

Material Didáctico Sistematizado.

Introducción a la Edición Digital Académica.

María Gimena del Rio Riande y Elena González Blanco García.

Cita:

María Gimena del Rio Riande y Elena González Blanco García (2016). *Introducción a la Edición Digital Académica*. Material Didáctico Sistematizado.

Dirección estable: <https://www.aacademica.org/gimena.delrio.riande/19>

ARK: <https://n2t.net/ark:/13683/pdea/FYS>



Esta obra está bajo una licencia de Creative Commons.
Para ver una copia de esta licencia, visite
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>.

Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite: <https://www.aacademica.org>.

Módulo 1. Introducción a la Edición digital académica

- 1.1. Breve introducción a las Humanidades Digitales y la Edición Digital
- 1.2. Centralidad de la metodología y práctica de la edición de textos en las Humanidades Digitales
- 1.3. Los grandes proyectos
- 1.4. Big data, big corpus y la necesidad de ediciones fiables
- 1.5. Bibliografía

0. Introducción de la introducción

Este primer módulo introductorio es apenas la piedra de toque, el puntapié inicial de una serie de módulos mucho más específicos, extensos y profundos que buscan guiarte en el aprendizaje y la práctica de las metodologías y herramientas que conforman el campo de la Edición digital académica.

Ten en cuenta que muchos de los temas que aquí avancemos, los verás en la práctica, al abordar el resto de los módulos.

Es importante que, más allá de la lectura progresiva y lineal del módulo, vuelvas sobre los hipervínculos marcados en el texto y profundices la lectura de este a través de ellos.

1.1. Breve introducción a las Humanidades Digitales y la Edición Digital: en el principio fue la *Humanist Computing*

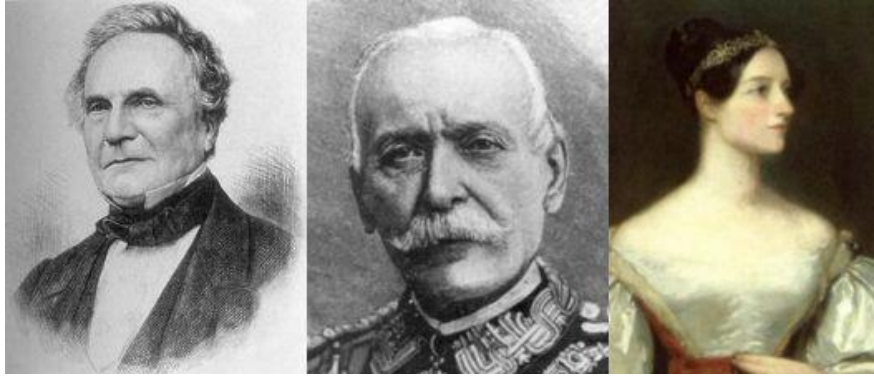
La primera vez que se pensó en que una máquina podía colaborar con la comprensión y funcionamiento de un objeto del mundo de las humanidades se remonta al ya mencionado siglo XIX, esa centuria de grandes datos y voluntad de sistematización (Robertson & Travaglia 2015). Fue Ada Lovelace, hija de Lord Byron, quien lo comprendió rápidamente. Augusta Ada King, Condesa de Lovelace (1815-1852), matemática, analista metafísica (en sus propias palabras), y escritora, e hija de Lord Byron, es considerada hoy la primera persona

que escribió un programa para una máquina, por ello cuenta con un día de celebración (13 de octubre) en el calendario de los programadores. Las notas que escribió para su traducción al inglés de las *Notions sur la machine analytique de Charles Babbage*, un informe realizado por Luigi Federico Menabrea (1842) sobre las disertaciones del matemático e inventor Charles Babbage, constituyen una perfecta teorización sobre cómo la tecnología computacional puede superar la instancia del pensamiento numérico y, a través de una máquina (en este caso, el *analytical engine* o motor analítico), llegar a aplicarse sobre múltiples objetos (Rio Riande 2015a).

Ada superaría las conclusiones del también matemático Menabrea al comprender que el razonamiento abstracto podía superar al pensamiento matemático y aplicarse al mundo de las Humanidades. En sus notas queda claro que este *analytical engine* no solo podía predecir respuestas, sino que era capaz de extender su funcionamiento, por ejemplo, a una máquina que, a través de sus algoritmos, trabajase en la composición automática de música:

Supposing, for instance, that the fundamental relations of pitched sounds in the science of harmony and of musical composition were susceptible of such expression and adaptations, the engine might compose elaborate and scientific pieces of music of any degree of complexity or extent (Lovelace en Menabrea, 1842, Nota A).

[Suponiendo, por ejemplo, que las relaciones fundamentales de los sonidos de cada tono en la ciencia de la armonía y de la composición musical fuesen susceptibles de dicha expresión y adaptaciones, el motor (analítico) podría componer piezas musicales elaboradas y científicas de cualquier grado de complejidad o extensión]



Charles Babbage-Luigi Federico Menabrea-Ada Lovelace

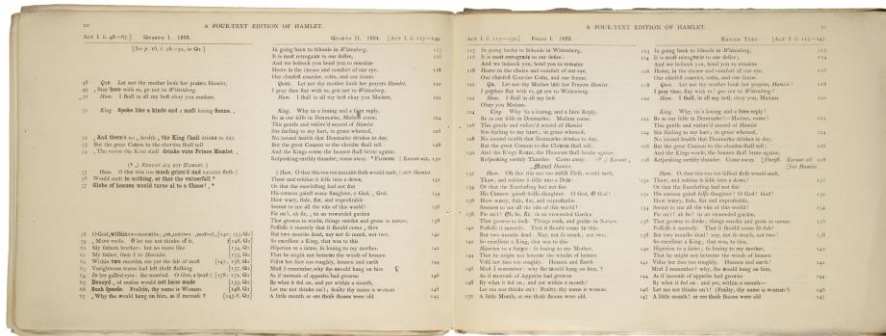


Ada en el webcomic de Sydney Padua

[2D Goggles, or the Thrilling Adventures of Lovelace and Babbage](#)

Para el mismo momento en el que Ada trabajaba en las notas a l analytical engine, hacia 1883 una joven llamada Teena Rochfort Smith creó un prototipo de una edición experimental comparada de Hamlet que, a día de hoy, sigue siendo el experimento impreso más complejo que se haya intentado sobre el *Hamlet* de Shakespeare. Incluso hoy en día las interfaces digitales son menos ambiciosos que lo Rochfort Smith previó, que empujó la tipografía victoriana y la tecnología de impresión de esos tiempos a sus límites máximos. Dada la complejidad única de *Hamlet* dentro de las obras de Shakespeare y la complejidad textual de Shakespeare en el canon de la literatura Inglesa, el prototipo de Rochfort Smith puede clasificarse entre los grandes experimentos en el diseño de interfaces humanas, incluyendo la *Biblia en seis versiones* de Orígenes, la *Biblia Políglota Complutense*, la edición de Robert Estienne 1550 del griego del *Nuevo Testamento*, la *New Variorum Shakespeare* iniciada por Howard Furness en la década de 1870, el hoy obsoleto

proyecto de la BBC llamado [Domesday Book](#), o las actuales [Juxta commons](#) o [Versioning Machine](#), un software para el cotejo de versiones, que visualiza la relaciones entre las variantes de los distintos textos de una misma obra literaria.



Ham. Oh that this too too solid Flesh, would melt,
Thaw, and resolve it selfe into a Dew :
Or that the Euerlasting had not fixt
His Cannon 'gainst Selfe-flaughter. O God, O God !

Imágenes de la edición prototipo del Hamlet de Teena Rochfort Smith

El experimento de Teena Rochfort, además de apuntar al cotejo de variantes en lo que una edición comparada podría ofrecer, también ponía el acento en lo paleográfico, buscando dar cuenta de los cambios de mano de copista o abreviaturas y otros alógrafos a través del uso de diferentes tipografías. Lamentablemente, N. Trübner & Co., la casa editorial de Londres con la que Teena pensaba sacar su libro, solo llegó a hacer una edición prototipo. El trabajo de cajista era inmenso, ya que la edición de Rochfort Smith requería de la colocación de diferentes tipografías en un mismo folio: Times New Roman, Clarendon, Sans-Serif, gótico, etc., en mayúsculas y minúsculas, y a veces la fusión de diferentes tipos en la misma palabra, y, a veces en combinación a través varios signos diacríticos. Así como el diseño de Charles Babbage para su *engine* superaba las capacidades de la mecánica victoriana, también lo hacía la edición de Rochfort Smith, que excedía las capacidades de la tipografía victoriana. El proyecto de Rochfort Smith

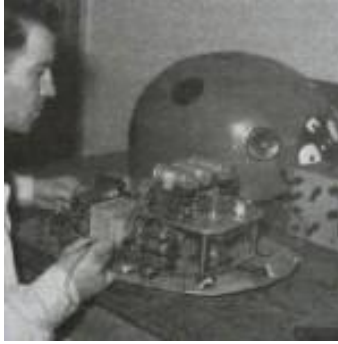
no fue más allá del prototipo que te mostramos aquí, creado por miembros de la New Shakespeare Society, y la única copia sobreviviente es la que hoy se conserva en la Biblioteca Folger.

