

Herramientas digitales para la edición de mapas y textos

Gimena del Rio Riande y Melisa Martí

Un campo científico llamado “Humanidades Digitales”

Las *Digital Humanities* son hoy día parte fundamental del discurso científico de la Academia norteamericana y europea. Una gran oferta de cursos de grado, posgrado, talleres, centros y laboratorios de Humanidades Digitales, así como publicaciones y revistas científicas, sostienen el campo y legitiman sus prácticas y metodologías. Y, aunque la primitiva *Humanities Computing* hoy se haya expandido a las *Digital Humanities*,¹ las aproximaciones y reflexiones desde el cómputo a las humanidades cuentan con una

1 Kirschenbaum (2010: 56) señala que fue John Unsworth, editor del seminal libro *A Companion to Digital Humanities*, quien decidió denominarlas así en el título de este volumen: “The real origin of that term [Digital Humanities] was in conversation with Andrew McNeillie, the original acquiring editor for the Blackwell Companion to Digital Humanities. We started talking with him about that book project in 2001, in April, and by the end of November we’d lined up contributors and were discussing the title, for the contract. Ray [Siemens] wanted ‘A Companion to Humanities Computing’ as that was the term commonly used at that point; the editorial and marketing folks at Blackwell wanted ‘Companion to Digitized Humanities’. I suggested ‘Companion to Digital Humanities’ to shift the emphasis away from simple digitization”.

larga trayectoria desde que en 1949 proyectos de investigación como el *Index Thomisticus* del Padre Busa e IBM² o la *Computerized Concordance to the Revised Standard Version of the Bible* de John W. Ellison y Remington utilizaran los avances de la informática de las tarjetas perforadas para estudiar grandes corpus.³

Este paisaje es muy diferente para la comunidad de habla hispana, donde el desarrollo y la aplicación de las Humanidades Digitales en el currículo universitario y la investigación resulta buen ejemplo de la dificultad de transposición de la disciplina (o línea de trabajo, método, campo o etiqueta), tal y como se fue construyendo dentro de los *English Departments*,⁴ y las herramientas y métodos de investigación. La historia cultural y académica de cada país, los condicionamientos socio-económicos, han moldeado de un modo completamente diferente, al norte y al sur, a las Humanidades Digitales y las *Digital Humanities*; y a un lado y al otro del Océano Atlántico, a las Humanidades Digitales que hablan español. De la mano de ello, uno de los grandes problemas que enfrenta la mayor parte de los países de América Latina y el Caribe, a la hora de posicionarse como práctica académica, es la inexistencia de fondos públicos que financien este tipo de investigación, la dificultad que enfrentan los investigadores para conformar grupos de trabajo interdisciplinarios, la necesidad de capacitación en el uso de herramientas digitales, y la debilidad y obsolescencia de las infraestructuras tecnológicas para la investigación.⁵

2 Accesible hoy en su última versión desde <http://www.corpusthomicum.org/it/index.age>.

3 del Río Riande (2015, 2016b).

4 Kirschenbaum (2010).

5 del Río Riande (2015).

Qué es Pelagios Commons - Qué es Recogito

Pelagios Commons es una comunidad y una infraestructura que trabaja con tecnologías semánticas como *Linked Open Geodata* para las humanidades.⁶ Su objetivo es fomentar mejores vínculos entre diferentes recursos en línea (fuentes relacionadas con datos históricos, geográficos, mapas) para documentar el pasado. *Pelagios Commons* ofrece a la comunidad de humanistas digitales herramientas y recursos online que, por un lado, no necesitan de conocimientos de programación y, por el otro, permiten un trabajo en la nube, con capacidad de exportación a diferentes formatos y plataformas. Más aún, *Pelagios Commons* acerca recursos que posibilitan un trabajo colaborativo y con garantías de autoría y preservación a largo plazo. Así, herramientas como *Recogito*⁷ o *Peripleo*,⁸ que tienen una curva de aprendizaje mínima y resultan sencillas de utilizar de forma autónoma, han demostrado ser de gran interés y utilidad para nuestra comunidad en Argentina desde que empezamos a trabajar con ellas, hace aproximadamente cuatro años.

