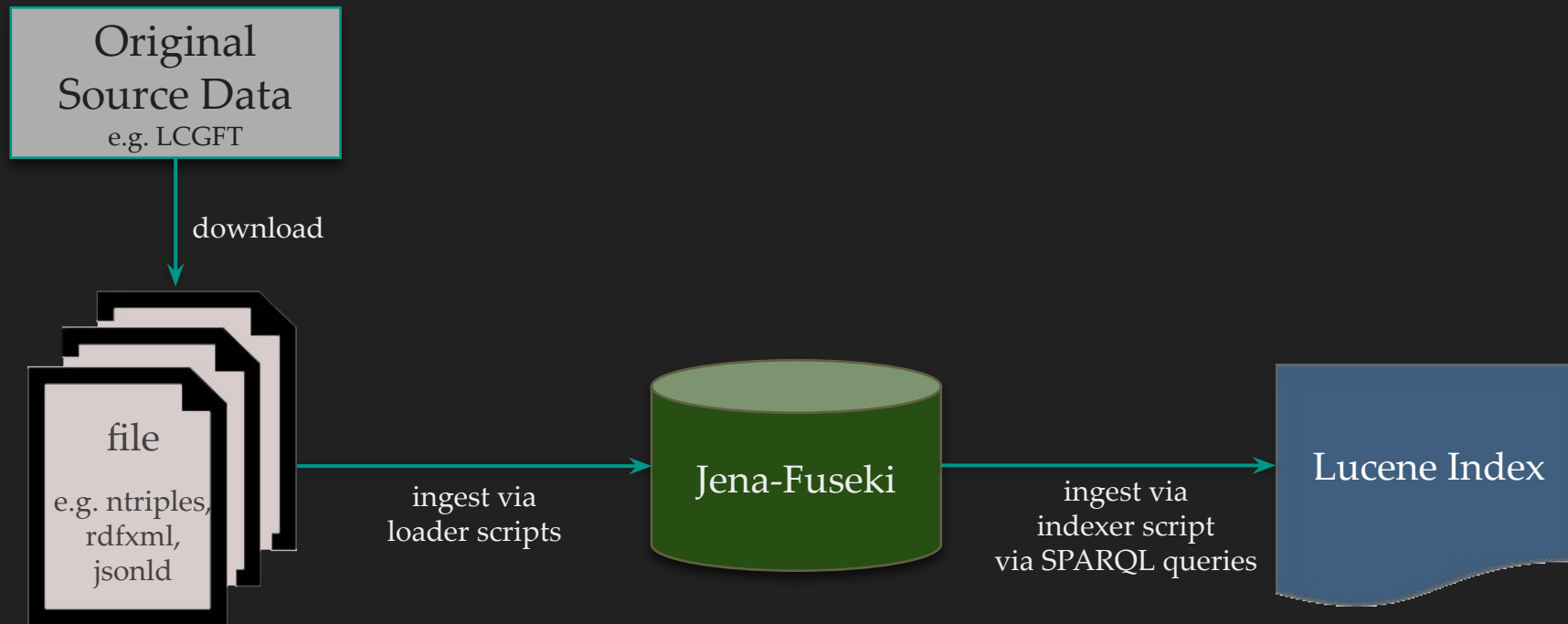
Why use a cache instead of direct access?

- Limited access by original data provider
 - Authority only provides a download of RDF
 - Authority supports reconcilable URI for term fetch, but no search API
- Search results do not include rank predicate
- Provide access to extended content in search results
- Control over indexing approach to improve accuracy of searches
- Control over when data changes

