

UC Berkeley

UC Berkeley Previously Published Works

Title

PhiloBiblon and the Wiki World: Sustainable Technology

Permalink

<https://escholarship.org/uc/item/4c8370pw>

ISBN

978-84-19077-35-6

Author

Faulhaber, Charles

Publication Date

2022

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NoDerivatives License, available at

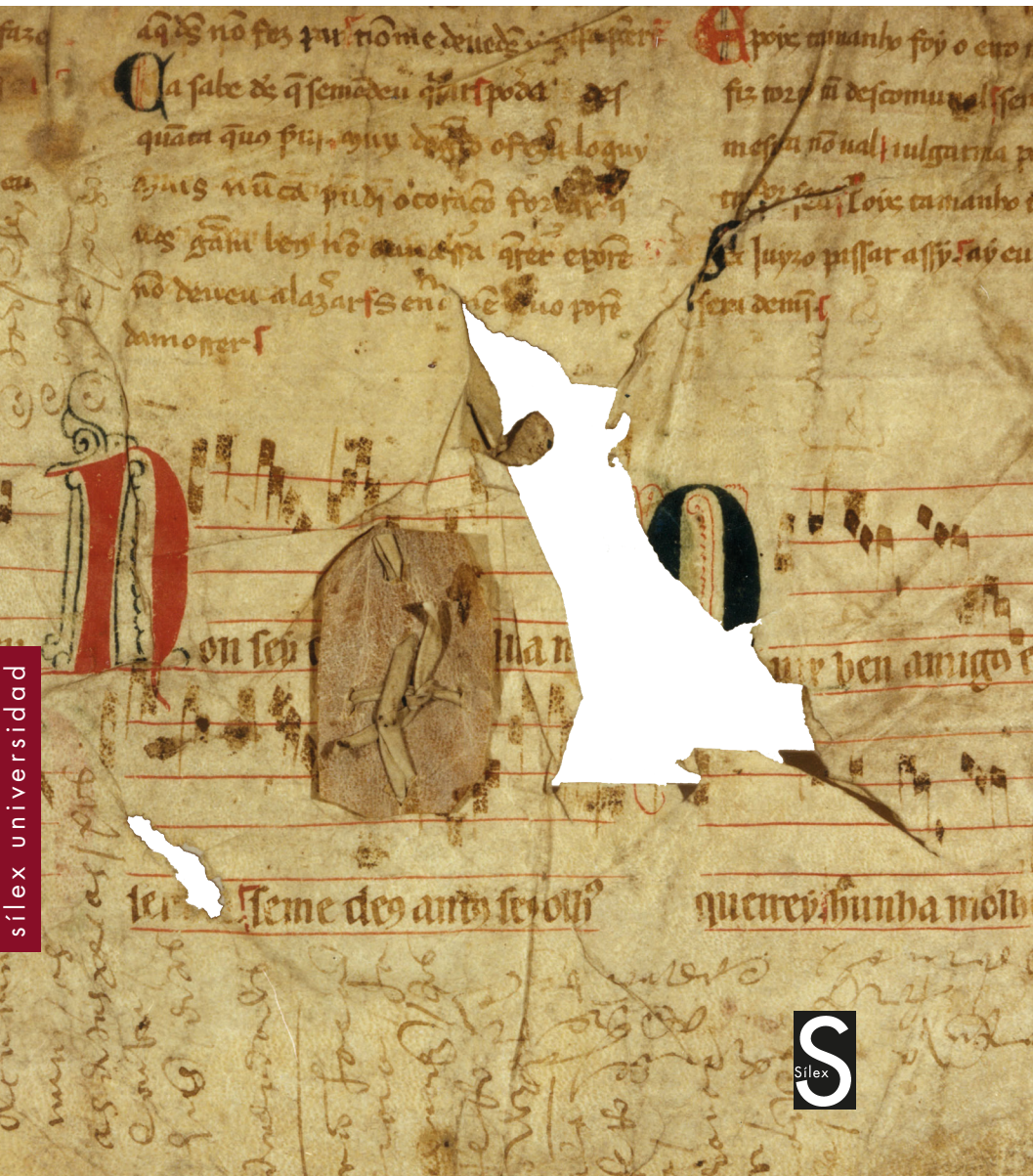
<https://creativecommons.org/licenses/by-nd/4.0/>

Peer reviewed

«Tenh'eu que mi fez el i mui gran ben»

ESTUDOS SOBRE CULTURA ESCRITA MEDIEVAL
DEDICADOS A HARVEY L. SHARRER

Ricardo Pichel (ed.)



«Tenheu que mi fez el i mui gran ben»
Estudos sobre cultura escrita medieval
dedicados a Harvey L. Sharrer

Ricardo Pichel
(ed.)



S
Silex

A publicación deste volume recibiu o apoio económico do proxecto “HERES. Patrimonio textual ibérico y novohispano. Recuperación y memoria” (CM/2018-T1/HUM-10230; UAH, 2019-2024), da Asociación Internacional de Estudos Galegos (AIEG) e do Instituto de Estudos Medievais da Universidade Nova de Lisboa (Proxecto Estratégico FCT UIDB/00749/2020).



© RICARDO PICHEL (ED.), 2022

© RESTO DE AUTORES, 2022

EDITOR: RAMIRO DOMÍNGUEZ HERNANZ

MAQUETADOR: RICARDO PICHEL

© Imagen de cubierta: *Pergaminho Sharrer*.

