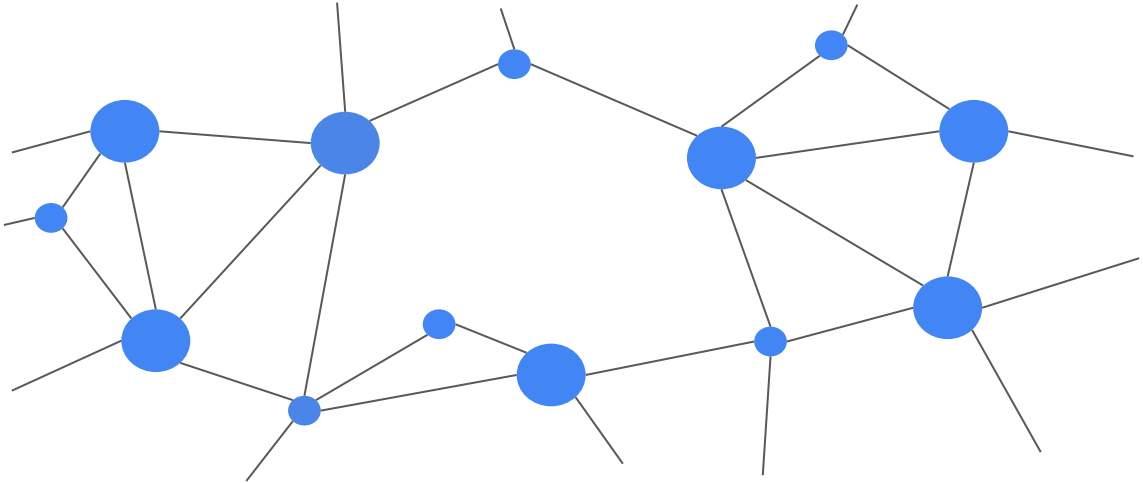
CLARIAH-CM: Workflows para la publicación de colecciones como datos

Gustavo Candela

31 de mayo de 2024



CENTRO BIBLIOTECA VIRTUAL MIGUEL DE CERVANTES

UNIVERSIDAD COMPLUTENSE MADRID

Universitat d'Alacant
Universidad de Alicante

Índice

- Sobre mi
- Introducción
- Collections as Data
- Actividad
- Desafíos y trabajo futuro
- Referencias

Índice

Gustavo Candela

- Desarrollador en la **BVMC** 2010-2023
- Profesor Ayudante Doctor
- @gus_candela
- Publicaciones
- gcandela@ua.es

https://www.cervantesvirtual.com/centro/

https://www.ua.es/

Introducción

La Biblioteca Virtual Miguel de Cervantes ha sido transformada recientemente en un centro de Humanidades Digitales.

https://www.cervantesvirtual.com/

Introducción

La BVMC forma parte de **CLARIAH-ES** y ha participado activamente en la red **INTELE**.

Introducción

El proyecto **BVMC Labs (data.cervantesvirtual.com)** tiene como objetivo la publicación y reutilización de colecciones digitales de forma creativa

- Herramientas y prototipos
- Publicaciones
- Tutoriales
- Congresos
- Ejemplos de uso

BIBLIOTECA VIRTUAL MIGUEL DE CERVANTES

http://hdl.handle.net/10045/110281

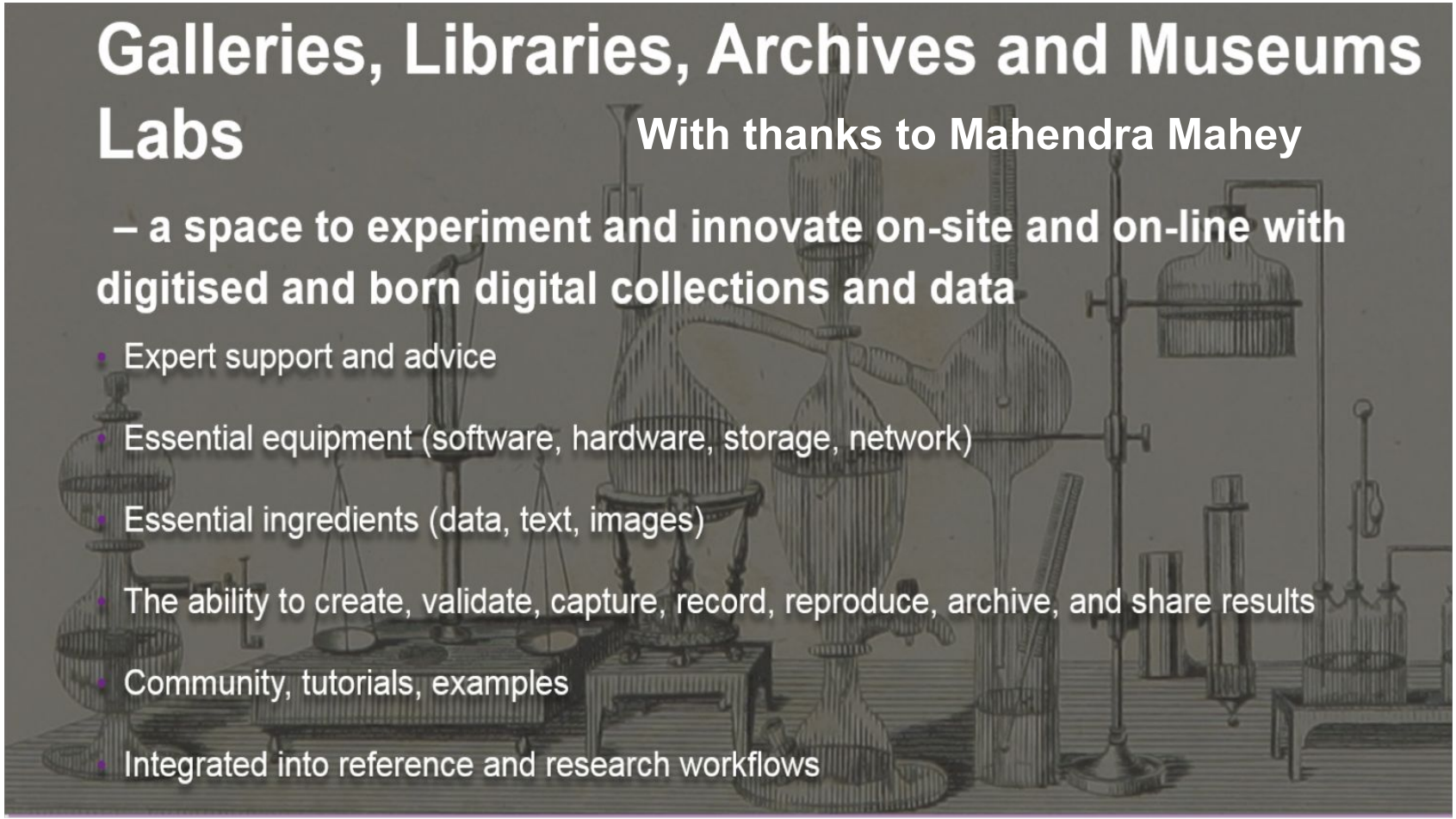
https://collectionsasdata.github.io/

Galleries, Libraries, Archives and Museums Labs

With thanks to Mahendra Mahey

– a space to experiment and innovate on-site and on-line with digitised and born digital collections and data

- Expert support and advice
- Essential equipment (software, hardware, storage, network)
- Essential ingredients (data, text, images)
- The ability to create, validate, capture, record, reproduce, archive, and share results
- Community, tutorials, examples
- Integrated into reference and research workflows



Introducción

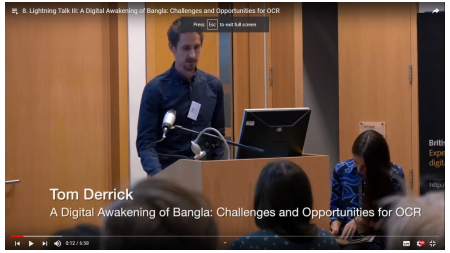
International GLAM Labs Community (https://glamlabs.io/)

- Library of Congress
- British Library
- Royal Danish Library
- National Library of the Netherlands
- National Library of Scotland
- Royal Library of Belgium
- Bibliotheca Alexandrina
- …

slack

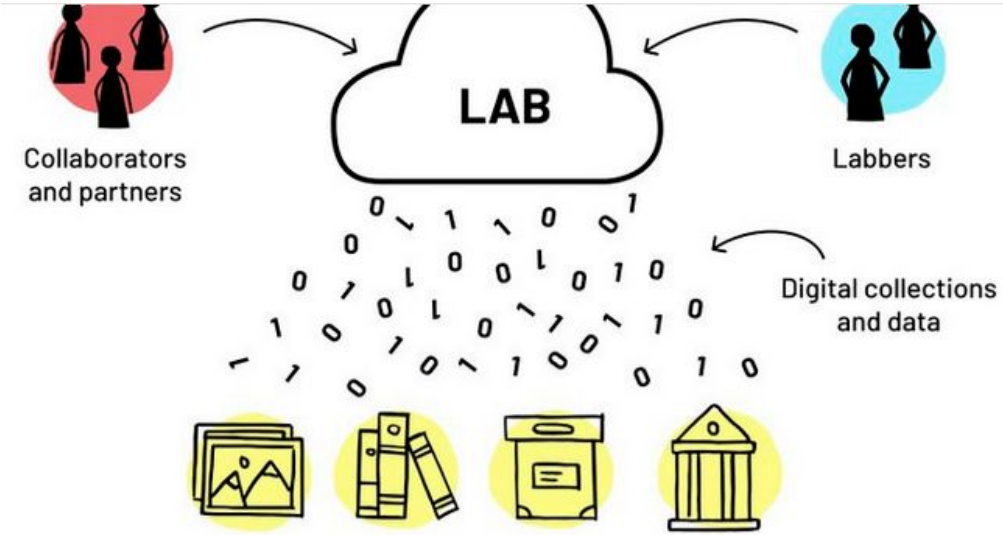
Jisc email subscription

It’s all about people!



