

La cultura de los datos o cómo traducir las ideas a datos

Susanna Allés Torrent
University of Miami

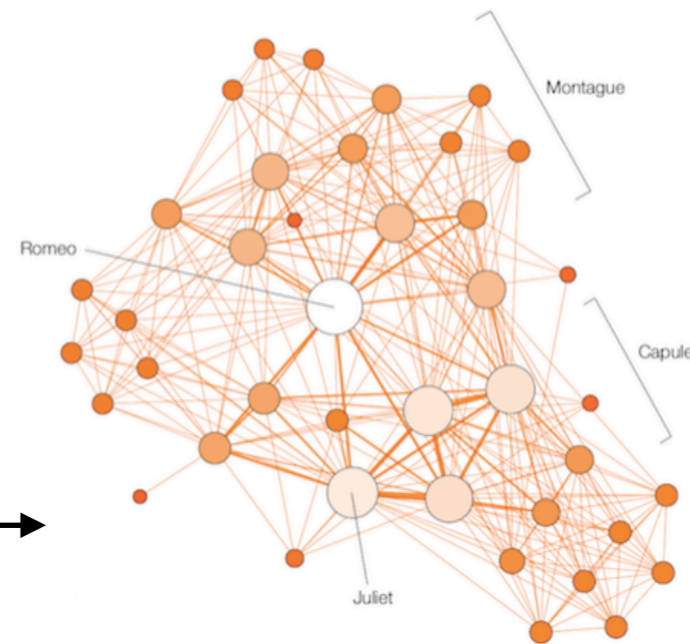
Rosario, 7-9 Noviembre 2018



I. Las humanidades digitales como disciplina

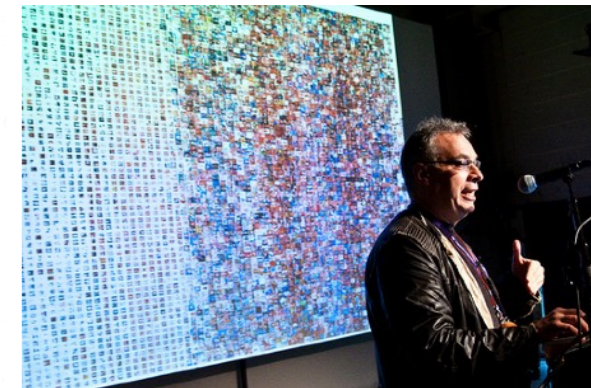
- Openness
- Collaboration
- Collegiality and Connectedness
- Diversity
- Experimentation

Lisa Spiro, “This Is Why We Fight”: Defining the Values of the Digital Humanities”, *Debates* 2012



ROMEO AND JULIET
Number of characters 41 | 37% Network density

Martin Grandjean (2015)



- **Permitir la manipulación y la explotación de los datos: gestionar, manipular, mapear, modelar,**
- Mejorar la enseñanza y el aprendizaje
- Compromiso social (“public engagement”)

Lisa Spiro, Why the Digital Humanities 2011



Gran diversidad en lo proyectos HD

EDICIONES DIGITALES

Homer Multitext
Vespasiano da Bisticci, Lettere
Vercelli Book
The World of Dante

BASES DE DATOS

Proyecto Boscán
Digital Scriptorium
Incunabula Short Catalogue

COLECCIONES DIGITALES

Europeana
Gallica

DIVERSIÓN

Periodic Table of Storytelling
Serendip-o-matic

NODOS DE INFORMACIÓN

Tapor
DiRT

DICCIONARIOS EN LÍNEA

Du Cange
Diccionario Griego Español

EXPOSICIONES VIRTUALES

Observatoire de Paris
Exhibitions

BIBLIOTECAS DIGITALES

Perseus Digital Library
Biblioteca Virtual Miguel de Cervantes
Women Writers Projects

CORPUS LINGÜÍSTICOS

CORDE
Corpus del español

MAPPING

Atlas of Early Printing
Mapping Colonial Americas
Publishing Project

VISUALIZACIONES

The Republic of Letters
Mapping Shakespeare's tragedies

LISTAS PROYECTOS HD

DH Commons
GO:DH Projects
TEI Projects

Una posible definición

Digital Humanities research and teaching takes place at the **intersection of digital technologies and humanities**. DH aims to produce and use applications and models that make possible new kinds of teaching and research, both in the humanities and in computer science (and its allied technologies). DH also studies the impact of these techniques on **cultural heritage, memory institutions, libraries, archives and digital culture**.

La investigación y la enseñanza en Humanidades Digitales tiene lugar en **la intersección entre las tecnologías digitales y las humanidades**. Las HD aspiran a producir y usar aplicaciones y modelos que hagan posible nuevos modos de **enseñar y de investigar**, tanto en las humanidades como en las ciencias de la computación (y sus tecnologías aliadas). Las HD también estudian el impacto de estas tecnologías en el campo del **patrimonio cultural, las instituciones, las bibliotecas, los archivos y la cultura digital**

👉 Melissa Terras y su equipo (UCL) apuestan por una *concepción* muy general, donde tanto el enfoque práctico como el enfoque teórico tienen cabida.

Evolución de las HD

1949 - 1970 INICIOS:

- Concordancias: **Roberto Busa** empieza el *Index Thomisticum*, con IBM.
- COCOA concordance program
- Análisis cuantitativos sobre el estilo de los autores
- *Translation Machine* (Traducción automática)

1970-1980 CONSOLIDACIÓN:

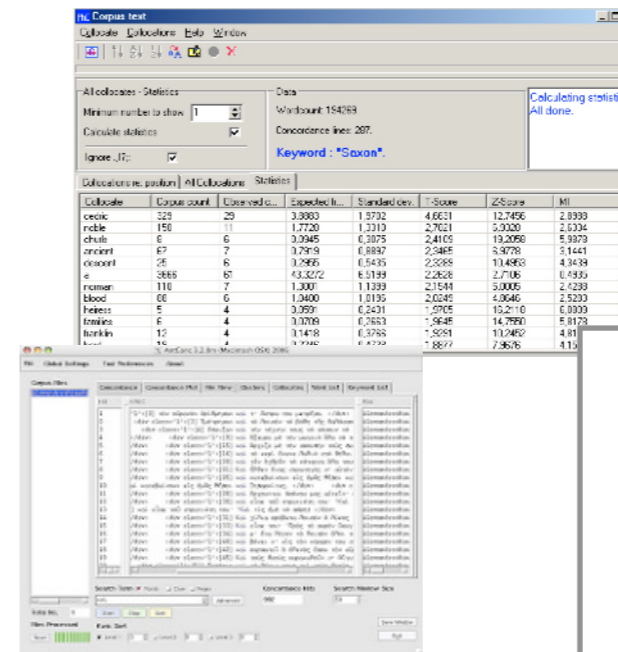
- Aparición de asociaciones internacionales, conferencias, revistas (*Literary and Linguistic Computing*)
- Énfasis en el **archivado y la preservación electrónica** de los textos: Oxford Concordance Program (OCP); Oxford Text Archive (OTA, 1976)
- *Thesaurus Linguae Graecae* (TKG)

1980 - 1990 NUEVOS DESARROLLOS:

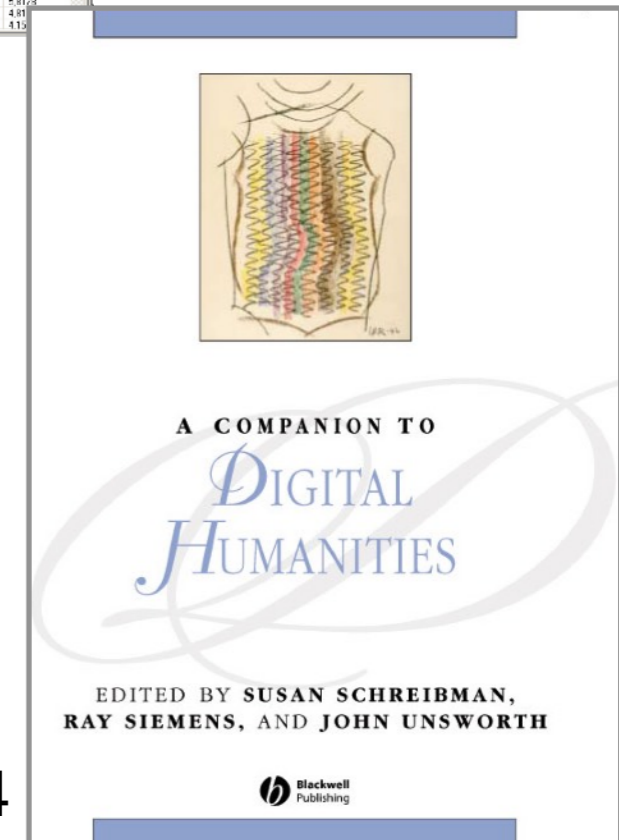
- Llegada del ordenador personal
- Programas OS DOS (las búsquedas remplazan las concordancias)
- Llegada del MAC: interfície gráfica
- **Text Encoding Initiative** 1986 (SGML)

1990-2004 - LA ERA DE INTERNET:

- Llegada del internet y exploradores como Mosaic (1993)
- Imágenes digitales (Historia del arte)
- Multimedia (Musicología, Estudios de cine)
- Geographic Information Systems (ArcGIS, 1999)
- Ediciones digitales académicas