Cartório Notarial de Lisboa, 7-A, caixa 1, maço 1, livro 3 (verso)

C/ San Gregorio, 8, 2, 2ª, 28004, Madrid

España

www.silexediciones.com

ISBN: 978-84-19077-35-6

Depósito Legal: M-28556-2022

Cualquier forma de reproducción, distribución, comunicación pública o transformación de esta obra solo puede ser realizada con la autorización de sus titulares, salvo excepción prevista por la ley. Diríjase a CEDRO (Centro Español de Derechos Reprográficos) si necesita fotocopiar o escanear algún fragmento de esta obra (www.conlicencia.com; 91 702 19 70 / 93 372 04 97)



Harvey L. Sharrer na Misión de Santa Bárbara, California (outubro de 2022)
(foto cortesía de Carmen Mejía Ruiz)

ÍNDICE

LIMIAR

<i>Harvey L. Sharrer e a memoria do patrimonio escrito medieval</i>	13
<i>Noticia bibliográfica</i>	21

ADRO

<i>BITAGAP (Bibliografía de Textos Antigos Galegos e Portugueses): um armazém da memória histórica</i> Harvey L. SHARRER	39
---	----

ESTUDOS

<i>BITECA (Bibliografía de Textos Catalans Antics): passat i futur</i> Gemma AVENOZA & Lourdes SORIANO	71
<i>O léxico afectivo nas Cantigas Santa María: preservación da memoria e construción da identidade galaico-portuguesa</i> María Jesús BOTANA VILAR	87
<i>Transcrición paleográfica da lírica galego-portuguesa: texto primario e recursos auxiliares</i> Mercedes BREA & Carmen de SANTIAGO GÓMEZ	103
<i>Livros de Direito na Idade Média Portuguesa: a biblioteca do Doutor Luís Martins (1451)</i> José DOMINGUES & Pedro PINTO	123
<i>Sobre un trobador “menor”: Nuno Treez</i> Leticia EIRÍN	151
<i>PhiloBiblon and the wiki world: sustainable technology</i> Charles B. FAULHABER	183
<i>En torno da segmentación do adverbio pronominal «i» nas cantigas</i> Manuel FERREIRO	201
<i>Nin santas, nin pecadoras; ou todo e máis. Da condición feminina ás voces e actuacións das mulleres na documentación medieval galega</i> Miguel GARCÍA-FERNÁNDEZ	225

<i>A corte literaria do conde de Barcelos: o ciclo dos privados</i>	
Raquel JABARES FERNÁNDEZ	285
<i>Crónicas universais a sul do Minho: a Historia Scholastica como alternativa à General Estoria?</i>	
Mariana LEITE	311
<i>O português e o galego no discurso lingüístico de Bernardo de Aldrete (1560-1641)</i>	
Ramón MARIÑO PAZ	327
<i>El Libro del infante don Pedro de Portugal. Ediciones españolas y portuguesas: problemas de transmisión y nuevas aportaciones</i>	
Carmen MEJÍA RUIZ & MaríaVictoria NAVAS SÁNCHEZ-ÉLEZ	357
<i>O «estrato tardío» da compilación trobadoresca. Subsídios para a súa delimitación e caracterización scriptolingüística</i>	
Henrique MONTEAGUDO	373
<i>Novos dados sobre a transmissão textual do Livro do Deão e do Livro do Conde D. Pedro: as linhagens dos Bragançons e dos Lanhoso</i>	
Filipe Alves MOREIRA	395
<i>El albañil herido, entre Alfonso X y Goya: estrategias para levitar en el vacío de la (pre)modernidad</i>	
José Manuel PEDROSA	417
<i>Sobre la tensó entre Afonso Sanchez y Vasco Martins de Resende y unas trovas escarninas contra don Cristóbal de Moura</i>	
Gimena del RIO RIANDE	467
<i>Quix e querrei e quero e direivos ùa ren: a sintaxe oracional no Pergamiño Sharrer</i>	
Alexandre RODRÍGUEZ GUERRA	479
<i>A mula voadora, o morto-vivo e os fantasmas entram num bar. Questões em torno de episódios naturais com seres sobrenaturais em textos quinhentistas portugueses</i>	
Maria de Lurdes ROSA	511
<i>Vida (e Milagres) de Santa Senhorinha de Basto. A late 16th/early 17th century copy (of Fr. Vasco Martins' 1403 text ?), in portuguese</i>	
Martha E. SCHAFFER	527
<i>A familia de Afonso X nas Cantigas de Santa Maria</i>	
Joseph T. SNOW	551

<i>Entre a corte e o mosteiro: traduzir a santidade nos séculos XIII-XIV</i> Cristina SOBRAL	567
Aparecerá porla presente escriptura até o tempo en que somos. <i>Ainda sobre a Crónica de Santa Maria de Íria</i> José António SOUTO CABO	591
COLOFÓN	
<i>Novos textos para o estudo da literatura medieval galego-portuguesa</i>	
Ricardo PICHEL & Miguel GARCÍA-FERNÁNDEZ & Xavier VARELA BARREIRO (coords.)	
<i>A recepción galego-portuguesa da Sexta Partida: dous novos fragmentos</i> Mariña ARBOR ALDEA & Miguel GARCÍA-FERNÁNDEZ & Ricardo PICHEL	647
<i>Unha nova ollada ao testemuño galego do Livro das Confissões</i> Arthur L.-F. ASKINS & Miguel GARCÍA-FERNÁNDEZ & Ricardo PICHEL	667
<i>A tradución galego-portuguesa da Crónica de Castela: o novo testemuño luxemburgués</i> José Carlos Ribeiro MIRANDA & Ricardo PICHEL	689
<i>Rezos romanceados en documentación de las dominicas de Lugo en el tránsito de la Edad Media a la Modernidad. Noticia del hallazgo y notas preliminares</i> Pablo S. OTERO PIÑEYRO MASEDA	707
Aqui se começan as rubricas deste liuro. <i>Dous testemuños galegos inéditos da Quinta Partida</i> Ricardo PICHEL & Harvey L. SHARRER	713
TABULA GRATULATORIA	751

PHILOBIBLON AND THE WIKI WORLD:
SUSTAINABLE TECHNOLOGY

Charles B. Faulhaber¹

I am delighted to have this opportunity to honor my good friend and University of California colleague Harvey Sharrer, whom I have known since the fall of 1967 when we met in the Biblioteca Nacional in Madrid. We were both there to work on our dissertations, Harvey on the legendary history of Britain in the *Libro de las bienandanzas y fortunas* of Lope García de Salazar and I on the influence of classical Latin rhetoric in early Castilian literature.