YouTube: https://bit.ly/2YMxXYj

International GLAM Labs Community



https://glamlabs.io/books/

https://blogs.bl.uk/digital-scholarship/2019/07/invitation-to-join-digital-cultural-heritage-innovation-labs-book-sprint-doha-qatar-23-27-september-.html

Introducción

Pasos BVMC

Nueva plataforma

2015 2016 2017 2018 2019 2020 2021 2022 2023

WIKIDATA

WIKI DATA CON

DATeCH
Digital Access to Textual Cultural Heritage

I Premio Aporta
El valor del dato para la Administración

BRITISH LIBRARY

International GLAM Labs Community

DET KGL. BIBLIOTEK
Royal Danish Library

International GLAM Labs Community

OPEN A GLAM LAB

WIKI DATA CON

Lancaster University

UNLOCKING COLONIAL ARCHIVE

Primer Premio
British Library Public Awards 2020
Runner Up
Teaching & Learning Award 2020
Special Commendation
Community Award 2020
BVMC.Labs
y el Equipo de Investigación en Bibliotecas de la Universidad de Alicante

INTELE

eadh

Coloquio Internacional
La Práctica Teatral Colaborativa en el Siglo XVII: Textos y Contextos Europeos

impact
digitisation.eu
centre of competence

International GLAM Labs Community

National Library of Scotland
Leabharlann Nàiseanta na h-Alba

COLLECTIONS AS DATA

مؤتمر المكتبات
Library Conference
7TH - 9TH NOV 2023

CLARIAH-EUS

CLARIAH-ES

BNE

europeana

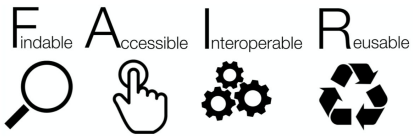
DARIAH-EU
Digital Research Infrastructure for the Arts and Humanities

Collections as Data

“Developing cultural heritage collections that support computationally-driven research and teaching.”

(*Always Already Computational: Collections as Data*: https://collectionsasdata.github.io)

Collections as Data



Focused on characteristics of data

Findable
Rich metadata and persistent identifier

Accessible
Data and metadata understandable by humans and computers

Interoperable
Metadata described using controlled vocabularies

Reusable
License & provenance information

CARE Principles for Indigenous Data Governance

Complement FAIR considering both people and purpose

Collective benefit
Inclusive and equitable outcomes

Authority to control
Recognizing rights and interests

Responsibility
For indigenous languages and worldviews

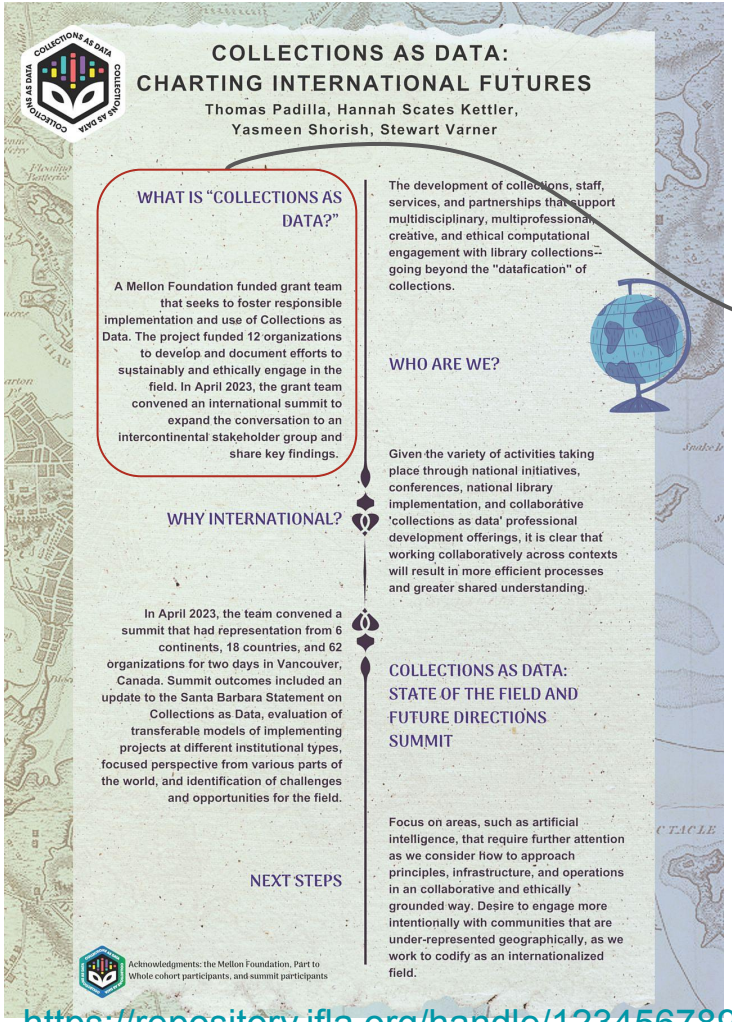
Ethics
Minimizing harm, and for justice and future use

Berners-Lee:

5 estrellas

"5-star Open Data"

Relacionado con FAIR, y abogando por el paradigma de los Datos Enlazados Abiertos (Linked Open Data)



WHAT IS "COLLECTIONS AS DATA?"

A Mellon Foundation funded grant team that seeks to foster responsible implementation and use of Collections as Data. The project funded 12 organizations to develop and document efforts to sustainably and ethically engage in the field. In April 2023, the grant team convened an international summit to expand the conversation to an intercontinental stakeholder group and share key findings.

Deliverables

1. Final Report - https://doi.org/10.5281/zenodo.3152935
2. Santa Barbara Statement on Collections as Data - https://doi.org/10.5281/zenodo.3066209
3. Facets - https://doi.org/10.5281/zenodo.3066240
4. Personas - https://doi.org/10.5281/zenodo.3066515
5. 50 Things - https://doi.org/10.5281/zenodo.3066237
6. Position Statements - https://doi.org/10.5281/zenodo.3066161
7. Methods - https://doi.org/10.5281/zenodo.3146756

https://repository.ifla.org/handle/123456789/3085

https://osf.io/mx6uk/wiki/home/

Collections as Data

Position Statements -> Collections as Data: State of the field and future directions

https://zenodo.org/doi/10.5281/zenodo.7888575

We ask that you write a brief position statement (1-2 pages) derived from direct or related experience salient to the scope of work described in Collections as Data…
https://collectionsasdata.github.io/part2whole/recap/

Summit Participants, Internet Archive Canada, 4/26/2023

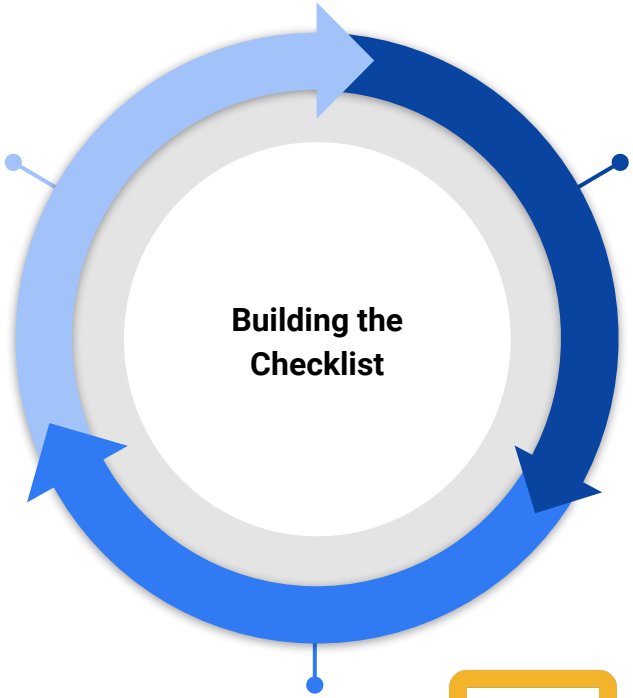
Vancouver Statement on Collections-as-Data
(translations to Spanish, Arabic, French, etc.)

International GLAM Labs Community

Literature review

Best practices, research articles, reports, etc.

...



Community feedback

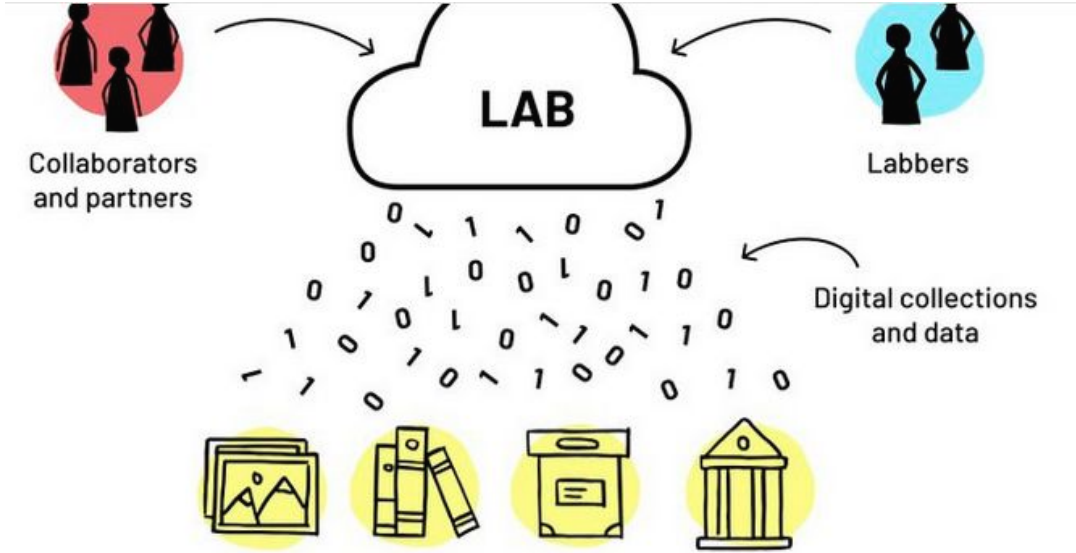
Questionnaire & webinar

With many thanks to Nele Gäbriels

Checklist

Items to assess

International GLAM Labs Community



Towards implementing Collections as Data in GLAM institutions, Tuesday, October 25th 2022, 15:30-16:45 CET.