2004



Las asociaciones HD:

ach

The Association for
Computers and the
Humanities

eadh EUROPEAN ASSOCIATION
FOR DIGITAL HUMANITIES



Asociación Argentina de
Humanidades Digitales

2013



Humanidades
Digitales
Hispánicas

2012

Red HD
RED DE HUMANIDADES DIGITALES

2011

HDU Asociación Uruguaya de Humanidades Digitales



RC
HD

RED COLOMBIANA
DE HUMANIDADES
DIGITALES

2017

Humanidades Digitales / Cuba



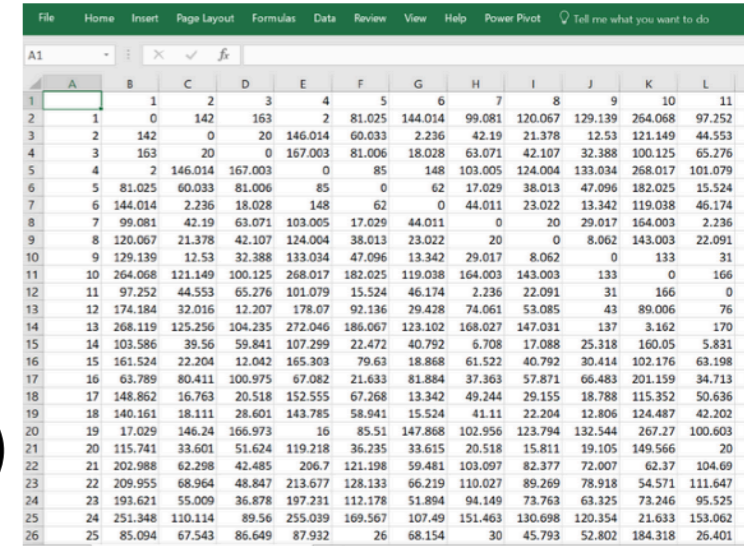
2017

II. La centralidad de los datos en HD

- Históricamente, los datos concebidos como información procesable constituyen el núcleo de la disciplina, desde Busa hasta hoy
- ¿Qué problemas afronta el humanista a la hora de trabajar con datos?
- Tendencia global e inevitable: Profesionales dedicados al “data management” (creación, recolección, estructuración, almacenaje, mantenimiento, venta).

¿Qué son los datos?

“Los datos en las humanidades podrían ser considerados una abstracción digital, construida selectivamente, y procesable por una máquina que respresenta algunos aspectos de un determinado objeto de investigación humanística” (Schöch 2013)



	A	B	C	D	E	F	G	H	I	J	K	L
1												
2	1	0	142	163	2	81.025	144.014	99.081	120.067	129.139	264.068	97.252
3	2	142	0	20	146.014	60.033	2.236	42.19	21.378	12.53	121.149	44.553
4	3	163	20	0	167.003	81.006	18.028	63.071	42.107	32.388	100.125	65.276
5	4	2	146.014	167.003	0	85	148	103.005	124.004	133.034	268.017	101.079
6	5	81.025	60.033	81.006	85	0	62	17.029	38.013	47.096	182.025	15.524
7	6	144.014	2.236	18.028	148	62	0	44.011	23.022	13.342	119.038	46.174
8	7	99.081	42.19	63.071	103.005	17.029	44.011	0	20	29.017	164.003	2.236
9	8	120.067	21.378	42.107	124.004	38.013	23.022	20	0	8.062	143.003	22.091
10	9	129.139	12.53	32.388	133.034	47.096	13.342	29.017	8.062	0	133	31
11	10	264.068	121.149	100.125	268.017	182.025	119.038	164.003	143.003	133	0	166
12	11	97.252	44.553	65.276	101.079	15.524	46.174	2.236	22.091	31	166	0
13	12	174.184	32.016	12.207	178.07	92.136	29.428	74.061	53.085	43	89.006	76
14	13	268.119	125.256	104.235	272.046	186.067	123.102	168.027	147.031	137	3.162	170
15	14	103.586	39.56	59.841	107.299	22.472	40.792	6.708	17.088	25.318	160.05	5.831
16	15	161.524	22.204	12.042	165.303	79.63	18.868	61.522	40.792	30.414	102.176	63.198
17	16	63.789	80.411	100.975	67.082	21.633	81.884	37.363	57.871	66.483	201.159	34.713
18	17	148.862	16.763	20.518	152.555	67.268	13.342	49.244	29.155	18.788	115.352	50.636
19	18	140.161	18.111	28.601	143.785	58.941	15.524	41.11	22.204	12.806	124.487	42.202
20	19	17.029	146.24	166.973	16	85.51	147.868	102.956	123.794	132.544	267.27	100.603
21	20	115.741	33.601	51.624	119.218	36.235	33.615	20.518	15.811	19.105	149.566	20
22	21	202.988	62.298	42.485	206.7	121.198	59.481	103.097	82.377	72.007	62.37	104.69
23	22	209.955	68.964	48.847	213.677	128.133	66.219	110.027	89.269	78.918	54.571	111.647
24	23	193.621	55.009	36.878	197.231	112.178	51.894	94.149	73.763	63.325	73.246	95.525
25	24	251.348	110.114	89.56	255.039	169.567	107.49	151.463	130.698	120.354	21.633	153.062
26	25	85.094	67.543	86.649	87.932	26	68.154	30	45.793	52.802	184.318	26.401

Tipos de datos:

- Datos discretos: anotación binaria de 0 y 1
- Estructuras lineares: (ej. hoja de cálculo)
- Estructura jerárquicas (ej. XML)
- Estructura multi-relational (ej. nodos de una red)

Organización: Datos estructurados (base de datos), no estructurados (texto plano) y semi-estructurados (XML files)

Naturaleza: Datos y Metadatos

```
<div type="verso">
  <pb facs="../images/Richter/06_2.jpg"
  <div type="mensaje">
    <opener>
      <salute>Queridísimo
      José María. </salute>
    </opener>
    <p>Desde Galicia (Bayona)
    saludo al ilustre conde
    de este mar y este
    caserío, hoy con más
    motivo que nunca
    pues este título está
    ya en pleno reino
    poético.
    <closer>
      <salute>Adiós José María.
      Aquí he venido con
      la Barraca.
      <salute>Un abrazo muy
      fuerte de tu siempre
    </signed>Federico</signed>
    </closer>
  </div>
</div>
```

Datos en las humanidades



TaDiRAH - Taxonomía sobre Actividades de investigación digital en humanidades

<Actividades de investigación> ▼

Almacenamiento

Enriquecimiento

Preservación

Meta-Actividades ▶

Captura ▶

Creación ▶

Análisis ▶

Interpretación ▶

Difusión ▶

<Objetos de estudio> ▶

<Técnicas de análisis> ▼

Alineación de secuencias

Almacenamiento

Análisis de clusters

Análisis de colocaciones

- Iniciativa para definir una taxonomía académica
- Vocabulario en español
- Adaptable a CMS, como Wordpress

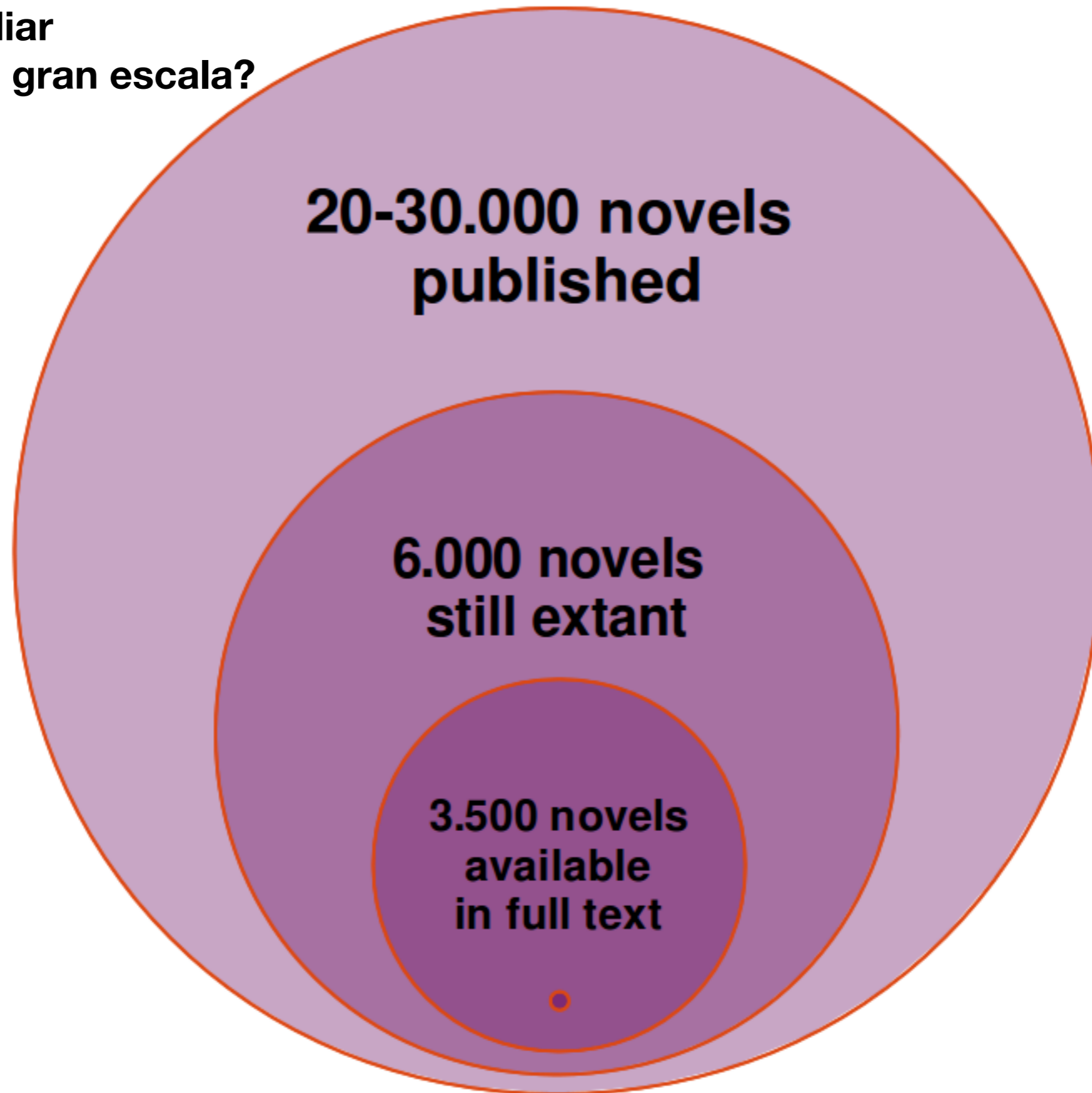
La dificultad de materializar los datos

- Claridad de conceptos humanísticos **vs** Complejidad de materializar las ideas a datos
- Dimensión interpretativa de los datos en humanidades
- Resistencia de los investigadores en estudios literarios y culturales a no hablar de “datos” cuando conciben sus temas de estudio (Schöch 2013)
- Falta de datos en humanidades.
- “La información que tenemos sobre el pasado es, en la mayoría de los casos, fragmentaria. Siempre hay grandes cantidades de datos que o simplemente faltan o que no podemos identificar.” (Marche 2012)
- A medida que aparecen más datos para la investigación en las humanidades, se facilita nuestra investigación.