Since then we have collaborated for over thirty-five years on PhiloBiblon, with Harvey focusing on the *Bibliografía de Textos Antigos Galegos e Portugueses* (BITAGAP) along with Berkeley professor and my colleague Arthur L-F. Askins, University of San Francisco professor Martha E. Schaffer, and numerous colleagues in Portugal.

In addition to his yeoman work on BITAGAP, Harvey has also provided a great deal of material for PhiloBiblon's two other databases, the *Bibliografía Española de Textos Antiguos* (BETA) and the *Bibliografía de Textos Antics Catalans, Valencians I Balears* (BITECA).

PhiloBiblon (PhiloBiblon 1997-21) is one of the oldest ongoing Digital Humanities projects in the U.S. and the oldest in the Hispanic world. Its immediate ancestor was the *Bibliography of Old Spanish Texts* (BOOST) (Cárdenas *et al.* 1975), an in-house database for the Dictionary of the Old Spanish Language project (UW-Madison, 1972-2001), sponsored by the National Endowment for

¹ <https://orcid.org/0000-0002-1361-1295>. University of California-Berkeley, Department of Spanish & Portuguese. Emeritus Professor. Contact: cbf@berkeley.edu.

the Humanities (NEH). Spun off in 1981 as an independent project, it has become the indispensable starting point for any serious research on medieval Iberian history, literature, or culture. In its more than 40 years of existence it has stimulated a renewed interest in manuscript studies and uncovered thousands of little known, unknown, and/or lost manuscripts and works: e.g., the only known MS of the Spanish masterpiece *Celestina*; the Portuguese version of John Gower's *Confessio amantis*; the 14th-c. Pergaminho Sharrer, discovered by our *homenajeado*, with works and music of seven poems by King Dinis of Portugal (1279-1325). *PhiloBiblon's* detailed descriptions go far beyond what is available in existing manuscript catalogs. Compare, for example, the difference between the Biblioteca Nacional de España's catalog record for VITR/4/6², a stunningly beautiful copy of the *Siete partidas* of Alfonso X the Learned, and the PhiloBiblon record (BETA manid 3373) with its detailed analysis of quire structure, pictorial elements, catchwords, rubrics, line fillers, ink types; history of the MS, and contents, with incipit and explicit of every major section of the text.

PhiloBiblon has been used as a cataloging resource by, among others, the library of the University of Madrid, the Royal Library in Madrid, the Ajuda library in Portugal, the National Libraries of Argentina, France, and Spain, and the Biblioteca Virtual Miguel de Cervantes (Spain's largest digital library).

There is no other printed or online resource that spans the gap between detailed description of primary sources and the historical context in which the cultures of the Iberian Middle Ages developed and flourished. By its very nature *PhiloBiblon* provides material for comparatists interested in other European and trans-Mediterranean cultures, both medieval and early modern, and thereby fosters studies of the complicated multicultural / multilingual hybridity of the Iberian and related European and Mediterranean worlds: representations and interactions of Christians, Muslims, and Jews, chains of translations and literary influences, gender, politics and religion,

² <http://catalogo.bne.es/uhtbin/cgiisirs/x/o/o/57/20/MSS.MICRO^2F8150/o/X1000389392?user_id=WEBSERVER> (accessed: 08/02/2021).

textual materiality, the prosopography of social classes... Moreover, PhiloBiblon makes research more efficient and therefore less costly. It saves scholars time and money as they begin research programs by making it possible to locate relevant materials or pose and answer exploratory questions in minutes or hours instead of days or weeks. Time is the most precious resource scholars have.

I have described PhiloBiblon's technical evolution from a flat-file dbms with ten fields (240 characters per field) in 1975 to the current system (Faulhaber 2016). The relational dbms (now Windows OpenInsight) that followed the flat-file dbms in 1987 is still used for data input: 1.246 data elements (more than any other MS description model), 98 controlled vocabularies (with +/-2000 properties), 30 data-entry forms, and 110 indexes in 10 related tables: MsEd (manuscripts and prints), Copies (of prints in MsEd), Uniform Title (works), Analytic (instances / witnesses), Geography (toponyms), Institutions, Persons, [holding] Libraries, Secondary references, Subject headings. The web version released in 1997 and enhanced with NEH support in 2015, has data entry and display capabilities for English, Spanish, Catalan, Galician, and Portuguese.

The current technology is both inelegant and inefficient. Data entry, entirely manual, is limited to a single user at a time. Regular uploading to the web (4-6 times a year) entails the export of XML files, validation of the resulting files against the PhiloBiblon schema with <Oxygen> XML Editor (SyncRO Soft 2002-21), correction of any errors, and the upload of the corrected files to the UCB Library server, where the California Digital Library's eXtensible Text Framework (XTF) runs a Perl (Perl 2002-21) script against the XML data files, exploding them into individual records, indexing them, and then converting them to HTML files on the fly as the result of an end user search.

This system, rife with potential for human error, has simply evolved over the past forty years in an attempt to keep up with hardware and, especially, software changes in an environment of constrained resources. Human labor has taken on tasks that in a well-financed system can and should be performed by software.

