

# **Collaborative Research Practices and Shared Infrastructures for Humanities Computing**

2<sup>nd</sup> AIUCD Annual Conference, AIUCD 2013  
Padua, Italy, 11-12 December 2013

Proceedings of Revised Papers

Maristella Agosti and Francesca Tomasi (Eds)

cleup

ASSOCIAZIONE per  
l'INFORMATICA UMANISTICA  
e la CULTURA DIGITALE



<http://www.umanisticadigitale.it>

Prima edizione: settembre 2014

ISBN 978 88 6787 260 2

CLEUP sc  
“Coop. Libreria Editrice Università di Padova”  
via G. Belzoni 118/3 – Padova (t. 049 8753496)  
[www.cleup.it](http://www.cleup.it) - [www.facebook.com/cleup](http://www.facebook.com/cleup)

© 2014 AIUCD

Tutti i diritti di traduzione, riproduzione e adattamento,  
totale o parziale, con qualsiasi mezzo (comprese  
le copie fotostatiche e i microfilm) sono riservati.

In copertina:  
Graphic Design: Massimo Malaguti – Scuola Italiana Design (elaborazione del logo  
di AIUCD).

# Organization

AIUCD 2013 was organized by the Information Management Group of the Department of Information Engineering of the University of Padua, Italy.

## Committees

### General Chair

*Dino Buzzetti*, Presidente AIUCD

### Program Chairs

*Maristella Agosti*, Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Padova

*Anna Maria Tammaro*, Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Parma

### Program Committee

*Fabio Ciotti*, Dipartimento Studi Umanistici, Università di Roma Tor Vergata

*Giorgio Maria Di Nunzio*, Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Padova

*Maurizio Lana*, Dipartimento di Studi Umanistici, Università del Piemonte Orientale

*Federico Meschini*, Dipartimento di Scienze Umanistiche, della Comunicazione e del Turismo, Università degli Studi della Tuscia

*Nicola Orio*, Dipartimento di Beni Culturali, Università degli Studi di Padova

*Nicola Palazzolo*, già ordinario nell'Università di Perugia

*Roberto Rosselli Del Turco*, Dipartimento di Studi Umanistici, Università di Torino

*Marco Rufino*, Fondazione Rinascimento Digitale, Firenze

*Francesca Tomasi*, Dipartimento di Filologia Classica e Italianistica, Università di Bologna

#### **Award Chair**

*Francesca Tomasi*, Dipartimento di Filologia Classica e Italianistica, Università di Bologna

#### **Local Committee**

*Debora Leoncini*, Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Padova

*Marta Manfioletti*, Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Padova

*Chiara Ponchia*, Dipartimento di Beni Culturali, Università degli Studi di Padova

*Gianmaria Silvello*, Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Padova

## Table of contents

PREFACE / PREFERAZIONE <i>Maristella Agosti, Francesca Tomasi</i>	11
<b>KEYNOTE</b>	
KEYNOTE ADDRESS / INTERVENTO INVITATO	
Toward a Computational Narratology <i>Jan Christoph Meister</i>	17
<b>INVITED</b>	
CONTRIBUTIONS FROM RESEARCH GROUPS AND CENTERS / CONTRIBUTI DI CENTRI E GRUPPI DI RICERCA	
Nuovi scenari per la ricerca in filosofia: i testi e gli strumenti del portale Daphnet <i>Michela Tardella, Cristina Marras</i>	39
Acquisizione e Creazione di Risorse Plurilingui per gli Studi di Filologia Classica in Ambienti Collaborativi <i>Federico Boschetti</i>	55
Da <i>Musisque Deoque</i> a <i>Memorata Poetis</i> . Le vie della ricerca intertestuale <i>Paolo Mastandrea, Luigi Tessarolo</i>	69

**PANELS**

DIGITAL RESOURCES AND NETWORK SERVICES FOR DIGITAL HUMANITIES  
 RESEARCH / RISORSE DIGITALI E SERVIZI DI RETE PER LA RICERCA  
 IN CAMPO UMANISTICO

*Digital humanities: difficoltà istituzionali e risposte infrastrutturali* 81

*Dino Buzzetti*

*Digital humanities e analisi dei testi* 89

*Paolo Mastandrea*

Infrastrutture e risorse digitali. L'esperienza dell'ILIESI 93

*Antonio Lamarra*

DH@ILC: linee di attività e ricerca 101

*Simonetta Montemagni*

THE DIGITAL LIBRARY TO SUPPORT THE COMPUTER HUMANIST /  
 LA BIBLIOTECA DIGITALE A SUPPORTO DELL'UMANISTA INFORMatico

Digital libraries and digital humanities scholars: community context,  
 workflow and collaboration 115

*Anna Maria Tammaro*

e-Infrastructures per le esigenze della ricerca 121

*Rossella Caffo*

(Formal) Models for systems, infrastructures, communities,  
 and cultures 129

*Nicola Ferro*

Biblioteche digitali e studi umanistici 135

*Maurizio Lana*

Some remarks about Museo Galileo's digital collections 143

*Stefano Casati, Fabrizio Butini*

**PAPERS**

DIGITAL PHILOLOGY / FILOLOGIA DIGITALE

L'Open Philology Project dell'Università di Lipsia. Per una filologia  
 'sostenibile' in un mondo globale 151

