

# Humanidades Digitais: novos desafios e oportunidades

(Digital Humanities: new challenges and opportunities)

Dália Guerreiro, Universidade de Évora/CIDEHUS, Portugal

José Luís Borbinha, IST / INESC-ID, Portugal

**Resumen:** *As Humanidades Digitais são uma área do conhecimento dinâmica e interdisciplinar, assente na conexão e articulação de domínios do conhecimento com diferentes metodologias de investigação. No âmbito das Humanidades Digitais, têm particular relevância as bibliotecas digitais construídas para os investigadores. Atualmente, existem dois tipos de bibliotecas digitais, as que apresentam o acervo em modo de imagem e as que apresentam a informação em modo texto. Estas últimas, se não forem devidamente estruturadas, não permitem a reutilização da informação através da programação informática. O modelo das bibliotecas digitais está em evolução e procura adaptar-se às atuais exigências dos especialistas e do público.*

**Palabras clave:** *humanidades digitais, bibliotecas digitais*

**Abstract:** *The Digital Humanities are a dynamic and interdisciplinary area of knowledge, based on the connection and articulation of knowledge's domains with different research methodologies. Within the Digital Humanities, digital libraries, built for researchers, are particularly relevant. Currently, there are two types of digital libraries, those presenting the collection in image mode and those presenting information in text mode. These ones, if not properly structured, do not allow the reuse of information through computer programming. The model of digital libraries is evolving and seeks to adapt to the current requirements of experts and the public.*

**Keywords:** *digital humanities; digital libraries*

## Para a definição do conceito

*Revista Internacional del Libro, Digitalización y Bibliotecas*

Volume 2, Número 2, 2014, <<http://sobrellibro.com>>, ISSN 2255-2871

© Common Ground. Dália Gerreiro, José Luís Borbinha. Todos os direitos reservados.

Permisos: [suporte@commongroundpublishing.com](mailto:suporte@commongroundpublishing.com)

**A**s Humanidades Digitais consolidam-se como um novo campo de saber e, como tal, apresentam os indicadores de um novo campo firmado através de numerosas associações, centros de estudos, departamentos, programas de ensino, revistas especializadas e congressos:

Existen varios indicadores de la consolidación de un nuevo campo de estudio. Entre ellos: la formación de asociaciones, organizaciones y centros especializados, la creación de programas académicos de enseñanza, la celebración de congresos y la publicación de revistas y libros especializados. El campo de las Humanidades Digitales se consolida cada vez más y existe a nivel internacional una comunidad importante que se identifica como “humanista digital. (Galina Russell, 2011a, p. 4)

Esta nova área de investigação e desenvolvimento, transversal a áreas como as humanidades e as ciências da informação e da computação (Schreibman, Siemens, & Unsworth, 2004), (Borgman, 2009) e (Evans & Rees, 2012), orienta uma parte substancial da pesquisa para a identificação, a definição e a descrição dos métodos, das boas práticas e das normas que o sustentam.

O alcance das humanidades digitais ultrapassa largamente a mera transferência do analógico para o meio digital, centrando-se no desafio epistemológico e na articulação com os conhecimentos e os métodos utilizados nas ciências humanas com o mundo digital.

As bibliotecas, na sua forma analógica, foram essenciais para o desenvolvimento das humanidades ao longo dos séculos, assumindo-se como instituições guardiãs da memória e do conhecimento que ia sendo produzindo. Atualmente, esse papel foi ampliado, pois, além de continuarem a ter funções de custódia, alargada ao mundo digital, também devem incluir a produção nos novos meios.

A biblioteca universitária tende a evoluir para uma configuração como centro de competências. “La biblioteca es una parte fundamental de la estructura universitaria y puede proporcionar la estabilidad e institucionalidad que se requiere para la manutención de estos recursos académicos.” (Galina Russell, 2011b, p. 7) Além da mediação, da organização e de todas as funções inerentes à gestão da biblioteca tradicional e digital, vai desenvolver outras atividades e competências, por exemplo, quer na área da publicação digital, dando apoio na criação de wikis, blogues, etc., quer nas questões inerentes ao direito de autor e à criação de normas e/ou boas práticas para os projetos de humanidades digitais.

## **Bibliotecas digitais**

A UNESCO aprovou, na 36ª Conferência Geral (de 25 de Outubro a 10 de Novembro de 2011, em Paris), o *Manifesto da IFLA/UNESCO sobre as bibliotecas digitais*. Neste texto, define-se que a biblioteca digital tem de ser constituída de acordo com as normas internacionais, ser persistente no tempo e manter os conteúdos devidamente organizados e estruturados, usando as novas tecnologias no acesso e na recuperação da informação. As bibliotecas digitais têm a missão de fornecer o acesso direto à informação.