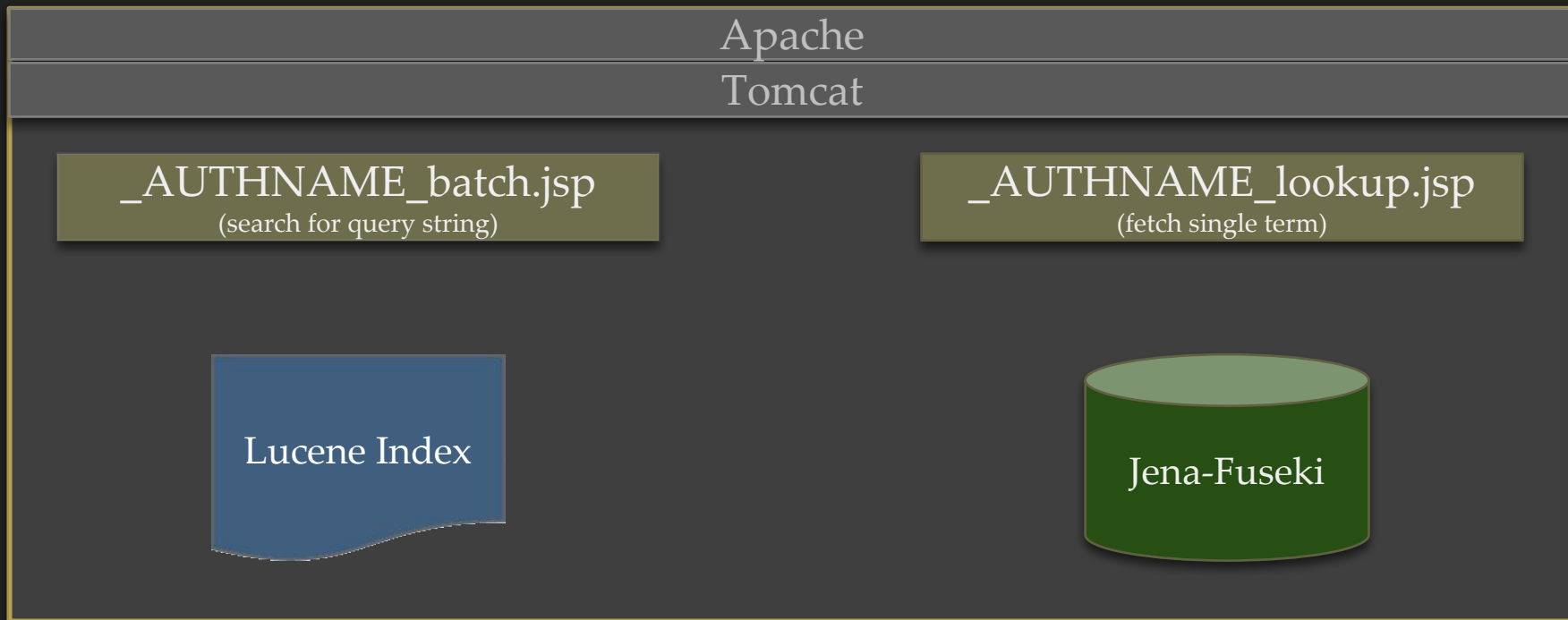
Challenges of caching data

- Equipment required to store the data locally
- Expertise in triplestores and scalability issues
- Synchronizing data updates
- Difficulty in providing real time updates
 - Original source doesn't provide access to data in real time

Cache Loading Process



Current Services Stack



Cache API Access – Fetch One

Normalization - Lookup Service

CURL request to `cache lookup.jsp`

RDF serialization

Apache
Tomcat

`_AUTHNAME_batch.jsp`
(search for query string)

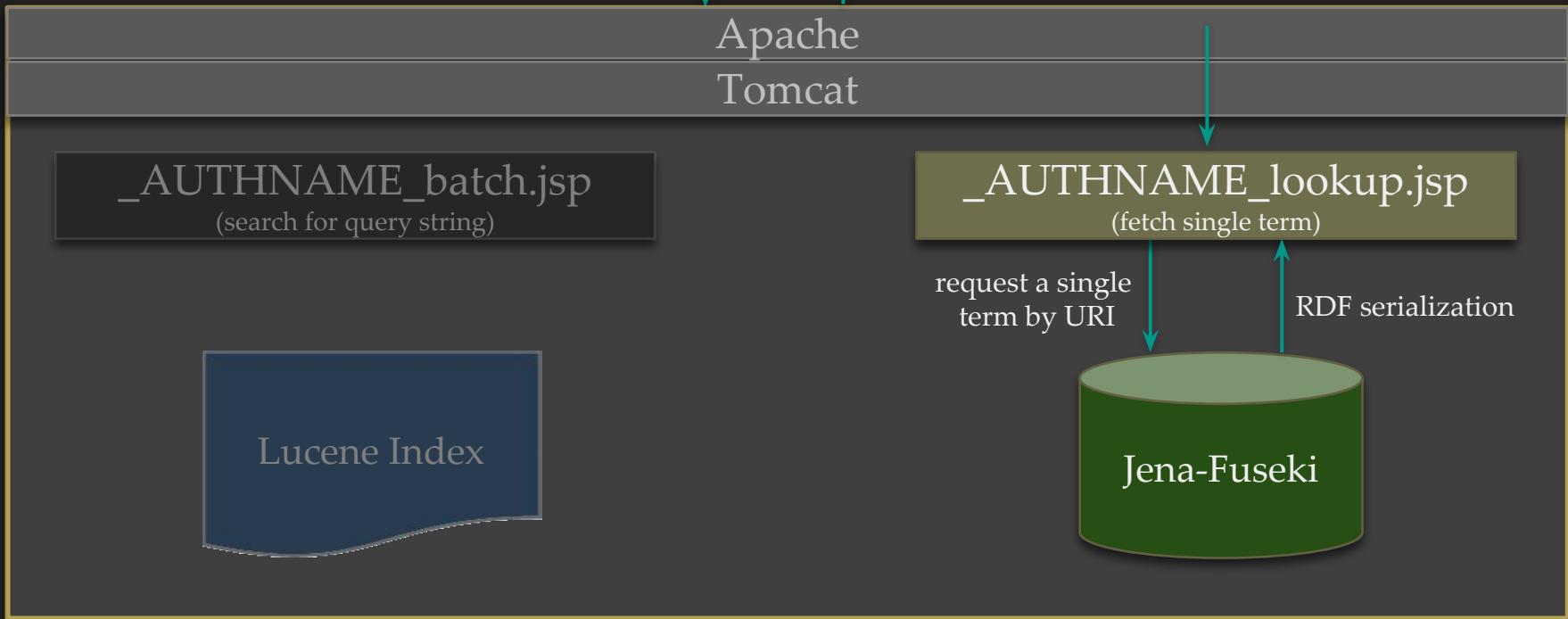
`_AUTHNAME_lookup.jsp`
(fetch single term)