Recogito, por ejemplo, es una plataforma de anotación y georreferenciación en línea que ofrece un espacio personal de trabajo donde se pueden cargar, recopilar y organizar materiales fuente —textos, imágenes y datos tabulares— y realizar y/o colaborar en anotaciones y tareas de georreferenciación. Por otra parte, permite que este tipo de trabajo sea visible en la web como producto textual anotado y/o para publicar los resultados de una investigación como dato abierto. En síntesis, *Recogito* es una herramienta semiautomática de anotación de referencias geográficas que,

6 Accesible en <http://commons.pelagios.org/>

7 Accesible en <https://recogito.pelagios.org/>

8 Accesible en <http://peripleo.pelagios.org/>

mediante el recurso de *geotagging*, identifica los topónimos presentes en un texto y permite etiquetarlos, para luego volcar la información en la plataforma de *Pelagios* y mapearla en un nomenclátor universal o *gazetteer* (este paso es habitualmente nombrado como de *georresolución* o *georesolving* a través de sistemas de información geográfica o GIS). Por lo tanto, *Recogito* brinda la oportunidad de reconocer lugares geográficos de textos literarios pretéritos, y a la vez ofrece la posibilidad de regularizar los topónimos antiguos, que han sido atravesados por variaciones lingüísticas y fonéticas, a lugares geográficos actuales.

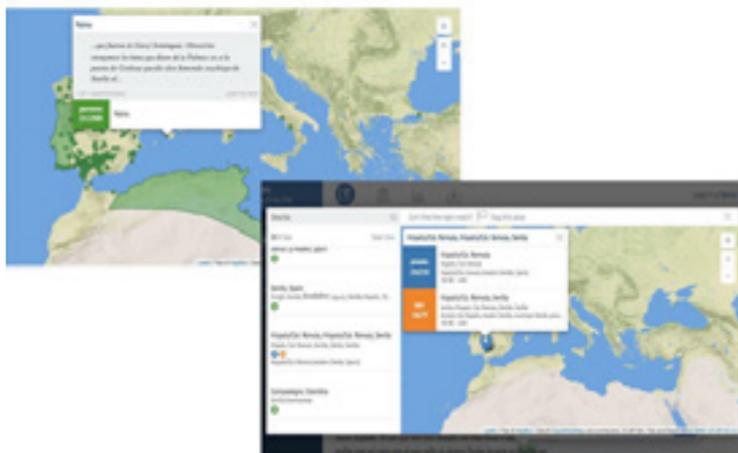


Figura 1. Marcado de una referencia geo-histórica con *Recogito*

Cabe destacar que, en estos últimos cuatro años, las herramientas y recursos de *Pelagios* fueron presentados y/o utilizados en varios eventos en Argentina y llamaron la atención de nuestra comunidad. Por ejemplo, *Recogito* se presentó en noviembre de 2014 en las I Jornadas de Humanidades Digitales de la Asociación Argentina de Humanidades

Digitales (AAHD);⁹ en el I Congreso Internacional de la AAHD en noviembre de 2016;¹⁰ en la serie de workshops organizados en el marco de la Semana de la Ciencia-Conicet o en el CAICYT en 2017;¹¹ en las III Jornadas de Ficcionalización y Narración “Un milenio de contar historias: los conceptos de ficcionalización y narración de la Antigüedad y el Medioevo” en ese mismo año;¹² y en el curso “Técnicas para la creación, enriquecimiento y análisis de textos digitales” en la New York University de Buenos Aires en marzo de 2018.¹³

Proyectos de Pelagios Commons en español: Medieval Iberia/Iberia Medieval y Pelagios South/Pelagios al Sur

El primer proyecto de *Pelagios Commons* en español fue llevado a cabo en conjunto con el equipo del Laboratorio de Innovación en Humanidades Digitales (LINHD) de la