Una biblioteca digital es una colección en línea de objetos digitales de buena calidad, creados o recopilados y administrados de conformidad con principios aceptados en el plano internacional para la creación de colecciones, y que se ponen a disposición de manera coherente y perdurable y con el respaldo de los servicios necesarios para que los usuarios puedan encontrar y utilizar esos recursos.

La misión de las bibliotecas digitales consiste en proporcionar acceso directo a recursos informativos, digitales y no digitales, de manera estructurada y fiable, para de esa manera vincular la tecnología de la información, la educación y la cultura en las bibliotecas actuales. (“Manifiesto de las IFLA/UNESCO Sobre las Bibliotecas Digitales | IFLA,” 2011)

As bibliotecas digitais surgiram na década de 1990, quando a tecnologia o permitiu. Até cerca do ano 2000, houve um período experimental, durante o qual foram sendo definidas as boas práticas para a digitalização. Como um dos grandes marcos dessa época, refira-se a Library of Congress<sup>1</sup> que, em 1994, iniciou o programa de digitalização e, em 1996, os investigadores de ciências sociais começaram a utilizar o termo *digital library*. (Cfr. Besser, 2002) Na maioria dos casos, as bibliotecas digitais foram criadas por bibliotecários, arquivistas (Cfr. Lucía Megías, 2010, p. 373) e informáticos. Um reflexo desta parceria ainda hoje é perceptível na pesquisa da informação que é feita essencialmente através dos elementos da ficha bibliográfica. Foi o início da disponibilização em linha do obras digitalizadas maioritariamente em formato imagem, ainda com carácter embrionário e experimental (Choudhury & Seaman, 2008).

A partir do ano 2000 procedeu-se à consolidação das bibliotecas digitais, com a criação de normas e boas práticas. Em 2005 começou a discussão acerca da Biblioteca Digital Europeia que, em 2008, foi colocada em linha com a designação de Europeana<sup>2</sup>.

---

<sup>1</sup> Disponível em: <http://www.loc.gov/library/libarch-digital.html>

<sup>2</sup> Disponível em: <http://www.europeana.eu/>

Desde o início deste processo de construção de bibliotecas digitais, distinguem-se duas tipologias:

- As de serviços básicos, que disponibilizam o acesso através da pesquisa bibliográfica e os acervos em formato imagem (PDF, JPG, PNG, etc.);
- As de serviços de valor acrescentado, que disponibilizam as obras em formatos estruturados (como, por exemplo, o texto transcrito e marcado), permitindo uma pesquisa mais elaborada, quer através da ficha bibliográfica ou dos metadados, quer em texto livre, e possuem ferramentas que facilitam a análise dos documentos.

As bibliotecas digitais vocacionadas para a investigação devem dar particular atenção à utilização e à reutilização da informação para a criação de novo conhecimento, sem esquecer a descrição e a estruturação do objeto digital e a conservação e a preservação da informação ao longo do tempo.

Atualmente, os investigadores na área das humanidades são também autores de projetos que visam a disponibilização de recursos digitais relacionados com as investigações em que estão envolvidos, ainda que “estos recursos se crean fuera del contexto formal editorial o bibliotecario, temas como la clasificación, registro en catálogos, formación de colecciones, notificación al lector, manejo de derechos, diseminación, mantenimiento y preservación a largo plazo, son frecuentemente pasados por alto.” (Galina Russell, 2011b, p. 3) Por isso, cabe ao bibliotecário a tarefa de auxiliar os investigadores em humanidades a criar produtos que, além de serem vantajosos para a investigação, cumpram as normas e as diretivas e fomentem a divulgação e a preservação.

As bibliotecas digitais, “cuja acessibilidade tornou mais rápida e menos árdua a actividade do investigador, que deixou de estar condicionado aos horários de arquivos e bibliotecas convencionais” (Gonçalves & Banza, 2013, p. 2) apresentam algumas vantagens face às bibliotecas convencionais. Por um lado, facultam o acesso a obras raras ou reservadas por motivos de preservação, por outro lado, permitem que vários leitores em sítios distintos consultem a mesma obra em simultâneo. Além disso, disponibilizam recursos e serviços adicionais que facilitam a consulta do exemplar em moldes vantajosos face à consulta presencial. Por tudo isso, aumentam a visibilidade das coleções, a sua difusão e respetivo acesso.

De acordo com os dados disponibilizados no relatório de janeiro de 2014 da ENUMERATE (Stroeker & Vogels, 2014), na Europa, apenas 11,62% das coleções das bibliotecas estão digitalizadas, ainda que os documentos sejam o bem patrimonial mais digitalizado (84%), e

apenas cerca de 34% das instituições disponibilizam a política de acesso aos acervos. A razão mais apontada para a constituição de coleções digitais é o apoio à investigação. Em contrapartida e apesar de a preservação digital ser uma preocupação recorrente em todas as bibliotecas, 48% das instituições não têm uma solução de preservação no tempo.