Desconociendo el trabajo de estas pioneras en las Humanidades Digitales y la Edición digital académica, un siglo después, hacia 1949, un joven jesuita recientemente doctorado, el Padre Roberto Busa, siguió intuitivamente este legado. Con el fin de sistematizar la cuantiosa y heterogénea información que había trabajado en su tesis doctoral sobre concordancias léxicas en los textos de Santo Tomás de Aquino, comenzó a pensar en un lematizador –que hoy puede verse en [Corpus Thomisticum](#)– con informáticos de IBM, quienes asimismo colaboraban, para ese momento, con una máquina que perseguía la traducción automática, la *Machine Translation* y la *Computerized Concordance to the Revised Standard Version of the Bible* de John W. Ellison, híbridos universitarios modelados desde los Estados Unidos de Norteamérica por IBM y Remington (Rio Riande, 2015b).

Cabe destacar que el proyecto del que hoy es considerado el *padre* de las Humanidades Digitales, el Index Thomisticus, sigue activo (tal y como es esperable para un proyecto en Humanidades Digitales) y fue elaborado en una primera fase en tarjetas perforadas (ya conocidas por Babbage), luego indexado en CD-Rom y, con posterioridad, digitalmente.



El Padre Busa y un ordenador IBM utilizado para el trabajo de lematización de la obra de Santo Tomás



Machine Translation IBM

Poco a poco, esta línea relacionada con la lingüística, la lengua y la literatura fue abriéndose paso en las universidades norteamericanas y noreuropeas desembocando en una disciplina que poco a poco pasó a denominarse *Humanist Computing (Informática Humanística)*. Así, menos de veinte años después del comienzo del proyecto del padre Busa, en 1965, se llevó a cabo el congreso *Computers for the Humanities?* en la Universidad de Yale y, apenas un año más tarde, comenzó la publicación de la primera revista académica —también norteamericana— sobre el tema, *Computers and the Humanities*, que apenas suspendió su publicación en 2004.

La década del 80 estuvo bastante centrada en la Lingüística. Susan Hockey iluminó prontamente el camino al publicar en 1980 *A Guide to Computer Applications in the Humanities* (London: Duckworth), y así, ya a mediados de los 90 la *Humanist Computing* era una tarea que, sin declararse abiertamente en los programas de estudios, muchos profesores e investigadores del área de *Humanities and Arts* de la Universidad de Virginia ponían en práctica en sus clases y proyectos. Hacia el año 1998 se discutió en el Senate de esa universidad la importancia de la tecnología informática para las Humanidades y si la *Humanist Computing* podía ser considerada una disciplina universitaria y todo terminó con la conformación de un [seminario permanente en Humanidades Digitales](#) en 2001 y, finalmente, con un [Master en](#)

[Humanidades Digitales](#) dirigido por quien así las bautizó en el año 2004, John Unsworth¹.

En los últimos quince años hemos asistido a una progresiva modificación en el campo que nació con la Informática Humanística, redefiniéndose así las actividades con relación a esta más allá de una práctica centrada en la informática 'aplicada a' las Humanidades, y hacia una que pretende producir y asimismo reflexionar/reflejarse en lo digital como continente y contenido.

Si quieres saber más sobre la *Humanist Computing*, te recomendamos leas este capítulo del *Companion to Digital Humanities*, [The History of Humanist Computing](#). Si quieres ver cómo el término *Digital Humanities* terminó definiendo el campo, mira [este gráfico](#) de Google Trends.

1.2. Centralidad de la metodología y práctica de la edición de textos en las Humanidades Digitales

La línea filológica de edición digital nació como parte de la Informática Humanística, y, siguiendo distintas escuelas, fue desarrollándose con gran éxito tanto en los países de Norteamérica como Europa.

¹ En su gran artículo [What is Digital Humanities and what it is doing in the English Departments?](#), obligada lectura de toda persona interesada en las Humanidades Digitales, Kirschenbaum señala que fue John Unsworth, editor del seminal libro [A Companion to Digital Humanities](#) (2004), quien decidió denominarlas así en el título de este volumen al que en breve volveremos a referirnos: «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». [El verdadero origen de este término [Humanidades Digitales] está en la conversación con Andrew McNeillie, el editor de la adquisición original para el "A Companion to Digital Humanities" de Blackwell. Comenzamos a hablar con él acerca del proyecto del libro en 2001, en abril, ya finales de noviembre habíamos ya reunido a los colaboradores y estábamos discutiendo el título, para el contrato. Ray [Siemens] quería "A Companion to Humanities Computing" como ese era el término utilizado habitualmente hasta ese momento; las editoriales y la gente de marketing de Blackwell quería "A Companion to Digitized Humanities". Yo sugerí "Companion to Digital Humanities" para alejarnos de la simple digitalización.]

Un proyecto pionero fue el de Deborah Parker, la edición de la *Divina Comedia* de Dante, parte de su tesis doctoral en la ya mencionada Universidad de Virginia, usando un marcado en [SGML](#) que hoy, en un proyecto a mayores, puede leerse en un entorno mucho más completo, [World of Dante](#).



The World of Dante. <http://www.worldofdante.org/>

En España, por ejemplo, la línea filológica de la Informática Humanística abrió el campo para muchos proyectos relacionados tanto con la lingüística, la bibliotecología y la edición. En el año 1994 Francisco Marcos Marín publicó en España el primer tratado sobre la aplicación de la tecnología a la investigación en Humanidades, *Informática y Humanidades* (Madrid: Gredos, Grandes Manuales). Su aproximación claramente filológica² y muy relacionada con los estudios sobre edición de textos y Edad Media sería rápidamente seguida por otros como José Manuel Lucía Megías ["Editar en Internet (che quanto piace il mondo è breve sogno)", *Incipit*, XVIII, 1998, 1-40]. Otra española, también emigrada a los países del norte de América, Estelle Irizarry, publicaría un trabajo relevante en su momento, [Informática y literatura:](#)

² Le seguiría *El comentario filológico con apoyo informático*, Madrid: Síntesis, 1996. |

[análisis de textos hispánicos en español](#), de 1997 (Barcelona: Proyecto A). En el resto de Europa, la conformación de tempranos centros de investigación como el *Literary and Linguistic Computing Centre* (LLCC) de la Universidad of Cambridge (1964) y su *Association for Literary and Linguistic Computing* (ALLC) (1973) venían apoyando este tipo de trabajo desde hacía muchos años (Vanhoutte, 2013: 129). En un aspecto más amplio, esto es lo que puede verse en el temario del Call for papers del próximo [Congreso de la ADHO](#) (la alianza internacional de Humanidades Digitales de la que hablaremos más adelante), que se celebrará en la ciudad de Cracovia, Polonia, en julio de este año:

→Aplicaciones informáticas en estudios literarios, lingüísticos, culturales o históricos, incluyendo literatura en formato electrónico, humanidades en el ámbito público o aspectos interdisciplinarios relacionados con la investigación académica actual.

El campo de la edición digital académica fue modelándose así en los distintos países, siguiendo las líneas filológicas más trabajadas (la crítica textual neolachmanniana en España, la edición genética en Francia), marcando el sesgo académico y universitario que nunca ha perdido y diferenciándose así de otras actividades relacionadas con la edición de textos y las Humanidades Digitales.

Por ello, la edición digital se ha autodenominado como *académica*, es decir, enmarcándose en disciplinas y prácticas universitarias tradicionales sobre edición de textos antiguos y modernos. En este sentido, creemos que es importante destacar que, quienes entienden que las Humanidades Digitales apuntan a un reformateo de las disciplinas académicas³, muchas veces abordan sesgadamente la edición digital académica, malinterpretándola. Bien es verdad que el fundacional [A Companion to Digital Humanities](#) (2004) se organiza a partir de la noción de disciplina universitaria (Historia del arte,

³ Como dice Piscitelli: "[El reformateo pedagógico son los Labs](#)".