request a single
term by URI

RDF serialization

Lucene Index

Jena-Fuseki



Cache API Access – Search for multiple (part 1)

Normalization - Lookup Service

CURL request to `cache batch.jsp`

Apache
Tomcat

`_AUTHNAME_batch.jsp`
(search for query string)

perform lucene
search

Lucene Index

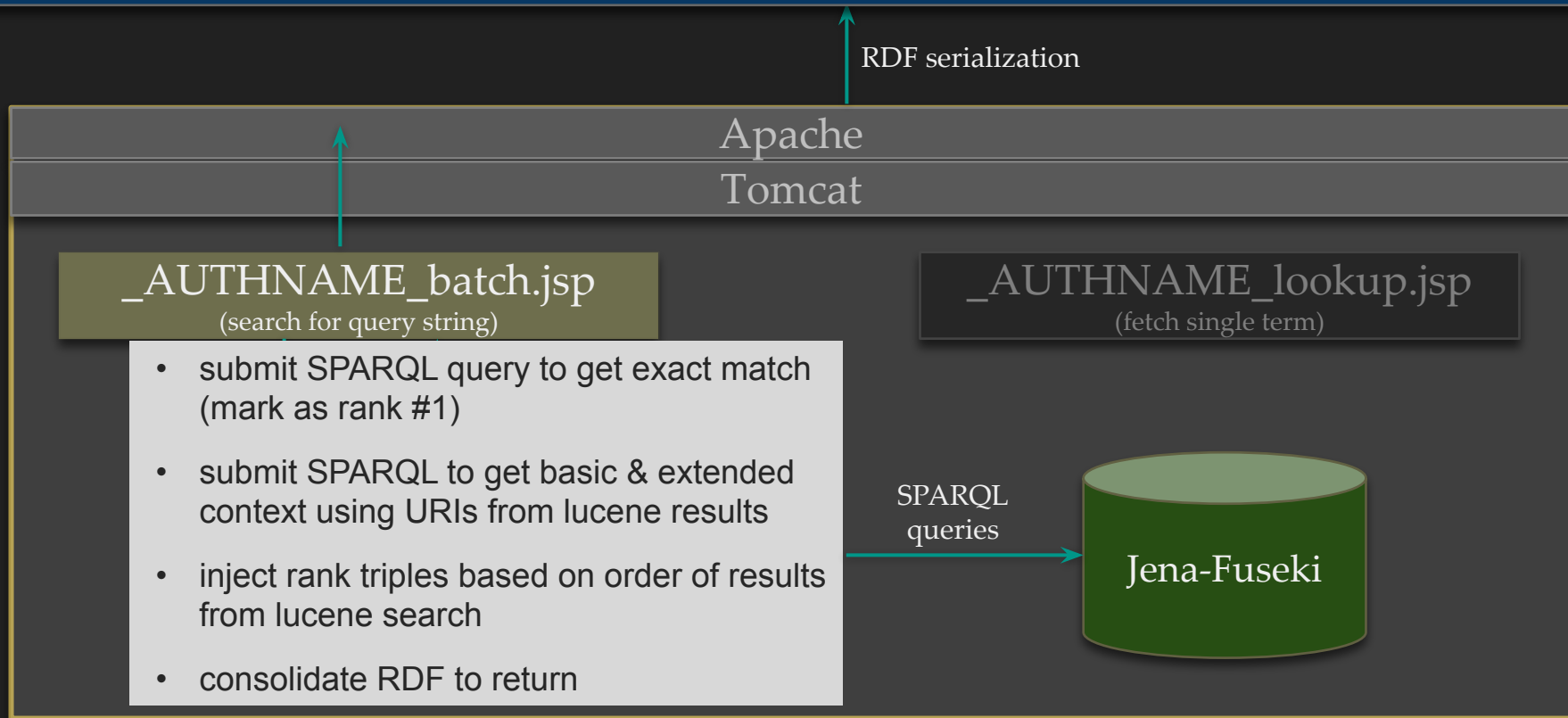
matching
results

`_AUTHNAME_lookup.jsp`
(fetch single term)