Nevertheless, PhiloBiblon continues to grow, from 966 records in 1975 to 269,935 records in 2006 to the current total of 327,957 as of the end of 2020. This growth speaks to the dedication and hard work of the members of PhiloBiblon's three research teams, one of whom, Gemma Avenzoa, we just lost in a terrible blow to our community (Faulhaber, Perea *et al.* 2021). While a small amount of this data was entered by grant-funded student assistants, the bulk of it was input without cost by the scholars involved in the project, like Gemma. Most funding has been devoted to technical enhancements and has come primarily from periodic NEH grants (PhiloBiblon 1997-2021 Acknowledgments³) supplemented by smaller grants from other organizations, e.g. Andrew W. Mellon Foundation, Gladys Krieble Delmas Foundation, Fundación Ignacio Larramendi). While enormously helpful, they have never been large enough or frequent enough to keep PhiloBiblon at the forefront—or as some colleagues call it, the bleeding edge—of technological change. In the last twenty years in particular lack of funding has made it increasingly difficult to keep up with the pace of technology.

The Humanities Computing—now Digital Humanities—world has conceived numerous projects for the description of medieval manuscripts. These in turn have led to formal data models such as the Text Encoding Initiatives (TEI) *P5 Guidelines to msdescription* (TEI 2020, Driscoll *et al.* 2005, Vertan and Reimers 2010), the European Data Model of the Europeana consortium (EDM) (Europeana 2015, Europeana 2016). One of the prototypes for TEI *msdescription* as well as one of the first projects (1996) with digitized images of medieval manuscripts was Digital Scriptorium (on which Faulhaber served as Co-PI). Perhaps the most interesting of the projects based on TEI *msdescription* is *e-codices*, digitized manuscripts from Swiss libraries, with extensive technical documentation. Other significant projects for cataloguing medieval manuscript are: Wellcome Arabic MSS online, Schoenberg Database of Manuscripts, TRAME (Text and Manuscript Transmission of the Middle Ages

³ <https://bancroft.berkeley.edu/philobiblon/acknowledgments_en.html> (accessed: 08/02/2021).

in Europe), Biblissima, West African Manuscript Data Base (hosted at Berkeley), and Manuscriptorium. Indexes dealing with other periods and based on other writing surfaces have also been created, e.g., CDLI (Cuneiform Digital Library Initiative 2021), Papyri.info.

We had initially hoped that we could simply port PhiloBiblon's records into an existing platform based on Linked Data (W3C 2015) and the Resource Description Format (W3C 2014) –LD/RDF–, but there is no current or proposed national or international project with a data model as complex as PhiloBiblon's; nor one that, in addition, goes beyond manuscripts to include other entities. Data entry, the single most difficult and demanding task in the creation of scholarly indexes, is generally manual. In *e-codices*, for example descriptive metadata is manually entered directly into a TEI-based XML schema (W3schools 2021), or indirectly, via Word files, then loaded into its Content Management System. Discovery and visualization capabilities in similar projects are, in most cases primitive, limited to works, named entities, languages, chronology, or toponyms. None of them have begun to grapple in any serious way with the use of LD/RDF for data input or export.

The Hill Museum and Manuscript Library (HMML, St. John's University) is an exception for its handling of exactly the same sort of practical problems PhiloBiblon must deal with, particularly distributed data entry in its vHMML Reading Room (*Manual for Metadata Entry in vHMML* 2016), and export of search results in multiple formats⁴.

PhiloBiblon staff have watched developments in the broader library world as well, starting with the Dublin Core Metadata Initiative (OCLC 1995-2021) and culminating, to date, with BIBFRAME 2.0 (Futornick & Kovari 2018), whose data model, still heavily MARC-dependent (Library of Congress 13/03/2020), cannot capture PhiloBiblon's complex data, nor would it be able to do so without extensive development.

⁴ We broached the possibility of folding *PhiloBiblon* into vHMML with Daniel Gullo, HMML's assistant director, but he informed us that to do this the vHMML data model would need to be considerably expanded.

The only viable long-term solution is to move PhiloBiblon to an existing platform that does have the strategic and tactical ability to keep up with technological change. Such a platform should be based on the Wiki world, which has grown up outside the other two and has now matured sufficiently to be a viable option, specifically MediaWiki (WikiMedia Foundation 26/05/2020), the technological underpinning of Wikipedia (WikiMedia Foundation 2001-2021), arguably the most visible and successful example ever of knowledge technology in service to humanity, which just celebrated its twentieth anniversary (WikiMedia Foundation 14/01/2021). Wikipedia depends, like PhiloBiblon, on unpaid and volunteer effort, in its case more than 280.000 editors who create and maintain more than 55 million articles in 300 different languages read 15 billion times a month. In its twenty years of life Wikipedia has spun off a host of other Wiki-based projects⁵. For the purposes of PhiloBiblon the most important of these are WikiData (WikiMedia Foundation 20/12/2019), the structured information database for all Wiki projects, and Wikibase (WikiMedia Foundation 10/02/2021), its underlying software base.

It was not possible to consider using MediaWiki with WikiData and Wikibase as a serious LD/RDF platform even a few years ago, but currently they are being studied by major libraries and consortia (OCLC Project Passage [Godby *et al.* 2019], Deutsch Nationalbibliothek GND+Wikibase [Fischer and Ohlig 2020], Bibliothèque nationale de France National Entities+Wikibase [Bibliothèque nationale de France 2020], Association of Research Libraries [2019]) with promising results (Thorsen 2019, Smith-Yoshimura 2020). Paraphrasing the BnF, Wikibase enables a major change of perspective with respect to bibliographic information as a graph of interconnected entities. It is devised for collaborative creation and maintenance of data, makes use of standard Web technologies, is based on open-source solutions, has a very active user community, and is inherently multilingual.

⁵ <<https://wikimediafoundation.org/our-work/wikimedia-projects/>> (accessed: 08/02/2021).

A MediaWiki platform would position PhiloBiblon to take advantage of current and future semantic web developments and decrease long-term sustainability costs. Moreover, PhiloBiblon on MediaWiki can serve as a model for low-cost light-weight database development for similar academic projects with limited resources. MediaWiki, however, was not designed as a platform for scholarly projects like PhiloBiblon. While moving to MediaWiki is necessary to allow PhiloBiblon to keep up with changes in the world-wide information technology environment, it is not sufficient. What is needed is a purpose-built platform based on MediaWiki.