Lexicografía, Arqueología, etc.) y que el otro volumen seminal para las Humanidades Digitales, [Debates in the Digital Humanities](#) (2012), contiene un artículo de Mathew Kirschenbaum que resume tanto el lugar como la problemática inserción de las Humanidades Digitales al apuntar con claridad a los departamentos universitarios, [What is Digital Humanities and what's it doing in the English Departments?](#)

Sin embargo, la línea de la edición digital académica es una de las que mejor da cuenta de cómo el software se apoderó de las prácticas universitarias (y de todas nuestras prácticas, como bien nos lo ha contado Manovich en su [Software takes command](#) (2008, 2013)⁴). Si bien es verdad que no todos los docentes, los departamentos o facultades han recibido estos cambios con una actitud positiva, en muchos casos se ha favorecido la edición digital académica a la hora de abordar tanto la problemática de edición como la edición en sí de ciertos textos antiguos. Solo por mencionar algunos casos con respecto a España. Tesis doctorales como la defendida por Gimena del Rio Riande en la Universidad Complutense de Madrid en 2010 acerca de la edición del *Cancionero del rey Don Denis*, donde se reflexiona acerca de una frustrada edición hipertextual, la edición del *Cancionero de Nuno Fernández Torneol*, hoy incluida en [MedDB](#), por Miguel Pousada Cruz en la Universidad de Santiago de Compostela en 2013, o las reflexiones sobre la edición digital de las [Soledades de Góngora](#) por Antonio Rojas Castro, entre otras, dan buena cuenta de la explotación

⁴ El libro de Manovich es verdadero ejemplo del fenómeno digital y del modo en el que la mutación permanente de la web afecta a los artefactos que lo pueblan. Acerca de las diferentes versiones de su libro desde el año 2008 dice Manovich en el copyleft a través del link que te ofrecemos: “One of the advantages of online distribution which I can control is that I don’t have to permanently fix the book’s contents. Like contemporary software and web services, the book can change as often as I like, with new “features” and “big fixes” added periodically. I plan to take advantage of these possibilities. From time to time, I will be adding new material and making changes and corrections to the text”. [Una de las ventajas de la distribución en línea que puedo controlar es que yo no tengo que fijar de modo permanente el contenido del libro. Al igual que los servicios de software y web contemporáneos, el libro puede cambiar tan a menudo como me guste, con nuevas “características” y “grandes soluciones”, añadidos periódicamente. Tengo la intención de aprovechar estas posibilidades. De vez en cuando, voy a añadir material nuevo y hacer cambios y correcciones en el texto.]

de esta línea en el ámbito de los textos medievales, especialmente, poéticos.

Así y todo, estas nuevas formas de trabajo han dado impulso a la creación de espacios colaborativos y de formación dentro de las universidades por fuera de la currícula académica. Por un lado, muchas de estas ediciones han sido incluidas en proyectos que contienen otros textos editados, como es el caso del cancionero de Torneol en la base de datos de la lírica profana gallego-portuguesa, MedDB, y han dado lugar a nuevas metodologías y prácticas de trabajo como las relacionadas con el crowdsourcing, fomentado, por ejemplo, por el proyecto [ReMetCa](#), Repertorio de la Métrica Medieval Castellana, desde hace varios años en España.

La mayor parte de estas ediciones ocupa a un abultado grupo de investigadores, colaboradores, estudiantes e informáticos y suele aglutinarse bajo el título de "proyecto académico", algo que el artículo [What Is Humanities Computing and what is not?](#) (2002) de John Unsworth deja claro, haciendo uso de la palabra *project* para nombrar los objetos relacionados con la práctica de las Humanidades Digitales.

Los laboratorios son los espacios que rápidamente hicieron bien uso de la explotación de la línea de edición digital. Así, en Estados Unidos, el [MetaLab](#) de Harvard, el [Scholar's Lab](#) de Virginia o el [Digital Innovation Lab](#) de North Carolina son precursores en la apertura de este tipo de espacios en la universidad. En Europa, el [Digilab](#) en Italia, el [DHD](#) en Alemania, el [IATEX](#) en la Universidad de las Palmas de Gran Canaria y nuestro [Laboratorio de Innovación en Humanidades Digitales](#) en la UNED en su oferta de curso de *Experto en Humanidades Digitales* y *Experto en Edición Digital*, gozando de una continuidad académica desde el primer curso de verano que organizara en 2014.

Asimismo, a día de hoy distintas universidades norteamericanas y europeas cuentan con carreras de grado, posgrado (master) y doctorado en Humanidades Digitales, que hacen hincapié en la edición digital académica. Por solo mencionar unas pocas, las de la

[Ecole Nationale des chartes](#), [King's College](#), la de la [Universidad de Victoria](#) o [Lethbridge](#) en Canadá y [Maynooth](#) en Irlanda, han explotado la edición de textos académicos utilizando la [Text Encoding Initiative](#) (TEI) y otras tecnologías de edición digital.

En España la oferta de posgrado en Humanidades Digitales se abrió en el año 2010 con un [Master en Humanidades Digitales](#) (MHD) en la Escuela Superior de Ingeniería Informática de la Universidad de Castilla-La Mancha que, lamentablemente, no prosperó y se cerró en el año 2010. La Universidad Autónoma de Barcelona, a través del Departamento de Filología Española, ofrece un [Master en Humanidades Digitales](#) al que califica como master propio. El grupo L.E.E.T.H.I montó un curso en Literatura Digital hace unos años en la Universidad Complutense de Madrid y hoy encauza este trabajo formativo desde un [Erasmus on Digital Literatures](#). La oferta del LINHD es, por el momento, la que goza de una continuidad académica en lo que hace a la edición digital académica desde el primer [curso de verano en 2014](#), este [curso de experto 2015](#), varios talleres formativos online, y los Expertos mencionados.

En Latinoamérica no existe, por el momento, una oferta de grado o posgrado en Humanidades Digitales y el campo de la edición digital académica es aún un espacio casi desconocido. Cabe destacar que la RedHD de México organiza una actividad extracurricular, el [SeminarioHD](#), en la Facultad de Filosofía y Letras de la UNAM y que desde esta institución se ha encarado el proyecto de edición digital de la [Biblioteca Digital del Pensamiento Novohispano](#). En Argentina, puede señalarse el proyecto hermano de ReMetCa, [Diálogo Medieval](#) y añadirse que el CAICYT del CONICET, que ofrece desde 2014 [cursos y conferencias](#) para investigadores en bibliotecas digitales, está trabajando en una biblioteca de herramientas digitales para la edición de textos en el marco del Proyecto [Metodologías en Herramientas Digitales para la Investigación en Humanidades](#).

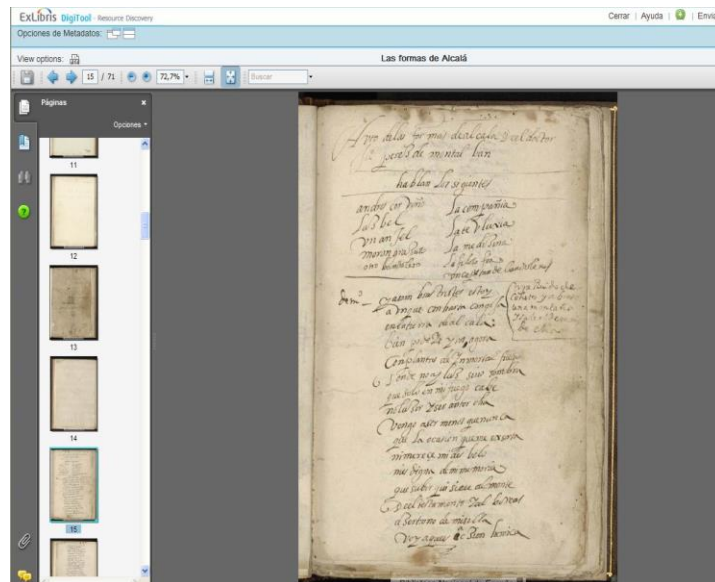
1.3. Los grandes proyectos: casos de estudio

Sería imposible enumerar en un simple apartado todos los proyectos en Humanidades Digitales o que, al menos, han significado un avance en la utilización de la tecnología digital aplicada a la investigación las Humanidades. A pesar de que en el sitio de [CenterNet](#) encontrarás un amplio listado de proyectos en Humanidades Digitales, y que la clasificación de estos proyectos es compleja (Lucía Megías: 2012) (¿bibliotecas digitales/virtuales, plataformas, repositorios, bases de datos?), y que Patrick Sahle nos ha recopilado una enorme cantidad de ediciones digitales en su [base de datos](#), queremos aquí acercarte algunos ejemplos que bien incorporaron la digitalización de obras, la edición digital, la constitución de una biblioteca en la web, etc. como parte de un proyecto de investigación. Verás que algunos están más cerca de ser proyectos de bibliotecas públicas, de consorcios que congregan grandes inversiones públicas y privadas para la difusión y preservación del patrimonio cultural, y que otros son iniciativas de grupos universitarios más pequeños o que, como en el caso de *The World of Dante*, comenzaron como una tesis doctoral. Creemos que no todos son proyectos en Humanidades Digitales, sino que debemos reflexionar aquí sobre nuevas definiciones para este tipo de actividades. Sin embargo, y aunque no podremos aquí dar cuenta de todos, valoramos de utilidad repasar brevemente la historia, contenido y metodología de algunos de estos proyectos relacionados con la disciplina filológica y la edición de textos.