@GLAM_labs

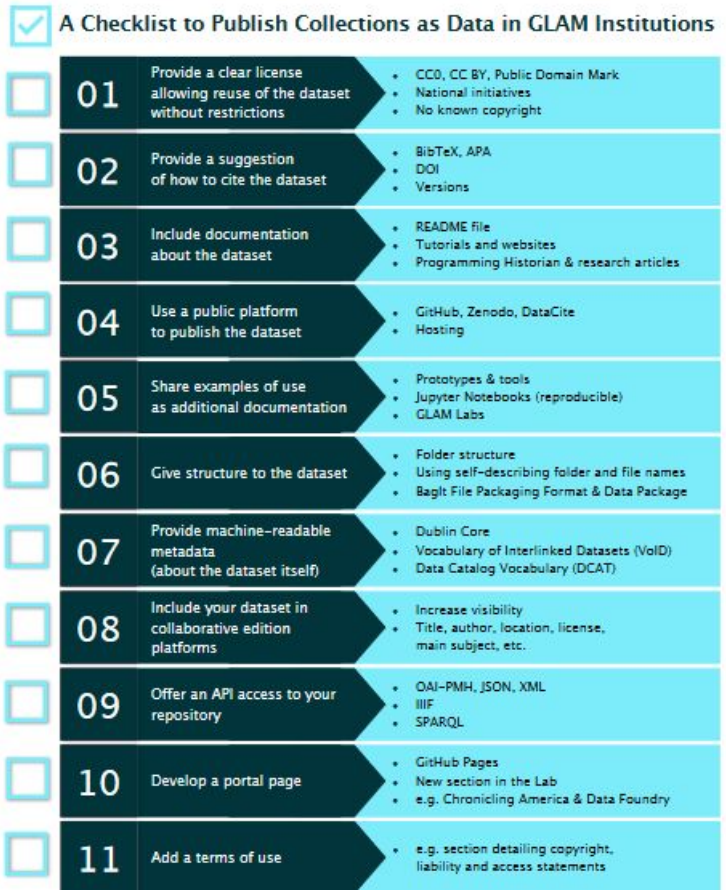
International GLAM Labs Community Meetups

Collections as Data

Publicación

A checklist to make available digital collections suitable for computational use

International GLAM Labs Community



https://doi.org/10.48550/arXiv.2304.02603

Global Knowledge, Memory and Communication

Inicio / Archivos / Vol. 18 (2024) / Bibliotecas y servicios de información y documentación

Collections as data: Acceso computacional a colecciones digitales

Gustavo Candela
Universidad de Alicante
https://orcid.org/0000-0001-6122-0777

DOI: https://doi.org/10.3145/thinkepi.2024.e18a06

Palabras clave: Collections as Data, Acceso computacional, Bibliotecas, GLAM

Resumen

Idioma
English
Español (España)

Información
Para lectores/as
Para bibliotecarios/as

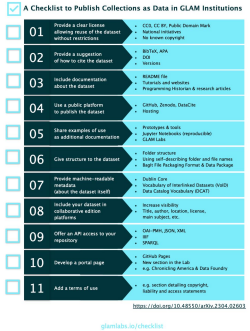
https://doi.org/10.3145/thinkepi.2024.e18a06

¿Por qué un workflow?

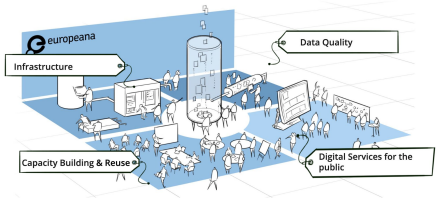
- Más práctico y sistemático
- Reduce barreras
- Adoptar buenas prácticas
- Desarrollo iterativo
- Más fácil de implementar

Automatización de la publicación de Collections as Data

Workflow - Proceso de creación



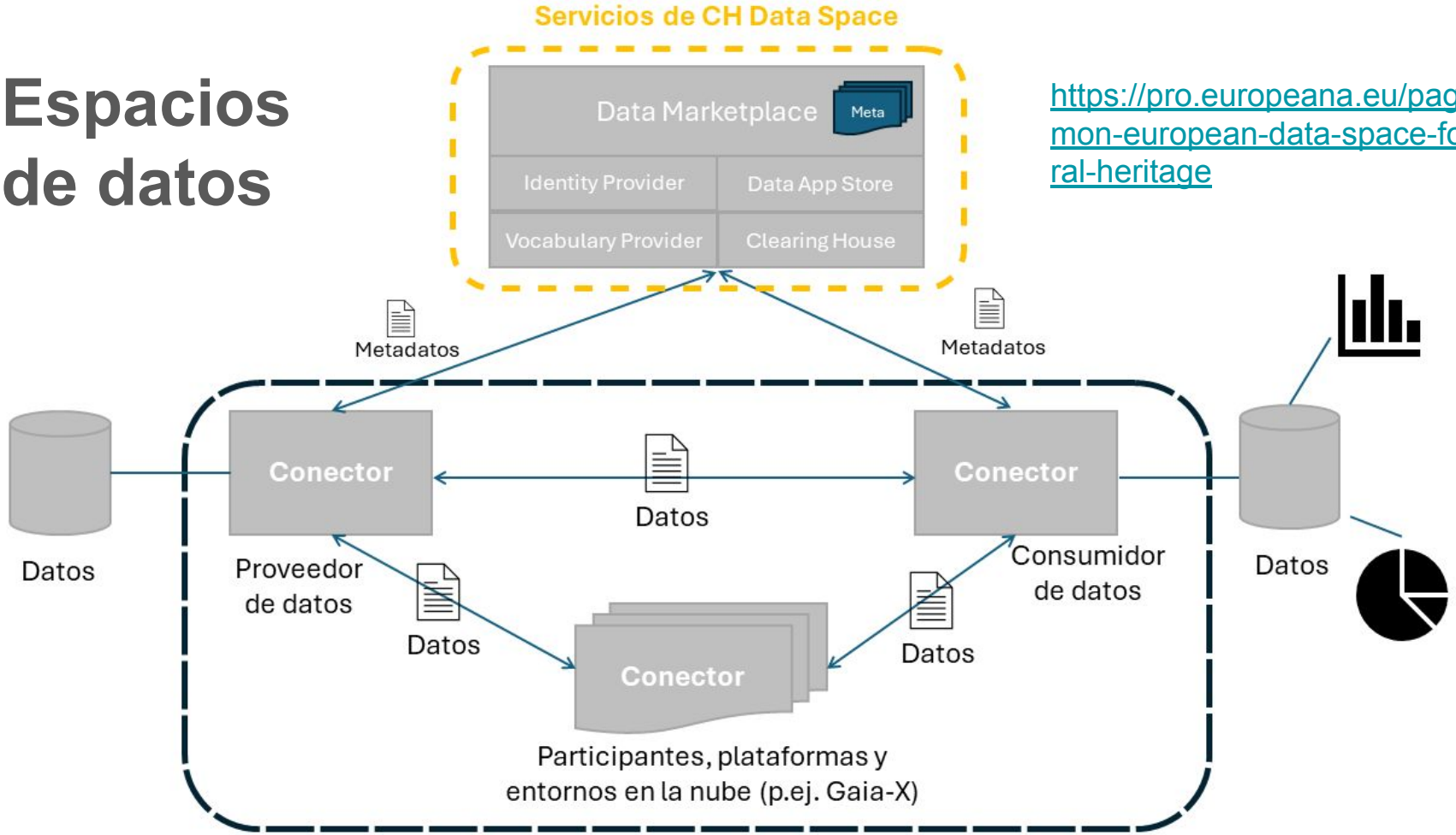
International GLAM Labs Community



European data space for cultural heritage

SSH Open Marketplace
Social Sciences and Humanities Open Marketplace

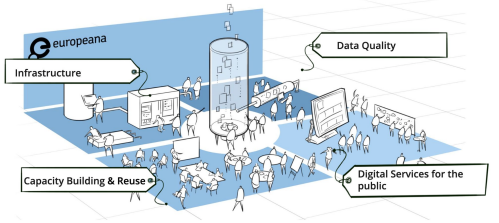
Espacios de datos

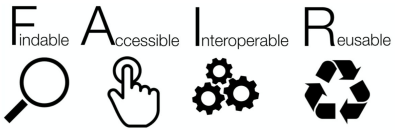


https://pro.europeana.eu/page/common-european-data-space-for-cultural-heritage

A *Collections as Data* Workflow on the SSH Open Marketplace

https://marketplace.sshopencloud.eu/workflow/l3JvP6





With many thanks to Sally Chambers & Alba Irollo

A *Collections as Data* Workflow - 10 pasos

1	Provide a clear license allowing reuse of the dataset without restrictions	Expand
2	Provide a suggested citation for the dataset so reusers are aware of how to cite it	Expand
3	Include documentation about the dataset	Expand
4	Use a public platform to make available the dataset for the public	
5	Share examples of use to demonstrate how the dataset can be reused	
6	Think about a structure for the dataset for a better understanding of how to reuse the content	Expand
7	Include machine-readable metadata about the content provided in the dataset	Expand
8	Use an existing collaborative-edition platform to include the information about the dataset	Expand
9	Provide the dataset by means of an existing API	Expand
10	Create a website to present and describe the dataset to encourage its reuse	Expand

https://marketplace.sshopencloud.eu/workflow/I3JvP6

Workflow steps (10)

