

30 Nov 2016

SWIB16: Bonn, Germany

Person Entities: Lessons learned by a data provider

John Chapman

Senior Product Manager, Metadata Services

Our focus for today...

- Why we did the pilot project
- How we built and provided entity data
- What did we learn?
- What should we do next?

Person Entity Lookup Pilot

Primary goal: improve access to entities via “API First” services

Small group, short timeframe, shut-off date

- Two Phases:
 - Phase 1: “Same As” identifier lookup
 - Phase 2: String matching for person names

Phase 1: “Same As” Service

- Based on VIAF matching algorithms
- A RESTful API
- Client requests include a known identifier
- For a match, a Person Entity URI and all other IDs returned

Phase 1: “Same As” Service

Lookup Identifier



Related Identifiers

<http://viaf.org/viaf/96994048>

http://dbpedia.org/resource/William_Shakespeare

<http://d-nb.info/gnd/118613723>

<http://vocab.getty.edu/ulan/500272240-agent>

<http://data.bnf.fr/ark:/12148/cb119246079#foaf:Person>

<http://alpha.bn.org.pl/record=a11579006>

<http://id.ndl.go.jp/auth/entity/00456207>

<http://libris.kb.se/resource/auth/198702>

<http://worldcat.org/entity/person/id/2643040000>

<http://id.loc.gov/authorities/names/n78095332>

<http://viaf.org/viaf/96994048>

<http://www.idref.fr/027136086/id>

<http://id.worldcat.org/fast/29048>

<http://www.wikidata.org/entity/Q692>

Phase 2: Search Service

- Text-based search
- Additional data supplied:
 - Preferred name
 - Other name forms (with language tags)
 - + Roles
 - + Topics
 - + Score

Roles, Topics, and Score were derived from WorldCat bibliographic data and the WorldCat Identities aggregation

[http://\[server\]/?q=Zadie&20Smith&wskey=\[YOUR_OCLC_SYMBOL\]](http://[server]/?q=Zadie&20Smith&wskey=[YOUR_OCLC_SYMBOL])

```
{
  {
    "uri": "http://worldcat.org/entity/person/id/2642331361",
    "defaultLabel": "Zadie Smith",
    "birthDate": "1975-10-25",
    "role": "Author",
    "topic": "College teachers",
    "score": "9222.581",
    "languageLabels": {"it-IT": "Zadie Smith", "ca-ES": "Zadie Smith", "no-NO": "Zadie
Smith", "pl-PL": "Zadie Smith", "ja-JP": "Zadie Smith", "es-ES": "Zadie Smith", "ar <snip>},
    "alternateNames": ["מית, זייד", "Смит, Зэди", "Zadi Smit", "Zadie SMITH", "זייד
מית", "Зеді Сміт", "जेडी ममिष", "زادی اسمیت", "Zadie Smith", "Зейди Смит", "查蒂·史密斯
", "سمیث، زادی", "ゼイデー・スミス", "Zadie Smithová"]  }
```

UI prototype

Francisco Goya

<http://worldcat.org/entity/person/id/2636494134>

Role: Author

Related Topic: Artists

Born: Mar 30, 1746

Died: Apr 16, 1828

Spanish painter and printmaker (1746–1828) [\[WikiData\]](#)

Francisco José de Goya y Lucientes (/ˈɡoiə/; Spanish: [franˈθisko xoˈse ðe ˈɣo.ja i luˈθjentes]; 30 March 1746 – 16 April 1828) was a Spanish romantic painter and printmaker regarded both as the last of the Old Masters and the first of the moderns. Goya was court painter to the Spanish Crown; throughout the Peninsular War he remained in Madrid, where he painted the portrait of Joseph Bonaparte, pretender to the Spanish throne, and documented the war in the masterpiece of studied ambiguity known as the Desastres de la Guerra. Through his works he was both a commentator on and chronicler of his era. The subversive imaginative element in his art, as well as his bold handling of paint, provided a model for the work of artists of later generations, notably Édouard Manet, Pablo Picasso and Francis Bacon. [\[DBPedia\]](#)



[\[Image Source\]](#)

Alternate Names

Франсиско Гойя

Փրանցիսկո Գոյա

Francisco José de Goya y
Lucientes

Franciscus a Goya et Lucientes

Same As Relationships

<http://d-nb.info/gnd/118541064>

<http://id.worldcat.org/fast/29789>

<http://www.idref.fr/027323765/id>

<http://data.bnf.fr/ark:/12148/cb11>

Family Relationships

Father:

José Benito de Goya y Franque

Mother:

Gracia de Lucientes y Salvador

Spouse:

Lessons learned

The Data Aggregator's View:

- Many sources available
 - **No single source is good at everything**
 - Quality varies by element type
- Data Aggregation is crucial
 - **Context at scale**
 - Weighting and scoring are crucial

Lessons learned

The Service Consumer's View:

- Workflow support should be worked into design
- Context is key for names
- Language support is important but labor-intensive and inexact
- Unsolved problem around sparse clusters

Lessons learned

The Combined View:

- Supporting workflows efficiently means rethinking ID creation
- Automation only gets us so far
- Need systems for enhancement – multiple levels to this
- Next steps will require us all

Where do we go from here?

- Continue starting (and ending) pilots and experiments
- Move from projects to production
- Commit to sustainable, persistent systems
- Consider positive and negative incentives
- Surface local expertise to build context

Working together

- More data allows for richer context
- A single aggregation will never be complete and comprehensive
- Focused experimentation is needed

- Let's continue to work together –VIAF, ISNI, WorldCat

Questions?

John Chapman

Senior Product Manager, Metadata Services

chapmanj@oclc.org

Special thanks to my colleagues:

Jeff Mixer

Stephan Schindehette

Bruce Washburn

**Because
what is
known must
be shared.SM**