1.3.1. Bibliotecas digitales

Empecemos por España: la [Biblioteca Digital Hispánica](#) de la Biblioteca Nacional de España (BNE) ofrece en la actualidad más de 10.000 obras organizadas en diferentes colecciones que van desde las

Bellas artes a la Medicina y acercan a cualquier investigador o interesado en el mundo una gran cantidad de testimonios manuscritos digitalizados:



Reproducción del manuscrito de *Las formas de Alcalá* en la Biblioteca Digital Hispánica de la BNE

Sin tratarse de un proyecto en Humanidades Digitales, la Biblioteca Digital Hispánica funciona como un buen ejemplo del modo en el que deberían aplicarse los metadatos a la hora de catalogar bibliográficamente una obra y asimismo de los modos de acercar al lector el material digitalizado en una interfaz simple y flexible. La Biblioteca Digital Hispánica trabaja así en la línea de [Gallica](#), uno de los proyectos informáticos pioneros y más influyentes puestos en marcha por la Bibliothèque Nationale de France (BNF). *Gallica* se define como una “bibliothèque encyclopédique et raisonnée”, y ofrece una enorme cantidad de colecciones de documentos libres de derechos digitalizados por la BNF y por colecciones de colaboradores públicos (bibliotecas, centros de investigación, universidades, etc.):



Gallica (BNF)

Un proyecto pionero en España que incorporó métodos de digitalización y un ordenamiento en el sentido de biblioteca digital apelando a materiales procedentes de diversas instituciones fue la [Biblioteca virtual Miguel de Cervantes](#). El trabajo constante de los últimos quince años (piensa en la *Biblioteca Americana*, la *Biblioteca Joan Lluís Vives*, la *Biblioteca Letras galegas*, la dedicada a la literatura infantil y juvenil, la nueva biblioteca sobre Hagiografía o las bibliotecas de autor, como la dedicada a Ausiàs March) la ha convertido en uno de los portales en español más visitados. Un problema de sus bibliotecas es que no están ordenadas ni sistematizadas siguiendo un mismo patrón o metodología. En algunos casos ofrecen registros bibliográficos, en otras, bibliografía crítica, ediciones, y un material muy recoge las digitalizaciones de los testimonios manuscritos que han transmitido la obra de un autor:

Biblioteca Virtual Miguel de Cervantes

Biblioteca Virtual Joan Lluís Vives > Biblioteques d'autor

MIGUEL DE CERVANTES
www.cervantesvirtual.com

Ausiàs March

CERCA RÀPIDA
Text
PRER
>> CERCA AVANÇADA

Ausiàs March > Pàgina principal

Biblioteca d'autor Ausiàs March

*Ffantasiant, Amor a mi descobre
los grans secrets c'als pus suptils amaga,
e mon jorn clar als hòmens és nit fosqua,
e visch de ço que persones no tasten.
Tant en Amor l'esperit meu contempla,
que par del tot forçat l'esperit: (no se han obtenido resultados)
car mos desigs no són trobats en home,
sinó en tal que la carn punt no'l torbe.*

(Ausiàs March, poema XVIII)

L'any 2009 es compleix el 550 aniversari de la mort del cavaller valencià Ausiàs March (1400-1459). Es tracta del poeta més emblemàtic de les lletres catalanvalencianes medievals, tant per la singularitat del seu cançoner, com per la notable influència que ha exercit sobre poetes posteriors i l'enorme difusió internacional que ha assolit a través d'un munt ingent de traduccions. El corpus poètic d'Ausiàs, que supera la tradició trobadoresca de la qual parteix, vehicula un discurs de forta càrrega moral el nucli del qual gira al voltant de l'experiència amorosa, de les contradiccions irresolubles d'aquesta i dels patiments que se'n deriven per al jo poètic.

Ultims continguts incorporats
Consulteu els continguts més recentment afegits a la biblioteca.

Biblioteca de autor Ausiàs March (Biblioteca Virtual Miguel de Cervantes)

Otros proyectos tempranos relacionados con la constitución de bibliotecas virtuales/digitales en universidades españolas son el de la [Biblioteca Digital de la Universitat de València](#) y la [Biblioteca Digital Dioscórides](#) de la Universidad Complutense de Madrid. Como decíamos, si bien estos proyectos no se declaran enmarcados dentro de las premisas de las Humanidades Digitales, no dejan de resultar buenos ejemplos del modo en el que pueden trabajarse en un proyecto de edición digital metadatos e imágenes.

1.3.2. Algunos ejemplos de ediciones digitales

Pongamos un ejemplo de uno de los proyectos en Humanidades Digitales más antiguos interesantes en la web, que ha rescatado las posibilidades de las ediciones sinópticas integrales en la pantalla del ordenador: [The Princeton Charette Project](#) ofrece las transcripciones de los ocho testimonios manuscritos conservados del *roman* de Chrètien de Troyes, *Le chevalier de la charrette*, del siglo XII, al que se le añade

también, entre otros materiales lexicográficos, el texto de la edición de Foulet- Uitti (1989), junto a la reproducción facsímil de todos ellos en una disposición sinóptica:

http://www.princeton.edu/~lancelot/new-traditional/

The Traditional Charrette Viewer | Manuscripts | Transcription Key | Princeton Charrette Project

Browse the Manuscript Images and Transcriptions

Color images of the manuscripts are accessible from the transcriptions (except MS I, not yet available).

MS ⇒ FU Lines □	A Chartilly, Musée Condé 472	C ("Guido"), Paris, Bibliothèque Nationale de France, fonds français 794	E Escorial, Real Monasterio de San Lorenzo M.B.21	F Paris, Bibliothèque Nationale de France, f. fr. 1450	G Princeton, Firestone Library, Garrett 125	I Paris, Bibliothèque de l'Institut de France 6138 (formerly 4676)	T Paris, Bibliothèque Nationale de France, f. fr. 12560	V Vatican, Biblioteca Vaticana, Regina 1725	U Text of the Foulet-Uitti edition
1-31	*	<u>C</u>	<u>E</u>	*	<u>G</u>	*	<u>I</u>	*	<u>U</u>
31-1000	<u>A</u>	<u>C</u>	<u>E</u>	*	<u>G</u>	*	<u>I</u>	<u>V</u>	<u>U</u>
1001-2000	<u>A</u>	<u>C</u>	<u>E</u>	*	<u>G</u>	<u>I</u>	<u>I</u>	<u>V</u>	<u>U</u>
2001-3000	<u>A</u>	<u>C</u>	<u>E</u>	*	<u>G</u>	*	<u>I</u>	<u>V</u>	<u>U</u>
3001-4000	<u>A</u>	<u>C</u>	<u>E</u>	*	<u>G</u>	<u>I</u>	<u>I</u>	<u>V</u>	<u>U</u>
4001-5000	<u>A</u>	<u>C</u>	<u>E</u>	*	*	<u>I</u>	<u>I</u>	<u>V</u>	<u>U</u>
5001-6000	<u>A</u>	<u>C</u>	<u>E</u>	<u>E</u>	*	*	<u>I</u>	<u>V</u>	<u>U</u>
6001-7000	<u>C</u>	*	*	<u>E</u>	*	*	<u>I</u>	<u>V</u>	<u>U</u>
7001-7134	*	<u>C</u>	*	<u>E</u>	*	*	<u>I</u>	<u>V</u>	*

* No lines
(folios lost or
destroyed,
lines omitted)

The Princeton Charrette Project: tabla de testimonios

La veta filológica del proyecto es innegable, y busca acercarse a la *recensio* de todos los testimonios, la transcripción de los mismos y la reproducción digital de los folios conservados:

Figura T6 :: Charrette | Home | Pages | Figures | Words | Episodes | XML

Mspage Search

GO Restablecer

Code	MS	Folio	Side	Mj. ltr	Mj. hght	Text
A-196-r	<input type="checkbox"/> A	<input type="checkbox"/> 1	<input type="checkbox"/> r	<input type="checkbox"/> A	<input type="checkbox"/> 1	(1) FU
A-196-v	<input type="checkbox"/> C	<input type="checkbox"/> 2	<input type="checkbox"/> v	<input type="checkbox"/> B	<input type="checkbox"/> 2	Line n.
A-197-r	<input type="checkbox"/> E	<input type="checkbox"/> 3		<input type="checkbox"/> C	<input type="checkbox"/> 3	
A-197-v	<input type="checkbox"/> F	<input type="checkbox"/> 4		<input type="checkbox"/> D	<input type="checkbox"/> 4	
A-198-r	<input type="checkbox"/> G	<input type="checkbox"/> 5		<input type="checkbox"/> E	<input type="checkbox"/> 5	
A-198-v	<input type="checkbox"/> I	<input type="checkbox"/> 6		<input type="checkbox"/> F	<input type="checkbox"/> 6	
A-199-r	<input type="checkbox"/> T	<input type="checkbox"/> 7		<input type="checkbox"/> H	<input type="checkbox"/> 8	
A-199-v	<input type="checkbox"/> V	<input type="checkbox"/> 8		<input type="checkbox"/> I		
A-200-r		<input type="checkbox"/> 9		<input type="checkbox"/> K		
A-200-v		<input type="checkbox"/> 10		<input type="checkbox"/> L		
A-201-r		<input type="checkbox"/> 11		<input type="checkbox"/> M		
A-201-v		<input type="checkbox"/> 12		<input type="checkbox"/> N		
A-202-r		<input type="checkbox"/> 13		<input type="checkbox"/> O		
A-202-v		<input type="checkbox"/> 14		<input type="checkbox"/> P		
A-203-r		<input type="checkbox"/> 15		<input type="checkbox"/> Q		
A-203-v		<input type="checkbox"/> 16		<input type="checkbox"/> S		
A-204-r		<input type="checkbox"/> 17		<input type="checkbox"/> T		
A-204-v		<input type="checkbox"/> 18		<input type="checkbox"/> U		
A-205-r		<input type="checkbox"/> 19		<input type="checkbox"/> V		
A-206-r		<input type="checkbox"/> 20				

Mspage Object

A-196-r



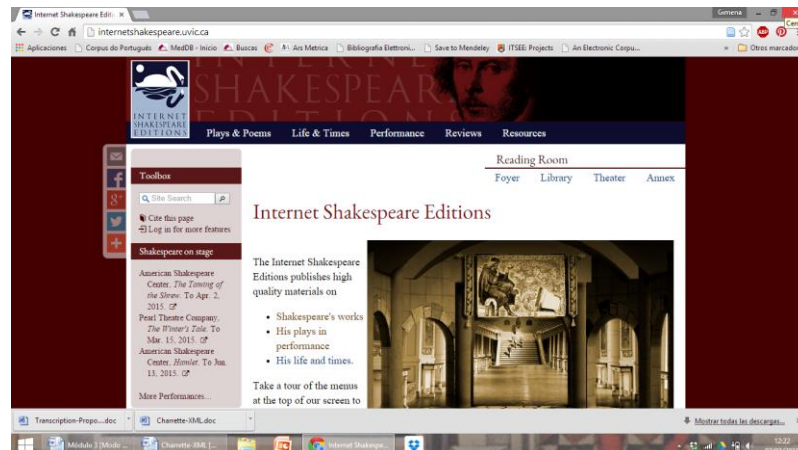
Majuscules: A (31), L (63), A (87), S (114), O (137), S (173)
 Line Span: 31 to 180 (FU)
 Transcription:
 31 (a,1): A un jor dunea}cen}{[apost]o s
 32 (a,2): fu uenu} deu[apost]} cart
 [apost]o[apost]s
 33 (a,3): liroi} .ar. et}tneueots
 34 (a,4): cort ml[apost]t rice achamolotas
 35 (a,5): tele comme au jor e}tuts
 36 (a,6): ap-hbar} mang[apost]
 ne}eremuts
 37 (a,7): liroi}punct d[apost]et-e }e}
 compaigno-hbar}s
 38 (a,8): ml[apost]t ot en}ale baro-
 hbar}s
 39 (a,9): et2 }ifu laroina en}anbles
 40 (a,10): S iot aueuc li ceme }anbles
 41 (a,11): Mai[apost]te bele dame
 cortoi} }es
 42 (a,12): B n-hbar} parlant enlangue
 francoi} }es
 43 (a,13): et2 Kell} qui ot }erui a}

Status Menu target: left :: Left target: right :: Right target: left

Como decíamos, se trata de un proyecto de larga data, que nació en 1990 y, aunque no ha sido demasiado actualizado en los últimos años (en contenido y en el diseño de su interfaz, algo anticuado para estos días), es una propuesta que no solo acerca imágenes de manuscritos (aunque las digitalizaciones no son todas, de igual calidad) y edición de los mismos, pero también la metodología de trabajo y la tecnología utilizada para montar la investigación. Así, por ejemplo, sabemos que el proyecto usa distintos [módulos de TEI](#) customizados por sus filólogos e informáticos para la edición de textos franceses antiguos y, como puede verse en la captura de pantalla de más arriba, el usuario tiene acceso a la codificación en XML-TEI de todos los textos.

[Shakespeare Internet Editions](#), coordinado por Michael Beste, es otro proyecto pionero en edición digital en el Reino Unido. Comenzado en 1996, se caracteriza por su desarrollo y actualización constante. Una de las últimas novedades del proyecto es la de su versión web optimizada para teléfonos móviles y tabletas. Al igual que el proyecto

de Princeton, se encuentra a mitad de camino entre una propuesta de edición y un archivo digital a mayores —formado por textos, versiones digitales de los facsímiles, pero también fragmentos de vídeo y audio de representaciones, así material crítico sobre cada obra— y se brinda al lector la posibilidad de explorar los textos de Shakespeare desde diferentes ediciones textuales de forma simultánea, añadiendo, en este caso, una variada cantidad de recursos multimedia. Diferencia a ambas propuestas el hecho de que, en este caso, no se ofrecen ediciones propias sino que se recogen trabajos editoriales anteriores, editados digitalmente. Tampoco se acerca una documentación clara y completa, en lo editorial y lo técnico, que permita al usuario adentrarse con rigurosidad académica en el sitio. Como decíamos antes, estas cuestiones son de sumo interés a la hora de evaluar los diferentes tipos de proyectos digitales que nos ofrece la web, más o menos cercanos a una disciplina académica como las Humanidades Digitales:



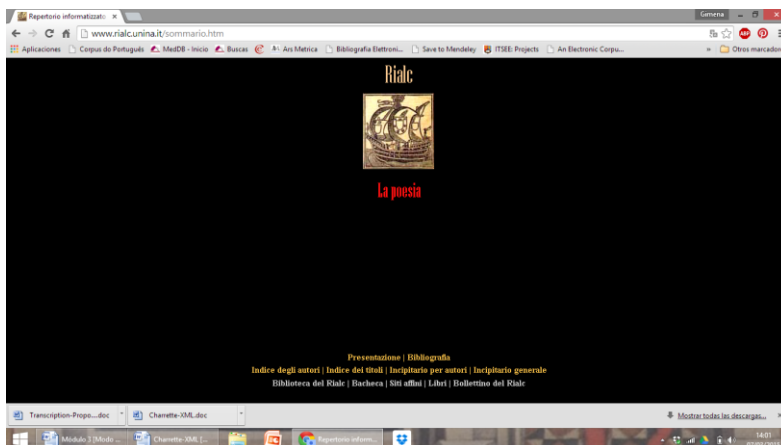
Internet Shakespeare editions

1.3.4. Repertorios textuales y métricos⁵

De clara raíz filológica es *Rialc*, el [Repertorio Informatizado de la Antigua Lírica Catalana](#), que funciona desde 1999 dirigido por Costanzo

⁵ Un trabajo más exhaustivo sobre este tipo de recursos lo encuentras en [La codificación informática del sistema poético medieval castellano, problemas y propuestas en la elaboración de un repertorio métrico digital: ReMetCa](#).

de Girolamo, desde la Università di Napoli Federico II y con apoyo ministerial. El proyecto se encuentra exhaustivamente [documentado](#) desde lo filológico, aunque no desde lo técnico:



RIALC

Otro tipo de repertorios filológicos que han encontrado un lugar en la web gracias a la utilización de bases de datos que permiten un trabajo a gran escala es el de los repertorios métricos. La clasificación precisa y expresada en unos pocos datos, sumada al objetivo de reagrupar material de acuerdo a unas determinadas coordenadas, encuentra en la estructura de la base de datos y en el espacio de la pantalla de ordenador, un modo mucho más completo de aprovechamiento de las coincidencias en las búsquedas métrico-rimáticas de los repertorios en papel.