*Monica Berti, Greta Franzini, Emily Franzini,*

*Giuseppe Celano, Gregory R. Crane*

Table of contents	9
A collaborative tool for philological research: experiments on Ferdinand de Saussure's manuscripts <i>Angelo Mario Del Grosso, Simone Marchi, Francesca Murano, Luca Pesini</i>	163
Edition Visualization Technology: a tool to publish digital editions <i>Raffaele Masotti, Julia Kenny</i>	177
Codifying the codex. The digital edition of the <i>Becerro Galicano</i> of San Millán <i>David Peterson</i>	187
DIGITAL CULTURAL HERITAGE / PATRIMONIO CULTURALE DIGITALE	
ASIt: Atlante Sintattico d'Italia: A linked open data geolinguistic web application <i>Giorgio Maria Di Nunzio, Jacopo Garzonio, Diego Pescarini</i>	197
The "Verbo-Visual Virtual" Platform for Digitizing and Navigating Cultural Heritage Collections <i>Alessandro Marchetti, Sara Tonelli, Roberto Sprugnoli</i>	205
<i>Dante</i> . A Web Application for the History of Art <i>Chiara Ponchia</i>	219
Digital Lightbox: a web-based visualization framework applied to paleographical research <i>Giancarlo Buomprisco</i>	229
Towards a shared methodology for audio preservation: Luciano Berio's private collection of sound recordings <i>Federica Bressan, Sergio Canazza</i>	237
Knowledge objects and bodies of knowledge: knowledge sharing platforms applied to international relations <i>Giuseppe Vitiello</i>	249
EDUCATIONAL APPROACHES / DIDATTICA	
Moodle as a collaborative platform for digital humanities <i>Giuseppe Fiorentino, Maria Accarino, Alessia Pierfederici, Daniela Rotelli</i>	261

Geostoria del quotidiano. Proposte per un'analisi automatica del testo letterario <i>Alessia Scacchi</i>	269
Managing Educational Information on University Websites: a proposal for Unibo.it <i>Federico Nanni</i>	279
Author index	287



# L'Open Philology Project dell'Università di Lipsia. Per una filologia 'sostenibile' in un mondo globale\*

Monica Berti<sup>1</sup>, Greta Franzini<sup>1</sup>, Emily Franzini<sup>1</sup>,  
Giuseppe G.A. Celano<sup>1</sup>, Gregory R. Crane<sup>1,2</sup>

<sup>1</sup>Humboldt Chair of Digital Humanities / Universität Leipzig, Leipzig, Germany  
{berti, franzini, efranzini, celano, crane}@informatik.uni-leipzig.de

<sup>2</sup>Perseus Project / Tufts University, Medford, MA, USA  
gregory.crane@tufts.edu

**Abstract:** Argomento di questo articolo è la presentazione dell'*Open Philology Project* della Humboldt Chair in Digital Humanities dell'Università di Lipsia. Il progetto nasce nell'ambito delle attività del Perseus Project della Tufts University e ha come scopo primario lo sviluppo di una collezione di risorse linguistiche greche e latine leggibili dalla macchina, la creazione di manuali dinamici basati su *corpora* annotati e l'avvio di nuove forme di pubblicazione riguardanti le lingue classiche, che possono includere sia annotazioni individuali che edizioni tradizionali integrate con dati elaborabili dalla macchina. L'*Open Philology Project* include tre componenti principali costituite dall'*Open Greek and Latin*, dall'*Historical Languages e-Learning Project*, e dall'*Open Access Publishing*.

**Parole chiave:** big data, OCR, e-Learning, greco, latino, didattica, publishing, business, treebanking, annotazione linguistica, riusi testuali.

## 1. Introduzione

L'*Open Philology Project* (OPP) della Humboldt Chair in Digital Humanities dell'Università di Lipsia aspira a riaffermare il ruolo e il valore della filologia nel senso più ampio del termine<sup>1</sup>. Due secoli fa, nella sua fondamentale opera di ripensamento degli studi classici, il filologo tedesco August Böckh definiva la filologia come *universae antiquitatis cognitio historica et philosophica* (Böckh 1858, 105; Id. 1877, 12). Prendendo spunto da questa affermazione, s'intende recuperare il significato originario della parola greca *philologia* (φιλολογία), la quale denota lo studio più vasto ed esaustivo

---

\* M. Agosti, F. Tomasi (Eds). *Collaborative Research Practices and Shared Infrastructures for Humanities Computing*. 2nd Aiucd Annual Conference, Aiucd 2013. CLEUP, Padova, 2014.