Such a platform exists: *FactGrid: a database for historians* (Simons 2021), which has a well-established set of standards, practices, and principles—the same ones used in Wikipedia—as well as over four years of practical experience with Wikibase. FactGrid, an incubator and collaboratory, is an integral part of Universität Erfurt's IT infrastructure; it is currently working with WikiMedia Deutschland and the Deutsche Nationalbibliothek on the GND (Integrated Authority File) database.

For PhiloBiblon's purposes, the single most important feature of Wikibase on FactGrid is its extremely flexible, low-cost, and light-weight approach to data modeling combined with the ease with which non-technical staff can modify the model. We believe that it can solve the problem of technological and financial sustainability for a low-budget all-volunteer project like PhiloBiblon.

The Wikibase data model is simple. It consists of RDF/LD “triplestores” (W3C 25/02/2014) based on entities (Q) and properties (P), each with a unique Wikibase id number. A triplestore, or triple, is the combination of a property P relating two items, e.g., Q1 and Q2. Each triple can be qualified with sub-statements and footnoted with references. Development of PhiloBiblon's Wikibase data model will consist first in identifying PhiloBiblon properties, currently instantiated in software links between records in related tables, with existing P# (e.g., P50 “written by” to relate records in the Person and the Uniform Title tables) and creating new ones if necessary. These P# will then be used to link Q# at the item level: Thus “*Libro de buen amor* (Q2283127) written by (P50) Juan Ruiz (Q434597)”. As part of

the process of cleaning up PhiloBiblon data before moving them to FactGrid, Wikibase/FactGrid Q# (in turn based on WikiData Q#) will be added to all PhiloBiblon entities already in Wikibase. The vast majority of PhiloBiblon's 327,000+ entities, however, do not have Wikibase Q#. They will be created automatically as part of the process of exporting the data from PhiloBiblon to Wikibase.

PhiloBiblon's existing relational model of entities and properties, developed by and for scholars, can be mapped into this Wikibase model of items and properties through the re-use and extension, if necessary, of existing vocabularies and ontologies (e.g., EDM, Pelagios [toponyms: Pelagios Network], PeriodO [historical periods: Rabinowitz *et al.* 2016], Shared Canvas [Sanderson *et al.* 2011]).

Checking WikiData for P# and Q# and creating the ones missing will be tedious and time-consuming but not difficult. Wiki Mix'n'Match (Wikimedia Foundation 23/12/2020) will be used to facilitate this process. Since it does not require a technical IT background, additional members of the PhiloBiblon teams will be pressed into service to supplement the academic staff and student assistants.

The existence of a robust Wiki and FactGrid community of users and developers accustomed to collaborative projects will help to resolve technical issues. Wikibase is compliant with international semantic web standards (LOD [W3C 2015], OWL [W3C 2012], SKOS [W3C 2004], RDF triples [W3C 25/02/2014]), currently has the capability of exporting to and importing from standard formats (CSV [Wikipedia 2020], JSON-LD [Ecma International 2017], XML [W3schools 2021]), and can have endpoints to (with appropriate authority) and from relevant external ontologies and vocabularies (e.g., VIAF [OCLC 2010-2021] and GND [named entities: Deutsche Nationalbibliothek], Getty Vocabularies [Harpring 2014], LC Linked Data Service [Library of Congress]) as well as internal controlled vocabularies.

We are confident that PhiloBiblon on FactGrid/Wikibase can achieve the goals set forth below. **Bolded items are already standard in Wikibase and will therefore not need to be developed for PhiloBiblon:**

- ¶ **A light-weight low-cost development process that minimizes the necessity for paid technical staff and takes advantage of the skills and content knowledge of volunteer academic staff.**
- ¶ **A Wikibase model to capture all of the information in PhiloBiblon: works, instances, and physical objects (manuscripts and prints), and personal, organizational, chronological, and geographical entities, based on and responsive to international standards and norms (FAIR [Findable, Accessible, Interoperative, Reusable (Wikipedia 19/01/2020)], LD/RDF, EDM, TEI, VIAF) and instantiated in stable, widely used Wikibase software.**
- ¶ **Easily customizable data model, with access points to both internal and external controlled vocabularies, GIS data, and projects such as Pelagios, PeriodO, and Shared Canvas.**
- ¶ **Open access and open source, built on free software and with the source code available under a Creative Commons license.**
- ¶ **Multilingual and Unicode compliant, both data and metadata.**
- ¶ **Modular, with data input, data query, and export packages.**
- ¶ **Data query screens built for each entity type and a branded PhiloBiblon web portal.**
- ¶ **Tools for record creation in a crowd-sourcing environment (cf. Ancient Lives [Williams *et al.* 2014]).**
- ¶ **Tools to add metadata and create controlled vocabularies.**
- ¶ **Tools to work with electronic texts and images.**
- ¶ **A report generator capable of exporting selected or complete data in multiple formats, e.g.: JSON-LD, RDF, XML, TEI msdescription, MARC (Library of Congress 13/03/2020), T_EX (Knuth 1984), CSV.**
- ¶ **Capable of linking to aggregators like Europeana.**
- ¶ **Harvestable via OAI-PMH protocol (OAI-PMH n.d.).**
- ¶ **Faceted and semantic search facilities.**
- ¶ **Tools to support the automated or semi-automated ingest of collection metadata based on LD/RDF records (e.g., endpoints to VIAF/LC). WikiMedia Deutschland is currently working on this.**

- ¶ Big Data modules for pattern visualization, e.g., prosopographical projection in relational (networks), chronological (timelines), and geographical (maps) perspectives. **FactGrid already has some of these tools and intends to develop others.**

The transformation of PhiloBiblon's technology will enable new research by opening its rich data universe to the world. PhiloBiblon will continue to stimulate interest in recovering the primary manuscript and printed sources of medieval Iberian culture by highlighting neglected works that remain unedited and unstudied and serve as a guide for editorial and digitization projects. Moreover, eventually the web site could serve as a portal for related Wikibase tools, e.g., analytical instruments for geo-chronological studies based on PhiloBiblon records and repertoires of watermarks like BITECA.net (Biteca.net 2018) or of paleographical or codicological "phenotypes".