Los primeros ejemplos de repertorios métricos digitales, como el de Horvath para la poesía húngara medieval, son buena muestra del modo en el que los repertorios métricos se vieron beneficiados por el método y la forma del mundo digital, como bien lo constata este cuadro comparativo entre el *Naetebus* en versión papel (1891) y de la base de datos contemporánea accesible desde la web en el [Nouveau Naetebus](#), la que se le han añadido otros *corpus* que completan el universo de la poesía narrativa francesa, como el trabajo de Långfors o Sinclair. El proyecto está dirigido por Levente Sélaf:

VIII. aaaa. 12 S.

1. La vie et l'histoire du mauvais riche home.

Devant l'uis au riche home le ladre s'arresta 25 Str.
Por la grant fain qu'il ot forment se dementa.

Hs. Paris, Nat. bibl. f. fr. 957 Bl. 118.

Mit den ersten 5 Zeilen angeführt 1848 von P. Paris, Mss. fr. VII, 339. — Über ein in paarweis reimenden Achtsilbfern verfasstes Livre de l'exemple du riche homme et du ladre vgl. van Hamel, Rencl. S. CXCIX und Rom. XVIII, 642.

Hs.: Anfang des 15. Jahrh.

Fragmento del repertorio en papel de Gotthold Naetebus, *Die nicht lyrischen Strophenformen des Altfranzösischen*

Concordance avec les répertoires:	Långfors : 90 Raynaud - Spanke : - Mölk - Wolfzettel : -
Incipit: Variantes de l'incipit:	Devant l'uis au riche homme le ladre trespassa Devant l'uis au riche homme le ladre trespassa, Devant l'uis au riche homme le ladre s'arresta,
Source:	Cambridge, Magdalene College, Coll. S. Pepys 1938 Paris, BNF fr. 957
Titre: Auteur:	(texte anonyme) Cambridge, Magdalene College, Coll. S. Pepys 1938
Auteurs secondaires: Versification:	texte entièrement en vers poème strophique poème isostrophique
Nombre des strophes:	28 Cambridge, Magdalene College, Coll. S. Pepys 1938 Paris, BNF fr. 957
Structure strophique:	12aaaa
Irrégularités métriques (numéro de la strophe - structure strophique):	8 - 12aa Paris, BNF fr. 957 Cambridge, Magdalene College, Coll. S. Pepys 1938 10 - 12aaa Paris, BNF fr. 957 Cambridge, Magdalene College, Coll. S. Pepys 1938 16 - 12aaa Paris, BNF fr. 957 Cambridge, Magdalene College, Coll. S. Pepys 1938

Le Nouveau Naetebus — [Répertoire des poèmes strophiques non-lyriques en langue française d'avant 1400](#)

Mientras que las búsquedas en papel limitan las coincidencias al material que vamos encontrando página a página, las realizadas en el formato digital arrojan datos que, a simple golpe de vista, pueden relacionarse entre sí. Por ejemplo, si buscásemos una estrofa amplísimamente explotada, como la cuarteta de alejandrinos monorrimos, en un repertorio en papel como el *Naetebus*, deberíamos escudriñar material a lo largo de más de una veintena de páginas. Por el contrario, esta búsqueda en el *Nouveau Naetebus* nos arroja todas las coincidencias a partir de incipits que se muestran en una única pantalla por la que el interesado se mueve, y de allí decide si

profundizar o no su búsqueda, a partir de las diferentes fichas que se ofrecen para cada poema. Para los interesados en las relaciones intertextuales y/o de contrafactura este primer paso de cotejo entre incipits que facilita la versión digital del repertorio resulta fundamental para emprender un análisis formal completo entre las diferentes piezas de ese corpus. Se suma a todo ello, como ventaja fundamental, el hecho de que las búsquedas son mucho más sencillas en el espacio digital, ya que, a través de búsquedas simples se puede acceder a distintos campos en función del tipo de búsqueda, mientras que la consulta del libro exigía un elevado conocimiento de estas materias para saber cómo estaba organizado, cómo buscar en los índices, qué significaban las abreviaturas y dónde se encontraban los poemas.

Por nombrar unos pocos ejemplos de este tipo de recursos, traemos aquí un proyecto pionero, que desde 1996 ha innovado en la edición digital de textos líricos su presentación junto a otros recursos (repertorio métrico, digitalización de imágenes) en la web. Se trata de [MedDB2, Base de Datos de la Lírica Gallego-Portuguesa](#). Mercedes Brea dirige este proyecto que se aloja en el Centro de Investigación Ramón Piñeiro de Santiago de Compostela. Como decíamos, MedDB fue pionera en ofrecer al investigador todo el corpus profano gallego-portugués acompañado de información bibliográfica y de los esquemas métrico-rimáticos del Repertorio *métrico de la lírica gallego-portuguesa* de Tavani, que aquí encontró, desde nuestro punto de vista, su máxima explotación.

MedDB — Base de datos da Lírica profana galego-portuguesa

Un repertorio también híbrido, aunque no textual es el de la base de datos de las [Cantigas de Santa María](#), dirigida por Stephen Parkinson en la Universidad de Oxford, que además de incluir una amplia información sobre poemas, manuscritos, milagros y distintas versiones de cada uno de los episodios incluye información métrico-rimática para cada pieza. Cabe destacarse que este recurso no permite el cotejo de los distintos esquemas métrico-rimáticos de las cantigas sino que estos datos son parte de fichas que se ofrecen para cada uno de los textos del rey Sabio de Castilla:

The Oxford Cantigas de Santa Maria Database

Cabe destacar que todos estos proyectos de investigación, buscan enmarcarse, desde su metodología en un trabajo que híbrida filólogos y técnicos, en las Humanidades Digitales. Pero, a pesar de que

dan una precisa presentación y justificación filológica de su contenido, hasta el momento no han acercado al usuario material relacionado con documentación técnica.

En este sentido, el proyecto [ReMetCa](#), Repertorio métrico digital de la poesía medieval castellana, ha declarado desde 2012, tanto en los artículos que más arriba mencionamos como en el sitio web del proyecto, los planteamientos filológicos y técnicos del proyecto, desde los fundamentos acerca de la utilización de TEI hasta la declaración de los campos buscables de nuestro prototipo. Este proceder encontró una justificación teórica en el trabajo de Patrick Sahle de 2014, en colaboración con Georg Vogeler y miembros del grupo IDE, [Criteria for Reviewing Scholarly Digital Editions, version 1.1](#). Allí los investigadores estipularon —con especial atención a las ediciones digitales académicas— los criterios que los proyectos enmarcados en las Humanidades Digitales debían acercar a sus usuarios. De allí, las preguntas que entendemos debemos formularnos a la hora de abordar un sitio web que contenga un proyecto de este tipo a la hora de reseñar sus características:

¿Cuál es el ámbito académico y /o disciplinario del objeto? ¿Cuenta con financiación gubernamental, institucional, privada? ¿Se trata de un objeto cerrado o en proceso? ¿Cómo se relaciona con otros recursos impresos o digitales, predecesores o proyectos similares? ¿Con qué fuentes y/o documentos, textos, herramientas u otros objetos se trabaja? ¿Qué aporta el objeto al estado actual de conocimientos sobre el tema? ¿Se trabaja con imágenes, transcripciones, textos completos, documentos digitalizados, ediciones propias, bibliografía, etc.) ¿Hay una descripción de las particularidades, los objetivos y/o métodos del objeto? ¿A qué campos de la investigación contribuye el objeto? ¿Cuál es/son el/los objetivo/s del objeto? ¿Se cumple con los objetivos propuestos? ¿Hay un enfoque metodológico claro?⁶

Puedes añadir algo aquí sobre el libro de la Pierazzo de ed. Genética, el trabajo de recopilación de SAhle

⁶ Material confeccionado por las autoras para la asignatura.

1.4. Big data y big corpus y la necesidad de una edición fiable

Alejandro Piscitelli trajo el siguiente ejemplo al [THATCamp Buenos Aires](#) de julio de 2013: un contundente gráfico elaborado a partir del trabajo señero de [Elizabeth Eisenstein](#) (1995) que mostraba [la diseminación de la imprenta en Europa en las últimas cinco décadas de 1400](#) y revelaba que su implantación tuvo lugar primero en las ciudades eminentemente comerciales, y sólo después en las universidades. El ejemplo, además de ser interesante para observar los movimientos exógenos y endógenos en las Humanidades Digitales y el afuera-adentro del mundo académico y sus interpenetraciones, obraba asimismo bien como muestra de un trabajo con un gran volumen de datos. Lev Manovich, años después del trabajo de Eisenstein, demostró que, disponiendo de un conjunto de herramientas bien elegidas es posible explorar [enormes colecciones de imágenes](#) de manga o de la revista *Time*.