<sup>1</sup> L'indirizzo del progetto è <http://www.dh.uni-leipzig.de/wo/projects/>

possibile delle testimonianze linguistiche al fine di promuovere una conoscenza approfondita dell'attività intellettuale prodotta dall'uomo. Nel caso specifico, l'OPP mira a concentrare l'attenzione sul greco e sul latino per quattro diversi motivi: 1) sono già disponibili in rete collezioni e strumenti dedicati a queste lingue; 2) esistono comunità di utenti particolarmente numerose (circa 35.000 utenti al mese accedono alle collezioni di fonti greche e latine della Perseus Digital Library<sup>2</sup>); 3) il progetto ha sede in Europa, il cui patrimonio culturale costituisce un bacino naturale per la creazione, lo sviluppo e la distribuzione di materiali pertinenti all'antichità greco-latina; 4) la città di Lipsia vanta una tradizione editoriale e libraria di prim'ordine nel campo della filologia classica – basti pensare alle edizioni critiche di testi greci e latini pubblicate dalla casa editrice Teubner<sup>3</sup> – e si pone dunque come spazio privilegiato per la ridefinizione della filologia nell'ambito degli studi di informatica umanistica.

L'OPP è stato concepito con la speranza di creare un modello applicabile anche allo studio di altre lingue storiche. Più in particolare, esso persegue tre obiettivi diversi ma strettamente connessi fra loro: 1) la creazione di una collezione di risorse linguistiche leggibili dalla macchina, le quali siano aperte, estensibili e riutilizzabili; 2) lo sviluppo di manuali dinamici basati su *corpora* annotati, che permettano di personalizzare il vocabolario e la grammatica dei testi esistenti e coinvolgere gli studiosi e gli studenti a produrre nuove annotazioni in maniera collaborativa; 3) la promozione di nuove forme di pubblicazione, che possono consistere sia in annotazioni individuali argomentate che in edizioni tradizionali integrate con dati elaborabili dalla macchina. Questi obiettivi sono definiti attraverso le tre componenti dell'OPP presentate qui di seguito: 1) *Open Greek and Latin Project*; 2) *Historical Languages e-Learning Project*; 3) *Open Access Publishing*.

## 2. Open Greek and Latin Project

L'*Open Greek and Latin Project* (OGL) si sta attualmente dedicando alla raccolta e alla scannerizzazione di edizioni di testi classici al fine di realizza-

<sup>2</sup> Il progetto OPP nasce nell'ambito delle attività del Perseus Project presso la Tufts University <http://www.perseus.tufts.edu/>

<sup>3</sup> Sulla *Bibliotheca scriptorum Graecorum et Romanorum Teubneriana* si veda la pagina della casa editrice De Gruyter <http://www.degruyter.com>

re la più grande biblioteca digitale in materia, contribuendo nel contempo all'arricchimento della collezione greca e latina di Google Books. In questo ambito l'OGL riveste anche un ruolo di consulenza sulla legge europea sul diritto d'autore, dato che redige una lista di edizioni europee che Google Books può digitalizzare, offrendo dunque una tutela contro eventuali cause legali<sup>4</sup>.

Tale raccolta, che è *open source* e *open access*, fornisce anzitutto immagini ricercabili di edizioni di testi classici libere dai vincoli del *copyright*, corredandole di traduzioni multilingue e codificandole secondo lo standard TEI XML (*subset* EpiDoc<sup>5</sup>). L'architettura dell'OGL è concepita per gestire e mettere a disposizione degli utenti edizioni e traduzioni diverse per ogni opera classica prodotta dall'antichità greco-latina, coprendo un arco di tempo che va dall'epoca arcaica al 600 d.C. Questa caratteristica distingue l'OGL dalla maggior parte dei *corpora* esistenti (i quali prevedono generalmente un'unica edizione per opera) e costituisce un presupposto imprescindibile sul quale fondare edizioni digitali che siano realmente critiche e multitestuali (sul concetto di 'multitesto' si veda Blackwell-Crane 2009). Per poter realizzare questo obiettivo, l'OGL ha avviato collaborazioni con istituzioni accademiche di altri paesi al fine di promuovere lo scambio di dati con progetti di respiro internazionale. Tra i paesi coinvolti si annoverano la Bulgaria (progetti *Romulus Bulgaricus* e *Theseus*<sup>6</sup>), la Croazia (Università di Zagabria, Dipartimento di Filologia Classica, progetto *Croala*<sup>7</sup>), la Georgia (Ivane Javakhishvili Tbilisi State University<sup>8</sup>), il Nebraska (progetto *Digital Athenaeus*<sup>9</sup>) e l'Italia (Università del Piemonte Orientale, progetto *digilibLT*<sup>10</sup>). Tale iniziativa vorrebbe naturalmente estendersi ad altri paesi europei, sperando di spostarsi anche su zone meno esplorate come l'Est Europeo e il Medio Oriente.