1	Provide a clear license and terms of use allowing reuse of the dataset without restrictions	Expand
2	Provide a suggested citation for the dataset so reusers are aware of how to cite it	Expand
3	Include documentation about the dataset	Expand
4	Use a public platform to make available the dataset for the public	Expand
5	Share examples of use to demonstrate how the dataset can be reused	
6	Think about a structure for the dataset for a better understanding of how to reuse the content	
7	Include machine-readable metadata about the content provided in the dataset	
8	Use an existing collaborative-edition platform to include the information about the dataset	
9	Provide the dataset by means of an existing API	
10	Create a website to present and describe the dataset to encourage its reuse	

SSH Open Marketplace
Social Sciences and Humanities Open Marketplace

¿Cómo usarlo?

Detalles

5 Share examples of use to demonstrate how the dataset can be reused — Collapse

Why Examples of use of the contents provided by a digital collection are useful to inspire researchers. For example, reproducible code provided by Jupyter Notebooks and prototypes help to better understand how to use the content provided by a dataset. Additional approaches include the definition of research scenarios to address a particular topic.

How For example, the Jupyter Notebooks to process the Europeana newspaper text resources have been designed to help getting started with the processing of historical texts (from Europeana Newspapers) using Natural Language Processing (NLP) tools. Additional examples of reuse based on Jupyter Notebooks are provided by the International GLAM Labs Community. Other approaches proposed by KBR, the Royal Library of Belgium's DATA-KBR-BE project are based on the definition of research scenarios focused on a particular topic such as mapping the publication of literature in Belgian newspapers across the first century of the Belgian nation state.

Related items (5)

- Reutilizar colecciones digitales: GLAM Labs
  Esta lección muestra cómo reutilizar colecciones digitales publicadas por instituciones de patrimonio cultural.
  Read more
- Reusing digital collections from GLAM institutions
  For some decades now, Galleries, Libraries, Archives and Museums (GLAM) institutions have published and provided access to information resources in digital format. Recently, innovative approaches have appeared such as th...
  Read more
- Jupyter notebooks for Europeana newspaper text resource processing with CLARIN NLP tools
  Binder These notebooks have been designed to help getting started with the processing historical text resources (from Europeana Newspapers) with
- Introduction to Jupyter Notebooks
  Jupyter notebooks provide an environment where you can freely combine human-readable narrative with computer-readable code. This lesson

https://marketplace.sshopencloud.eu/workflow/I3JvP6

europeana research

DARIAH-EU

TRAINING WORKSHOP

A workflow to publish Collections as Data

13 February 2024 | Online

Funded by the European Union

Image: Exposing Online the European Cultural Heritage: The impact of Cultural Heritage on the Digital Transformation of The Society. Photo by Sebastiaan ter Burg, CC BY 4.0

¿Cuál de estos elementos es más importante para tu trabajo?

1. Incluir licencia y términos de uso
2. Incluir una sugerencia de cita
3. Documentación sobre el conjunto de datos
4. Utilizar una plataforma para publicar el conjunto de datos
5. Proveer ejemplos de reutilización

¿Cuál de estos pasos es más complejo de aplicar en tu organización?

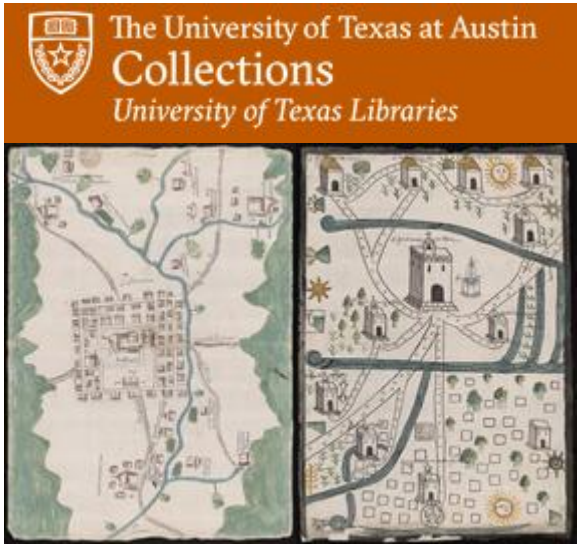
1. Dar estructura a la colección de datos
2. Proporcionar metadatos legibles por máquina
3. Utilizar una plataforma de edición colaborativa
4. Utilizar un API para publicar la colección
5. Crear una web para la colección de datos

¿Dónde podemos buscar colecciones digitales?

SSH Open Marketplace

zenodo

International GLAM Labs Community



CARIBBEAN NEWSPAPERS

Collections as Data

europeana pro
https://pro.europeana.eu/index.php/page/harvesting-and-downloads

Tímarit.is
https://timarit.is/

SAAM
https://americanart.si.edu/about/lod

Data Foundry
Data collections from the National Library of Scotland
https://doi.org/10.5281/zenodo.8051036

LIBRARY LABS
LIBRARY OF CONGRESS
https://id.loc.gov

KUNGL. BIBLIOTEKET
National Library of Sweden
https://libris.kb.se/sparql

16 40 THE NATIONAL LIBRARY OF FINLAND
https://data.nationallibrary.fi

BRITISH LIBRARY
https://bnb.data.bl.uk

DATOS·BNE·es
https://datos.bne.es/

BnF Data
https://data.bnf.fr/

KB LAB
https://data.bibliotheken.nl

CENTRO BIBLIOTECA VIRTUAL MIGUEL DE CERVANTES
https://data.cervantesvirtual.com/

Bibliothèque nationale du Luxembourg
Open Data
https://data.bnl.lu/

DEUTSCHE NATIONAL BIBLIOTHEK
https://www.dnb.de/EN/lds

Österreichische Nationalbibliothek
https://labs.onb.ac.at/en/dataset/lod

Collections as Data (nacional)

Diferente contenido y formato

BIBLIOTECA DIGITAL HISPÁNICA
BIBLIOTECA NACIONAL DE ESPAÑA

DATOS·BNE·es

BNE Hemeroteca Digital
Biblioteca Nacional de España

datos.gob.es
reutiliza la información pública

CENTRO BIBLIOTECA VIRTUAL
MIGUEL DE CERVANTES

AYUNTAMIENTO DE
TOLEDO

ARCHIVO HISTÓRICO DIGITAL
DE LA BIBLIOTECA DE LA UPM

DOI 10.5281/zenodo.3634442

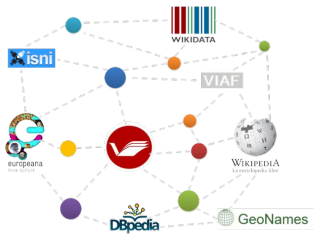
Corpus of Spanish Golden-Age Sonnets

MNEMOSINE

Biblioteca Virtual Miguel de Cervantes & RDA

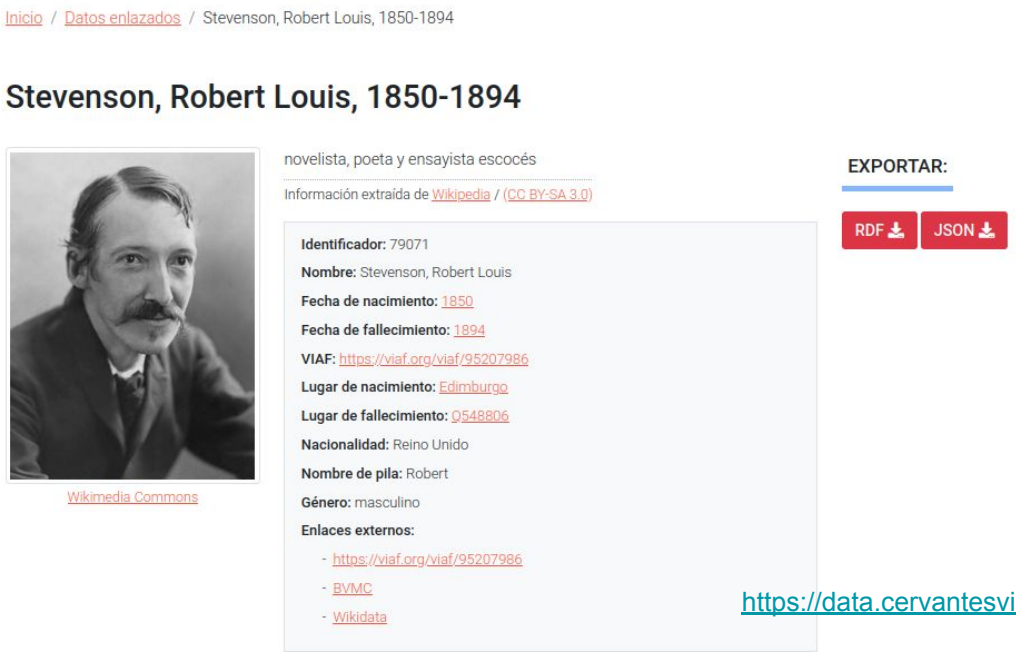
En 2015 la BVMC transformó el catálogo a RDA para facilitar su **reutilización**