## **Bibliotecas digitais: estado da arte**

As bibliotecas digitais, cujo objetivo já não é apenas a disponibilização de imagens em linha, como acontecia nos primeiros projetos, tendem a tornar-se mais flexíveis. Atualmente, exige-se que as bibliotecas digitais forneçam serviços personalizados e que integrem ferramentas úteis e adequadas para grupos específicos de utilizadores, como os investigadores e os professores.

Não obstante, alguns autores (Rydberg-Cox, 2006) defendem uma contenção na disponibilização de ferramentas, considerando que estas se podem tornar desajustadas ou supérfluas, acabando por não ser usadas e, mesmo, tornar-se prejudiciais à consulta da obra. Por outro lado, recomendam a criação e o desenvolvimento de boas páginas de ajuda, a realização de ações de esclarecimento e formação e a produção de campanhas de divulgação e publicidade, com o objetivo de potenciar a utilização das bibliotecas de forma rentável e adequada.

Algumas bibliotecas digitais, como a Perseus<sup>3</sup>, o Internet Archive<sup>4</sup> e as Bibliothèques Virtuelles Humanistes<sup>5</sup>, assumiram como princípios orientadores a criação de um produto que fosse acessível, de fácil consulta e cujos resultados fossem reutilizados tanto por investigadores como pelo leitor comum.

### ***Perseus Digital Library***

O projeto Perseus Digital Library, iniciado em 1987 e sediado no departamento de estudos clássicos na Tufts University, nos Estados Unidos da América, foi construído de forma colaborativa entre universidades europeias (Consiglio Nazionale delle Ricerche, Itália; University of Roma Tor Vergata, Itália e University of Leipzig, Alemanha) e americanas (Center for Hellenic Studies, College of the Holy Cross, Furman University, Tufts University, University

---

<sup>3</sup> Disponível em: <http://www.perseus.tufts.edu/hopper/>

<sup>4</sup> O Internet Archive, foi criado para guardar o arquivo das páginas da Internet. Atualmente essa componente mantém-se, mas possui outras capacidades, podendo definir-se como uma biblioteca digital. Disponível em: <https://archive.org/>

<sup>5</sup> Disponível em: <http://www.bvh.univ-tours.fr/>

of Massachusetts - Amherst, University of Missouri at Kansas City, dos Estados Unidos da América e Mount Allison University, Canadá).<sup>6</sup> Congrega várias instituições de diferentes nacionalidades e especialidades que vão desde os editores digitais, a especialistas em grego e árabe, em edições multilíngue, em autores fragmentados, em tratamentos de dados, etc. Constitui, por isso, um dos mais consistentes exemplos de projeto colaborativo e multidisciplinar.

Figura 1 Perseus Digital Library - página de acolhimento

**PERSEUS DIGITAL LIBRARY**  
GREGORY R. CRANE, EDITOR-IN-CHIEF  
TUFTS UNIVERSITY

Home Collections/Texts Perseus Catalog Research Grants Open Source About Help

**Welcome to Perseus 4.0, also known as the Perseus Hopper.**  
Read more on the [Perseus version history](#).  
New to Perseus? [Click here](#) for a short tutorial.

**Perseus Digital Library Updates**

- **29/5/2014** [Student Commentaries Published in Perseus](#)  
*One goal of the Perseids Project is to engage humanities students in research and enable them to contribute to the open access ...*
- **2/5/2014** [New Version of Ancient Greek OCR Released](#)  
*Visiting research scholar at Perseus, Nick White has released a new version of Ancient Greek OCR, free software to accurately ...*
- **21/3/2014** [Announcing the Perseus Lexical Inventory - an open linked data set.](#)  
*Many different linguistic services and tools are dependent on lexical information as it is commonly found in Latin and Greek ...*
- **10/3/2014** [Pelagios used in Tufts Classes](#)  
*Pelagios, Pleiades, and Perseids workshop took place at week-long hackathon On Monday, March 3, students in Marie-Claire ...*
- **5/3/2014** [Tufts Reimagines Humanities Using Tools of the Digital Age](#)  
*Read about Perseids, the Bodin Project, Corpora and Digital Humanities at Tufts in the article, Tufts Reimagines Humanities ...*
- **16/12/2013** [Announcing the Leipzig Open Fragmentary Texts Series \(LOFTS\)](#)  
*The Humboldt Chair of Digital Humanities at the University of Leipzig is pleased to announce a new effort within the Open ...*