Lo sforzo intrapreso dall'OGL comporta un lavoro di inserimento di dati e l'uso di tecnologie OCR per arricchire un *corpus* potenzialmente già esistente, che sia aperto e sufficientemente ampio da includere i circa 100.000.000

<sup>4</sup> Per quanto riguarda le leggi sul *copyright* vigenti in diversi paesi si può consultare la voce Wikipedia [http://en.wikipedia.org/wiki/List\\_of\\_countries%27\\_copyright\\_lengths](http://en.wikipedia.org/wiki/List_of_countries%27_copyright_lengths)

<sup>5</sup> <http://sourceforge.net/p/epidoc/wiki/Home/>

<sup>6</sup> Si vedano rispettivamente <http://romulus-bg.net> e <http://theseus.proclassics.org>

<sup>7</sup> <http://www.tei-c.org/Activities/Projects/cr02.xml>

<sup>8</sup> <http://www.tsu.edu.ge/en/>

<sup>9</sup> <http://www.dh.uni-leipzig.de/wo/open-philology-project/digital-athenaeus/>

<sup>10</sup> <http://digiliblt.lett.unipmn.it>

di parole prodotte dai primordi della classicità sino al VII secolo d.C. A questo riguardo OGL ha firmato contratti con due aziende in grado di produrre tale mole di lavoro. Il primo contratto è stato firmato con la compagnia francese Jouve, la quale si dedica alla digitalizzazione e, dove necessario, all'inserimento manuale dei dati del *Corpus Scriptorum Ecclesiasticorum Latinorum* (CSEL) e dei primi cinquanta volumi della *Patrologia Latina*<sup>11</sup>. Inoltre, alla luce della collaborazione italo-tedesca, Jouve si occuperà della digitalizzazione di volumi destinati all'arricchimento di *digilibLT*, la biblioteca digitale dei testi latini tardoantichi dell'Università del Piemonte Orientale di Vercelli<sup>12</sup>. Il secondo contratto è stato firmato con Digital Divide Data (DDD), un'azienda americana con filiali in Laos e Cambogia<sup>13</sup>. DDD si occupa di digitalizzare i volumi 51-122 della *Patrologia Latina* e di altre opere greche, tra le quali quelle di Ateneo, Filone Alessandrino, Libanio, i commenti greci ad Aristotele (*Commentaria in Aristotelem Graeca*) e, in un prossimo futuro, la *Patrologia Graeca*<sup>14</sup>. L'intento, infatti, è quello di testare entrambi i *workflows* e, qualora portassero a buoni risultati, rinnovare i contratti per produrre edizioni elettroniche di Eschilo, della raccolta dei frammenti degli storici romani (*Historicorum Romanorum Reliquiae*) e di qualsiasi altro autore fosse richiesto dai collaboratori dell'OGL.

Il *workflow* di queste attività prevede che il gruppo di ricerca dell'Università di Lipsia gestisca l'*input* e verifichi la validità del prodotto finale, mentre le compagnie con le quali sono stati stipulati i contratti si occupano della parte tecnica e meccanica del progetto. L'organizzazione del lavoro può essere riassunta nel modo seguente:

Università di Lipsia – Ogni autore o volume o serie di volumi deve essere codificata secondo la struttura dell'edizione di riferimento. Questa necessità comporta la creazione di *templates* molteplici che riflettono la diversità delle edizioni, pur rimanendo sempre compatibili con le specifiche di EpiDoc e

<sup>11</sup> Per digitalizzazione si intende il riconoscimento ottico dei caratteri (OCR), la correzione dell'*output* dell'OCR, nonché la codifica in EpiDoc XML. Per informazioni sul gruppo Jouve si veda <http://www.jouve.com/>

<sup>12</sup> Vd. n. 10.

<sup>13</sup> <http://www.digitaldividedata.org>

<sup>14</sup> La decisione di dividere i volumi della *Patrologia Latina* fra Jouve e DDD è stata dettata dal desiderio di paragonare due differenti *workflows* e *outputs*. I risultati prodotti permetteranno di scegliere il procedimento migliore in termini di metodo/qualità/prezzo.

in particolare con la classe di marcatori (*tags*) CITE-friendly<sup>15</sup>. Il gruppo di lavoro dell'Università di Lipsia si occupa di analizzare la struttura di ogni edizione e di ricavarne un documento descrittivo con *template* allegato da inoltrare alle aziende ingaggiate per il lavoro. Esso, inoltre, si occupa anche di scaricare e fornire le scansioni esistenti delle suddette edizioni in formato TIFF, PNG o Jp2<sup>16</sup>. Queste immagini vengono correttamente catalogate e caricate su un *server* che contribuisce alla creazione del *corpus Open Greek and Latin*. Per quanto concerne le collaborazioni, è responsabilità dell'ente collaboratore fornire al gruppo di lavoro dell'Università di Lipsia le immagini necessarie. Sebbene la correzione degli errori prodotti dall'OCR venga effettuata dalle due aziende Jouve e DDD, lo strumento che queste ultime utilizzano per svolgere tale compito è stato sviluppato dall'Università di Lipsia. Nello specifico, il *Proofreader* (cfr. fig. 1) ottimizza uno strumento sviluppato da Bruce Robertson e Federico Boschetti e permette di allineare l'*output* dell'OCR a edizioni conosciute e consentire correzioni semi-automatiche tramite un'interfaccia semplice e intuitiva (Boschetti *et al.* 2009; cfr. inoltre Manmatha-Feng 2006 e Bryant *et al.* 2010)<sup>17</sup>. Il gruppo di lavoro di Lipsia si occupa infine di supervisionare il *workflow* e assicurarsi che i termini e le scadenze previste siano rispettati.

Jouve e DDD – Come si è detto, le due aziende si occupano di 'OCRizzare' le immagini fornite dall'Università di Lipsia, di correggere eventuali errori utilizzando lo strumento loro fornito e di codificare il testo secondo le specifiche EpiDoc che sono state predisposte. Eventuali commenti, problemi e richieste vengono gestite tramite posta elettronica e videoconferenze a scadenza regolare.