Virtual libraries like Europeana provide minimal access points, usually a subset of the fifteen fields of the Dublin Core⁶ model or similar generic metadata; yet scholars need much more. Users of PhiloBiblon must be able to search, for example, for "manuscripts written in the second half of the 15th century in gatherings of 8 leaves", or "Franciscans active in Castile between 1450 and 1475", thus enabling more precise codicological studies or prosopographical analysis of a key intellectual community.

A low-maintenance open-access open-source technology platform is thus crucial for PhiloBiblon, a project with constrained resources and a non-technical volunteer staff. Once the basic data model is in place on FactGrid, enhancements to properties can be created *ad hoc* by that same staff, just as they now create records for new entities.

PhiloBiblon has thrived since 1981 because of the sustained dedication of its original staff (Faulhaber, Askins, Sharrer, Beltran) as well as of younger scholars (Mariña Arbor, Gemma Avenzoa [†], Álvaro Bustos, Antonio Cortijo, Ralph DiFranco, José Luis Gonzalo,

⁶ <<https://dublincore.org/>> (accessed: 08/02/2021).

Filipe Alves Moreira, María Morrás, Óscar Perea, Pedro Pinto, Lurdes Rosa, Martha E. Schaffer, Cristina Sobral, Lourdes Soriano), many of whom began work on the project as graduate students and now, as leaders in their own institutions, are well-positioned to assume larger roles in PhiloBiblon. PhiloBiblon seminars during the summers of 2015-2021, organized primarily by Avenozza and Perea, are preparing the next generation of scholars as well.

The UCB Library has provided robust support for PhiloBiblon since 1997 and will continue to do so. Should FactGrid cease to exist, the Library will assume control of the existing data records and software and host its own Wikibase site. If necessary, it is prepared to migrate both Wikibase and records as a fundamental scholarly resource to a successor institution, e.g., the Universitat Pompeu Fabra in Barcelona, which currently hosts a mirror site and with which Berkeley has an MOU. All textual data and metadata belong to the U. of California but are freely available for re-use under an appropriate Creative Commons license. The UCB Library will host legacy PhiloBiblon as long as necessary, take responsibility for minor software maintenance, and serve as back-up for FactGrid PhiloBiblon.

When will all this happen? We are currently prototyping PhiloBiblon on FactGrid during the 2021-2022 academic year, with a follow-on implementation grant to start –if the application is successful– in 2023. Since a logical and conceptual data model already exists in the PhiloBiblon schema⁷, we have focused during this prototyping period on mapping it to the Wikibase model. In the implementation grant, development of front-end query and report screens will proceed in parallel with the mapping process, with each aspect informing the other. After an initial set-up period devoted to basic wireframe modeling, the selection of test records from PhiloBiblon, and the establishment of procedures for mapping the PhiloBiblon data schema to Wikibase, the team will map sequentially each of PhiloBiblon's ten relational tables to Wikibase, while a front-end

⁷ <<https://sunsite3.berkeley.edu/NewPhiloBiblon/schema/documentation/>> (accessed: 08/02/2021).

developer will design and code prototype data input, query, and report screens. Work on each table will take advantage of lessons learned in the preceding phase, from the experience of the OCLC Passage⁸ project, and from work going on in parallel on similar issues by WikiMedia Deutschland and FactGrid. The lengthiest part of this process will be mapping PhiloBiblon's 98 existing controlled vocabularies, with properties describing relationships within and among entities, to some of the 700 existing Wikibase P# and creating new ones as necessary. Because core academic staff are intimately familiar with PhiloBiblon's design and data, they can carry out much of this work in consort with student assistants and PhiloBiblon colleagues. This work will also give academic staff an intimate knowledge of Wikibase's protocols and data entry protocols.

Wikipedia is arguably the most visible and successful example ever of knowledge technology in service to humanity, from K-12 students onward. We believe that PhiloBiblon on Wikibase can share that visibility and provide the same kind of service and collaborative experience to anyone interested in the Iberian Middle Ages. At a moment when the Middle Ages are being redefined in a global context, PhiloBiblon on Wikibase stands to become a model instrument for research and one of the crucial tools for understanding and enriching studies and resources across the medieval world's disciplines, languages, and cultures.

The National Endowment for the Humanities found these arguments compelling. In 2021 it awarded a grant (PW-277550-21) for "PhiloBiblon: From Siloed Databases to Linked Open Data via Wikibase: Proof of Concept". Although we are only about two-thirds of the way through this pilot project, we have proved to our satisfaction that FactGrid on Wikibase can serve as a viable technological platform for PhiloBiblon. In the summer of 2022 we shall request a two-year implementation grant to move PhiloBiblon fully onto that platform.

⁸ <<https://www.oclc.org/research/publications/2019/oclcresearch-creating-library-linked-data-with-wikibase-project-passage.html>> (accessed: 08/02/2021).