9 Sitio web y actas del congreso: <https://www.aacademica.org/jornadasaahd>.

10 Sitio web del congreso: <https://www.aacademica.org/aahd.congreso>. Las Actas fueron publicadas en abril de 2018.

11 Con el fin de abrir el proyecto a la comunidad y capacitar a los interesados en el uso de las herramientas de *Pelagios Commons*, Gimena del Río y Romina De León dictaron en 2017 dos talleres. El primero, el 5 de septiembre de 2017, en el marco de la Semana de la Ciencia, en el CAICYT de Conicet: “Taller de Georreferenciación, Humanidades Digitales y relatos rioplatenses en los Siglos XVI y XVII”. Aquí una breve entrada sobre la experiencia: <http://www.caicyt-conicet.gov.ar/micrositios/hd/?p=1043>. El segundo taller, fue dictado el 12 de octubre HD CAICYT y presentó una nueva edición del “Taller de Georreferenciación, Humanidades Digitales y relatos rioplatenses en los Siglos XVI y XVII” especialmente pensada para historiadores de la Facultad de Filosofía y Letras, Instituto Ambrosetti e IDECU: <http://www.caicyt-conicet.gov.ar/micrositios/hd/?p=1072>.

12 Este taller fue dictado el día 24 de noviembre por Melisa Martí y Gimena del Río en el marco de las III Jornadas de ficcionalización y narración “Un milenio de contar historias: los conceptos de ficcionalización y narración de la Antigüedad y el Medioevo”, en la Facultad de Filosofía y Letras de la Universidad de Buenos Aires. Las slides del taller las pueden encontrar aquí: <https://es.slideshare.net/GimenaDelRioRiande/taller-pelagios-jornadas-ficcionalizacin-ffyl-uba>

13 Véase: <http://www.caicyt-conicet.gov.ar/micrositios/hd/?p=1212>

UNED (<http://linhd.es/>) y el Instituto de Investigaciones Bibliográficas y Crítica textual (IIBICRIT) de Conicet en el proyecto *Medieval Iberia* (*Pelagios Microgrants* 2016, dirigido por la Dra. Gimena del Rio).¹⁴

En la primera fase del proyecto, el equipo de *Mediaeval Iberia* trabajó en:

- 1) Una edición digital revisada de los *Documentos Castellanos* de Alfonso X (en *Textos y Concordancias Electrónicas de Documentos Castellanos de Alfonso X* por M^a Teresa Herrera, M^a Nieves Sánchez, M^a Estela González de Fauve y M^a Purificación Zabía. Madison: Seminario Hispánico de Medieval Studies, 1999).¹⁵
- 2) Un nomenclátor específico en Pleiades¹⁶ para el mundo ibérico medieval basado en los lugares de los *Documentos Castellanos* de Alfonso X.
- 3) Un laboratorio virtual *Recogito*. Esta experiencia, que desarrollamos entre España y Argentina (LINHD y Conicet), se organizó para etiquetar los lugares de los *Documentos Castellanos* de Alfonso X.
- 4) Una herramienta basada en Procesamiento del Lenguaje Natural (NLP) para la localización y etiquetado de lugares en textos en español (MEGHISTa) fue desarrollada por María Luisa Díez Platas (LINHD)¹⁷.

14 Véase: <http://commons.pelagios.org/2016/10/mediaeval-iberia-through-pelagios-commons/>

15 Estos textos fueron amablemente compartidos por Francisco Gago Jover y Javier Pueyo de la Biblioteca Digital de Textos del Inglés Antiguo (BIDTEA, Seminario Hispano de Estudios Medievales-Universidad de Wisconsin-Madison) (<http://www.hispanicseminary.org/manual-es.htm>).