<sup>15</sup> Sulla CTS/CITE Architecture sviluppata dall'*Homer Multitext Project* per la codifica dei manoscritti omerici si veda <http://www.homermultitext.org/hmt-doc/cite/>

<sup>16</sup> Le biblioteche digitali di riferimento sono Archive.org (<https://archive.org/details/texts>), HathiTrust Digital Library (<http://www.hathitrust.org/>) e Deutsche Digitale Bibliothek (<https://www.deutsche-digitale-bibliothek.de/>). Le scansioni vengono scaricate e convertite nel formato richiesto da Jouve e DDD in maniera semi-automatica con strumenti *ad hoc* sviluppati dal gruppo di lavoro dell'Università di Lipsia. Ogni scansione necessita anche una corretta catalogazione in quanto le biblioteche digitali di riferimento spesso presentano *meta-data* errati. Un esempio è la *Patrologia Graeca* sotto la quale risultano essere stati erroneamente catalogati molti volumi, che richiedono pertanto un ulteriore controllo manuale.

<sup>17</sup> Il *Proofreader* è stato sviluppato e ottimizzato da Frederik Baumgardt (Università di Lipsia), Bruce Robertson (Mount Allison University) e Federico Boschetti (CNR Pisa). Su questi strumenti si vedano <https://github.com/CoPhi> e <http://hml.mta.ca/rigaudon>

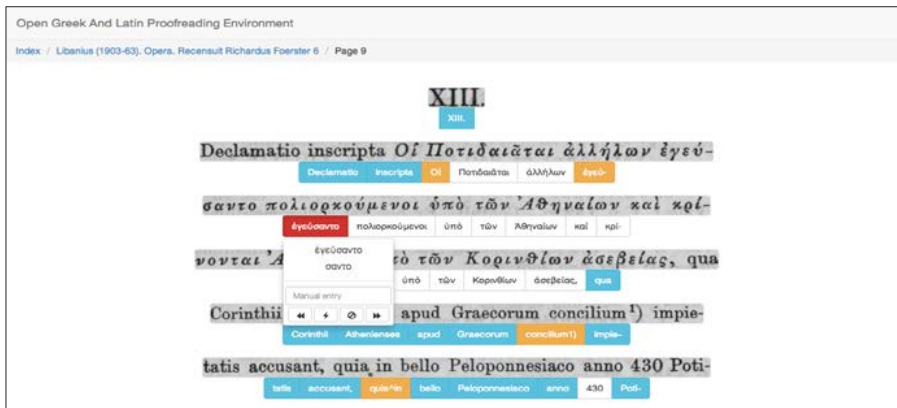


Fig. 1. *Proofreader*: strumento di correzione dell'*output* OCR.

Un terzo contratto, ancora in fase di definizione, includerà la scansione di edizioni pubblicate tra il 1922 e il 1985 e che sono dunque ancora soggette ai vincoli del *copyright*. Parte del *workflow* della biblioteca responsabile (tedesca per questioni di logistica) comporterà la rimozione degli apparati critici e delle note, fornendo all'Università di Lipsia solo il testo latino o greco dell'opera curata dall'editore. Tali scansioni verranno aggiunte al sistema digitale bibliotecario tedesco per permettere a terzi di usufruirne. Qualora la biblioteca in questione non disponesse dei libri necessari, sarà cura della biblioteca dell'Università di Lipsia fornirgliene copia<sup>18</sup>.

Un progetto che verrà avviato in futuro vedrà anche la partecipazione degli utenti al lavoro di digitalizzazione, come sta avvenendo ora mediante il coinvolgimento degli studenti dei corsi di filologia digitale organizzati presso l'Università di Lipsia e degli studenti Erasmus<sup>19</sup>. Il gruppo di lavoro di Lipsia sta infatti sviluppando un processo computazionale integrato con un sistema di pianificazione e notifica, che fornirà una visione sequenziale del progresso dei lavori dell'OGL e faciliterà i contributi esterni, per esempio da parte di ricercatori e studenti e di tutti coloro che sono interessati all'iniziativa. Vista la natura pubblica e aperta del progetto OGL, il *workflow* dell'OCR è stato progettato con interfacce che permettono agli utenti di partecipare

<sup>18</sup> La biblioteca di Lipsia, per quanto attrezzata in termini di digitalizzazione, non dispone ancora delle risorse umane necessarie per svolgere questo lavoro in tempi brevi.

<sup>19</sup> <http://www.dh.uni-leipzig.de/wo/courses/>

al lavoro di digitalizzazione. Sviluppato sulla base dell'Oracle Grid Engine, il *workflow* consiste di tre componenti principali: 1) un nucleo (*core*) intercambiabile di uno dei tre motori OCR (Gamera, Tesseract, OCropus); 2) un livello di ottimizzazione sviluppato da Bruce Robertson e Federico Boschetti; 3) un modulo per allineare l'*output* dell'OCR a edizioni conosciute e consentire correzioni semi-automatiche.