**Popular Texts**

- Caesar, *Galic War* (English, Latin)
- Catullus, *Carmine* (English, Latin)
- Cicero, *In Catilinam I* (English, Latin)
- Vergil, *Aeneid* (English, Latin)
- Herodotus, *Histories* (English, Greek)
- Homer, *Odyssey* (English, Greek)
- Plato, *Republic* (English, Greek)
- Tom Martin, *Overview of Classical Greek History from Mycenae to Alexander* (English)

**Art and Archaeology**

Aegina, Temple of Aphaia      Silver obol from Athens

Esta biblioteca digital desconstrói o livro físico, mas mantém a relação do texto com os parágrafos e com a página da edição impressa que lhe deu origem. No que se refere aos periódicos, mantém a mesma lógica, perdendo a localização da notícia na página, os separadores e as variantes das fontes tipográficas. As obras são também disponibilizadas em linha na Biblioteca do Congresso<sup>7</sup>. Quando foi constituída tinha o objetivo de disponibilizar em linha o conteúdo de um conjunto de fontes primárias e secundárias que fundamentassem a história, a literatura e a cultura clássicas, em grego e latim, com a respetiva tradução para inglês. Posteriormente, a coleção foi aumentada e hoje possui uma seção de arte e arqueologia (moedas, vasos, esculturas, edifícios) e outras coleções: textos em árabe; textos em alemão; documentos relativos ao séc. XIX americano; literatura inglesa da Renascença; poesia italiana do séc. XII à primeira metade do séc. XVI; e o periódico *Times Dispatch*, de Richmond.

Todas as operações de pesquisa e recuperação da informação são executadas de forma amigável e transparente para o utilizador, o que torna o Perseus um dos projetos mais completos

<sup>6</sup> <http://www.perseus.tufts.edu/hopper/about/who>

<sup>7</sup> <http://chroniclingamerica.loc.gov/lccn/sn84024738/>

e exemplares no âmbito das humanidades digitais. “The rise of huge collections and new analytical methods is certainly important because it can provide new tools with which scholars can see what later ages had to say about Greek and Latin sources” (Crane, 2011, p. 18). Além disso, este projeto desenvolveu um conjunto significativo de competências adicionais, como a estruturação do conhecimento através de marcadores (“tag”), o multilinguismo, a aplicação extensiva do TEI-compliant XML<sup>8</sup> (Text Encoding Initiative – compliant Extensible Markup Language), a otimização dos algoritmos de pesquisa e de reconhecimento de caracteres (OCR), etc. (Bamman, Babeu, & Crane, 2010).

O acesso às obras realiza-se através da navegação pelas páginas das várias coleções ou da pesquisa alargada ao conteúdo dos documentos. Nos textos apresentados, as pessoas/personagens e os locais mencionados estão referenciadas através de hiperligações. Para os textos em versão bilingue, são indicadas as opções de tradução. As obras foram codificadas em TEI.

O Perseus permitiu a otimização dos algoritmos de pesquisa e de reconhecimento de caracteres (OCR), contribuindo para a definição de um conjunto de boas práticas aplicáveis a projetos idênticos. Constitui um *corpus* de informação legível pelo homem ou pelo computador. Como a informação está estruturada com metadados normalizados, é possível criar algoritmos para realizar outras “leituras” sobre os textos e criar novo conhecimento.

A informação é fornecida em três camadas:

- Conhecimento legível: imagens digitalizadas de objetos, lugares, inscrições, páginas impressas, informações geográficas e outras representações digitais de objetos e espaços. Nesta camada, acede-se diretamente ao objeto digitalizado.
- Conhecimento acionável: registros de catálogo, artigos de enciclopédia, verbetes e outras fontes de informação estruturadas.
- Conhecimento gerado: novos conhecimentos gerados através da análise de sistemas de informação automatizados. A máquina de conhecimento acionável permite, por exemplo, ver uma entrada de dicionário ou encontrar traduções pré-existentes para uma citação em latim ou grego.

---

<sup>8</sup> TEI - Text Encoding Initiative. O TEI foi criado em 1987 por um consórcio das associações académicas, Association for Computers and the Humanities (ACH), Association for Computational Linguistics (ACL) e Association for Literary and Linguistic Computing (ALLC), que o propuseram como o resultado da pesquisa efetuada para desenvolver um esquema de metadados que respondesse cabalmente à colocação em linha de textos digitais, nos seus múltiplos formatos.

O projeto está acessível a partir da Universidade de Tufts, mas possui cópias (“*mirrors*”) nas Universidades de Oxford e Berlim.