- Basado en la Web **Semántica y Linked Open Data**
- **Modelado de datos** siguiendo el vocabulario RDA Registry
- **Enriquecimiento** con repositorios externos como **Wikidata**
- Disponible en SPARQL API



Biblioteca Virtual Miguel de Cervantes

La BVMC y Wikidata están conectadas = metadatos enriquecidos

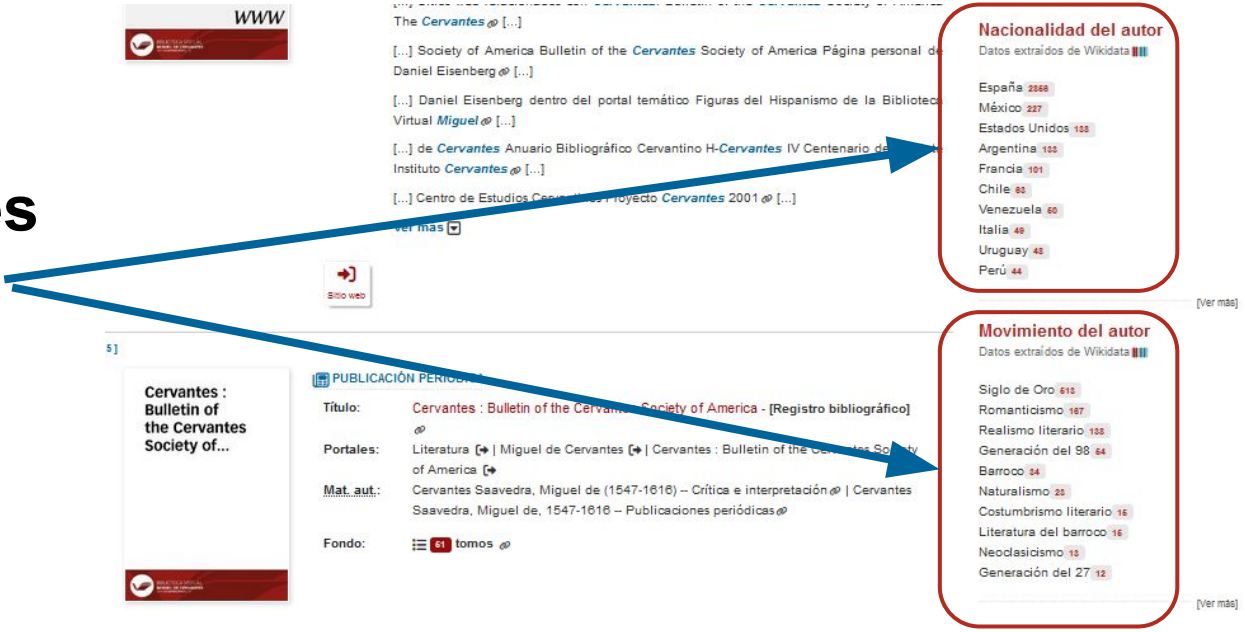


https://data.cervantesvirtual.com/person/79071

Biblioteca Virtual Miguel de Cervantes

La BVMC y Wikidata están conectadas = metadatos enriquecidos

Nuevas opciones para explorar el catálogo



https://www.cervantesvirtual.com/buscador/?q=miguel+de+cervantes

LOD data.cervantesvirtual.com (Q111396572)

Linked Open Data from Biblioteca Virtual Miguel de Cervantes edit

In more languages
Configure

Language	Label	Description	Also known as
English	LOD data.cervantesvirtual.com	Linked Open Data from Biblioteca Virtual Miguel de Cervantes	
Spanish	LOD data.cervantesvirtual.com	Proyecto de datos abiertos de la Biblioteca Virtual Miguel de Cervantes	
Catalan	No label defined	No description defined	
Galician	No label defined	No description defined	

Statements

instance of	open data portal 0 references	edit + add reference + add value

Biblioteca Virtual Miguel de Cervantes: aplicando el workflow

https://data.cervantesvirtual.com

CENTRO BIBLIOTECA VIRTUAL
MIGUEL DE CERVANTES

Workflow steps (10)

SSH Open Marketplace
Social Sciences and Humanities Open Marketplace

1 | **Provide a clear license and terms of use allowing reuse of the dataset without restrictions**

¿Bajo qué licencia se han publicado los datos? ¿Y términos de uso?

CONDICIONES DE USO

El sitio web data.cervantesvirtual.com está dirigido por la Biblioteca Virtual Miguel de Cervantes y da acceso a un amplio elenco de patrimonio cultural digitalizado de toda Europa y del resto del mundo. Este material lo proporcionan un gran número de instituciones y organizaciones. La Biblioteca Virtual Miguel de Cervantes trabaja para que todos los recursos de este sitio web estén disponibles para su reutilización en otros espacios. Es por ello que, todos los metadatos (información textual sobre el patrimonio cultural digitalizado) en esta web, se publican sin ningún tipo de restricción. La mayoría de los demás materiales, como las previsualizaciones del patrimonio cultural digitalizado, están claramente etiquetados indicando si son reutilizables y en qué condiciones. Todos los metadatos disponibles en data.cervantesvirtual.com se publican sin restricciones, bajo las condiciones de Creative Commons CC0 1.0 Universal Public Domain Dedication. Sin embargo, si reutiliza los datos publicados por la Biblioteca Virtual Miguel de Cervantes, le recomendamos que siga las pautas de uso de esta para metadatos y cite la fuente de procedencia siempre que sea posible.

https://data.cervantesvirtual.com/condiciones-de-uso/

2 | **Provide a suggested citation for the dataset so reusers are aware of how to cite it**

¿Cómo puedo citarlo?

REFERENCIAS

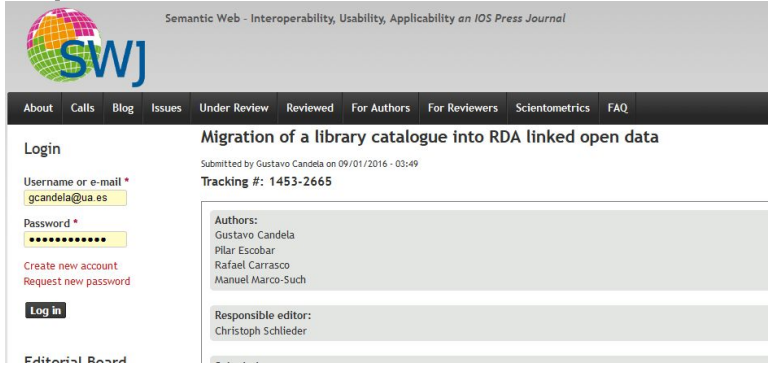
- Candela, G., Escobar, P., Carrasco, R. y Marco-Such, M. (2020). Evaluating the quality of linked open data in digital libraries. *Journal of Information Science*. https://doi.org/10.1177/0165551520930951
- Candela, G., Escobar, P., Carrasco, R. y Marco-Such, M. (28-29 de octubre de 2019). *Evaluating the quality of linked open data in digital libraries*. WikidataCon. Berlín, Alemania. https://www.wikidata.org/wiki/Wikidata:WikidataCon_2019/Program/Sessions/Libraries_panel
- Candela, G., Escobar, P., Carrasco, R. y Marco-Such, M. (2019). A linked open data framework to enhance the discoverability and impact of culture heritage. *Journal of Information Science*, 45(6), 756-766. https://doi.org/10.1177/0165551518812658
- Candela, G., Escobar, P., Carrasco, R. y Marco-Such, M. (2018). Migration of a Library Catalogue into RDA Linked Open Data. *Semantic Web Journal*, 9(4), 481-491. http://dx.doi.org/10.3233/SW-170274 [Versión *preprint*]

https://data.cervantesvirtual.com/datos-enlazados

CENTRO BIBLIOTECA VIRTUAL MIGUEL DE CERVANTES

3 | **Include documentation about the dataset**

¿Cómo proporcionar documentación?



CENTRO BIBLIOTECA VIRTUAL MIGUEL DE CERVANTES

Workflow steps (10)

SSH Open Marketplace

5 | **Share examples of use to demonstrate how the dataset can be reused**

¿Cómo proporcionar ejemplos de uso?

Jupyter Notebooks

Esta sección introduce varios proyectos basados en Jupyter Notebooks y el acceso computacional a colecciones digitales realizados en el marco de la Biblioteca Virtual Miguel de Cervantes e inspirados por la Comunidad Internacional GLAM Labs, el libro Open a GLAM Lab, Collections as Data y el proyecto GLAM Workbench. Este trabajo ha sido premiado por la Biblioteca Británica y presentado en varias charlas y conferencias como Research Libraries UK (RLUK), European Library Automation Group y la Universidad de Lancaster.

Todas las colecciones están publicadas en abierto, utilizan el lenguaje de programación Python y pueden ser ejecutadas en la nube sin necesidad de instalar ningún *software*, únicamente utilizando un navegador. Los notebooks han sido creados con la herramienta Jupyter.

Con el objetivo de facilitar la reutilización y el acceso, los notebooks han sido clasificados según el nivel de dificultad *introducción* o *avanzado* teniendo en cuenta las librerías de código utilizadas y la complejidad para reproducir los resultados.