Internet Archive

¿Cómo estudiar la literatura a gran escala?



Digitization and availability of British Nineteenth-Century Novels. Image published under a CC-BY license.

**Ejemplos: ¿cómo traducir
una idea a datos?**

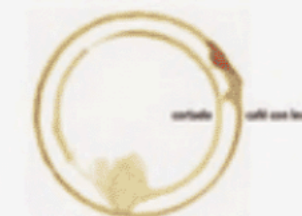
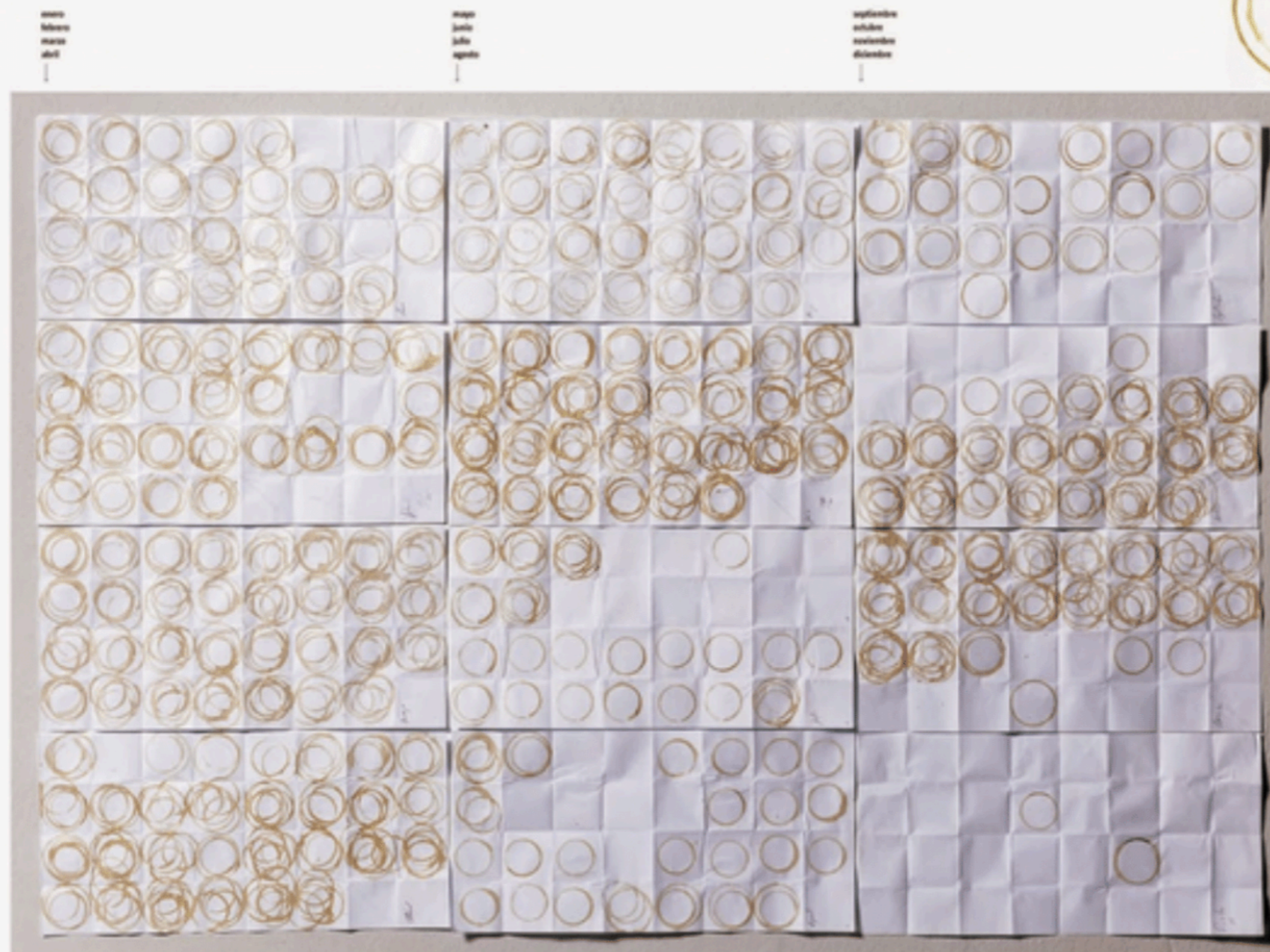
Documental: Jaime Serra III

Jaime Serra Pidal Periodista y artista multidisciplinar. En los últimos 25 años ha dirigido la infografía de algunos de los principales medios de Europa, EE.UU. y Latinoamérica. En su trabajo artístico utiliza los métodos y el lenguaje propio de la infografía científica para tratar temáticas pertenecientes a su propio mundo personal. Su obra se ha expuesto en Estados Unidos, Argentina, México y España. Reconocido por la prestigiosa Society for News Design como el infógrafo más influyente del periodo 1990-2013. Es, desde octubre del 2013, responsable de las infografías de 'La Vanguardia'.

Tres historias verdaderas

Los datos son materia prima, en bruto no tienen utilidad. Deben ser procesados y transcritos a un lenguaje visual para ser comprendidos. La herramienta que dibuja la información que contienen es la infografía. Una herramienta de comunicación habitualmente asociada a la divulgación científica pero que, como cualquier otra herramienta de comunicación humana, puede narrar hechos o verdades, realidad o ficción. Los datos colectivos, recogidos con las actuales tecnologías capaces de predecir de la intermediación del individuo, son objetivos. Precisamente por eso no pueden aportar nada sobre el individuo: no pueden ser verdad. La verdad contiene un componente de interpretación que permite que cada individuo disponga de la suya propia. Son los datos aplicados a lo que hace una persona concreta, en un momento dado concreto, sobre un tema concreto lo que puede resultar revelador para otro individuo. Porque en la intimidad todos los seres humanos, por analogía u oposición, nos encontramos.

PATROCINADO POR

Café a diario

La relación que mantengo con el café es dual: por una parte puedo llegar a consumirlo de modo compulsivo, casi como una adicción, por otra parte la reacción de mi organismo a los lacteos -no me gusta solo- me obliga a reducir o abandonar por completo su consumo durante días.

Durante el año 2013 cuantifiqué las tazas de café que tomaba a partir de las manchas que dejaban sobre hojas de papel.

Del número total de tazas existen dos tipologías: café cortado y café con leche, usando un modelo de taza de diferentes diámetros para cada caso. Con el fin de otorgarle mayor exactitud a la dotificación, utilicé las mismas tazas todo el año.

Para la visualización utilicé una hoja de 25 cm x 50 cm para cada mes. Como debía transportar las hojas a todas partes era necesario doblarlas. La dimensión de la hoja elegida permitía que, doblada cinco veces sobre su mitad y nuevamente desplegada, dibujase treinta y dos cuadrados de 8 cm cada uno, una forma ligeramente mayor que el diámetro de la taza de café con leche. A modo de calendario, cada cuadrado corresponde a un día, reservando el último para el nombre del mes y otras anotaciones.

<http://visualoop.com/28792/portfolio-of-the-week-jaime-serra>

Alberto Cairo • University of Miami • www.thefunctionalart.com • Twitter: @albertocairo

Writing Lives

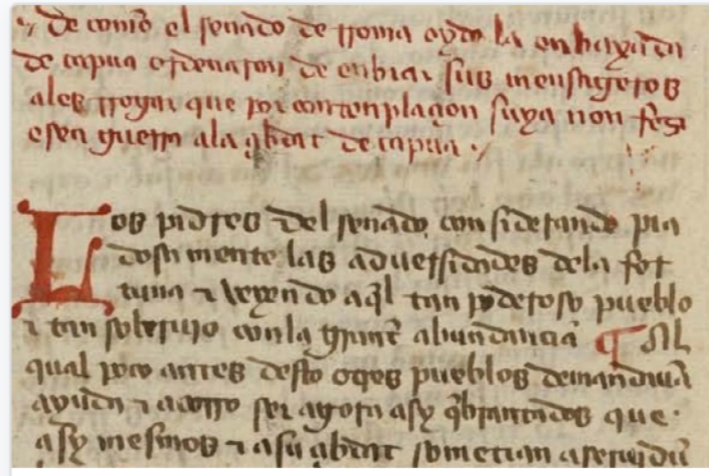
A Biographical Corpus of Medieval and Early Modern Iberia



Biographers / Authors

Access

Total:



Corpus of Biographees

Access

Total:

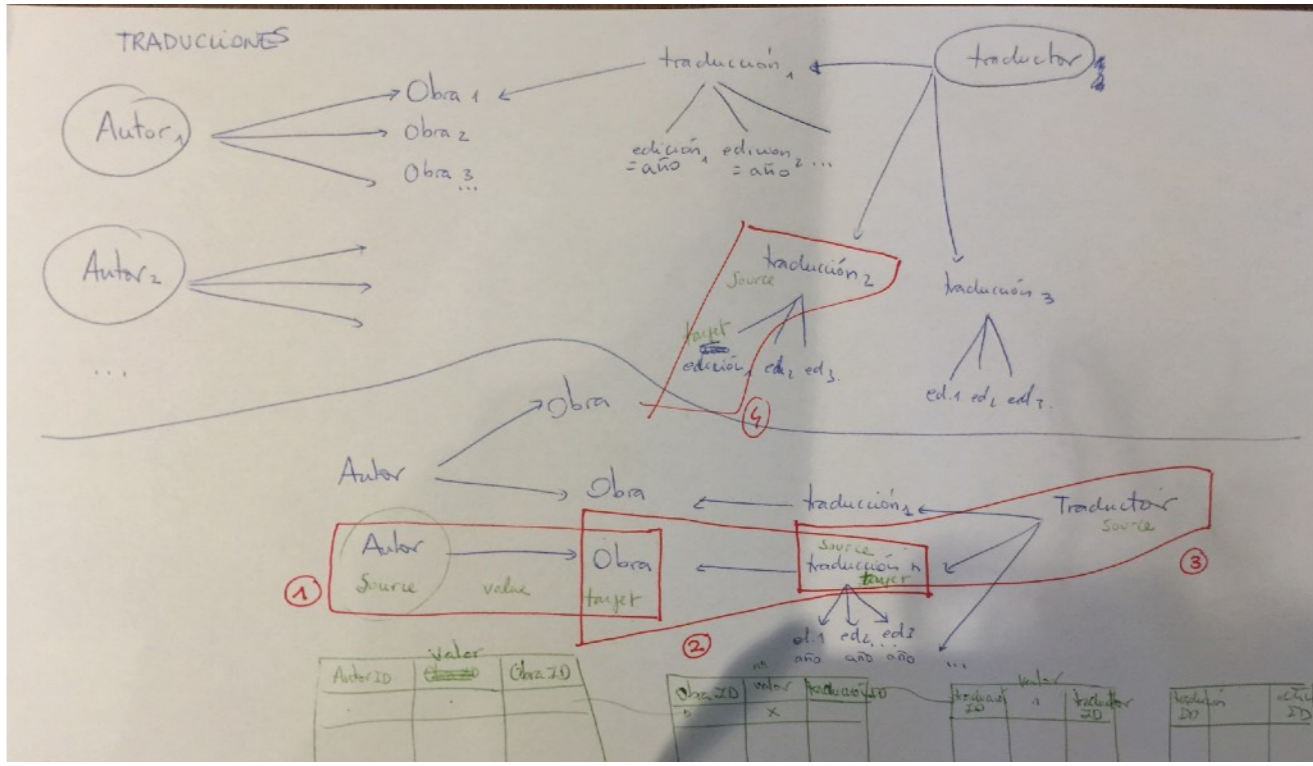
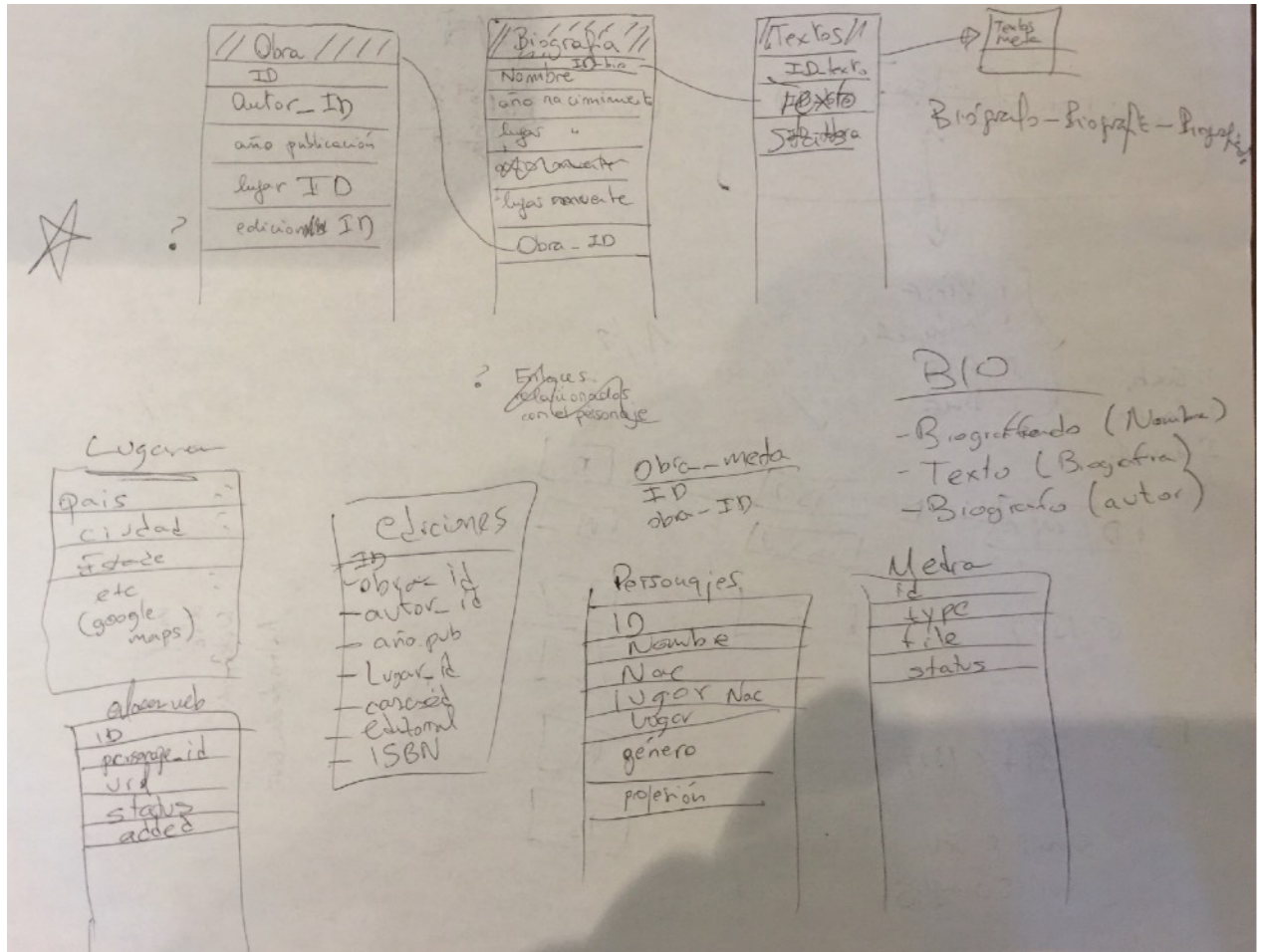
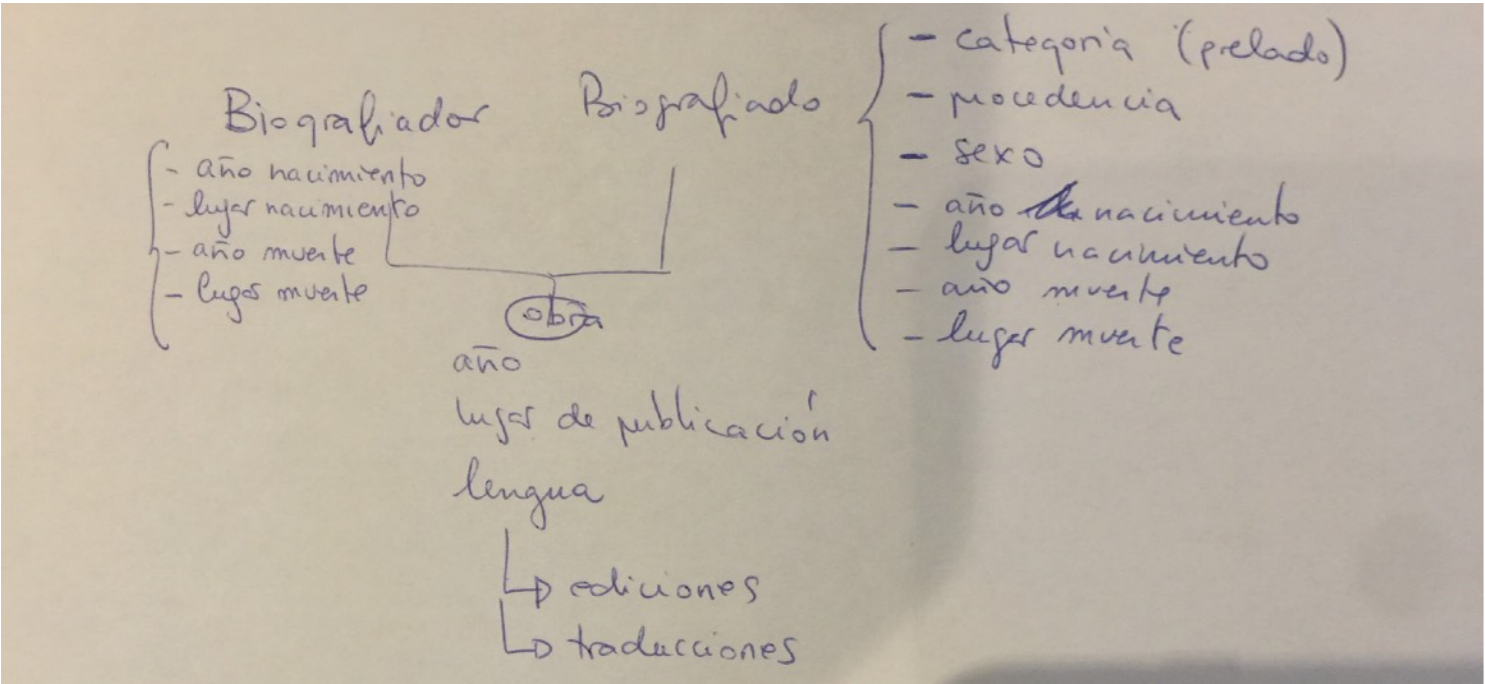