I dati prodotti da questo processo vengono codificati secondo le specifiche EpiDoc, le quali forniscono un tipo di marcatura standardizzata, ma non ristretta, e compatibile con i testi dell'OGL. Tale codifica viene realizzata in parallelo alla conversione in EpiDoc dei *file* della Perseus Digital Library. La possibilità di disporre dei testi della Perseus DL e dell'OGL in formato EpiDoc faciliterà lo scambio e il collegamento dei dati con le collezioni di documenti epigrafici e papirologici che sono già stati codificati in questo modo e con tutte quelle altre banche dati che sono attualmente in fase di conversione, come per esempio EAGLE (*Europeana Network of Ancient Greek and Latin Epigraphy*)<sup>20</sup>.

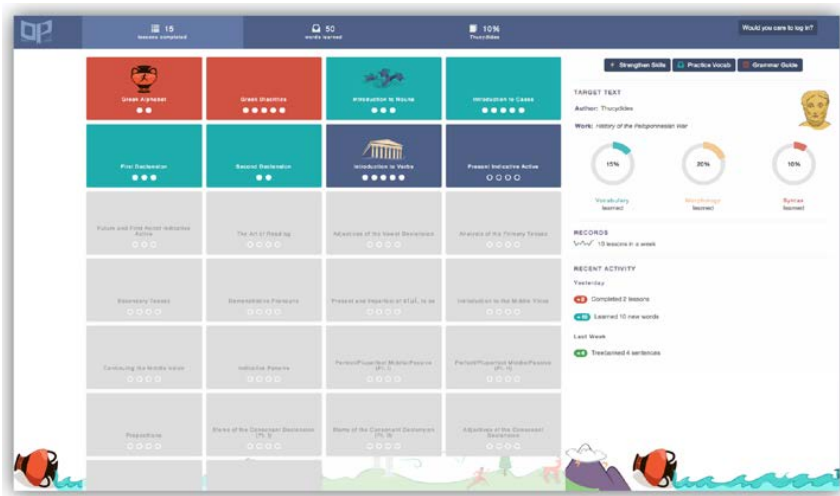
### 3. Historical Languages e-Learning Project

Un'ulteriore componente dell'OPP è rappresentata dall'*Historical Languages e-Learning Project*, il cui obiettivo è quello di realizzare un sistema per l'apprendimento delle lingue storiche in ambiente digitale. Questo sistema permette di selezionare frasi che abbiano una certa morfosintassi e/o un certo lessico, sulla base degli interessi specifici del discente o del docente che vuole impiegare questo sistema per insegnare le lingue classiche. Il testo selezionato per il *pilot* del progetto è una sezione del primo libro della *Guerra del Peloponneso* di Tucidide nota come *Pentecontaetia* (Thuc. 1.89-118). Lo scopo del *pilot* è quello di insegnare alcuni aspetti della lingua greca sia tramite l'uso della piattaforma *e-Learning* sia tramite un'attiva partecipazione di annotazione al testo. Dal successo del *pilot* dipenderà l'espansione del progetto per includere altri testi.

Per suscitare la curiosità del pubblico interessato e fidelizzare gli utenti, i creatori del *pilot* lavorano anche all'estetica della piattaforma e all'organizzazione del contenuto (cfr. fig. 2). L'intento è che il materiale sia disposto in modo intuitivo, divertente e incoraggiante, e che sia permesso all'utente di

<sup>20</sup> <http://www.eagle-network.eu/>





**Fig. 2.** *Historical Languages e-Learning Project*, Homepage.

imparare e partecipare, qualsiasi sia la sua conoscenza della lingua greca. La piattaforma *e-Learning* offrirà ulteriori vantaggi: la possibilità di iscriversi alla piattaforma (tramite registrazione e *log in*) e quindi la scelta di abbandonare e riprendere l'apprendimento a piacere, la possibilità di visualizzare il proprio percorso di apprendimento e una cronologia delle proprie annotazioni al testo, e infine la possibilità di esercitare la propria conoscenza linguistica tramite esercizi basati direttamente sul testo greco. La piattaforma offrirà inoltre la possibilità di scegliere la lingua moderna così che gli utenti possano imparare il greco con strumenti tradotti nella propria lingua madre (che non sia necessariamente l'inglese).

Motore del progetto è l'annotazione morfosintattica. I testi greci e latini sono annotati semi-automaticamente per la morfologia utilizzando il *tagger* Morpheus sviluppato dal Perseus Project, il quale restituisce un testo con l'analisi morfologica di ogni parola<sup>21</sup>. Nel caso di più analisi possibili, spetta all'annotatore decidere quale sia quella corretta sulla base del contesto. L'annotazione morfologica costituisce la base per l'annotazione sintattica che viene eseguita manualmente. Attraverso l'interfaccia grafica offerta da

<sup>21</sup> <http://wiki.digitalclassicist.org/Morpheus>



Alpheios<sup>22</sup>, l'annotatore costruisce un albero sintattico secondo delle *guidelines* che si ispirano a quelle adottate per la *Prague Dependency Treebank 2.0*<sup>23</sup>.