GLAM JUPYTER NOTEBOOKS

Colección premiada por la Biblioteca

REUTILIZAR COLECCIONES DIGITALES: GLAM LABS

IDENTIFICAR TÉRMINOS EMERGENTES

Compartir Pantalla: Linked Open Da

Proyecto presentado en la actividad "Compartir pantalla" con la Asociaci

launch binder

DOI 10.5281/zenodo.10123566

HDH Humanidades Digitales Hispánicas

BIBLIOTECA VIRTUAL MIGUEL DE CERVANTES

Universitat d'Alacant Universidad de Alicante

Exploring collections as data with Jupyter Notebooks

https://glamlabs.io/

https://data.cervantesvirtual.com/notebooks

https://github.com/hibernator11/hdh-compartir-pantalla-2023

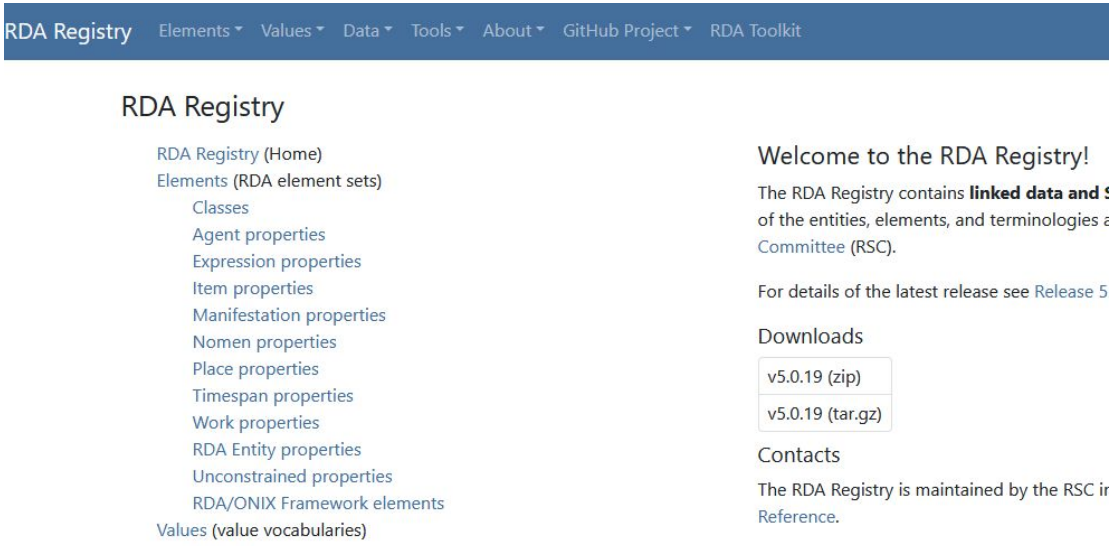
CENTRO BIBLIOTECA VIRTUAL MIGUEL DE CERVANTES

Workflow steps (10)

SSH Open Marketplace

7 | **Include machine-readable metadata about the content provided in the dataset**

¿Qué vocabulario se utiliza para describir los metadatos?



https://www.rdaregistry.info/

CENTRO BIBLIOTECA VIRTUAL MIGUEL DE CERVANTES

Workflow steps (10)

SSH Open Marketplace

8	Use an existing collaborative-edition platform to include the information about the dataset

¿Entidad en Wikidata?

WIKIDATA

LOD data.cervantesvirtual.com (Q111396572)

Linked Open Data from Biblioteca Virtual Miguel de Cervantes — edit

In more languages
Configure

Language	Label	Description	Also known as
English	LOD data.cervantesvirtual.com	Linked Open Data from Biblioteca Virtual Miguel de Cervantes	
Spanish	LOD data.cervantesvirtual.com	Proyecto de datos abiertos de la Biblioteca Virtual Miguel de Cervantes	
Catalan	No label defined	No description defined	
Galician	No label defined	No description defined	

Statements

instance of	open data portal	edit
	0 references	+ add reference
		+ add value

SSH Open Marketplace
Social Sciences and Humanities Open Marketplace

https://www.wikidata.org/wiki/Q111396572

CENTRO BIBLIOTECA VIRTUAL
MIGUEL DE CERVANTES

Workflow steps (10)

SSH Open Marketplace
Social Sciences and Humanities Open Marketplace

9 | Provide the dataset by means of an existing API

https://data.cervantesvirtual.com/sparql

CENTRO BIBLIOTECA VIRTUAL
MIGUEL DE CERVANTES

Workflow steps (10)

SSH Open Marketplace
Social Sciences and Humanities Open Marketplace

10 | **Create a website to present and describe the dataset to encourage its reuse**

https://data.cervantesvirtual.com

GitHub pages?

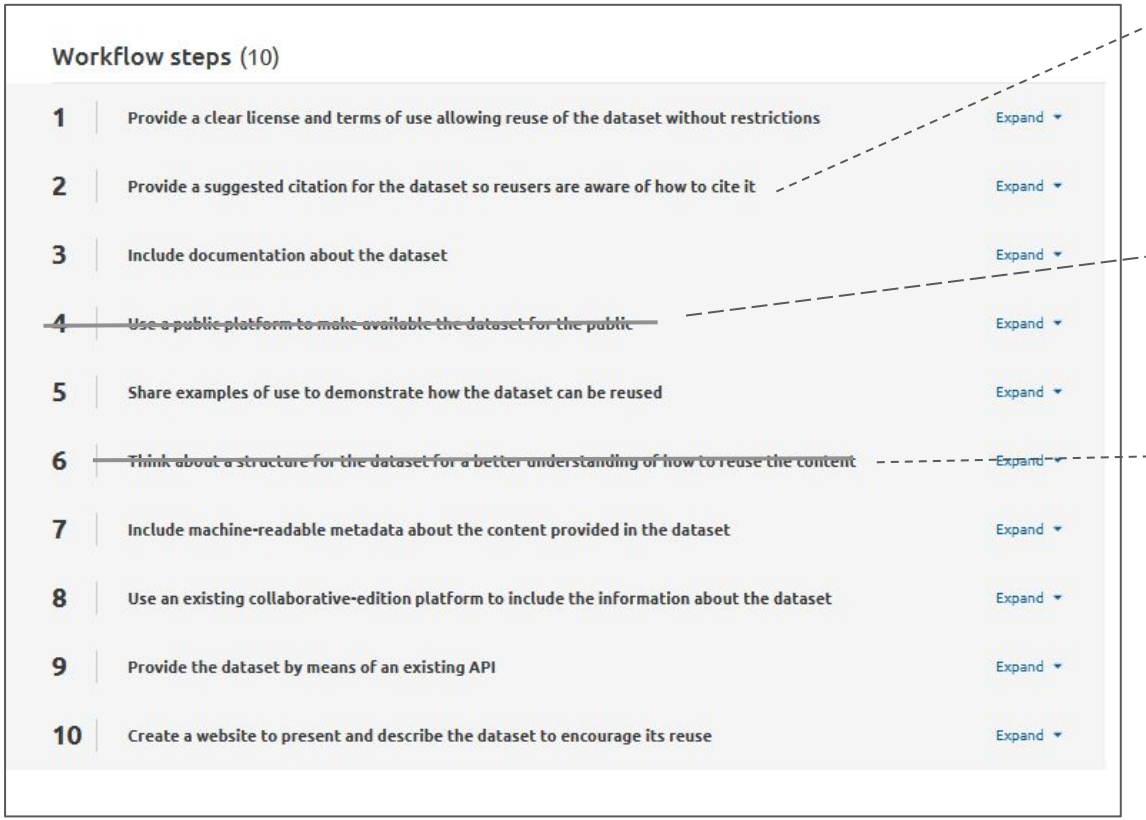
GitHub Pages

Websites for you and your projects.

https://pages.github.com/

CENTRO BIBLIOTECA VIRTUAL
MIGUEL DE CERVANTES

Resumen de la evaluación



¿cita sugerida?

¿Zenodo (DOI), GitHub, institutional repositories?

No aplica

¿Dudas o preguntas?

https://data.cervantesvirtual.com

Actividad
¿Cómo usar el workflow propuesto?
¿Dificultades?¿Por dónde empezar?
Propuestas del público

Uso de Jupyter Notebooks y Wikidata

CENTRO BIBLIOTECA VIRTUAL MIGUEL DE CERVANTES

UNIVERSIDAD COMPLUTENSE MADRID

Universitat d'Alacant Universidad de Alicante

Desafíos

- Las instituciones GLAM pueden jugar un papel relevante en IA
  - https://www.rluk.ac.uk/digital-shift-manifesto/
- Adopción de Collections as data por parte de las instituciones más pequeñas
  - Publicar el texto, no solo metadatos
  - Diferentes versiones de la colección digital
  - Uso de licencias abiertas
  - Ética y términos de uso (CARE principles)

impact digitisation.eu centre of competence

RLUK Research Libraries UK

Desafíos

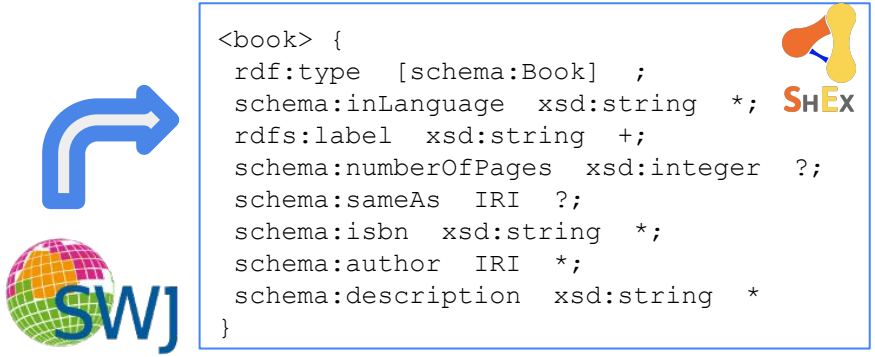
Diversidad en el uso de vocabularios controlados para definir los metadatos