REFERENCES⁹

- Ancient Lives: <<https://www.ancientlives.org/>>.
- Association of Research Libraries (2019): “ARL White Paper on Wikidata. Opportunities and Recommendations”, <<https://www.arl.org/wp-content/uploads/2019/04/2019.04.18-ARL-white-paper-on-Wikidata.pdf>>.
- BETA = Charles B. Faulhaber (dir.) (1997-): *Bibliografía Española de Textos Antiguos*. Berkeley: The Bancroft Library, University of California, <https://bancroft.berkeley.edu/philobiblon/beta_es.html>.
- Biblioteca Virtual Miguel de Cervantes, <<http://www.cervantesvirtual.com/>>.
- Bibliothèque national de France (2020): “French national entities file (FNE): Project overview” [blog], <<https://www.transition-bibliographique.fr/fne/french-national-entities-file/>>.
- Biblissima: Bois, Kevin, Eduard Frunzeanu, and Régis Robineau (co-ords.): *Biblissima*, <<https://portail.biblissima.fr/en/>>.
- BITAGAP = Arthur L.-F. Askins, Harvey L. Sharrer e Martha E. Schaffer (dirs.) (1997-): *Bibliografía de Textos Antigos Galegos e Portugueses*. Berkeley: The Bancroft Library, University of California, <https://bancroft.berkeley.edu/philobiblon/bitagap_po.html>.
- BITECA = Gemma Avenozza, Lourdes Soriano Robles e Vicenç Beltran (dirs.): *Bibliografía de Textos Antics Catalans, Valencians i Balears*. Berkeley: The Bancroft Library, University of California, <https://bancroft.berkeley.edu/philobiblon/biteca_ga.html>.
- Biteca.net: Avenozza, Gemma, Lourdes Soriano, and Elena Cañizares (2018): *Biteca.net: filigranes*, <http://stel.ub.edu/biteca/html/ca/filigranes_intro.html>.
- California Digital Library: *eXtensible Text Framework*, <<https://xtf.cdlib.org/>>.
- Cárdenas, Anthony; John J. Nitti, and Jean Gilkison (comps.) (1975): *Bibliography of Old Spanish Texts*. Madison: Hispanic Seminary of Medieval Studies, Ltd.
- CDLI: *Cuneiform Digital Library Initiative* (2021), <<https://cdli.ucla.edu/>>.
- Deutsche Nationalbibliothek: *The Integrated Authority File (GND)*, <https://www.dnb.de/EN/Professionell/Standardisierung/GND/gnd_node.html>.

⁹ All on-line materials were seen the week of February 8, 2021.

- Digital Scriptorium, <<https://digital-scriptorium.org/>>.
- Driscoll, Matthew; Merrilee Proffitt, and David Birnbaum (2005): “Manuscript Description”. In C. M. Sperberg-McQueen, and Lou Burnard (eds.). *TEI PS: Guidelines for Electronic Text Encoding and Interchange*, <<https://tei-c.org/release/doc/tei-p5-doc/en/html/MS.html>>.
- Ecma International (2017): *ECMA-404: The JSON data interchange syntax*, <<https://www.ecma-international.org/publications-and-standards/standards/ecma-404/>>.
- e-codices: *Virtual Manuscript Library of Switzerland*, <<https://digital-scriptorium.org/>>.
- Europeana (2015): *Digitised Manuscripts to Europeana - Final Public Report* (1st February 2012 - 31st January 2015, <http://141.20.126.149/files/Public_Final_Report_150630.pdf>.
- (2016): *Definition of the Europeana Data Model v5.2.7*, <https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Definition_v5.2.7_042016.pdf>.
- FactGrid:PhiloBiblon: Prototype website on FactGrid: A database for Historians at the Forschungszentrums Gotha der Universität Erfurt, <<https://database.factgrid.de/wiki/FactGrid:PhiloBiblon>>.
- Faulhaber, Charles B. (2016): “PhiloBiblon and the Semantic Web. Notes for a Future History”. In María Jesús Lacarra, and Nuria Aranda Garda (eds.). *La literatura medieval hispánica en la imprenta (1475-1600)*. Valencia: Publicacions de la Universitat de Valencia (Colección Parnaseo, 28), pp. 75-93, <<https://escholarship.org/uc/item/2wr4s1tn>>.
- Faulhaber, Charles B.; Óscar Perea, *et al.* (2021): “Gemma Avenozza (1959-2021) *In Memoriam*” [blog], <<https://update.lib.berkeley.edu/2021/01/24/avenozza/>>.
- Fischer, Barbara K., and Jens Ohlig (2020): “Report ‘GND meets Wikibase’ 2” [blog], <<https://wiki.dnb.de/pages/viewpage.action?pageId=167019461>>.
- Futornik, Michelle, and Jason Kovari (2018): *The Art & Rare Materials BIBFRAME Ontology Extension*. Part of Linked Data for Production (LD4P), <<https://wiki.lyrasis.org/display/LD4P/Rare+Materials+Ontology+Extension>>.
- Godby, Jean, *et al.* (2019): “Creating Library Linked Data with Wikibase: Lessons Learned from Project Passage”. *OCLC Research* (Dublin, OH), <<https://doi.org/10.25333/faq3-axo8>>.

