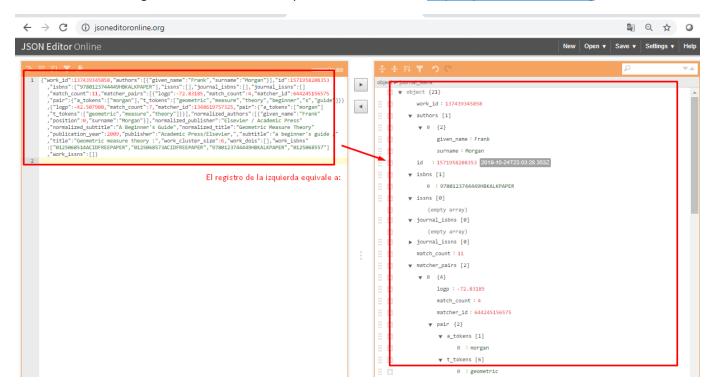# Estructura de archivos Syllabus

Fecha: septiembre 4 de 2019

- Fuente datos. https://app.box.com/s/o6x5l94hpbjjsq4xp9z237ffx0astj4h Estructura original de archivos en formato .json
- Ruta con archivos actualizados (no .json). Sharepoint Biblioteca Alfonso Borrero Cabal, S.J.,

  Documentos > Tecnología > Documentación técnica SI > Datos Syllabus - Desarrollo de Colecciones
- Carpetas. Tres: 'catalog.json', 'matches.json' y 'syllabi.json'.
- Estructura actualizada de archivos: .csv separados por comas, formato UTF-8.
- Validación registros. Se recomienda copiar la información en https://jsoneditoronline.org

- Desgloce. Ver los siguientes apartados.

1. Carpeta 'catalog.json': 99 archivos

- Ruta Sharepoint con archivos carpeta 'catalog.json'.
- Ejemplo registro nativo:
  {"work_id":137439345858,"authors":[{"given_name":"Frank","surname":"Morgan"}],"id":15719 58208353,"isbns":["9780123744449HBKALKPAPER"],"issns":[],"journal_isbns":[],"journal_issns": [],"match_count":11,"matcher_pairs":[{"logp":-72.83185,"match_count":4,"matcher_id":644245156575,"pair":{"a_tokens":["morgan"],"t_toke

ns":["geometric","measure","theory","beginner","s","guide"]}},{"logp":-42.507908,"match_count":7,"matcher_id":1348619757325,"pair":{"a_tokens":["morgan"],"t_tokens":["geometric","measure","theory"]}}],"normalized_authors":[{"given_name":"Frank","position":0,"surname":"Morgan"}],"normalized_publisher":"Elsevier / Academic Press","normalized_subtitle":"A Beginner's Guide","normalized_title":"Geometric Measure Theory","publication_year":2009,"publisher":"Academic Press/Elsevier,","subtitle":"a beginner's guide /","title":"Geometric measure theory :","work_cluster_size":6,"work_dois":[],"work_isbns":["0125068514ACIDFREEPAPER","0125068573ACIDFREEPAPER","9780123744449HBKALKPAPER","0125068557"],"work_issns":[]}

- Ejemplo de registro comprensible:

```
□ ▼ object {21}
   □       work_id : 137439345858
   □   ▼ authors [1]
   □       ▼ 0   {2}
   □             given_name : Frank
   □             surname : Morgan
   □       id    : 1571958208353  2019-10-24T23:03:28.353Z
   □   ▼ isbns [1]
   □         0 : 9780123744449HBKALKPAPER
   □   ▼ issns [0]
   □         (empty array)
   □   ▼ journal_isbns [0]
   □         (empty array)
   □   ► journal_issns [0]
   □       match_count : 11
   □   ▼ matcher_pairs [2]
   □       ▼ 0   {4}
   □             logp : -72.83185
   □             match_count : 4
   □             matcher_id : 644245156575
   □         ▼ pair {2}
   □           ▼ a_tokens [1]
   □                 0 : morgan
   □           ▼ t_tokens [6]
   □                 0 : geometric
   □                 1 : measure
```

```
                    2  : theory
                    3  : beginner
                    4  : s
                    5  : guide
        ▼ 1   {4}
            logp : -42.507908
            match_count : 7
            matcher_id : 1348619757325  2012-09-26T00:35:57.325Z
            ▼ pair  {2}
                ▼ a_tokens [1]
                    0  : morgan
                ▼ t_tokens [3]
                    0  : geometric
                    1  : measure
                    2  : theory
    ▼ normalized_authors [1]
        ▼ 0   {3}
            given_name : Frank
            position : 0
            surname : Morgan
        normalized_publisher : Elsevier / Academic Press
        normalized_subtitle : A Beginner's Guide
        normalized_title : Geometric Measure Theory
        publication_year : 2009
        publisher : Academic Press/Elsevier,
        subtitle : a beginner's guide /

        title : Geometric measure theory :
        work_cluster_size : 6
    ▼ work_dois [0]
            (empty array)
    ▼ work_isbns [4]
            0  : 0125068514ACIDFREEPAPER
            1  : 0125068573ACIDFREEPAPER
            2  : 9780123744449HBKALKPAPER
            3  : 0125068557
    ▼ work_issns [0]
            (empty array)
```