Numa apreciação global, pode dizer-se que esta biblioteca digital serve os fins muito específicos relacionados com a colocação em linha de textos gregos e latinos e respetiva tradução, cumprindo os objetivos inicialmente traçados. Em contrapartida, como modelo a ser implemento nas bibliotecas digitais patrimoniais, é complexo e requer um grande investimento em termos quer financeiros, quer humanos. Provavelmente não se justifica a aplicação do TEI a todas as obras, mas reconhece-se a importância da sua aplicação a obras que sejam consideradas fontes primárias para os vários domínios do conhecimento.

### ***Internet Archive***

O Internet Archive foi fundado em 1996 com a missão de preservar e fornecer o acesso gratuito e universal ao conhecimento. Inclui textos (monografias, periódicos, documentação de arquivo, etc.) registos audiovisuais, material informático (programas, jogos, etc.) e o arquivo de páginas web desde 1996. É financiada por várias instituições, como Alexa Internet, HP Computer, Library of Congress, LizardTech, National Science Foundation, Prelinger Archives, Sloan Foundation e Kahle/Austin Foundation, e contribuidores particulares.

Um dos serviços com mais notoriedade é o Wayback Machine, que atualmente fornece acesso a mais de quatrocentos mil milhões de páginas web arquivadas num processo segundo as regras robotx.txt (<http://www.robotstxt.org/>).

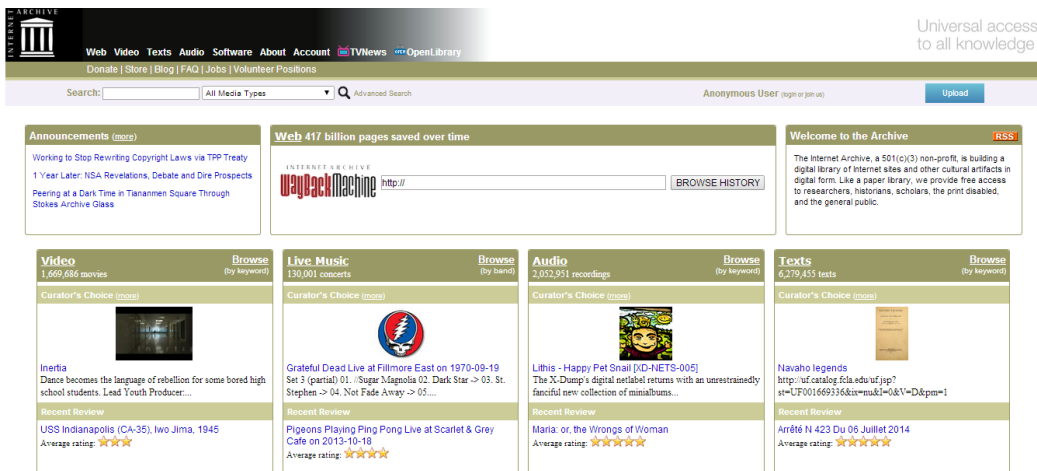
Além do arquivo de páginas web, o Internet Archive disponibiliza também mais de seis milhões de textos em linha, de forma gratuita e universal, provenientes da sua atividade de digitalização e do contributo dos parceiros de projeto: Library of Congress, Harvard University, the New York Public Library, Smithsonian Institution, The Getty Research Institute, University of California, University of Toronto, Biodiversity Heritage Library, Boston Library Consortium, C.A.R.L.I., Johns Hopkins University, Allen County Public Library, Lyrasis, Massachusetts Institute of technology, State Library of Massachusetts, e mais de mil instituições que aderiram ao Open Content Alliance OCA.<sup>9</sup>

---

<sup>9</sup> The Open Content Alliance (OCA) is a collaborative effort of a group of cultural, technology, nonprofit, and governmental organizations from around the world that helps build a permanent archive of multilingual digitized text and multimedia material. <http://www.opencontentalliance.org/about/>



Figura 2 Internet Archive - página de acolhimento



O acesso às obras é feito através da pesquisa no próprio portal ou pela navegação nas seguintes coleções: American Libraries, Canadian Libraries, Universal Library, Community Texts, Project Gutenberg, Children's Library, Biodiversity Heritage Library, Additional Collections. A pesquisa no portal é efetuada, quer através dos elementos bibliográficos da obra<sup>10</sup>, quer através do conteúdo textual dos documentos ou dos metadados adicionados.<sup>11</sup>

As obras são disponibilizadas em múltiplos formatos: leitura em linha HTML, PDF, PDF com OCR, ePUB, Kindle (Mobi), Daisy, texto e DjVu. O formato ePub é um formato textual aberto e normalizado que tem a vantagem de permitir a leitura dos documentos em todos os tipos de ecrã.

As opções de leitura em linha e em PDF apresentam um nível de qualidade muito razoável. Em contrapartida, as versões baseadas no OCR, sobretudo em obras mais antigas, têm um desempenho bastante inferior com um elevado índice de erro no reconhecimento dos caracteres. Porém, dado que apresenta sempre uma versão fac-similada das obras, é possível comparar a versão com OCR com o original e corrigir os erros existentes.