16 Disponible en <http://pleiades.stoa.org>

17 Disponible en <https://github.com/MLuisaDiezPlatas/MEGHISTa-tool>

De todos modos algunos problemas importantes, relacionados con la idea de *cronotopo* bajtiniano,¹⁸ tuvieron que abordarse antes de comenzar este trabajo. Los topónimos ibéricos medievales no fueron fáciles de identificar y localizar por varias razones. En primer lugar, no existe un nomenclátor español / ibérico medieval para identificar lugares geográficos de fuentes literarias y sus equivalentes en la geografía actual. Además, también es necesario regularizar nombres y términos, ya que los textos que sobreviven en la Edad Media contienen muchas variantes fonéticas y lingüísticas que no permiten que los sistemas de tecnologías de reconocimiento de entidades nombradas (Named Entity Recognition-NER) los reconozcan automáticamente. Finalmente, algunos lugares que aparecen en textos literarios fueron inventados por autores medievales o nombrados de una manera diferente a lo largo de los siglos.

Después de probar diferentes herramientas del Procesamiento del Lenguaje Natural (*Natural Language Processing-NLP*) para obtener automáticamente entidades geográficas con sistemas NER, como Freeling,¹⁹ se desarrolló un marco para implementar las diferentes fases definidas en un procedimiento de extracción de conocimiento.

Medieval Iberia resultó ser una actividad de aprendizaje de las herramientas de *Pelagios Commons* para muchos investigadores españoles y latinoamericanos. Como segunda instancia, iniciamos un proyecto que pudiera relacionar a España y América Latina a través de la historia, la literatura, los mapas y los recursos de *Pelagios Commons*.

Pelagios Commons al Sur/Pelagios South fue el segundo proyecto de *Pelagios Commons* en español y pretendió centrarse en la región del Río de la Plata y sus textos. *Pelagios al Sur/*

18 Bajtin (1989).

19 Véase <http://nlp.lsi.upc.edu/freeling/>

Pelagios South (Resource Development Grant 2017, dirigido por la Dra. Gimena del Rio) propuso extender las herramientas de *Pelagios Commons* a otras fronteras espaciales y geográficas, más allá del mundo antiguo o medieval y de la geografía euro-asiática.

Como es sabido, el sur del continente americano fue un destino bastante tardío para los españoles; sin embargo, de alguna manera, ha sido *visitado* y descrito por los ojos de muchas partes del mundo. Por ello, para nuestro proyecto pusimos principal atención en un corpus de textos que pertenece a los siglos XVI-XVII y que describe, desde una mirada particular y por vez primera en *letra dura*, el sur del continente americano, sobre todo lo relacionado con las llamadas *tierras del Río de la Plata*. En un principio, delimitamos para nuestro proyecto los siguientes textos:

- » Comentarios (de Pero Hernández) a Alvar Núñez Cabeza de Vaca (1ª ed. 1542/1555)
- » Viaje al Río de la Plata - Ulrico Schmidl (1ª ed. 1567)
- » La Argentina - Martín del Barco Centenera (1ª ed. 1602)
- » La Argentina - manuscrita - Ruy Díaz de Guzmán (1ª ed. 1612)
- » Un Informe de un viaje por el Río de la Plata y desde allí por tierra a Perú...- Acarete du Biscay (1ª ed. 1672)

Entendimos que estos textos inaugurales, a mitad de camino entre la crónica y el relato de ficción, podían interesar a historiadores, filólogos, etcétera. El primer paso en *Pelagios South* fue el de buscar todos estos textos en una versión reutilizable que no violara derechos de autor y estuviese en acceso abierto. Una vez identificados, los textos se limpiaron y pasaron a formato .txt. Luego, se subieron a la plataforma de *Recogito* con sus metadatos revisados y las licencias *Creative Commons* pertinentes.

