Improving access to Culture Heritage:
data.cervantesvirtual.com

Gustavo Candela Romero

gustavo.candela@cervantesvirtual.com
Index

1. Introduction
2. Keys to success
3. Wikidata and BVMC
4. Future work
5. References

data.cervantesvirtual.com
1. Introduction

- In February 2015, the first release of the linked open data website based on RDA and FRBR data.cervantesvirtual.com was launched. The project intends to promote data sharing, interoperability, data re-use and dissemination of best practices.

- Starting from scratch, the project is no longer a demo, but a rich source of lessons learnt in order to stimulate innovative types of projects.
1. Introduction

Step by step

Marc21 → FRBR (Relational Database) → RDA (RDF Repository) → ?

Stanford Prize for Innovation in Research Libraries

TPDL 2015
IODC 2016
Semantic Web Journal

SPIRL

RDA
RESOURCE DESCRIPTION & ACCESS
1. Introduction

data.cervantesvirtual.com in 2017
1. Introduction

Why we did it?

- Open our traditional catalog to the world for human and computers.
- Provide a public interface for querying the dataset according to international recommendations (SPARQL).
- Establish relationships and links with broadly used data sets such as VIAF and Wikidata.
- Improve the catalog and promote reuse.
2. Keys to success

Preprocessing of sources

- Since some fields are required (for example, field 245 containing the title) while some others are optional or user-defined, the homogeneity of the data across libraries cannot be guaranteed. Furthermore, the content of a field can be expressed with different conventions, in different languages, or it may contain typos.
- These features represent a challenge when MARC21 records must be shared between libraries.
2. Keys to success

Preprocessing of sources

● However, further refinements are needed for the recognition and extraction of implicit relationships expressed in natural language, such as geographic locations and dates.

- En Sevilla, : en la imprenta de Joseph Padrino ..., [entre 1748 y 1775]
- Sevilla, : por Thomas Lopez de Haro ..., , 1679
- [Sevilla : s.n., 1760]
- Impreso en Sevilla : por Juan Francisco de Blas..., 1693
- Hispali :, Antonius Martinez, Alfonso de Portu et Bartholomaeus Segura, 1477
2. Keys to success

Reuse of vocabularies (RDA and FRBR)
www.rdaregistry.info/
# 2. Keys to success

## Identify access points

<table>
<thead>
<tr>
<th>Entity</th>
<th>URI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person</td>
<td><a href="http://data.cervantesvirtual.com/person/%7Bid%7D">http://data.cervantesvirtual.com/person/{id}</a></td>
</tr>
<tr>
<td>Corporate Body</td>
<td><a href="http://data.cervantesvirtual.com/corporatebody/%7Bid%7D">http://data.cervantesvirtual.com/corporatebody/{id}</a></td>
</tr>
<tr>
<td>Family</td>
<td><a href="http://data.cervantesvirtual.com/family/%7Bid%7D">http://data.cervantesvirtual.com/family/{id}</a></td>
</tr>
<tr>
<td>Work</td>
<td><a href="http://data.cervantesvirtual.com/work/%7Bid%7D">http://data.cervantesvirtual.com/work/{id}</a></td>
</tr>
<tr>
<td>Expression</td>
<td><a href="http://data.cervantesvirtual.com/expression/%7Bid%7D">http://data.cervantesvirtual.com/expression/{id}</a></td>
</tr>
<tr>
<td>Manifestation</td>
<td><a href="http://data.cervantesvirtual.com/manifestation/%7Bid%7D">http://data.cervantesvirtual.com/manifestation/{id}</a></td>
</tr>
<tr>
<td>Country</td>
<td><a href="http://data.cervantesvirtual.com/country/%7Bid%7D">http://data.cervantesvirtual.com/country/{id}</a></td>
</tr>
<tr>
<td>Date</td>
<td><a href="http://data.cervantesvirtual.com/date/%7Bid%7D">http://data.cervantesvirtual.com/date/{id}</a></td>
</tr>
<tr>
<td>Language</td>
<td><a href="http://data.cervantesvirtual.com/language/%7Bid%7D">http://data.cervantesvirtual.com/language/{id}</a></td>
</tr>
</tbody>
</table>
2. Keys to success

Metadata enrichment

- Wikidata
- DBpedia
- Wikipedia
- VIAF
- GeoNames
- ISNI
- Latin Place Names
- RBMS/BSC Latin Place Names File
- BNF (Bibliothèque nationale de France)
- BNE (Biblioteca Nacional de España)
- res (Research and Education Space)
2. Keys to success

Increase visibility and promote interoperability

- Social Media (Facebook and Twitter)
- Conferences
- Sharing data
- SEO techniques
- Technology blog
- Github profile
- Encouraging students at the university
3. Wikidata and BVMC

Wikidata properties

- [https://www.wikidata.org/wiki/Property:P2799](https://www.wikidata.org/wiki/Property:P2799) BVMC Person id (~7000 links)
- [https://www.wikidata.org/wiki/Property:P3976](https://www.wikidata.org/wiki/Property:P3976) BVMC Work id (~150 links)

Some examples of possible additional properties:

- BVMC Journal id
- BVMC Location id
- BVMC Date id
- BVMC Manuscript id
3. Wikidata and BVMC

Reusing Wikidata

- Wordpress plugin for querying wikidata authors of literary movements.

```php
$result = $sparql->query(
    'SELECT * WHERE {
    OPTIONAL {?writer wdt:P18 ?image } .
    OPTIONAL {?writer wdt:P569 ?_date_of_birth .
        BIND( year(?_date_of_birth) as ?birthYear )} .
    OPTIONAL {?writer wdt:P570 ?_date_of_death.
        BIND( year(?_date_of_death) as ?deathYear )} .
    FILTER ( lang(?label) = "es" )} .
} ORDER BY ?label '.
'LIMIT '. $numresults
```
3. Wikidata and BVMC

Reusing Wikidata

- Elasticsearch application example for BVMC linked authors.
- [https://github.com/fbvmc/elasticsearch-tutorial](https://github.com/fbvmc/elasticsearch-tutorial)
4. Future work

Keeping on exploring and innovating

BVMC Repository

Crowdsourcing

Natural language

Metadata Enrichment

Exploitation of digital content

I Still Haven't Found What I'm Looking For...
4. References

- [http://data.cervantesvirtual.com](http://data.cervantesvirtual.com)
- [http://data.cervantesvirtual.com/geosearch](http://data.cervantesvirtual.com/geosearch)
- [galatea.cervantesvirtual.com](http://galatea.cervantesvirtual.com)
- [SPARQL endpoint](http://galatea.cervantesvirtual.com)
- Migration of a library catalogue into RDA linked open data, Semantic Web Journal, 2017 [online](http://galatea.cervantesvirtual.com)
- Transformation of a Library Catalogue into RDA Linked Open Data. [TPDL 2015](http://galatea.cervantesvirtual.com)
- Semantic Enrichment on Cultural Heritage collections: A case study using geographic information. doi>[10.1145/3078081.3078090](https://doi.org/10.1145/3078081.3078090)
- [http://www.rdaregistry.info/](http://www.rdaregistry.info/)
- [https://www.ifla.org/publications/functional-requirements-for-bibliographic-records](https://www.ifla.org/publications/functional-requirements-for-bibliographic-records)
Thank You