Jena-Fuseki

Cache API Access – Search for multiple (part 2)

Normalization - Lookup Service



A Simple Lookup Request

```
<sparql:query var="result" endpoint="{id4}" resultType="triple">
  SELECT DISTINCT ?p ?o WHERE {
    ?s ?p ?o
    <c:if test="{not empty param.lang}">
      FILTER(!isLiteral(?o) || lang(?o) = "" || langMatches(lang(?o), "{param.lang}"))
    </c:if>
  }
  <sparql:parameter var="s" value="{param.uri}" type="IRI" />
</sparql:query>

<c:forEach items="{result.rows}" var="row" varStatus="rowCounter">
  <${param.uri}> ${row.p} ${row.o} .
</c:forEach>
```

A Typical Complex Lookup Request

```
<sparql:construct var="graph" endpoint="{d4}">
  CONSTRUCT { ?s ?p ?o . ?o ?q ?r . ?t gn:name ?u . ?v gn:name ?w . ?x gn:name ?y } WHERE {
    ?s ?p ?o .
    OPTIONAL {
      ?o ?q ?r
      FILTER (isBlank(?o))
      <c:if test="{not empty param.lang}">
        FILTER(!isLiteral(?r) || lang(?r) = "" || langMatches(lang(?r), "{param.lang}"))
      </c:if>
    }
    OPTIONAL {
      ?s gn:parentCountry ?t .
      ?t gn:name ?u
    }
    OPTIONAL {
      ?s gn:parentADM1 ?v .
      ?v gn:name ?w
    }
    OPTIONAL {
      ?s gn:parentADM2 ?x .
      ?x gn:name ?y
    }
  }
```

A Typical Complex Lookup Request, Part 2

```
<sparql:query var="result" graph="{graph}" resultType="triple">  
  SELECT ?s ?p ?o WHERE {  
    ?s ?p ?o .  
  } ORDER BY ?s ?p  
</sparql:query>
```

```
<c:forEach items="{result.rows}" var="row" varStatus="rowCounter">  
  {row.s} {row.p} {row.o} .  
</c:forEach>
```


Tuning “Knobs”

- At the Lucene index level (i.e. matching entities and the order in which they are returned)
 - Tokenization of input query
 - Relative weighting of terms
 - Semantics of query (disjunction vs. conjunction, etc.)
- At the Fuseki endpoint level (i.e., the shape of the response graph for each matching URI)
 - SPARQL query (or queries) can be extended as needed for complex contextual requirements