<sup>10</sup> A obra é catalogada no formato MARC (Machine Readable Cataloging) e com o DublinCore. Vd [https://archive.org/about/faqs.php#Texts\\_and\\_Books](https://archive.org/about/faqs.php#Texts_and_Books)

<sup>11</sup> A obra é digitalizada e adicionam o registo MARC e o Dublin Core, em XML.

### Bibliothèques Virtuelles Humanistes

O projeto Bibliothèques Virtuelles Humanistes é um projeto francês, criado, em 2002, pelo Centre d'Études Supérieures de la Renaissance - CERS, com a colaboração do Institut de recherche et d'histoire des textes - IRHT secção de humanismo.

Figura 3 Bibliothèques Virtuelles Humanistes - página de acolhimento



Disponibiliza a informação em quatro secções; o Fac-simile, o Epistemon, manuscritos e iconografia.

A Fac-simile corresponde às imagens de obras renascentistas digitalizadas pelos parceiros do projeto. Pode aceder-se às imagens dos documentos através da pesquisa simples (título, autor, data, edição, cota), da pesquisa avançada (os vários campos da ficha bibliográfica), ou das listas disponíveis por autor ou título, sendo possível fazer a ordenação por autor, título e data.

- Para facilitar a navegação na obra, estão disponíveis várias formas de acesso ao conteúdo:
- Pela *feuilletage*, que permite folhear a obra, mantendo a indicação do número das páginas da obra;
- Pelo *chemin de fer*, que mostra todas as imagens em miniatura;

- Pelo *sommaire de l'ouvrage*, que indica, através do sumário da obra, a parte da obra que está a ser visionada, referindo o número da página respetiva;
- Pelas *images*, que navega através das ilustrações existentes na obra.

Por norma, a obra também está disponível em PDF, permitindo descarregá-lo na íntegra ou parcialmente. O PDF total possui três modalidades: a preto e branco de alta definição para impressão; a cores, com OCR não corrigido e não visível; a cores, sem OCR.

A *Epistemon, corpus* de textos da Renascença, é uma base de dados textual, com a transcrição dos documentos para XML-TEI. A pesquisa é realizada através dos elementos da ficha bibliográfica ou através de palavra no texto. Nem todos os textos estão ainda completamente tratados, sendo que alguns estão disponíveis em HTML, com OCR mas sem a codificação em TEI. Depois de selecionar a obra pretendida, acede-se a uma página com informações acerca da obra e, em seguida, à página de rosto exibida em modo imagem e texto. A navegação é feita página a página ou através dos verbetes do sumário.

A secção dos manuscritos, atualmente em construção, destina-se à disponibilização de documentos séc. XV e XVI transcritos e codificados para TEI.

A secção de iconografia disponibiliza de forma autónoma as ilustrações das várias obras. A pesquisa realiza-se através do Iconclass<sup>12</sup>. A referência à obra de origem mantém-se anexa à imagem. Nesta secção, são também disponibilizadas bases de dados de letras capitulares, de marcas dos impressores e de retratos, fazendo sempre a referência à obra de origem.

O processo está devidamente documentado, nomeadamente, no que se refere às opções de codificação em TEI e às políticas de transcrição.

Em 24 de abril de 2014, este projeto ganhou o prémio Succeed<sup>13</sup>, atribuído ao projeto de digitalização de documentos históricos.

## Apreciação

Nenhuma das três bibliotecas digitais acima descritas possui ferramentas que permitam a consulta em simultâneo do texto e da imagem da obra, apesar de se reconhecer a sua importância

---

<sup>12</sup> Sistema de classificação de imagens, disponível em <http://www.iconclass.org/>

<sup>13</sup> <http://succeed-project.eu/succeed-awards>

para a confirmação e verificação das opções tomadas durante a conversão para texto e a codificação para TEI.

A codificação em TEI permite a leitura do texto, tanto pelo homem como pelos computadores, replica a estrutura física da obra, como a indicação dos parágrafos e das secções da obra, e permite a exportação para formatos portáteis como o e-Pub. Porém, nenhuma das bibliotecas que utilizam o TEI oferece a possibilidade de exportação em vários formatos.

A pesquisa efetuada nestas bibliotecas é alargada ao conteúdo dos textos. No entanto, em qualquer dos casos, é exigida uma correspondência perfeita entre o termo a pesquisar e a palavra no texto, dado que não permitem a pesquisa por termos semelhantes, nem sugerem termos alternativos.

Por fim, a exportação dos dados bibliográficos das obras para os gestores de referências também não é eficiente em nenhum dos casos, exigindo uma correção manual posterior.