- Harpring, Patricia (ed.) (2014): *The Getty Vocabularies and Linked Open Data. Introduction and Editorial Perspective*, <https://www.getty.edu/research/tools/vocabularies/cidoc_getty_vocab_lo_d.pdf>.
- Hill Museum & Manuscript Library (2016): *Manual for Metadata Entry in vHMML Reading Room. Manuscript Description*, <<https://docs.google.com/document/d/1cqqrOOj111LVGcyiHCJEDNuPr1sON4ozk9vmVSE4NU/edit#>>.
- Knuth, Donald E. (1984): *The TeXbook*. Reading, MA: Addison-Wesley, <<https://www-cs-faculty.stanford.edu/~knuth/abcde.html>>
- Library of Congress: *ID.LOC.GOV - Linked Data Service*, <<https://id.loc.gov/>>.
- (2016): *Overview of the BIBFRAME 2.0 Model*, <<https://www.loc.gov/bibframe/docs/bibframe2-model.html>>.
- (13/03/2020): *MARC Standards*, <<https://www.loc.gov/marc/>>.
- OAI-PMH: Open Archives Initiative. Protocol for Metadata Harvesting: *Interoperability through Metadata Exchange*, <<https://www.openarchives.org/pmh/>>.
- OCLC (09/02/2021): *Dublin Core Metadata Initiative*, <<http://dublincore.org/>>.
- (2010-2021): *VIAF: The Virtual International Authority File*, <<https://viaf.org/>>.
- Manuscriptorium: *Digital Library of Written Cultural Heritage*, <<http://www.manuscriptorium.com/>>.
- Papyri.info: <<http://papyri.info/>>.
- Pelagios Network: <<https://pelagios.org/>>.
- PeriodO. A gazetteer of periods for linking and visualizing data: *Guide*, <<https://perio.do/guide/>>.
- Perl.org (2020-2021): *About Perl*, <<https://www.perl.org/about.html>>.
- PhiloBiblon: Faulhaber, Charles B. (dir.) (1997-2022): *PhiloBiblon*. University of California-Berkeley, The Bancroft Library, <<https://bancroft.berkeley.edu/philobiblon/>>.
- Rabinowitz, Adam, *et al.* (2016): “Making Sense of the Ways We Make Sense of the Past: The PeriodO Project”. *Bulletin of the Institute of Classical Studies*, 59/2, pp. 42-55, <<https://onlinelibrary.wiley.com/doi/full/10.1111/j.2041-5370.2016.12037.x>>.
- Revelation Software (2007-2021): *OpenInsight*, <<https://www.revelation.com/>>.

- Sanderson, Robert, *et al.* (2011): “Shared Canvas: A Collaborative Model for Medieval Manuscript Layout Dissemination”, *arXiv*, <<https://arxiv.org/abs/1104.2925>>.
- Schoenberg Institute of Manuscript Studies: *Schoenberg Database of Manuscripts*, <<https://sdbm.library.upenn.edu/>>.
- Simons, Olaf (2021): “FactGrid: A database for historians”, <https://database.factgrid.de/wiki/Main_Page>.
- Smith-Yoshimura, Karen (2020): “Experimentations with Wikidata/Wikibase” [blog], <<https://hangingtogether.org/?p=8002>>.
- SyncRO Soft (2002-21): *oXygen XML editor*, <https://www.oxygenxml.com/about_us.html>.
- TEI: Text Encoding Initiative (2020): *P5. Manuscript Description*, <<https://tei-c.org/release/doc/tei-p5-doc/en/html/MS.html>>.
- Thorsen, Hilary (2019): “Wikibase Report for LD4P2” [10 pp. Prepared for a meeting of staff of the *Linked Data for Production Phase 2* project, 17/09/2019, revised 16/12/2019], <https://docs.google.com/document/d/1sOPWK9Ps_dC23kIOi3yI_SiyIxaewhMaUqIjI_SQJWA/edit#heading=h.eew2nj3wbgbag>.
- TRAME: Fondazione Ezio Franceschini (2013): *Text and Manuscript Transmission of the Middle Ages in Europe*, <<http://trame.fefonlus.it/trame/index.html>>.
- Vertan, Cristina, and Stefanie Reimers (2012): “A TEI-based application for Editing Manuscript Descriptions”. *Journal of the Text Encoding Initiative*, 2, <<https://journals.openedition.org/jtei/392>>.
- W3C: World Wide Web Consortium (2004): *SKOS Simple Knowledge Organization System*, <<https://www.w3.org/2004/02/skos/>>.
- (2012): *Web Ontology Language (OWL)*, <<https://www.w3.org/OWL/>>.
- (25/02/2014): *RDF 1.1. Concepts and Abstract Syntax*, <<https://www.w3.org/TR/rdf11-concepts/>>.
- (2014): *Resource Description Framework (RDF)*, <<https://www.w3.org/RDF/>>.
- (2015): *Linked Data*, <<https://www.w3.org/standards/semanticweb/data>>.
- W3schools (2021): “Introduction to XML”, <https://www.w3schools.com/xml/xml_whatIs.asp>.
- Wellcome Library: *Wellcome Arabic MSS online*, <<http://wamcp.bibalex.org/>>.

- West African Manuscript Data Base: <<https://waamd.lib.berkeley.edu/home>>.
- [Wikimedia Foundation]: *Wikibase Project*, <<https://wikiba.se/>>.
- Wikimedia Foundation: *Wikimedia Projects*, <<https://wikimediafoundation.org/our-work/wikimedia-projects/>>.
- (2001-21): *Wikipedia*, <<https://www.wikipedia.org/>>.
- (30/12/2019): *Wikidata*, <https://www.wikidata.org/wiki/Wikidata:Main_Page>.
- (26/05/2020): *MediaWiki*, <<https://www.mediawiki.org/wiki/MediaWiki>>.
- (23/12/2020): *Mix'n'Match / Manual*, <<https://meta.wikimedia.org/wiki/Mix%27n%27match/Manual>>.
- (14/01/2021): *Wikipedia 20th anniversary*, <<https://wikimediafoundation.org/news/2021/01/14/wikipedia-celebrates-20-years/>>.
- (10/02/2021): *Wikibase*, <<https://www.mediawiki.org/wiki/Wikibase>>.
- Wikipedia (19/01/2021): *FAIR data*, <https://en.wikipedia.org/wiki/FAIR_data>.
- (30/12/2020): “Comma-separated values”, <https://en.wikipedia.org/wiki/Comma-separated_values>.
- Williams, Alex C., *et al.* (2014): “A computational pipeline for crowd-sourced transcriptions of Ancient Greek papyrus fragments”. In *2014 IEEE International Conference on Big Data (Big Data)*. Washington, DC: IEEE: pp. 100-105, <<https://ieeexplore.ieee.org/document/7004460?reload=true&tp=&arnumber=7004460&queryText%3Dwilliam%20alex%20pridemore=>>>.