A User Interface for the Lookup Service

Check Status

- Check Connection Status only
 Check Accuracy only
 Run all checks

Connection Checks

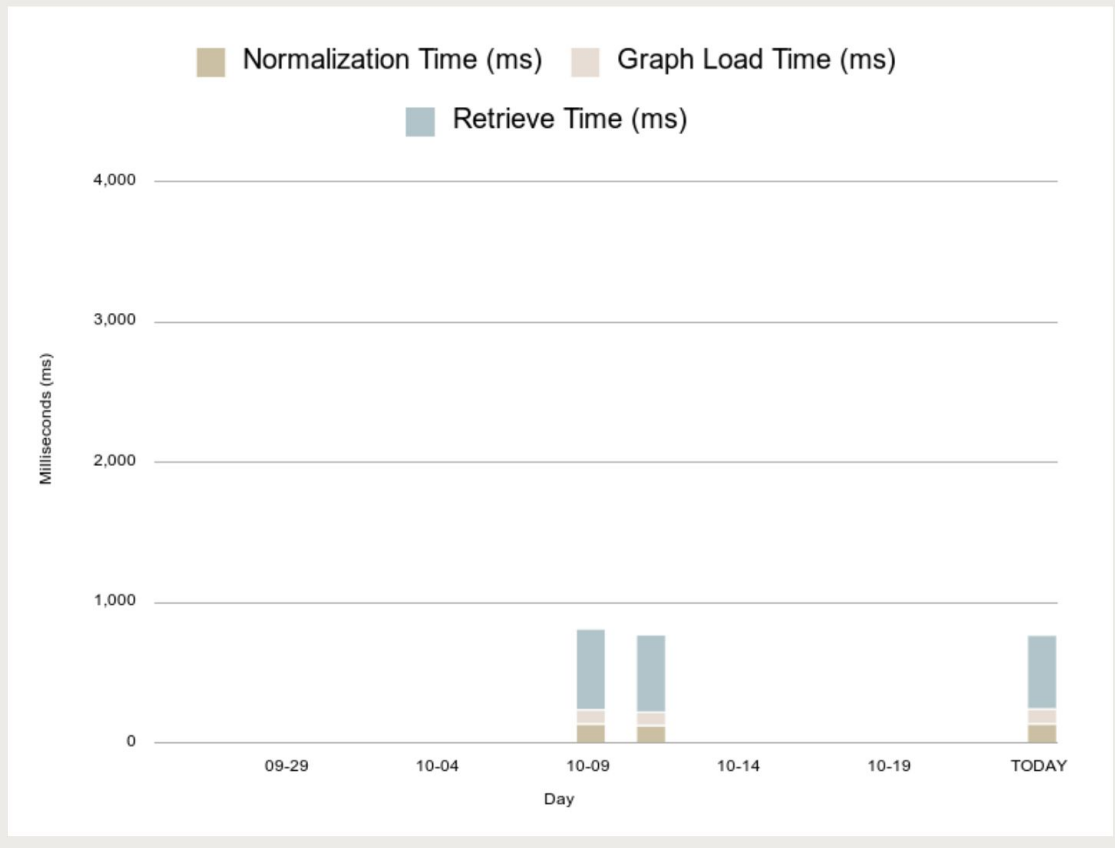
Status	Subauthority	Service	Action	Sample URL	Errors
GEONAMES_DIRECT					
√		direct	term	/authorities/fetch/linked_data/geonames_direct?uri=http%3A%2F%2Fsws%2Egeonames%2Eorg%2F261707%2F	
√		direct	search	/authorities/search/linked_data/geonames_direct?q=ithaca&maxRecords=4	

GETTY_AAT_LD4L_CACHE

[Day](#) **Month** [Year](#)

All Requests [Search](#) [Fetch](#)

Performance data for the last 30 days.



Performance Graph

monthly data

all requests

Performance Graph

day data

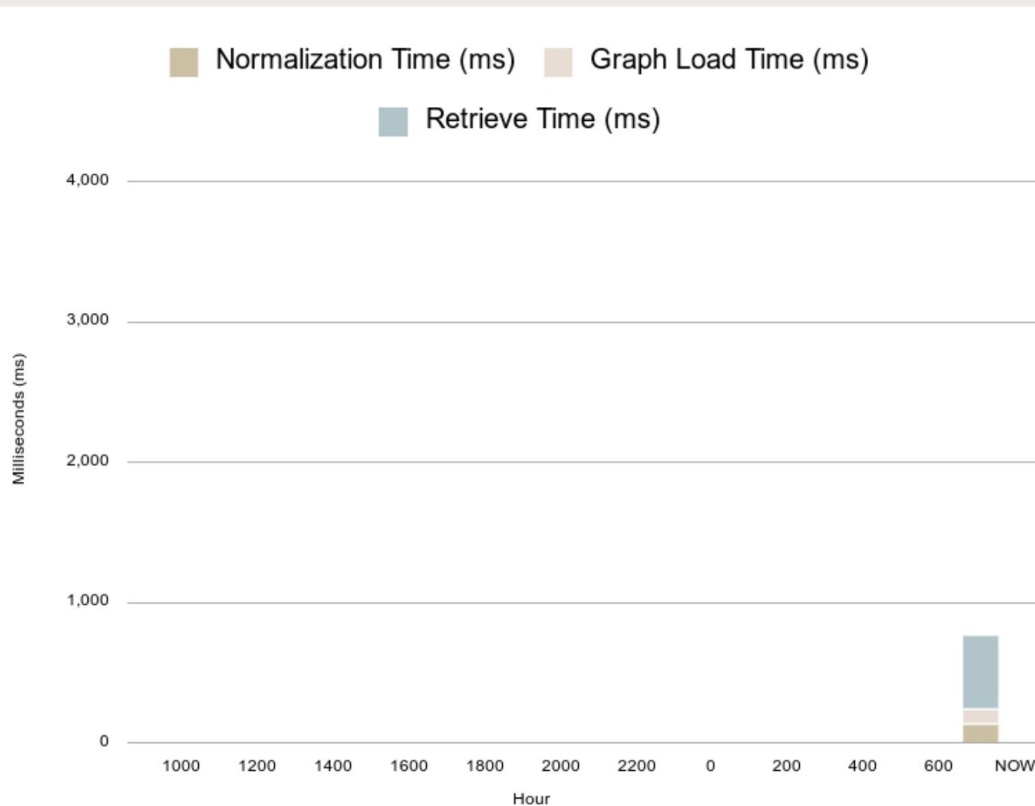
all requests

GETTY_AAT_LD4L_CACHE

Day Month Year

All Requests Search Fetch

Performance data for the last 24 hours.