Table 1. Overview of LOD repositories made available by CH institutions

Institution	URL	Vocabulary
Biblioteca Virtual Miguel de Cervantes	https://data.cervantesvirtual.com/sparql	RDA
British Library	https://bl.natbib-lod.org	BIBFRAME
Europeana	https://sparql.europeana.eu	EDM
Getty Research Institute	https://data.getty.edu/vocab/sparql	CIDOC-CRM
Library of Congress	https://id.loc.gov	BIBFRAME
National Library of Finland	https://data.nationallibrary.fi/bib/sparql	Schema.org
National Library of France	https://data.bnf.fr/sparql	FRBR
National Library of Scotland	https://data.nls.uk/tools/jupyter-notebooks/semantic-web/	BIBFRAME, schema.org
National Library of Spain	https://datos.bne.es/sparql	FRBR
National Library of the Netherlands	http://data.bibliotheken.nl/sparql	Schema.org
Rijksmuseum	https://data.rijksmuseum.nl	EDM
Smithsonian American Art Museum	http://edan.si.edu/saam/sparql	CIDOC-CRM
Zeri Photo Archive	http://data.fondazionezeri.unibo.it/	CIDOC-CRM
Yale Collections Discovery	https://lux.collections.yale.edu	CIDOC-CRM

Desafíos

Métodos para evaluar la **calidad de datos** en instituciones GLAM



A Shape Expression approach for assessing the quality of Linked Open Data in libraries

https://doi.org/10.3233/SW-210441

?

W3C SPARQL

Evaluating the quality of linked open data in digital libraries

https://doi.org/10.1177/0165551520930951

JASIST JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY

SHEX sheXer

An automatic data quality approach to assess semantic data from cultural heritage institutions

https://doi.org/10.1002/asi.24761

Desafíos

Reutilización - Jupyter Notebooks

An approach to **assess the quality** of Jupyter projects published by GLAM institutions

- A method to assess Jupyter notebook projects
- Results of the evaluation of Jupyter Notebooks projects
- Best practice and guidelines

Dimension	Criterion
Understandability	Using literate programming features
	Including additional documentation and guidelines
	Naming of the notebooks
	Storing cell output
	Audience/intended use
	Provisioning of metadata
Availability	License
Efficiency	Size
Traceability	Versioning
Portability	Providing dependencies
Recoverability	Providing citation information
	Last run date
Credibility	Trustworthiness on project level

Candela, G., Chambers,S., & Sherratt, T. (2023). An approach to assess the quality of Jupyter projects published by GLAM institutions. Journal of the Association for Information Science and Technology, 1–15. https://doi.org/10.1002/asi.24835

Desafíos

Calidad OCR


LA ABEJA.

```
Xlílll. 4s3.

llflrit S 10 lic jllili0 lic 1 S 34.

I 1) C II fll'1.0S.

precia dc lc inicrlcion cn Mcllevado cl pcriúd leo i
b lct vciattcc tnicritvrci. Pcr nn mcc co rc.

L^

..

P cio dc lc inicrícton cn lii pro-, vinclct franco d
Par nn mcc sg rs

ESTE OEBIODIt O SALE TODOS LOS DIAS KSCEPTO LOS LUíXK
ctl Rea\onza Xagürl un s Á cctc tic D p p crri llisin

ADI'El\TErCIA.
Hnbíeiidose tíignedo S. llf. Ió Iteittu Gobcrttedoru
ste cu eonoeímiettto tle lss señtores suscritores g d
tssntos dísneusurmi su heuenoíetteío ul neriótheo sun
```

impact
digitisation.eu
centre of competence

Impact White Paper

https://hemerotecadigital.bne.es/hd/es/viewer?id=121b4e8c-3012-4e82-8679-60fe7e9b2829

Desafíos

Inteligencia Artificial

KB nationale bibliotheek

AI companies do not ask permission to import these data and are not transparent about how these data are used. *KB management team member Martijn Kleppe*

KB restricts access to collections for training commercial AI

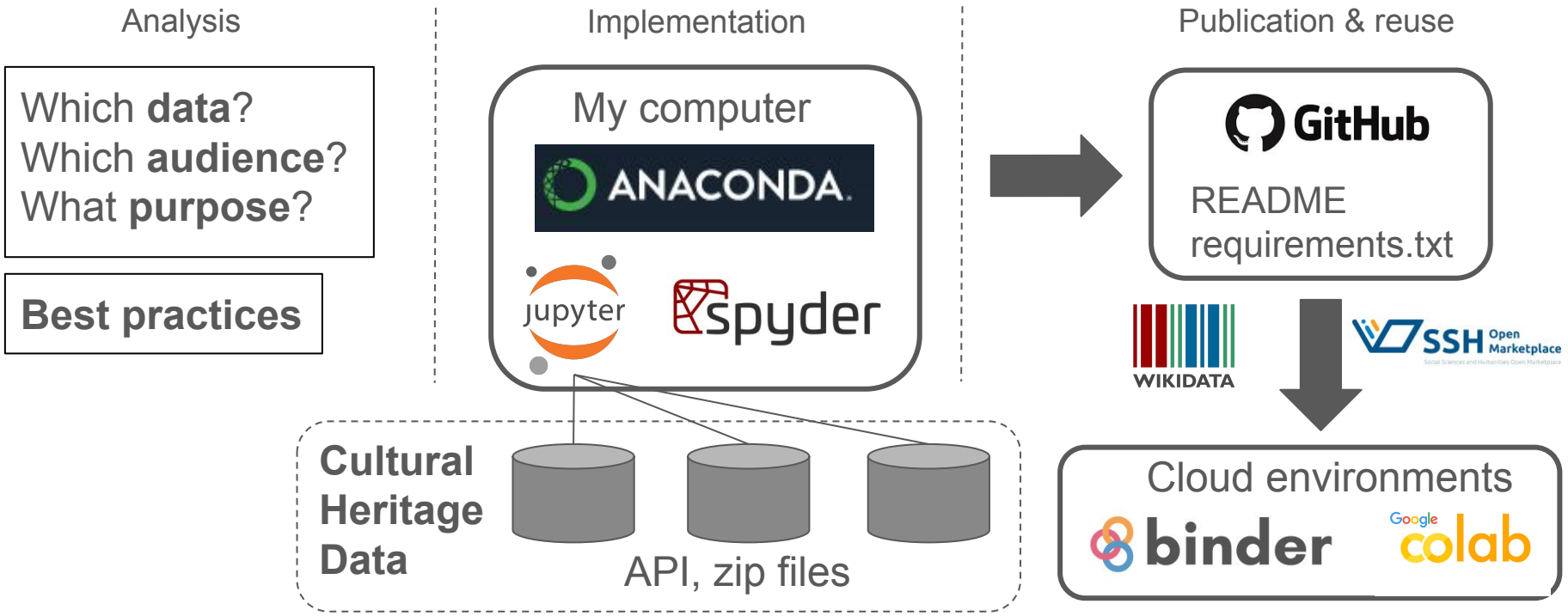
The KB does not want commercial companies to use digital resources without permission for training AI. That goes against the AI principles established by the KB. The KB has since taken measures to restrict this use and will be releasing a statement on this today.

https://www.kb.nl/en/nieuws/kb-restricts-access-to-collections-for-training-commercial-ai

Trabajo futuro

DHNB Digital Humanities in the Nordic & Baltic Countries

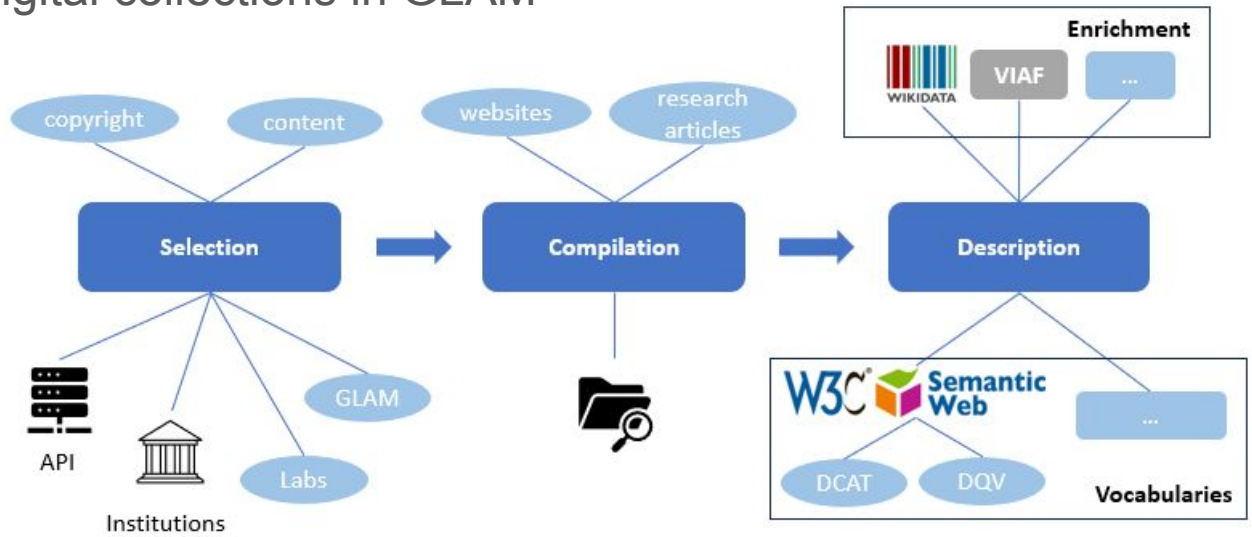
Creating a Jupyter Notebook collection from scratch