Portraits

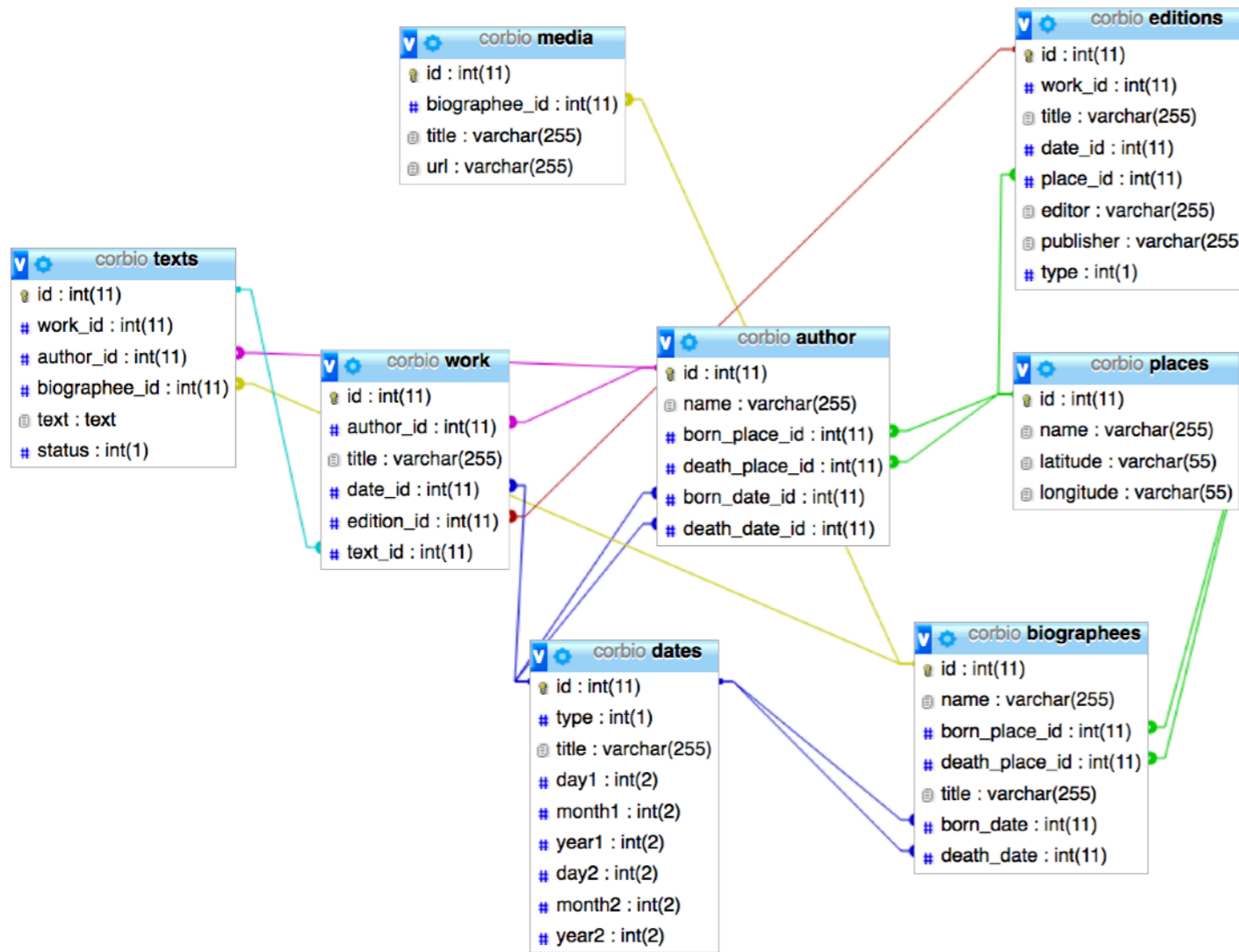
Access

Total:

Desde los primeros diseños...



A la implementación de la DB...



phpMyAdmin interface showing the database structure for 'corbio'.

Table	Action	Rows	Type	Collation	Size	Overhead
author	Browse Structure Search Insert Empty Drop	2	InnoDB	latin1_swedish_ci	112 KiB	-
biographees	Browse Structure Search Insert Empty Drop	54	InnoDB	latin1_swedish_ci	80 KiB	-
dates	Browse Structure Search Insert Empty Drop	100	InnoDB	latin1_swedish_ci	48 KiB	-
editions	Browse Structure Search Insert Empty Drop	0	InnoDB	latin1_swedish_ci	48 KiB	-
media	Browse Structure Search Insert Empty Drop	0	InnoDB	latin1_swedish_ci	32 KiB	-
places	Browse Structure Search Insert Empty Drop	5	InnoDB	latin1_swedish_ci	32 KiB	-
texts	Browse Structure Search Insert Empty Drop	0	InnoDB	latin1_swedish_ci	64 KiB	-
work	Browse Structure Search Insert Empty Drop	0	InnoDB	latin1_swedish_ci	80 KiB	-
8 tables	Sum	161	InnoDB	latin1_swedish_ci	496 KiB	0 B

Críticas y matices a los “datos”

Fetichismo de los datos que esconde una falta de base teórica.

Grandes subvenciones a proyectos que, en campos como la minería de datos textuales, no han hecho más que constatar lo obvio (Andrew Smith, “The Promise of Digital Humanities”, *The Past Speaks*. 21-08-2011)

Recelo ante el aparente empirismo de la investigación basada en “datos”, pues obvia el contexto del objeto. (Schöch 2013)

Del lat. *datum* “lo que se da” “lo que se ha dado”. Johanna Data propone el término “Capta”, “lo que se ha recogido”. (Drucker 2011)

Los datos son siempre “proporcionados” (capta), manufacturados y creados (Owens 2011)

IV. Big Data”

¿Existen los “big data” en las humanidades?

¿Problema? La investigación humanística ha buscado siempre la exclusividad, el individualismo, lo que sobresale

Definiciones:

“Storage of **large volume of data**; and analyzing this data quantitatively and visually to find **patterns, establish laws and predict conduct.**”

(Ward & Barker, 2013)

“Big data refers to things one can do at a **large scale** that cannot be done at a smaller one, to extract new insights or create new forms of value.”

(Mayer-Schönberger & Cukier 2013)

Doug Laney establece 3 **calidades: volume, velocity and variety.** A la que se ha añadido otra 4^a: **Veracity.**

(NIST, Gartner, IBM).

Big data is a term describing the storage and analysis of large and or complex data sets using a series of techniques including, but not limited to: NoSQL, MapReduce and machine learning. (Ward & Barker 2013)

Big Data vs Smart Data

SMALL DATA	BIG DATA
Datos (semi-)estructurados	Datos no estructurados
Datos explícitos y enriquecidos (markup, annotations and metadata)	“Raw data” (datos sin formato)
Clean Data datos “limpios”	“Messy Data”
Small (requiere la intervención humana, tiempo)	Big
Importancia del modelado de datos (esquemas, base de datos relacionales)	NoSQL data

Only **smart big data** enables intelligent quantitative methods. (Schöch 2013)

¿Big Data en español?

Real Academia Española - Corpus Diacrónico del Español (CORDE)

Consulta:

Criterios de selección:

Autor:

Obras:

Cronológicos:

Medio: (Todos) Libros Periódicos Revistas Miscelánea Orales

Geográficos: (Todos) Argentina Bolivia Chile Colombia Costa Rica

Tema: (Todos) 11.- Lírica 12.- Narrativa 121.- Breve 1211.- Relato breve tradicional 1212.- Relato breve culto

Buscar Limpiar

Consulta CREA [Nómina de autores y obras](#) [Lista de frecuencias](#) [Cómo citar el CORPUS](#) [Ayuda.](#)

ISSN 2340-5651 Corpus diacrónico del español

The CLIGS textbox

DOI: [10.5281/zenodo.597430](https://doi.org/10.5281/zenodo.597430)

This repository contains several text collections made available by the CLIGS junior research group (see <http://cligs.hypotheses.org>). As of May 2018, these are:

- *Collection of 19th Century Spanish-American Novels (1880-1916)*, edited by Ulrike Henny-Krahmer: 24 novels
- *Corpus of Spanish Novels from 1880-1940*, edited by José Calvo Tello: 39 novels
- *Collection de nouvelles françaises du dix-neuvième siècle*, edited by Stefanie Popp: 28 novellas
- *Collection de romans français du dix-neuvième siècle*, edited by Christof Schöch and Stefanie Popp: 36 novels.
- *Collection de pièces de théâtre français du dix-septième siècle*, edited by Christof Schöch: 100 plays.
- *Collection of 19th Century Portuguese Novels (1840-1910)*, edited by Ulrike Henny-Krahmer: 30 novels
- *Collection of Italian Short Stories and Novellas (1880s-1920s)*, edited by Ulrike Henny-Krahmer: 90 short stories and novellas
- *Corpus of Spanish Short Stories from 1880-1940*, edited by José Calvo Tello: 20 collections of short stories, 302 single short stories
- *Collection of Italian Novels (1850 and 1915)*, edited by Katrin Betz and Christof Schöch: 21 novels

ADSO

Distant reading approach to Golden Age Spanish Sonnets

■ AgustínDeSalazarYTorres	Sonetos de Salazar y Torres
■ AntonioEnriquezGomez	Sonetos de Enriquez Gómez
■ AntonioHurtadoDeMendoza	Sonetos de Antonio Hurtado de Mendoza
■ ArgensolaBartolome	Sonetos de Bartolomé Leonardo de Argensola
■ ArgensolaLupercio	Sonetos de Lupercio Leonardo de Argensola
■ BaltasarDeAlcazar	Sonetos de Baltasar de Alcázar
■ BernardinoDeRebolledo	Sonetos de Bernardino de Rebolledo
■ BocangelyUnzueta	Sonetos de Gabriel Bocángel y Unzueta
■ CancerYVelasco	Sonetos de Jerónimo de Cáncer y Velasco
■ Cervantes	Sonetos de Miguel de Cervantes
■ CristobalDeVirues	Sonetos de Cristóbal de Virués
■ DiegoHurtadoDeMendoza	Sonetos de Diego Hurtado de Mendoza
■ DiegoRamirezPagan	Sonetos de Diego Ramírez Pagán
■ DiegoXimenezAyllon	Sonetos de Diego Ximénez de Ayllón
■ EstebanManuelDeVillegas	Sonetos de Esteban Manuel de Villegas
■ FcodeBorja	Sonetos de Francisco de Borja y Aragón