Performance Graph

year data

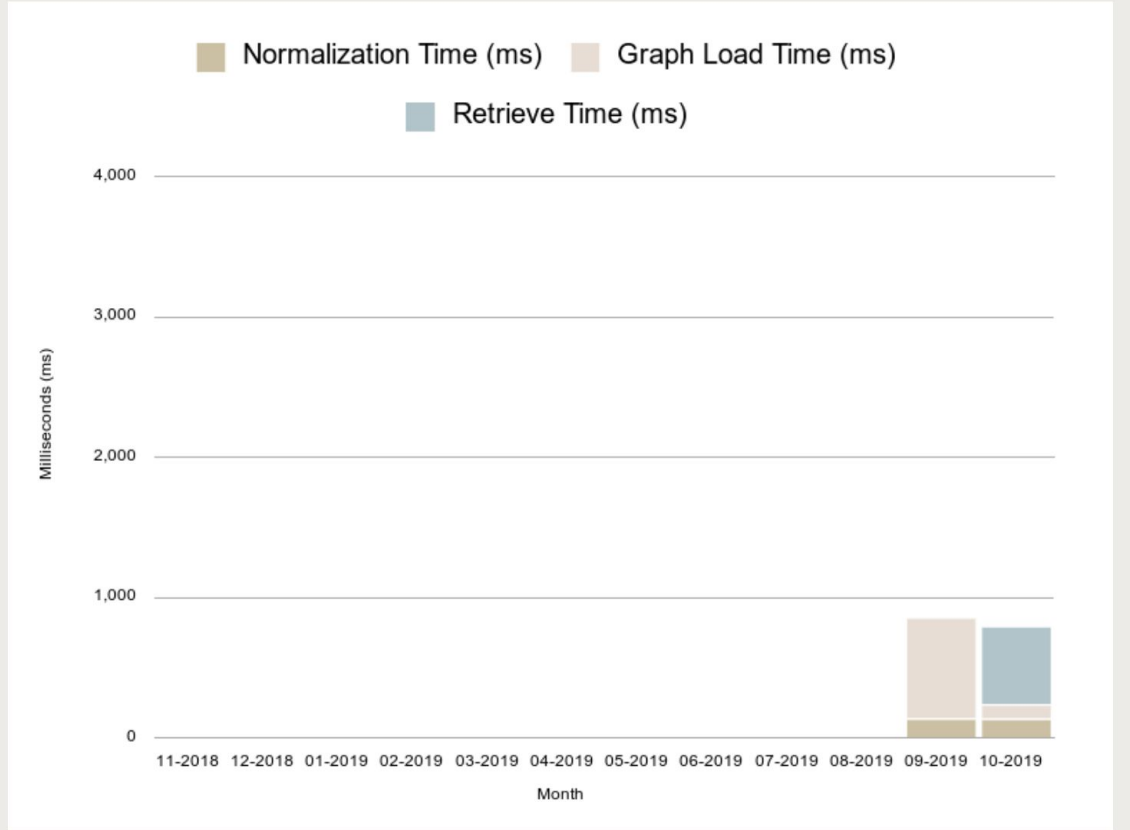
all requests

GETTY_AAT_LD4L_CACHE

Day Month Year

All Requests Search Fetch

Performance data for the last 12 months.