MOTI. Museum of the image

El mismo [Manovich](#) nos recuerda en un artículo suyo que en el mes de junio de 2008 la revista *Wired*, en el volumen especial *The petabyte age*, afirmaba que:

Our ability to capture, warehouse, and understand massive amounts of data is changing science, medicine, business, and technology. As our collection of facts and figures grows, so will the opportunity to find answers to fundamental questions.

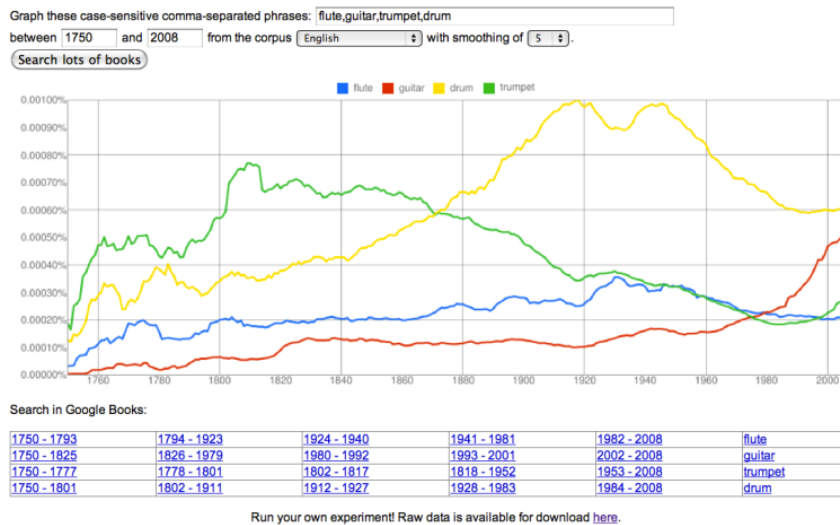
[Nuestra capacidad para capturar, almacenar y comprender enormes cantidades de datos está cambiando la ciencia, la medicina, los negocios y la tecnología. A medida que nuestra colección de datos y cifras crece, también lo hace la oportunidad de encontrar respuestas a preguntas fundamentales.]

Estas enormes cantidades de datos a las que Manovich se refiere son ese fenómeno del Big Data que hoy puebla la web y del que hablan desde políticos, especialistas en Marketing o Humanistas. Según Wikipedia, el término [Big Data](#) hace referencia a “conjuntos de datos que crecen tan rápidamente que no pueden ser manipulados por las herramientas de gestión de bases de datos tradicionales”. Este [Prezi](#) también te cuenta de qué se trata. Pero el concepto de Big Data se extiende más allá de lo que estamos acostumbrados a entender por “grandes volúmenes de información”. Siempre que buscamos en *Google*, enviemos un email, escribimos una entrada en *Facebook*, usamos un teléfono móvil, una tarjeta de crédito, dejamos detrás de nosotros una enorme cantidad de datos, huellas digitales y registros que ofrecen una información muy valiosa. Así, esta masa heterogénea de Big Data se está convirtiendo en el punto de inflexión de quienes —siendo esos “quienes” *Google*, el Banco Santander, la BNE o un proyecto de investigación— consiguen obtener ventaja procesando correctamente uno de sus principales activos: la información que se genera. Consecuentemente, el tamaño no es el único problema al que

nos enfrentamos cuando abordamos el estudio sistemático de las grandes masas de datos que nos ofrece la web o los productos digitalizados: más allá de almacenar los datos, es necesario consultar, gestionar y analizar correctamente toda esta información. Programadores, científicos y también humanistas se encuentran regularmente ante problemas generados por el no siempre acertado almacenamiento y procesamiento de big data. Y esto repercute, claro está, en la relevancia de los productos académicos derivados de la investigación (llámese artículo científico, ponencia, libro). Un caso sumamente interesante para el ámbito de las Humanidades es el que plantea Francesc Llorens en [este artículo](#) dedicado a investigar los alcances y límites de un análisis cultural e histórico con herramientas como Google Ngram viewer en el ámbito académico.



Proceso de optimización y validación de los datos



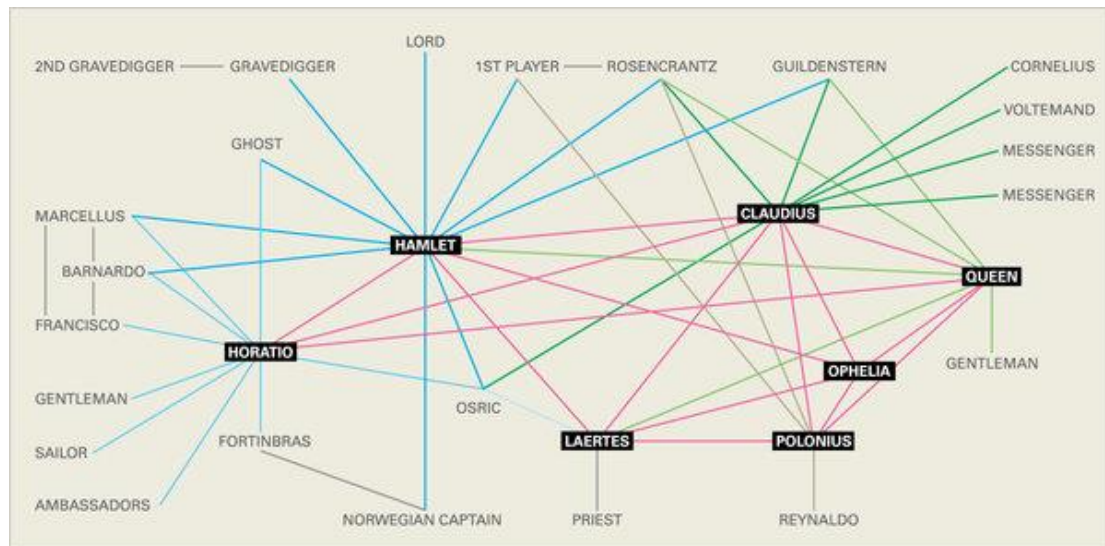
Google Ngram Viewer

Lo primero que salta a la vista es que los datos por sí mismos son incapaces de producir un beneficio. Solo las herramientas acertadas, y los humanos que sepan leer y revisar lo que estas nos acerquen podrán hacer buen uso de las grandes masas de *datos* y transformarlos en *textos*. Es por esto que además de la tecnología necesaria para analizar tal volumen de datos, es necesario desarrollar metodologías y competencias que no solo surfeen las superficies de los datos y sus estadísticas.

Dentro del ámbito de la literatura y su análisis, quien ha profundizado sobre la aplicación de técnicas como el data mining⁷, la visualización de información y la analítica visual es el ya mencionado Franco Moretti. Desde su [Lab en la Universidad de Stanford](#), Moretti ha desarrollado la teoría de la lectura distante (2013) pero antes había trabajado sobre otros interesantes proyectos como el [Atlas de la literatura europea](#) o el maravilloso [Gráficos, mapas, árboles](#), lectura obligatoria antes de que sigas con estas líneas. Los tres, creemos,

⁷ En el sentido de análisis automático o semi-automático de grandes cantidades de datos para extraer patrones hasta ese momento desconocidos, como grupos de registros de datos (análisis clúster) que den lugar al descubrimiento de anomalías, derivaciones, etc. Puedes ver más información, por ejemplo, en los temas que desarrolla el [Data Mining Program](#) de la Universidad de Florida.

relacionan perfectamente el trabajo con grandes datos y con grandes *corpus* textuales y problematizan nuestra lectura filológica sobre el dato. Empecemos por el concepto de *lectura distante* (*distant reading*) de Moretti, que podría semejarse a un plano general en el cine. ¿Qué hace que debamos hoy pensar en este tipo de lectura abarcativa y a gran escala? pues el crecimiento exponencial de la cantidad de información, la proliferación de diferentes soportes de lectura y herramientas de producción, y almacenamiento y circulación de estos datos. Todos ellos demandan un cambio en el modo de leer que heredamos de la imprenta, y al que Moretti denomina “lectura cercana” (*close reading*), algo similar a un *close-up* en una película. La idea de Moretti es simple: las grandes escalas a las que nos enfrenta el medio digital hace que necesitemos cuantificar la literatura y leerla en base a tres nuevas disposiciones provenientes de otras disciplinas científicas: la historia cuantitativa, la geografía y la teoría evolutiva. Así, historia cuantitativa, la geografía y la teoría evolutiva se transforman en los gráficos, los mapas y los árboles del trabajo que más arriba mencionábamos. A partir de ellos el investigador de Stanford realiza tres tipos de operaciones: periodizar (y mostrarlo en gráficos) y describir ciclos, localizar (en mapas) y agrupar o “clusterizar” (en árboles) relaciones entre personajes, palabras de alta frecuencia de aparición que permiten identificar los sentimientos, colores, lugares mayormente referenciados por todas las novelas de un período u otros que funcionan en la periferia por alguna razón específica. De este modo, la metodología de la lectura distante permite hacer una lectura de la totalidad de la producción literaria de un período dado, y no simplemente de una parte (que puede ser la mejor, la consagrada, pero que no da cabal cuenta de todo el período).