Trabajo futuro

Uso de vocabularios controlados para describir catálogos

A **Linked Open Data framework** based on **Data Catalog Vocabulary** to describe digital collections in GLAM



Categories of metadata

- Distribution
- Quality
- Provenance
- Examples of use
- Composition
- Terms of use
- Transparency
- Collecting process
- Preprocessing

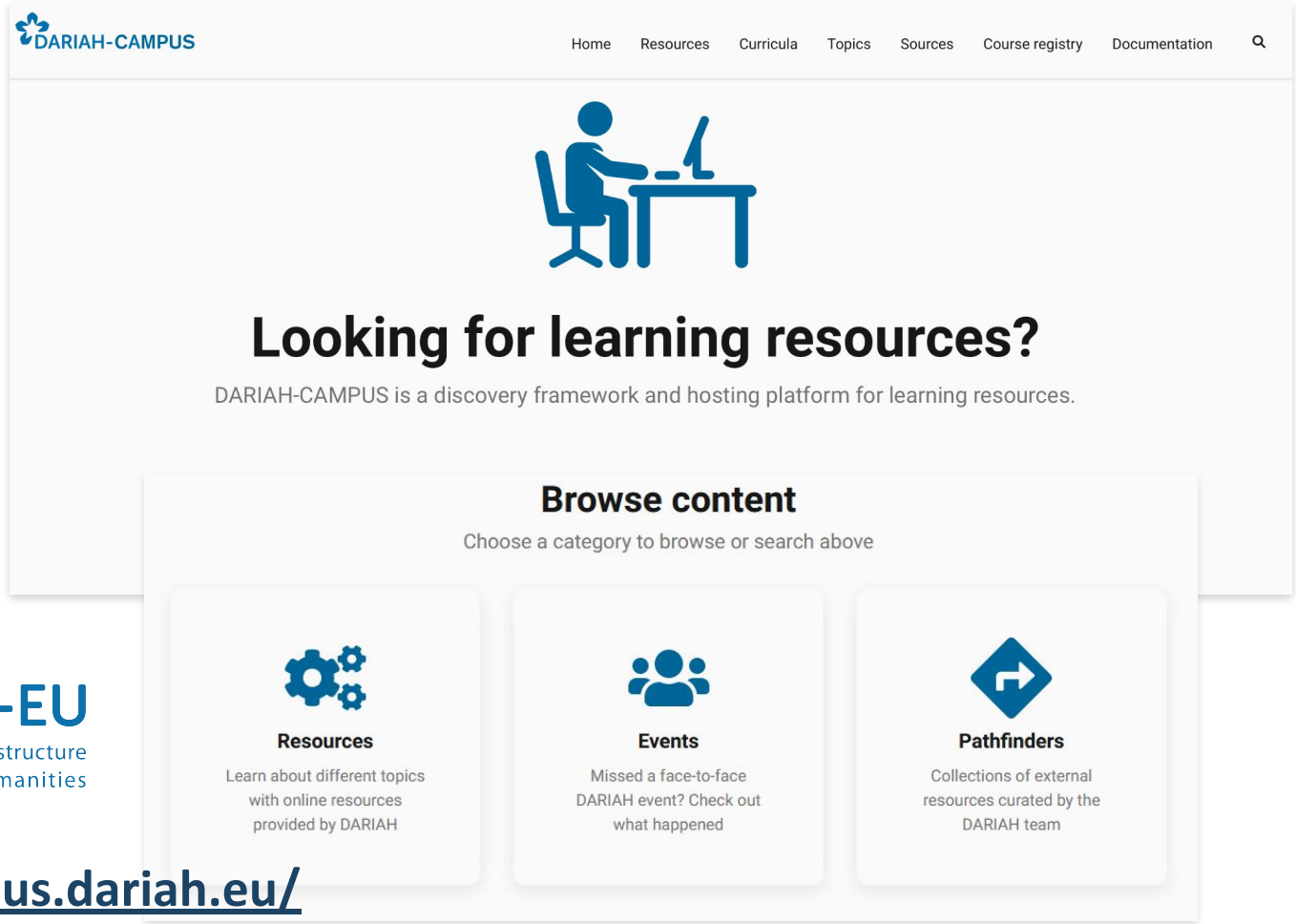
https://github.com/hibernator11/dcat-glam-catalog

DARIAH Annual Event 2024

June 18, 2024 - June 21, 2024

Workflows: Digital Methods for Reproducible Research Practices in the Arts and Humanities

https://annualevent.dariah.eu/



DARIAH-EU
Digital Research Infrastructure for the Arts and Humanities

https://campus.dariah.eu/

Ciencia abierta, datasets, EOSC

https://oscars-project.eu/oscars-first-open-call

Estancias en instituciones europeas

https://atrium-research.eu/

Referencias

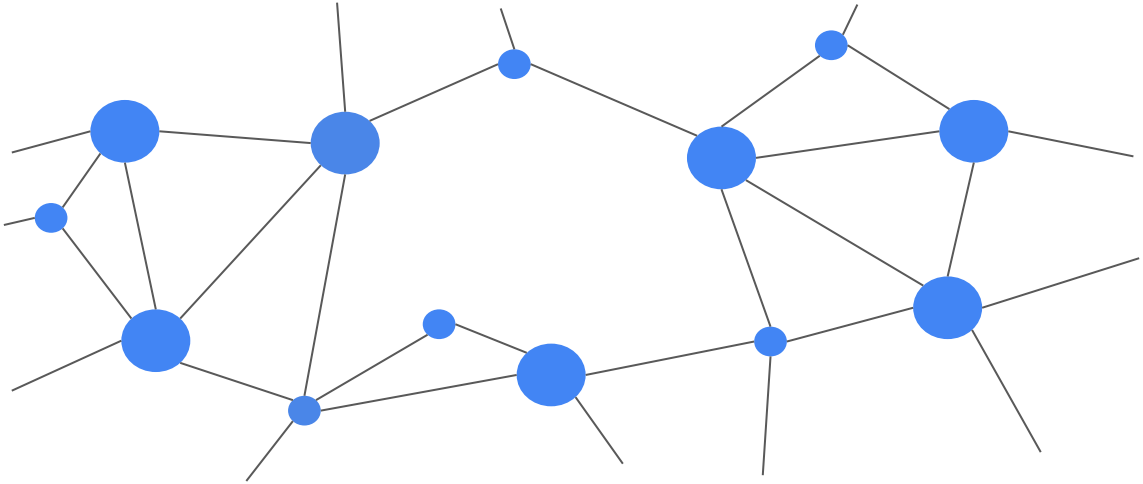
- Asociación Humanidades Digitales Hispánicas y data.cervantesvirtual.com
- Collections as Data: State of the Field and Future Directions
- Impact Centre of competence: Sharing and Sustaining Digitisation Knowledge
- International GLAM Labs Community
- Lancaster University: Unlocking the Colonial Archive
  - https://github.com/hibernator11/UCA-relacionesgeograficas
- National Library of Scotland
  - https://data.nls.uk/projects/the-national-librarians-research-fellowship-in-digital-scholarship-2022-23/
  - https://data.nls.uk/tools/jupyter-notebooks/semantic-web/
- OCLC Research
- Tim Sherratt. GLAM Workbench
- https://www.kbr.be/en/projects/data-kbr-be
- https://web.ua.es/es/actualidad-universitaria/2023/junio2023/12-16/la-ua-acoge-a-los-impulsores-de-la-red-estrategica-clariah-es.html

Referencias

- Mahey, M. et al. Open a GLAM Lab. International GLAM Labs Community, Book Sprint, 2019. https://doi.org/10.21428/16ac48ec.f54af6ae
- Candela, G., Chambers,S., & Sherratt, T. (2023). An approach to assess the quality of Jupyter projects published by GLAM institutions. Journal of the Association for Information Science and Technology, 1–15. https://doi.org/10.1002/asi.24835
- Candela, G. An automatic data quality approach to assess semantic data from cultural heritage institutions. J. Assoc. Inf. Sci. Technol. 74(7): 866-878 (2023). https://doi.org/10.1002/asi.24761
- Candela, G., Escobar, P., Carrasco, R. and Marco-Such, M. Evaluating the quality of linked open data in digital libraries. J. Inf. Sci. 48(1): 21-43 (2022). https://doi.org/10.1177/0165551520930951
- Candela, G. Towards a semantic approach in GLAM Labs: The case of the Data Foundry at the National Library of Scotland.Journal of Information Science. (2023) Online first. https://doi.org/10.1177/01655515231174386
- Candela, G., Pereda, J., Sáez, D., Escobar, P., Sánchez, A., Villa-Torres, A., Palacios, A., McDonough, K. and Murrieta-Flores, P. 2023. An ontological approach for unlocking the Colonial Archive. J. Comput. Cult. Herit. Just Accepted (April 2023). https://doi.org/10.1145/3594727
- Candela, G., Gabriëls, N., Chambers, S., Dobreva, M., Ames, S., Ferriter, M., Fitzgerald, N., Harbo, V., Hofmann, K., Holownia, O., Irollo, A., Mahey, M., Manchester, E., Pham, T.-A., Potter, A. and Van Keer, E. (2023), "A checklist to publish collections as data in GLAM institutions", Global Knowledge, Memory and Communication, Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/GKMC-06-2023-0195

CLARIAH-CM: Workflows para la publicación de colecciones como datos

¡Muchas gracias!



CENTRO BIBLIOTECA VIRTUAL MIGUEL DE CERVANTES

UNIVERSIDAD COMPLUTENSE MADRID

Universitat d'Alacant
Universidad de Alicante