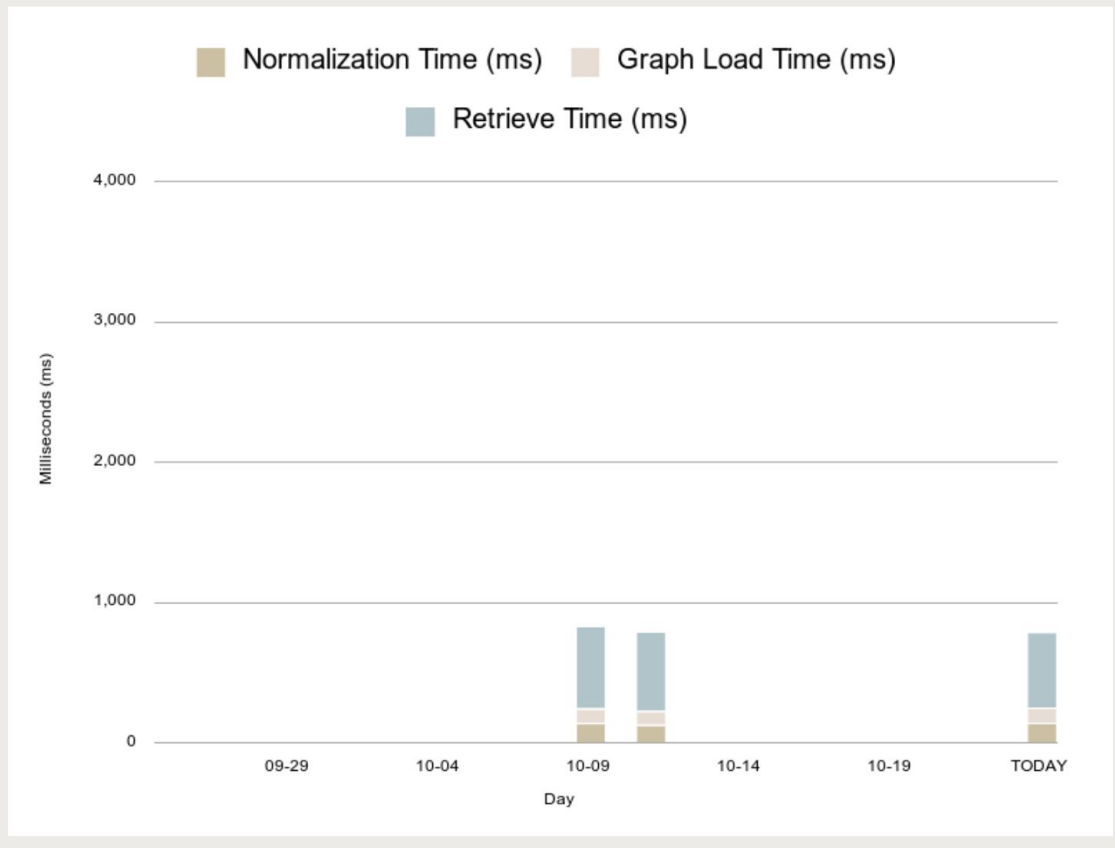


GETTY_AAT_LD4L_CACHE

[Day](#) **Month** [Year](#)

[All Requests](#) **Search** [Fetch](#)

Performance data for the last 30 days.