2. Carpeta 'matches.json': 100 archivos

- [Ruta Sharepoint con archivos carpeta 'matches.json'.](#)
- Ejemplo registro nativo:

{"a_tokens":["sibelius"],"doc_id":429728657835782,"doc_work_order":1,"id":1022202468716,"logp":-27.51484,"m1":{"field":"Title","i1":1781,"i2":1781},"m2":{"field":"Author","i1":1790,"i2":1790},"matcher_id":386547214758,"pvalid":0.5162439942359924,"t_tokens":["selections"],"work_id":146028981013}

- Ejemplo de registro comprensible:

```
object ▶ t_tokens ▶
  ☐ ▼ object {11}
⋮ ☐       work_id : 146028981013
⋮ ☐       id     : 1022202468716  2002-05-24T01:07:48.716Z
⋮ ☐     ▼ a_tokens [1]
⋮ ☐         0 : sibelius
⋮ ☐       doc_id : 429728657835782  +015587-07-24T07:37:15.782Z
⋮ ☐       doc_work_order : 1
⋮ ☐       logp : -27.51484
⋮ ☐     ▼ m1    {3}
⋮ ☐         field : Title
⋮ ☐         i1    : 1781
⋮ ☐         i2    : 1781
⋮ ☐     ▼ m2    {3}
⋮ ☐         field : Author
⋮ ☐         i1    : 1790
⋮ ☐         i2    : 1790
⋮ ☐       matcher_id : 386547214758
⋮ ☐       pvalid : 0.5162439942359924
⋮ ☐     ▼ t_tokens [1]
⋮ ☐         0 : selections
```

3. Carpeta 'syllabi.json': 100 archivos

- [Ruta Sharepoint con archivos carpeta 'syllabi.json'.](#)
- Ejemplo registro nativo:

{"BASIC2015":8,"CONTROL":1,"HBCU":false,"TRIBAL":false,"UGPROFILE2015":3,"UNITID":"203155","WOMENS":false,"city":"Nelsonville","country_code":"US","country_name":"United States","field_code":"52.01/52.02/52.04/52.05/52.06/52.07/52.08/52.09/52.10/52.11/52.12/52

.13/52.15/52.16/52.17/52.18/52.19/52.20/52.21/52.99","field_name":"Business","field_score": 0.958269501035709,"four_year":false,"graduate":false,"grid_id":"grid.431645.1","id":40726597 8876245,"institution_id":11158325035011,"institution_name":"Hocking College","institution_url":"www.hocking.edu","research":false,"state_code":"US-OH","state_name":"Ohio","syllabus_probability":0.968961877045258,"two_year":true,"year":2 012}

- Ejemplo de registro comprensible:


```
object ▶ t_tokens ▶
   ▼ object {26}
       id       : 407265978876245  +014875-09-29T14:34:36.245Z
       BASIC2015 : 8
       CONTROL : 1
       HBCU  :  false
       TRIBAL :  false
       UGPROFILE2015 : 3
       UNITID : 203155
       WOMENS :  false
       city : Nelsonville
       country_code : US
       country_name : United States
       field_code : 52.01/52.02/52.04/52.05/52.06/52.07/52.08/52.09/52.10/52.11/52.12/52.13/52.15/52.16/52.17/52.1
                    8/52.19/52.20/52.21/52.99
       field_name : Business
       field_score : 0.958269501035709
       four_year :  false
       graduate :  false
       grid_id : grid.431645.1
       institution_id : 11158325035011  2323-08-06T06:43:55.011Z
       institution_name : Hocking College
       institution_url : www.hocking.edu
       research :  false
       state_code : US-OH
       state_name : Ohio
       syllabus_probability : 0.968961877045258
       two_year :  true
       year : 2012
```