Colección de Textos Caballerescos Hispánicos



Digital Library of Old Spanish Texts

IMPACT-es diachronic corpus

Produced by: [Universitat d'Alacant/University of Alicante](#) and [Fundación Biblioteca Virtual Miguel de Cervantes](#)

IV. El arduo proceso de trabajar con datos informáticos...

<Actividades de investigación> ▼

- 3 Almacenamiento
- 2 Enriquecimiento
- 6 Preservación

Meta-Actividades ▶

- 1a Captura ▶
- 1b Creación ▶
- 4 Análisis ▶
- 5 Interpretación ▶

Difusión ▶

TaDiRAH - Taxonomía sobre Actividades de investigación digital en humanidades

1. Captura o creación de datos

- Tienen implicaciones teóricas, metodológicas, sociales y técnicas (Borgman 2015)
- Cada uno de los datos que recogemos pone dilemas de interpretación diferentes.
- A diferencia de los datos en ciencias, en humanidades los datos son complejos, con identidades superpuestas, y abiertos a interpretación (Levi 2013)

la naturaleza *vaga* (“fuzzy”) de los datos humanísticos; el hecho que la mayoría de humanistas trabaje con recursos textuales, visuales o audio que no se prestan a abstracciones geométricas o matemáticas; y el persistente aislamiento disciplinar ha impedido **colaboraciones serias entre humanistas y** geógrafos, urbanistas, arquitectos... y otros profesionales en el campo de los datos espaciales

Presner & Shepard 2017

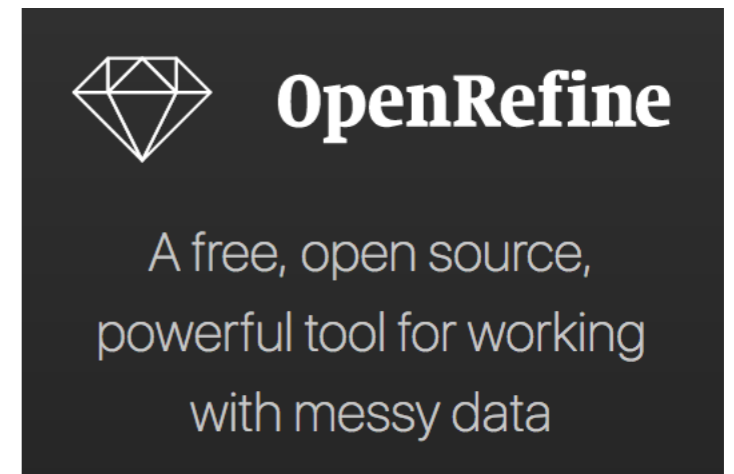
Modelado de datos

- La importancia del modelado de los datos condiciona el resultado de los datos
- El proceso de modelado está influenciado por elementos contextuales como hipótesis de base, supuestos teóricos, métodos científicos, formatos y tecnologías.
(Rojas Castro)

2. Limpiar los datos

Algunos ejemplos:

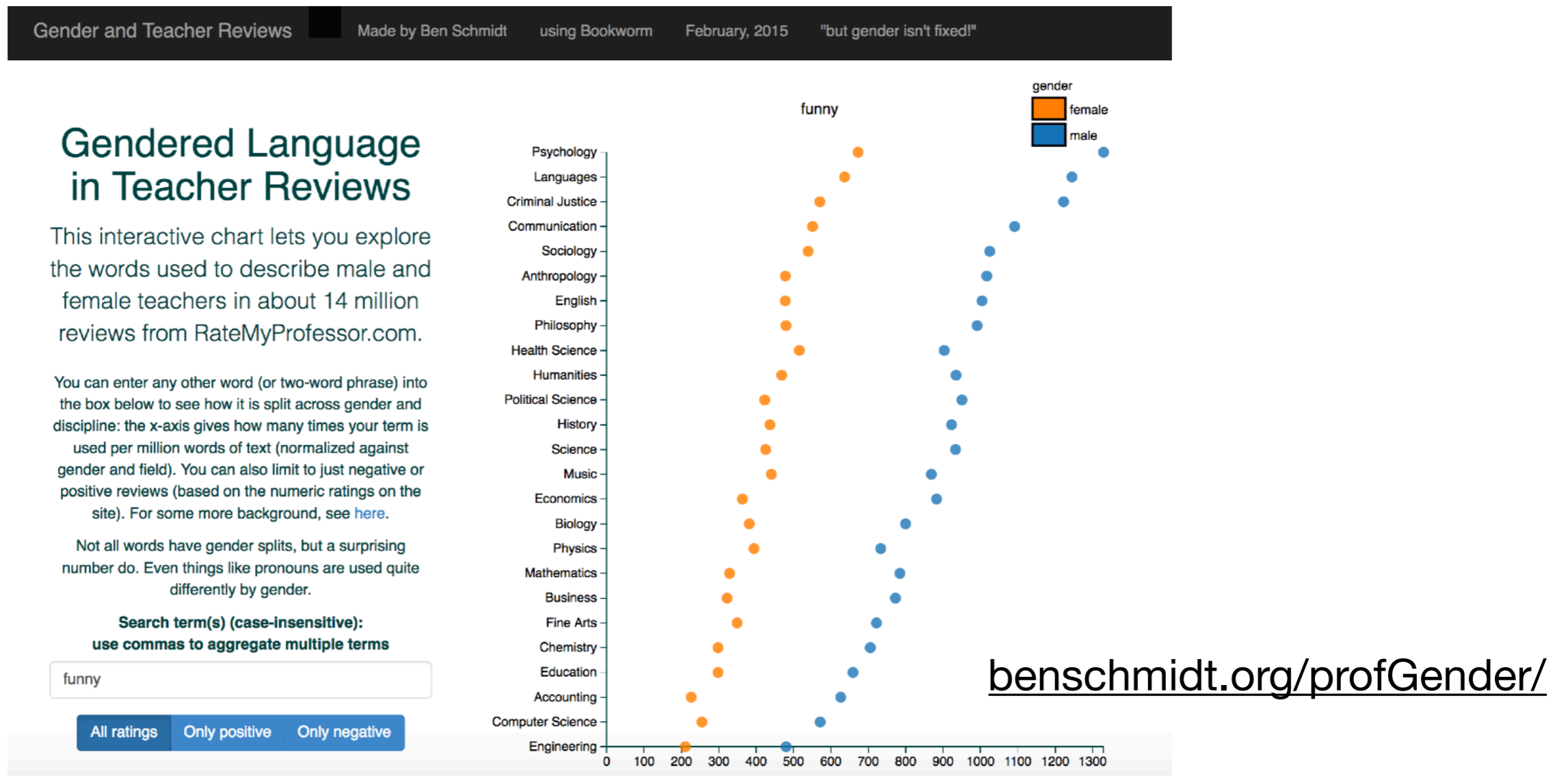
- Fechas
- Estandarización de abreviaturas y ortografías
- Limpiar un OCR
- Unificar geodatos (ej. todas las direcciones iguales)
- Preparación de un corpus (dividir textos en ficheros diferentes y asociar metadatos, etc.)



Más info: [DH at Berkeley](#)

3. Análisis de los datos

Minería de textos (Text Mining): Analisis cuantitativos y estadísticos (vs Close reading)



+ Procesamiento del Lenguaje Natural

3. Análisis de los datos

Análisis cuantitativos y estadísticos (vs Close reading)