Performance Graph

monthly data

search requests

Performance Graph

monthly data

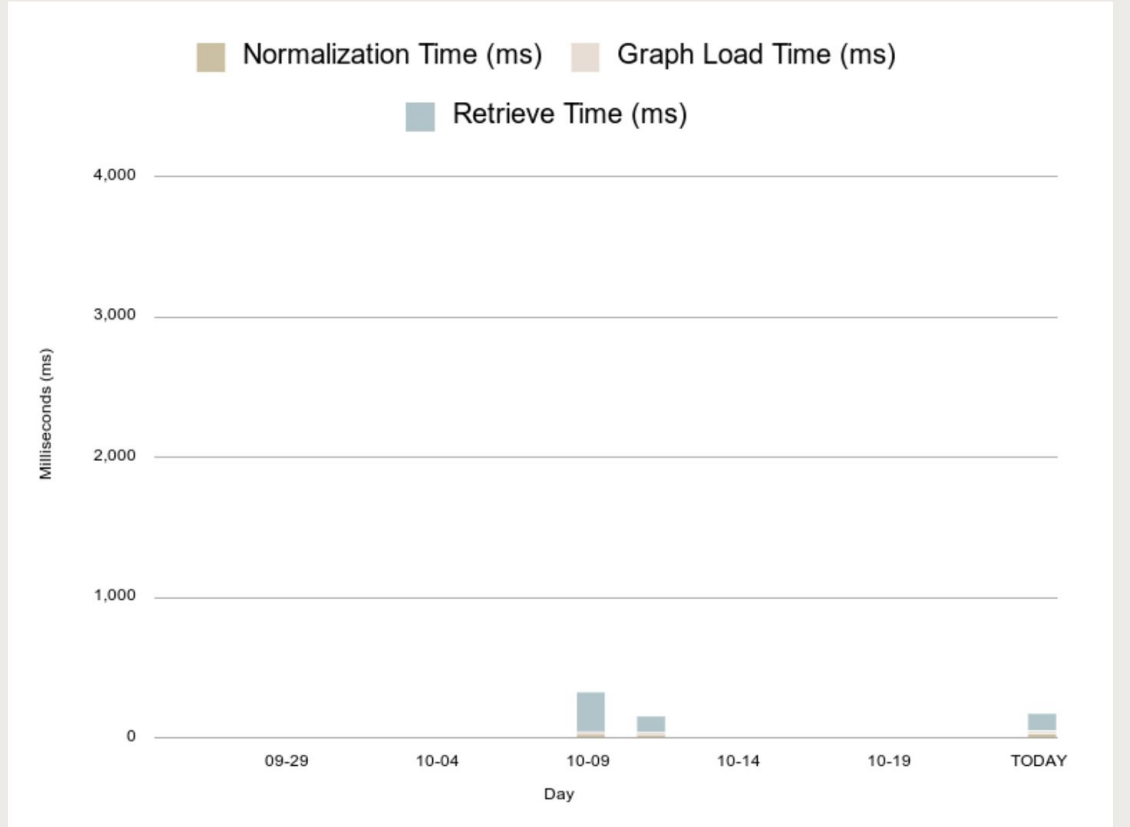
fetch requests

GETTY_AAT_LD4L_CACHE

Day **Month** Year

All Requests Search **Fetch**

Performance data for the last 30 days.

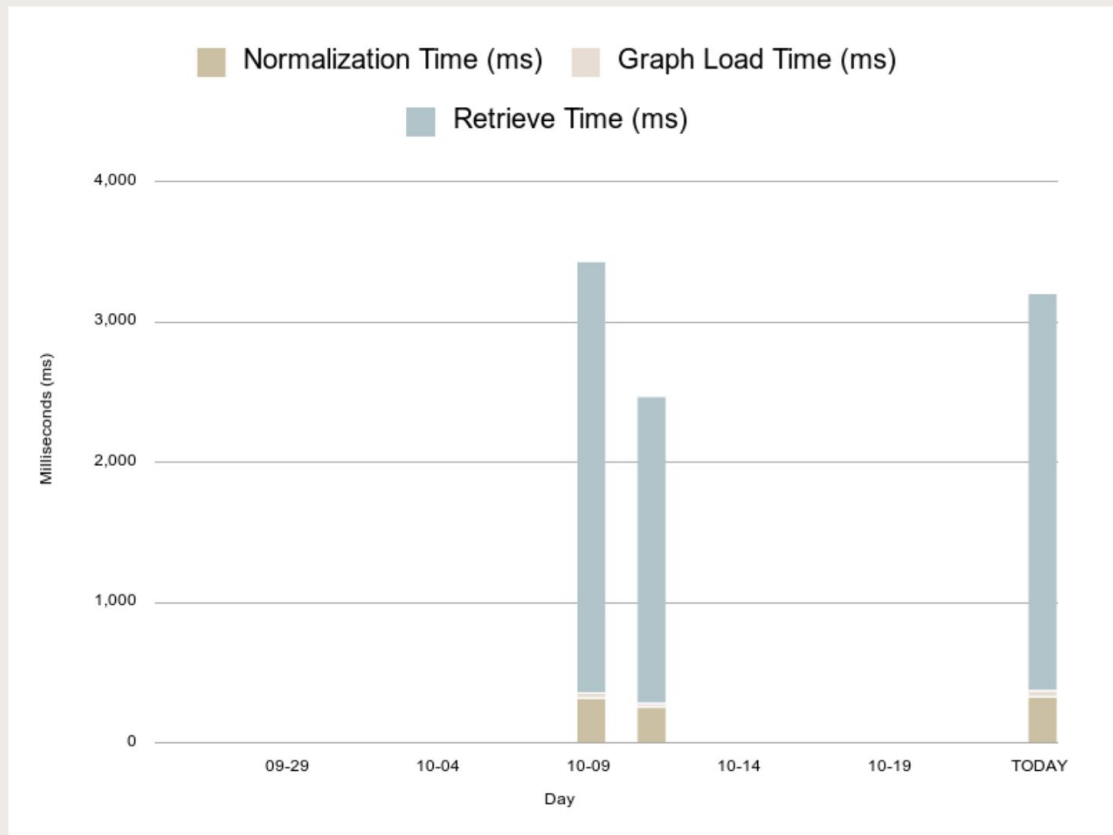


SHAREVDE_CORNELL_LD4L_CACHE

Day Month Year

All Requests Search Fetch

Performance data for the last 30 days.



Performance Graph

Identifying poorly performing authorities

Real Code! (Just add data...)

To set up a normalization lookup service...

- qa_server - https://github.com/ld4p/qa_server (start here)
- QA - https://github.com/samvera/questioning_authority ([documentation](#))
- linked data authorities - https://github.com/LD4P/linked_data_authorities

To set up a cache system...

- ld4l_services web app - https://github.com/eichmann/ld4l_services
- Lucene Tag Library - <https://github.com/eichmann/LuceneTagLib>
- SPARQL Tag Library - <https://github.com/eichmann/SPARQLTagLib>

<https://tinyurl.com/y2ubatf4>

Try out our server: <https://lookup.ld4l.org>

Exploration only. Not for production use beyond LD4P project.