La *Ancient Greek and Latin Dependency Treebank* del Persues Project conta circa 400.000 parole<sup>24</sup>. Al momento è in corso una revisione tesa ad arricchire l'annotazione con l'aggiunta di glosse secondo lo schema delle *Leipzig Glossing Rules*<sup>25</sup>, al fine di promuovere un tipo di analisi standard per la morfologia di ogni parola. L'annotazione conterrà inoltre riferimenti alla grammatica greca dello Smyth (1920) per coniugare il sapere della grammatica tradizionale con quello della *Functional Generative Description* della *treebank* di Praga.

#### 4. Open Access Publishing

Uno degli obiettivi principali dell'OPP consiste nella creazione di un nuovo modello di edizioni scientifiche native digitali. Questo obiettivo è attualmente perseguito mediante l'implementazione di Perseids, che è una piattaforma collaborativa della Perseus DL sviluppata mediante la personalizzazione di risorse *open source* create per annotare fonti classiche codificate secondo lo standard TEI XML (per una descrizione della piattaforma e di diversi progetti ad essa connessi si veda Almas-Beaulieu 2013)<sup>26</sup>. Perseids è un ambiente condiviso dove gli utenti possono editare, tradurre e commentare diverse tipologie di fonti antiche, comprese le iscrizioni e i manoscritti. L'obiettivo di Perseids è duplice, perché mira sia alla pubblicazione di edizioni scientifiche che allo sviluppo di risorse didattiche per gli studenti dei corsi universitari:

1) Per quanto riguarda la comunità scientifica, uno dei principali modelli di pubblicazione all'interno di Perseids è il *Fragmentary Texts Editor* (FTE), che ha la funzione di produrre annotazioni complesse concernenti opere conservate solo attraverso citazioni e riusi in testi coevi o posteriori (Almas-Berti 2013a; Eadd. 2013b; Almas *et al.* 2013)<sup>27</sup>. A tal fine Perseids

<sup>22</sup> <http://alpheios.net/>

<sup>23</sup> <http://ufal.mff.cuni.cz/pdt2.0/>

<sup>24</sup> <http://nlp.perseus.tufts.edu/syntax/treebank/index.html>

<sup>25</sup> <http://www.eva.mpg.de/lingua/resources/glossing-rules.php>

<sup>26</sup> Perseids è disponibile al seguente indirizzo ed è liberamente accessibile <http://sites.tufts.edu/perseids/>

<sup>27</sup> Per una *demo* dell'FTE si veda [http://perseids.org/sites/berti\\_demo/](http://perseids.org/sites/berti_demo/). Il codice sorgente è disponibile al seguente indirizzo <https://github.com/PerseusDL/lci-demo>

utilizza diversi metodi di *in-line* e *stand-off* markup combinando lo standard TEI XML e la CTS/CITE Architecture con altri *data model*, quali l'Open Annotation Collaboration (OAC), il Systematic Assertion Model (SAM) e il W3C Provenance Model (Almas *et al.* 2013). Parallelamente al *Fragmentary Texts Editor*, la cattedra di informatica umanistica dell'Università di Lipsia sta avviando il *Leipzig Open Fragmentary Texts Series* (LOFTS), il cui obiettivo è la realizzazione di nuove edizioni native digitali di autori frammentari<sup>28</sup>. Il progetto è supportato dal Perseus Project e avrà come sede di pubblicazione il Center for Hellenic Studies<sup>29</sup>. Il primo sforzo nell'ambito di questa iniziativa è la digitalizzazione dei cinque volumi dei *Fragmenta Historicorum Graecorum* pubblicati da Karl Müller tra il 1841 e il 1870 (progetto *Digital Fragmenta Historicorum Graecorum* (DFHG)), i quali costituiscono la prima opera monumentale di raccolta dei frammenti degli storici greci e rappresentano un ottimo punto di partenza per contribuire alla realizzazione di edizioni digitali in materia<sup>30</sup>.

2) I risultati che Perseids mira a produrre non riguardano soltanto gli studiosi ma anche gli studenti, i quali hanno l'opportunità di lavorare direttamente sui documenti originali e contribuire ai risultati della comunità scientifica. Questo tipo di attività è svolto in parallelo presso la Tufts University e l'Università di Lipsia, la quale ha avviato una serie di corsi di filologia digitale. Attraverso questi corsi gli studenti apprendono come trattare diverse forme di organizzazione del sapere scientifico sviluppate dalla cultura della stampa, come le edizioni critiche, i lessici, le enciclopedie, i commentari, gli indici e le grammatiche. Gli studenti hanno inoltre l'opportunità di concentrarsi su temi particolarmente complessi, come l'annotazione linguistica delle fonti storiche, la rappresentazione delle fonti frammentarie e dei riusi testuali, o l'allineamento linguistico dei testi.