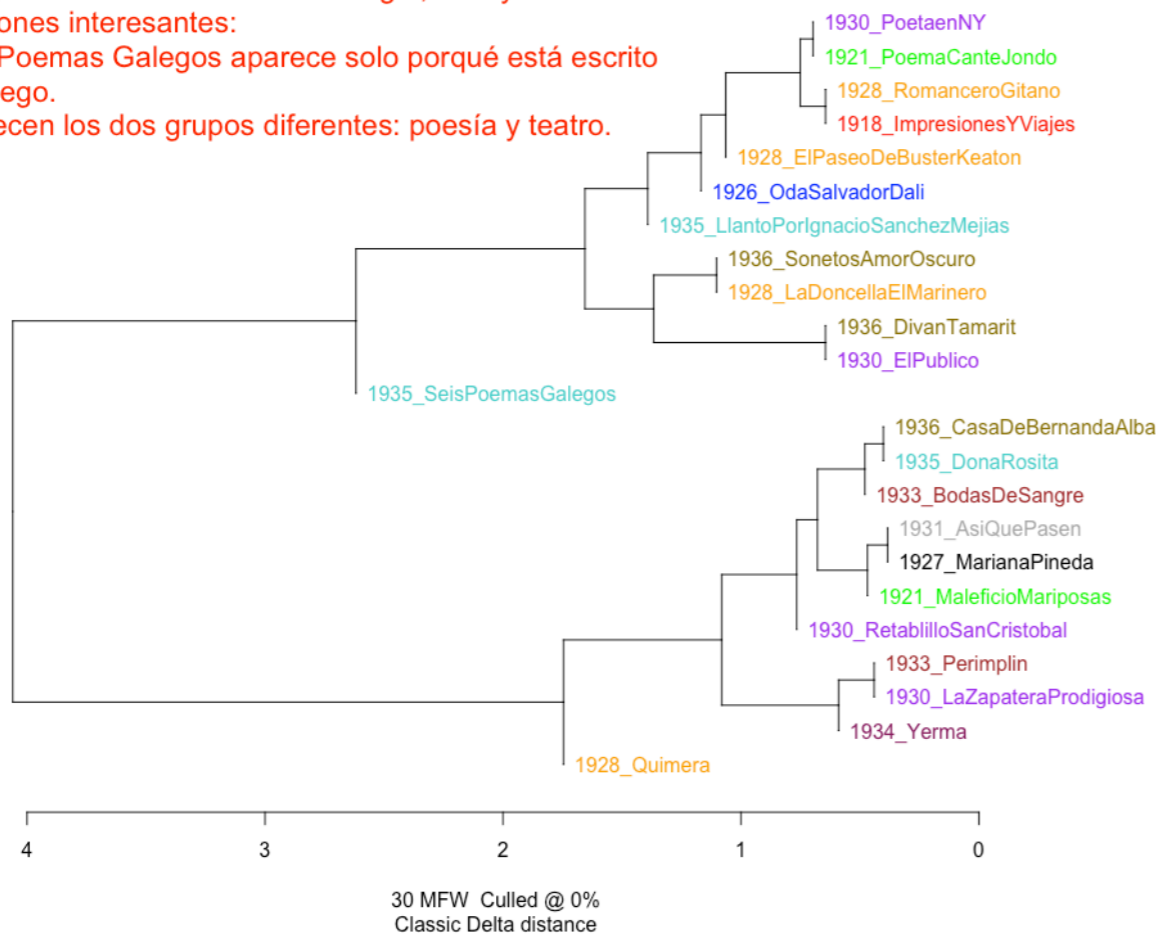
Estilometría

Lorca_Corpus
Cluster Analysis

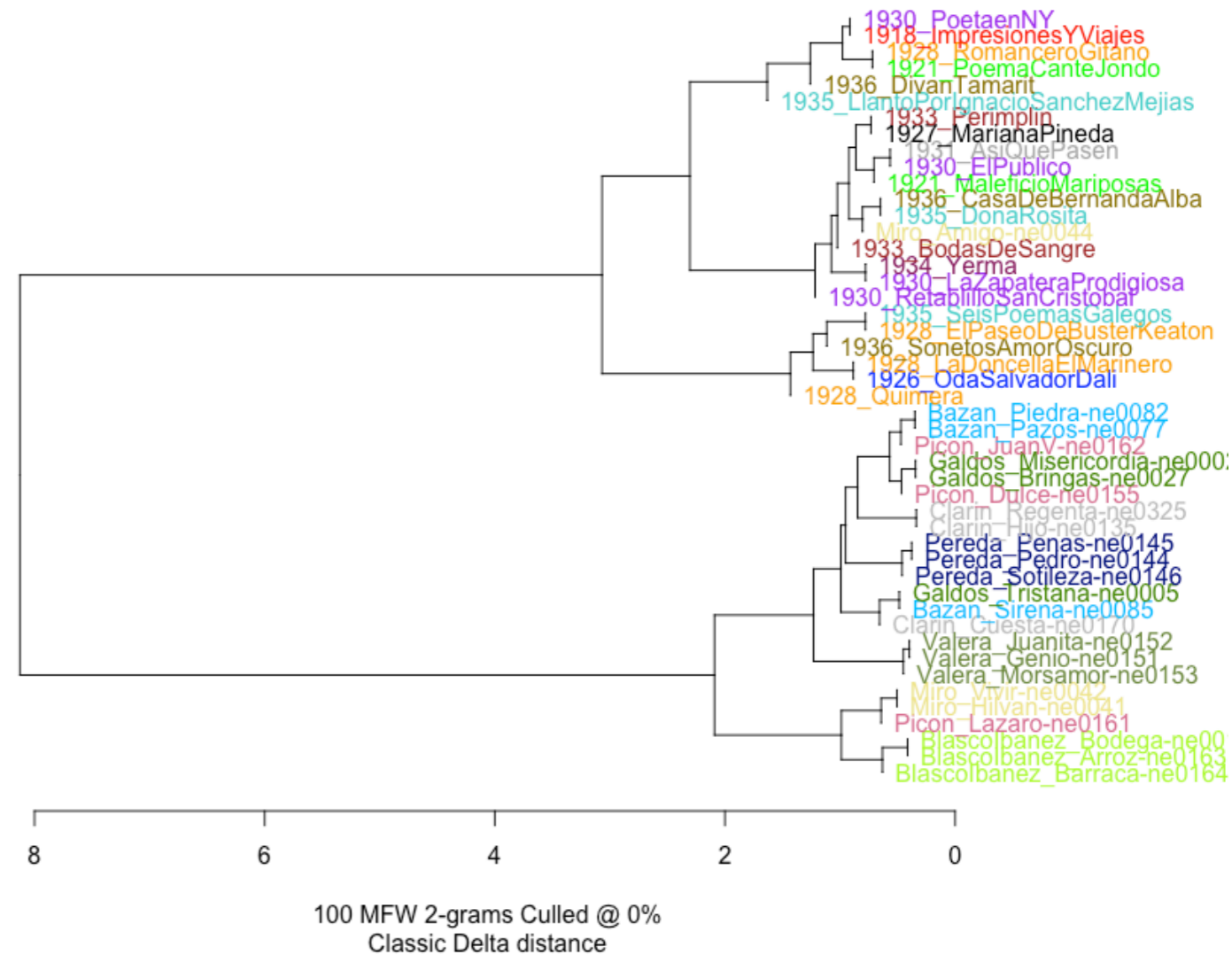
``stylo()` > `Features` > `MFW Settings`, Min. y Max. 30`

Cuestiones interesantes:

- Seis Poemas Galegos aparece solo porque está escrito en gallego.
- Aparecen los dos grupos diferentes: poesía y teatro.

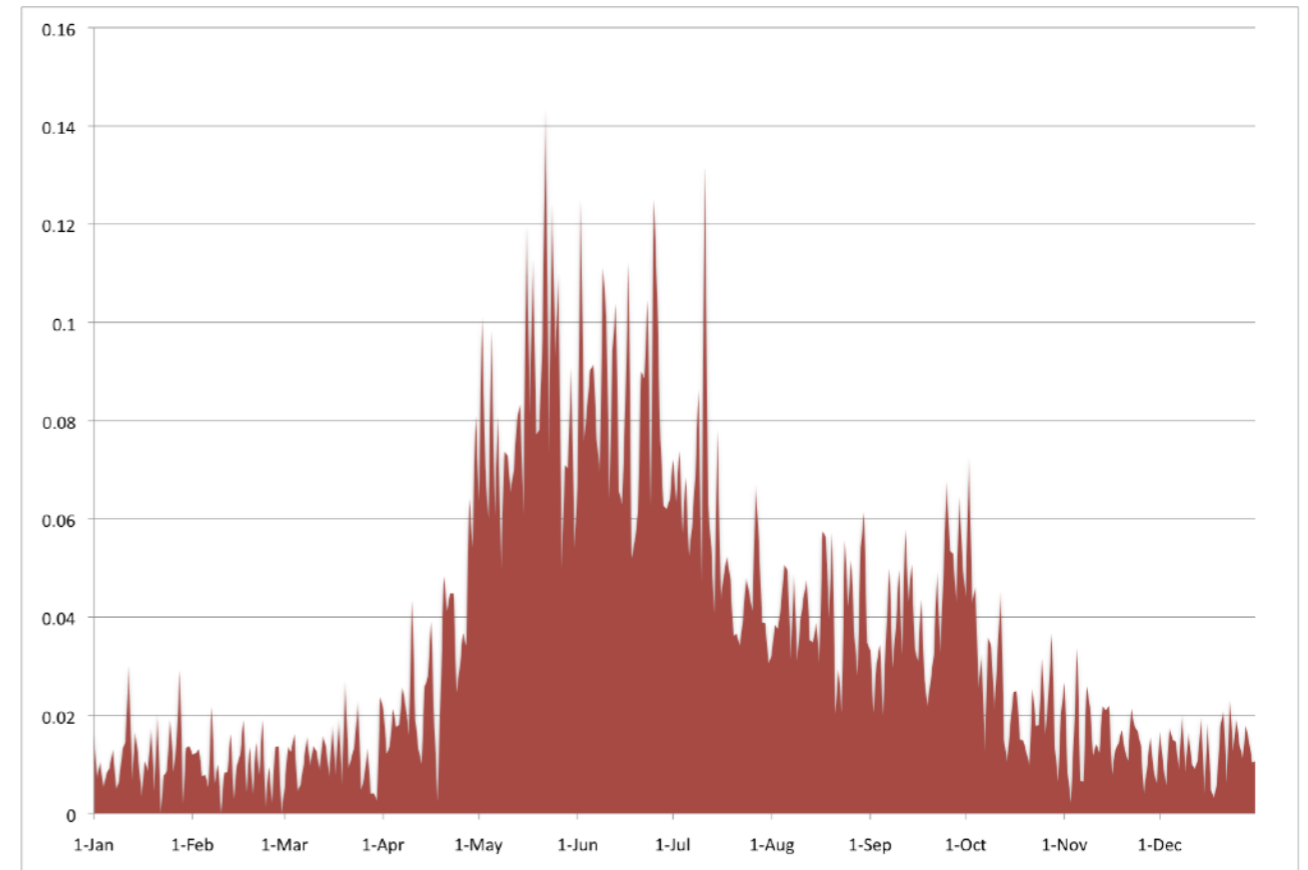
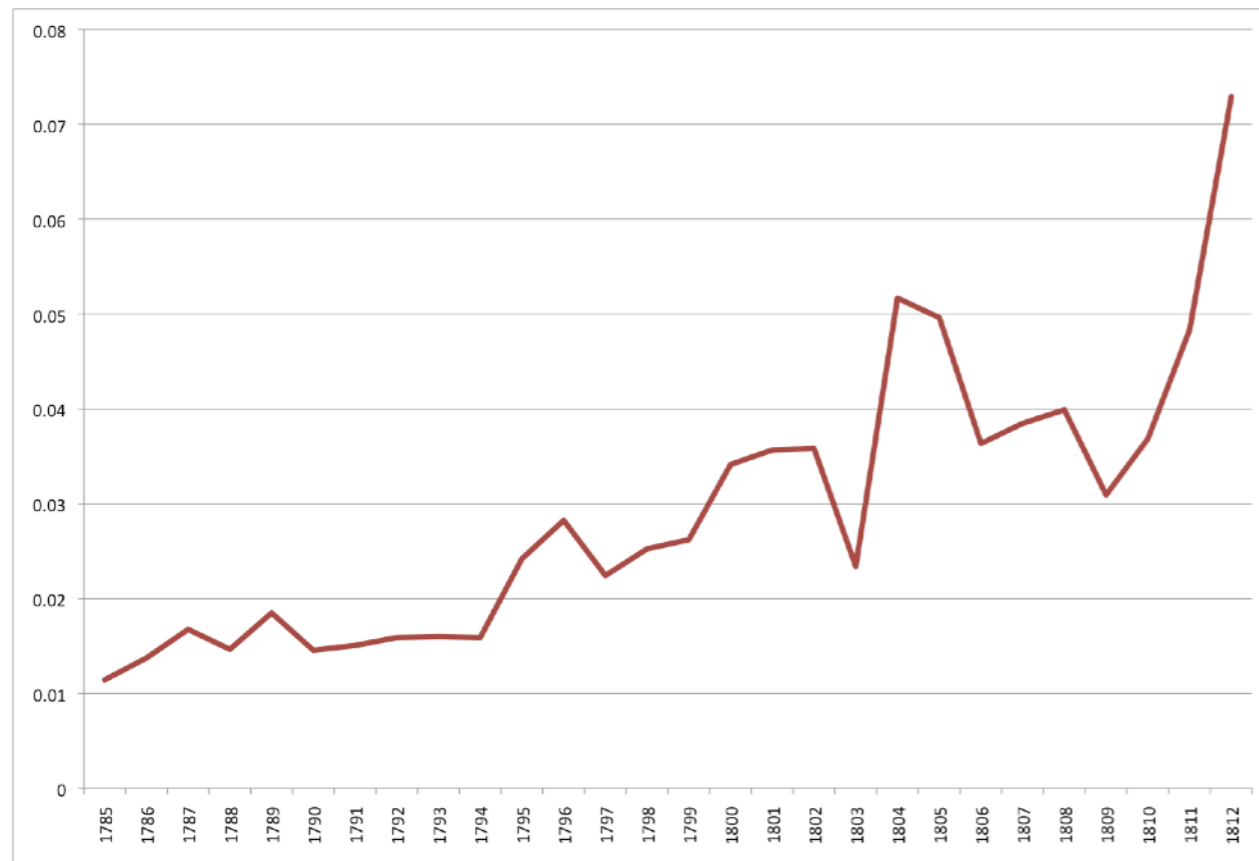