Tomado del [LitLab](#) de la U. de Stanford.

Cuando realizó el *Atlas de la literatura europea*, Moretti se hizo la siguiente pregunta: ¿puede un crítico desarrollar cabalmente si durante el siglo XIX se publicaron en Europa entre 20.000 y 30.000 novelas? ¿Cuántas podría leer un crítico de ese total? ¿Qué queda dentro y qué queda fuera de ese corpus? La aplicación de lo que él llama objetos artificiales (los gráficos, los mapas y los árboles), que proceden de otras ciencias lejanas a la Literatura y que son el resultado de un proceso de abstracción, nos permite tomar distancia del texto como objeto monumental y casi totémico y ver las relaciones entre diferentes textos: ¿qué une a determinados textos?, ¿qué regularidades se encuentran en determinados patrones que pueden surgir del análisis de estos textos?

En el caso de nuestro proyecto, [ReMetCa](#) (Repertorio Métrico digital de la Poesía Medieval Castellana), una herramienta digital online de acceso abierto diseñada para realizar búsquedas complejas simultáneas en los patrones métricos y rimáticos de toda la poesía castellana medieval (que podríamos balizar entre fines del siglo XII y principios del XVI), los conceptos de edición digital académica y una gran cantidad de datos trabajaron de la mano desde un comienzo. El proyecto de investigación que sustenta ReMetCa tiene como objetivo

integrar los estudios tradicionales en Filología (especialmente los que respecta al estudio de la métrica) con las Humanidades Digitales, revisar y clasificar el corpus castellano en un marco digital de híbrido que incorpore la edición de textos poéticos con etiquetas TEI-XML en una base de datos relacional que funciona en conjunto con un vocabulario controlado en medieval Poesía castellana. La base de datos, por ende, trabaja con un enorme volumen de datos, no solo de textos poéticos, sino de fórmulas métrico-rimáticas, etc.

Un caso interesante de estudio que ilustra la necesidad de la lectura distante en las nuevas aproximaciones a la literatura fue nuestro enfoque digital para el *Cancionero de Baena*, un gran libro de canciones que contiene casi seiscientos poemas transcritos y compilados en la primera mitad del siglo XV por Juan Alfonso de Baena, escribano de la corte del rey Juan II Jornadas de Castilla. Los datos recuperados del etiquetado y clasificación de los patrones métricos y rimáticos de una gran parte de este —relativo al corpus de los *antiguiores* o los poetas mayores, y los que compuso sus textos sobre todo en la segunda mitad del siglo XIV y principios del XV— nos llevaron a algunos resultados interesantes en el área de los Estudios Hispánicos dedicados a la métrica. Por un lado, descubrimos que los *antiguiores* componen una gran parte de sus poemas utilizando un patrón casi desconocido por sus predecesores, los trovadores gallego-portugueses: la octava octosílabos (isométrica o heterométrica). Además, fuimos capaces de identificar algunos grupos o ciclos de poemas compuestos sobre la base de la imitación o contrafactura métrica y rima (→ 4x8 @ 8). Pasamos así de intuiciones a certezas, de la comparación de pequeños grupos de textos al bosque textual de los cancioneros, y asimismo, de lo descriptivo a la dimensión connotativa.

Captura de pantalla del prototipo del buscador de ReMetCa.

Editar textos fiables, donde se sopesa, por ejemplo, en este caso, tanto texto como forma, resulta fundamental para más tarde poder recuperar datos certeros. Por eso, queremos cerrar este apartado parafraseando a Jordi Torres i Viñals, de quien recomendamos toda la lectura de [Del Cloud computing a la Big Data](#): “Parece que fue Albert Einstein quien dijo que ‘la información no es conocimiento’. Cuánta razón tenía, los datos necesitan ser analizados para que se les pueda extraer el valor que contienen” (2012: 29).

Puedes referir luego a tu ppt? Así cerramos

1.5. Bibliografía citada y recomendada

FLANDERS, J. (2009) [The Productive Unease of 21st-century Digital Scholarship](#), *Digital Humanities Quarterly*, Summer, Volume 3, Number 3.

- GOLD, M. (2012) [Debates in Digital Humanities](#), Minnesota, University of Minnesota Press.
- GONZÁLEZ BLANCO-GARCÍA, E. Actualidad de las Humanidades Digitales y un ejemplo de ensamblaje poético en la red: ReMetCa, Cuadernos Hispanoamericanos, 761, noviembre, 2013 (en el sitio de la autora en Academia.edu)
- GONZÁLEZ-BLANCO GARCÍA, Elena; MARTÍNEZ CANTÓN, Clara; MARTOS, María Dolores – RIO RIANDE, M^a. Gimena del: “Una propuesta de integración del sistema de formularios de bases de datos MYSQL con etiquetado TEI: ReMetCa, Repertorio digital de la métrica medieval castellana”, en López Poza Sagrario y Pena Sueiro, Nieves (eds.), *Humanidades Digitales: desafíos, logros y perspectivas de futuro*, Janus [en línea], Anexo 1, 2014, pp. 209-219. ISSN 2254-7290.
<http://www.janusdigital.es/anexos/contribucion.htm?id=19>
- HUTCHINS, John: [“The first public demonstration of machine translation: the Georgetown-IBM system, 7th January 1954”](#) (2006).
- IRIZARRY, E. (1997) *Informática y literatura: análisis de textos hispánicos*, Barcelona, Proyecto A Ediciones.
- JOCKERS, M. (2013) *Macroanalysis: Digital Methods and Literary History*. University of Illinois Press.
- JOCKERS, M. & FLANDERS, J. (2013) [A matter of scale](#)
- KIRSCHENBAUM, M. (2010) [“What is Digital Humanities and What is Doing in English Departments?”](#), *ADE Bulletin* 150, pp. 1-7.
- LATOUR, B. Opening Plenary, Digital Humanities 2014 (DH2014).
<http://dh2014.org/videos/opening-night-bruno-latour/>
- LUCÍA MEGÍAS, J. M. (1998) “Editar en Internet (che quanto piace il mondo è breve sogno)”, *Incipit XVIII*, pp. 1-40.
- LLORENS, F. (2013) [Google ngram viewer y la analítica cultural ¿Pueden las herramientas de análisis algorítmico de la cultura confirmar hipótesis sobre hechos históricos?](#)
- LLORENS, F. [Low-intensity Philosophy](#) (blog en español).

- MARCOS MARÍN, F. (1986) "Metodología informática para la edición de textos" *Incipit 6*, pp. 185-197.
- MARCOS MARÍN, F. (1994) *Informática y Humanidades*, Madrid, Gredos, Grandes Manuales.
- MORETTI, F. *Distant Reading*. London: Verso, 2013.
- RIO RIANDE, Gimena del. 2015a. ["Humanidades Digitales. Mito, actualidad y condiciones de posibilidad en España y América Latina"](#). *ArtyHum*, monográfico 1.
- RIO RIANDE, Gimena del. 2015b. [¿De qué hablamos cuando hablamos de Humanidades Digitales?](#) Blog Docentes en línea
- SCHREIBMAN, S.; SIEMENS, R. & UNSWORTH, J. (2004) [A Companion to Digital Humanities](#), Oxford, Blackwell.
- SPENCE, Paul, ["Centros y fronteras: el panorama internacional de las humanidades digitales"](#).
- SPENCE, P.; GONZÁLEZ-BLANCO, E. (2014) "A historical perspective on the digital humanities in Spain", en: *The Status Quo of Digital Humanities in Spain*, H-Soz-Kult, <http://www.hsozkult.de/debate/id/diskussionen-2449>.
- TALEB, N. [Beware the Big Errors of 'Big Data'](#)
- TERRAS, M.; NYHAN, J. & VANHOUTTE, E. (eds.) (2013) *Defining Digital Humanities. A Reader*, Farnham, Ashgate Publishing.
- TORRES Y VIÑALS, J. [Del cloud computing a la Big Data](#).
- UNSWORTH, J. [What Is Humanities Computing and what is not?](#) (2002).
- VANHOUTTE, E. [The mind tool](#)
- VANHOUTTE, E. (2013) ["The Gates of Hell. History and Definition of Digital | Humanities | Computing"](#), en: *Defining Digital Humanities. A Reader*, M. Terras, J. Nyhan, J. & E. Vanhoutte (eds.), Farnham, Ashgate Publishing, pp. 119-156.