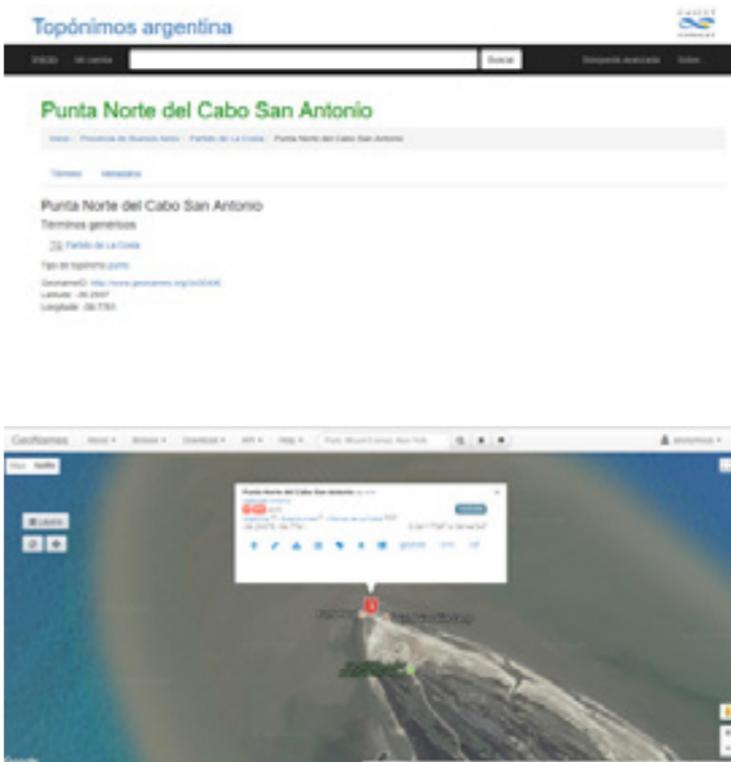


Figura 2. Portada del extenso poema *La Argentina*, de Martín del Barco Centenera, en edición de 1665.

Asimismo, creamos un *Vocabulario Controlado de lugares de la República Argentina de Geonames en TemaTres*.²⁰ *TemaTres* es una herramienta web para gestionar representaciones de conocimiento que permite administrar vocabularios controlados (glosarios, taxonomías, tesauros, listas) en servidores de código abierto. Registramos, por el momento, los nombres de los espacios o topónimos y las características geográficas existentes en el entorno sociogeográfico de los siglos XVI y XVII en este territorio. El tener los lugares señalados en este vocabulario controlado con su identificador en *Geonames* nos fue muy útil para algunas búsquedas en la interfaz de *Recogito*. Dado que la mayor parte de nuestros

²⁰ Accesible en <http://vocabularios.caicyt.gov.ar/geoar/vocab/index.php>

lugares no aparecían en los *gazetteers*, necesitábamos un modo de “engañar” a *Recogito* en las búsquedas. El primer ejemplo que se nos presentó es un topónimo denominado por Ruy Díaz de Guzmán en el texto de *La Argentina manuscrita* como *Cabo Blanco*, lugar que actualmente se denomina *Cabo San Antonio*. *Cabo Blanco* no es una referencia rastreable en *Recogito*, pero sí lo es *Cabo de San Antonio*, aunque solo en el vocabulario controlado o en *Geonames*. Una solución aquí es buscar *Cabo de San Antonio* por su número en *Geonames* en la interfaz de *Recogito*.





Figuras 3, 4 y 5. Proceso de marca de un topónimo, desde la información volcada desde *Geonames* a *TemaTres* y posterior búsqueda y geolocalización en *Recogito*.

Para proceder con nuestro trabajo, realizamos un examen detenido de los tipos de marcas que buscábamos. Muchas veces, el movimiento colonizador es descrito en estos textos a partir del nacimiento o desembocadura de un río. Entendamos que aquí no hay una construcción del espacio al *modo europeo*, no hay una noción de ciudad o de *marca civilizatoria* occidental. Esta teorización sobre las marcas trajo complejidad al marcado de los textos, más allá de las evidentes dificultades con la georreferenciación. Aquí citamos algunas recurrentes apariciones de ríos y otros accidentes geográficos, como ser *Río de la Plata*, *Cordillera de los Andes*, *Cataratas del Iguazú*, que podemos observar en las referencias de *OpenStreetMaps* pero que, evidentemente, no son reconocidos por *Recogito*.²¹

21 Para las marcas de lugares, personas y eventos, relevamos una bibliografía que compartimos desde un grupo de Zotero y utilizamos para referenciar cada anotación: https://www.zotero.org/groups/1669951/hdcaicyt-pelagios_al_sur