## **Conclusão**

A web 2.0 trouxe um novo conceito de utilização à internet. Hoje, pretende-se que o utilizador tenha mais interação e participação até na escolha dos conteúdos. As bibliotecas e, em especial, as digitais, têm de se adaptar à nova realidade, tornando-se mais interativas. Atualmente, já não é suficiente disponibilizar vastas massas documentais em formatos de imagem ou, mesmo, em PDF com OCR. As formas tradicionais de pesquisa e de recuperação da informação revelam-se eventualmente insuficientes.

O processo de codificação em TEI implica a quebra da ligação à obra física, ao tipo de letra, à grelha de impressão, ao tipo de papel, encadernação, etc., o que não invalida o reconhecimento das enormes vantagens na reutilização da informação, ou seja, dos algoritmos que se podem construir para a análise dos textos. Por outro lado, a utilização da codificação em Unicode (UFT-8) permite representar corretamente línguas mais complexas, como por exemplo as que usam diacríticos.

Assim, mais do que pensar em alterar os esquemas de apresentação, há que desenvolver a investigação no sentido da disponibilização das obras (objeto e conteúdo) de forma útil, acessível e adequada, tanto para o investigador, como para o leitor comum ou ocasional. Neste momento de charneira, o objetivo da investigação é ajustar o modelo anterior a um novo paradigma mais dinâmico e interativo.

Preconiza-se, no estado atual dos conhecimentos e das expectativas e exigências que têm vindo a ser formuladas, um modelo de biblioteca digital em que o leitor tivesse acesso à obra, simultaneamente, em formato imagem e texto, permitindo a pesquisa nos elementos bibliográficos e no conteúdo da obra, e também que o conteúdo das obras textuais fosse codificado em TEI, permitindo a utilização de programas para a sua análise e a exportação em múltiplos formatos de leitura.

## Agradecimentos

Agradeço ao CIDEHUS (Centro Interdisciplinar de História, Culturas e Sociedades) da Universidade de Évora, a disponibilidade para custear a minha participação no XII Congreso Internacional de Nuevas Tendencias en Humanidades em Madrid.

Esta investigação está a ser financiada pela FCT através da bolsa de formação avançada SFRH / BD / 82229 / 2011.





## REFERENCIAS

- Bamman, D., Babeu, A., & Crane, G. (2010). Multilingual Alignment and Projection. Medford. Retrieved from <http://www.perseus.tufts.edu/publications/jcd127-bamman.pdf>
- Besser, H. (2002). Moving from isolated digital collections to interoperable digital libraries. *First Monday*, 7(6). Retrieved from <http://www.firstmonday.org/ojs/index.php/fm/article/view/958/879>
- Borgman, C. L. (2009). The digital future is now: a call to action for the humanities. *Digital Humanities Quarterly (DHQ)*, 3(4), 1–30. Retrieved from <http://www.digitalhumanities.org/dhq/vol/3/4/000077/000077.html>
- Choudhury, G. S., & Seaman, D. (2008). The Virtual Library Mass : Virtual Library Collections. In S. Schreibman & R. G. Siemens (Eds.), *A companion to digital literary studies* (pp. 1–9). Oxford: Blackwell. Retrieved from [http://nora.lis.uiuc.edu:3030/companion/view?docId=blackwell/9781405148641/9781405148641.xml&chunk.id=ss1-6-10&toc.depth=1&toc.id=ss1-6-10&brand=9781405148641\\_brand](http://nora.lis.uiuc.edu:3030/companion/view?docId=blackwell/9781405148641/9781405148641.xml&chunk.id=ss1-6-10&toc.depth=1&toc.id=ss1-6-10&brand=9781405148641_brand)
- Crane, G. (2011). *White Paper Report*. Medford. Retrieved from <http://hdl.handle.net/10427/75558>
- Evans, L., & Rees, S. (2012). An interpretation of digital humanities. In D. M. Berry (Ed.), *Understanding digital humanities* (pp. 21–41). Houndmills, Basingstoke, Hampshire: Palgrave Macmillan.
- Galina Russell, I. (2011a). ¿Qué Son las Humanidades Digitales? *Revista Digital Universitaria*, 12(7). Retrieved from <http://www.revista.unam.mx/vol.12/num7/art68/>
- Galina Russell, I. (2011b). El papel de las bibliotecas en las humanidades digitales. In *World Library and Information Congress : 77th IFLA General Conference and Assembly*. San Juan: IFLA. Retrieved from [http://www.paginaspersonales.unam.mx/files/638/Publica\\_20110923192631.pdf](http://www.paginaspersonales.unam.mx/files/638/Publica_20110923192631.pdf)
- Gonçalves, M. F., & Banza, A. P. (2013). In limine. In M. F. Gonçalves & A. P. Banza (Eds.), *Património Textual e Humanidades Digitais: da antiga à nova Filologia*. Évora: CIDEHUS. Retrieved from <http://dspace.uevora.pt/rdpc/handle/10174/10468>
- Lucía Megías, J. M. (2010). De las bibliotecas digitales a las plataformas de conocimiento (notas sobre el futuro del texto en la era digital). In M. Arbor Aldea & A. Fernández Guiadanes (Eds.), *Estudos de edición crítica e lírica galego-portuguesa* (pp. 369–401). Santiago de Compostela: Universidade de Santiago de Compostela. Retrieved from <http://eprints.ucm.es/10767/>
- Manifiesto de las IFLA/UNESCO Sobre las Bibliotecas Digitales | IFLA. (2011). Retrieved June 18, 2014, from <http://www.ifla.org/ES/publications/manifiesto-de-las-ifla-unesco-sobre-las-bibliotecas-digitales>
- Rydberg-Cox, J. A. (2006). *Digital libraries and the challenges of digital humanities* (p. 103). Oxford: Chandos. Retrieved from [http://www.worldcat.org/title/digital-libraries-and-the-challenges-of-digital-humanities/oclc/61529070&referer=brief\\_results](http://www.worldcat.org/title/digital-libraries-and-the-challenges-of-digital-humanities/oclc/61529070&referer=brief_results)
- Schreibman, S., Siemens, R., & Unsworth, J. (2004). The digital humanities and humanities computing: an introduction. In S. Schreibman, R. G. Siemens, & J. Unsworth (Eds.), *A Companion to Digital Humanities*. Malden, Oxford: MA, Blackwel. Retrieved from <http://www.digitalhumanities.org/companion/view?docId=blackwell/9781405103213/9781405103213.xml&chunk.id=ss1-1-3&toc.depth=1&toc.id=ss1-1-3&brand=default>