SPA322_Intro_DH
Cluster Analysis



3. Análisis de los datos

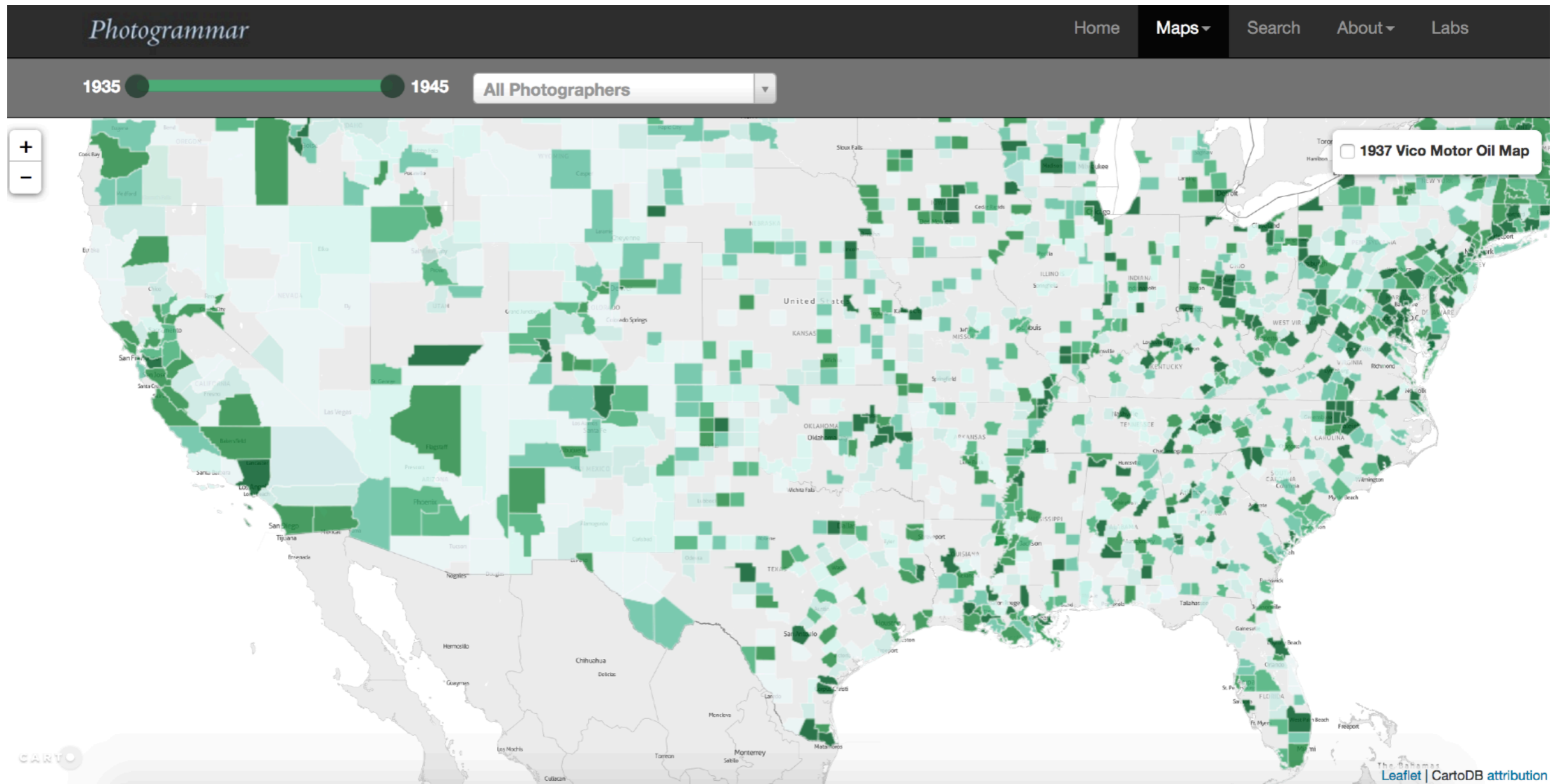
Topic modelling



Cameron Blevins, [“Topic Modeling Martha Ballard’s Diary”](#) (2010)



No solo de textos vive el hombre...



an open-source JavaScript library
for mobile-friendly interactive maps



<http://photogrammar.yale.edu/>

4. Publicación y almacenamiento

Los datos utilizados en la investigación deberían:

- ser publicados en abierto para que puedan ser consultados y contrastados
- ofrecerse en formatos planos para ser reutilizados
- ser albergados en repositorios institucionales

GitHub



.txt

Para acabar...

- Los datos constituyen las evidencias epistemológicas de nuestras investigaciones
- Nos ayudan a pensar crítica y creativamente sobre las humanidades (Trevor Muñoz. *Humanities Data*, 2014).

el análisis espacial y la visualización no consisten en producir ilustraciones y mapas para comunicar cosas que has descubierto a través de otros medios. Se trata de un **medio de hacer investigación**; genera preguntas que de otra manera no se habrían formulado, revela relaciones históricas que de otra manera habrían pasado desapercibidas, y debilitan, o corroboran, hechos sobre los cuales construimos nuestras propias versiones del pasado.

(White 2010)

Bibliografía

- Barlow, Mike (2013). *The Culture of Big Data*. Sebastopol, CA: O'Reilly.
- Christine Borgman (2009). "The Digital Future Is Now: A Call to Action for the Humanities." *Digital Humanities Quarterly* 3.4.
- (2010). *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. Cambridge & London: MIT Press.
- Laney, Doug (2001). "3D-Data Management: Controlling Data: Volume, Velocity and Variety"; republished at: <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>
- Drucker, Johanna. "Graphical Approaches to the Digital Humanities." In *A New Companion to Digital Humanities*, London: Blackweel, 2004, pp. 238-250
- Mayer-Schönberger, Viktor and Cukier, Kenneth (2013). *Big Data: a Revolution That Will Transform How We Live, Work, and Think*. Boston: Houghton Mifflin Harcourt.
- Marche, Stephen (2012). "Literature Is not Data: Against Digital Humanities." *Angeles Review of Books*, Oct. 28, 2012.
- Owens, Trevor (2011). "Defining Data for Humanists: Text, Artifact, Information or Evidence?", *Journal of Digital Humanities* 1.1
- Presner, Todd, and David Shepard. "Mapping the Geospatial Turn." In *A New Companion to Digital Humanities*, ed. by S. Schreibman, et al. Oxford: Wiley Black 2016, pp. 201-212
- Rojas Castro, Antonio. (2017) "Big Data in the Digital Humanities. New Conversations in the Global Academic Context." *Smart Culture. Analysis of digital trends*.
- Schöch, Christof. "Big? Smart? Clean? Messy? Data Humanities." *Journal of Digital Humanities*, 2.3. 2013
- Ward, Jonathan, and Barker, Adam (2013). "Undefined By Data: A Survey of Big Data Definitions". <https://arxiv.org/abs/1309.5821>
- White, R. (2010). What is spatial history? <http://web.stanford.edu/group/spatialhistory/cgi-bin/site/pub.php?id=29>

¡Muchas gracias!