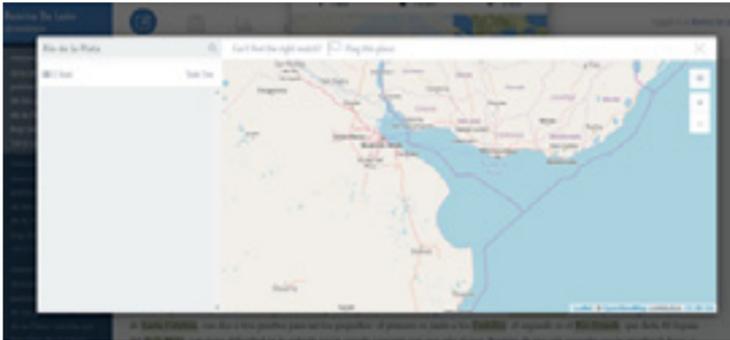


Figura 6. Dificultad en la marca de río en *Recogito*.



Figura 7. Uso de *Zotero* como repositorio bibliográfico del proyecto.

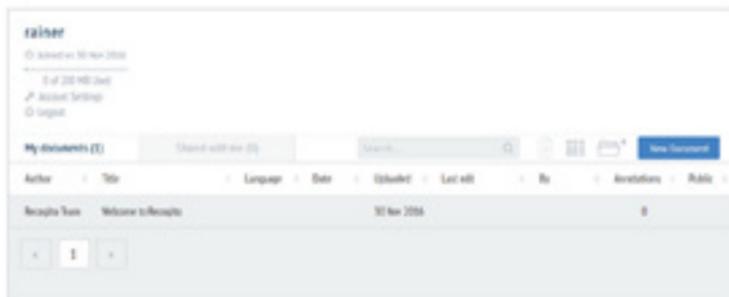
¿Cómo utilizar *Recogito*?

Recogito es una herramienta de anotación y georreferenciación intuitiva y con una curva de aprendizaje mínima. Aquí, un breve tutorial de los pocos pasos que muy pronto nos permiten, de modo individual o colectivo, enriquecer desde las marcas de lugar, evento y persona, los textos que se inserten en la plataforma:

- » Como primer paso debemos crearnos un usuario en <http://recogito.pelagios.org/>
- » En segundo lugar, debemos cargar en *New Document* en formato *.txt o *tei-xml.
- » A continuación, deberán cargarse los metadatos del archivo con el que se trabaje (título, autor, editor, fecha, etcétera).

Step 1. Upload a Document

Documents are used to organize your materials in Recogito. Each document can consist of multiple files. But for now, we are just going to upload a document consisting of a single text file.



Click **New Document**.

Figura 8. Interfaz interna de *Recogito*.

Para comenzar con el marcado, debemos abrir nuestro archivo y, una vez en el texto, debemos simplemente seleccionar una palabra o frase. Las opciones de marcado con las que cuenta *Recogito* son: *place*, *person* o *event*.²²

22 Puede consultarse, el tutorial en español de *Recogito* que las autoras y varios de los miembros de *Pelagios South* realizaron en el marco del grupo (SIG) de Multilingüismo de *Pelagios Commons*: <http://recogito.pelagios.org/help/es/tutorial>

Como puede apreciarse aquí, el trabajo con *Recogito* es de suma utilidad a la hora de estudiar el alcance del conocimiento de la geografía mundial en un momento determinado.

Para finalizar

Las Humanidades Digitales buscan generar un nuevo conocimiento a través del uso de herramientas digitales y de metodologías propias aplicadas a grandes corpus. En la intersección de lo humanístico con lo digital, de la lectura cercana y la distante,²³ entre lo cuantitativo y lo cualitativo, está el aporte de las Humanidades Digitales al trabajo académico y científico. Creemos que, por un lado, los recursos y herramientas de *Pelagios Commons* pueden definirse desde este lugar de intersección enriquecida y que, asimismo, se transformarán en elementos clave de las Humanidades Digitales latinoamericanas, por su naturaleza intuitiva y por la posibilidad de trabajo colaborativo y en acceso abierto.

23 Moretti (2013).