Stroeker, N., & Vogels, R. (2014). *Survey report on digitisation in european cultural heritage institutions* (pp. 1–54). Retrieved from <http://www.enumerate.eu/fileadmin/ENUMERATE/documents/ENUMERATE-Digitisation-Survey-2014.pdf>

## **SOBRE EL AUTOR**

**Dália Guerreiro:** Doutoranda em Ciências da Informação e da Documentação pela Universidade de Évora; mestre em Estudos de Informação e Bibliotecas Digitais, pelo ISCTE-IUL; pós-graduada em Ciências da Informação e da Documentação, variante Bibliotecas, pelo ISLA – Universidade Europeia; licenciada em Física, pela Faculdade de Ciências da Universidade de Lisboa. É membro integrado do Centro Interdisciplinar de História, Cultura e Sociedades da Universidade de Évora/Fundação para a Ciência e Tecnologia, como investigadora na Linha 3 – Bibliotecas, Literacias e Informação no Sul (CIDEHUS-UÉ/FCT – LIBIS). É bolsista da Fundação para a Ciência e Tecnologia. Integrou a equipa que iniciou e desenvolveu o projeto da biblioteca digital na Biblioteca Nacional e, na DigiCult -Produções Digitais, de que é sócia, tem realizado a metacodificação e edição digital para as bibliotecas digitais da Universidade de Lisboa e da Universidade de Coimbra, entre outras.

**José Borbinha:** é Professor Associado do Departamento de Informática do IST e Investigador do Grupo de Sistemas de Informação do INESC-ID. Tem desenvolvido atividade de engenharia e investigação em Sistemas de Informação, em particular em requisitos, modelação, desenho, e integração de sistemas, com ênfase especial em normalização e conformidade. Tem-se focado em contextos preservação digital, em bibliotecas e arquivos digitais. Foi Diretor dos Serviços de Inovação e Desenvolvimento da Biblioteca Nacional, onde lançou a Iniciativa da Biblioteca Nacional Digital. Tem participado e coordenado projetos nacionais e internacionais sobre bibliotecas e arquivos digitais. Foi consultor de programas da Comissão Europeia e da NSF – National Science Foundation (Estados Unidos). Tem sido ainda membro regular dos Comités de Programa das conferências JCDL, ECDL/TPDL, ICADL, RCDL e EIPUB. É membro do Comité Editorial do IJDL – International Journal of Digital Libraires. É membro da Ordem dos Engenheiros (colégios de Engenharia Eletrotécnica e de Engenharia Informática), IEEE - Institute of Electrical and Electronic Engineers, ACM - Association for Computer Machinery, INCOSE - International Council on Systems Engineering, e BAD - Associação Portuguesa de Bibliotecários, Arquivistas e Documentalistas. Foi membro fundador do IEEE TCDL – Technical Committee on Digital Libraries, e "chair" eleito entre 2008 e 2010.