<sup>28</sup> <http://www.dh.uni-leipzig.de/wo/open-philology-project/the-leipzig-open-fragmentary-texts-series-lofts/>

<sup>29</sup> <http://chs.harvard.edu/>

<sup>30</sup> Per una descrizione del progetto si veda <http://www.dh.uni-leipzig.de/wo/open-philology-project/the-leipzig-open-fragmentary-texts-series-lofts/digital-fragmenta-historicorum-graecorum-dfhg-project/>. La pagina contiene un collegamento alle *guidelines* sviluppate dal gruppo di lavoro dell'Università di Lipsia per la codifica dei volumi secondo lo standard EpiDoc e un catalogo degli oltre 600 autori frammentari pubblicati dal Müller nei cinque volumi dei *FHG*. Le linee guida, oltre a fornire uno strumento per tutti coloro che collaborano al progetto, contribuiscono allo sviluppo generale delle *guidelines* di EpiDoc (<http://www.stoa.org/epidoc/gl/latest/>), mentre il catalogo degli autori contribuisce allo sviluppo e all'arricchimento del Perseus Catalog (<http://catalog.perseus.org/>).

## 5. Open Data Revenue Models e Open Philology Publishing

A supporto dell'attività scientifica sopra descritta, l'OPP intende sviluppare un *business plan* per creare strategie che permettano di sostenere economicamente il progetto e renderlo in futuro autonomo da investimenti esterni. Dato che l'OPP è per definizione basato su un modello di accesso libero e gratuito, la parte più complessa consiste nello sviluppare modelli che consentano il sostenimento di una piattaforma di apprendimento aperta e gratuita tramite l'aggiunta di servizi sofisticati a pagamento. Il principio base è quello di creare un'alternativa all'attuale monopolio della produzione del sapere, la cui fruizione è molto costosa per l'utente, favorendo un accesso gratuito, il quale sia però arricchito di servizi addizionali a basso costo per apprendere, analizzare e contribuire ad una massa di dati complessi in costante crescita. Il progetto intende fornire strumenti destinati a studiosi e studenti, oltre che alle scuole e in generale al pubblico interessato. I servizi offerti copriranno diverse aree, dai servizi informatici per l'*e-Learning*, ai libri di testo interattivi, ai sistemi di valutazione e di *ePortfolio*.

## 6. Bibliografia

- Almas B., Beaulieu M.C. (2013). *Developing a New Integrated Editing Platform for Source Documents in Classics*. «Literary and Linguistic Computing», vol. 28, no 4, pp. 493-503. URL=<http://llc.oxfordjournals.org/content/28/4/493.abstract>. [ultima visita 3.3.2014].
- Almas B., Berti M. (2013a). *Perseids Collaborative Platform for Annotating Text Re-Uses of Fragmentary Authors*. In F. Tomasi, F. Vitali, a. c. di, *DH-Case 2013. Collaborative Annotations in Shared Environments: metadata, vocabularies and techniques in the Digital Humanities*, ACM, art. no. 7. URL=<http://dl.acm.org/citation.cfm?id=2517986>. [ultima visita 3.3.2014].
- Almas B., Berti M. (2013b). *The Linked Fragment: TEI and the Encoding of Text Re-uses of Lost Authors*. In F. Ciotti, A. Ciula, a. c. di, *The Linked TEI: Text Encoding in the Web. TEI Conference and Members Meeting 2013*. Università Roma La Sapienza, pp. 12-16. URL=<http://digilab2.let.uniroma1.it/teiconf2013/wp-content/uploads/2013/09/book-abstracts.pdf>. [ultima visita 3.3.2014].
- Almas B., Berti M., Choudhury S., Dubin D., Senseney M., Wickett K.M. (2013). *Representing Humanities Research Data Using Complementary Provenance Models*. Poster presentato al *Building Global Partnerships - RDA Second Plenary Meeting in Washington DC, 16-18 September 2013*.

- URL=[http://www.fragmentarytexts.org/wp-content/uploads/2013/09/LTH\\_RDAPoster\\_2013.pdf](http://www.fragmentarytexts.org/wp-content/uploads/2013/09/LTH_RDAPoster_2013.pdf). [ultima visita 3.3.2014].
- Blackwell C., Crane C. (2009). *Cyberinfrastructure, the Scaife Digital Library and Classics in a Digital Age*. «Digital Humanities Quarterly», vol. 3, no 1. URL=<http://www.digitalhumanities.org/dhq/vol/003/1/000035/000035.html>. [ultima visita 3.03.2014].
- Böckh A. (1858). *Gesammelte kleine Schriften*, vol 1. Druck und Verlag von B.G. Teubner.
- Böckh A. (1877). *Encyklopädie und Methodologie der philologischen Wissenschaften*. Druck und Verlag von B.G. Teubner.
- Boschetti F., Romanello M., Babeu A., Bamman D., Crane G. (2009). *Improving OCR Accuracy for Classical Critical Editions*. In Agosti M. et al., a. c. di, *Research and Advanced Technology for Digital Libraries*, vol. 5714, Springer-Verlag, pp. 156-167.  
URL=[http://link.springer.com/chapter/10.1007%2F978-3-642-04346-8\\_17](http://link.springer.com/chapter/10.1007%2F978-3-642-04346-8_17). [ultima visita 3.3.2014].
- Bryant M., Blanke T., Hedges M., Palmer R. (2010). *Open Source Historical OCR: The OCRopodium Project*. In Lalmas M., Jose J., Rauber A., Sebastiani F., Frommholz I., a. c. di, *Research and Advanced Technology for Digital Libraries*, vol. 6273, Springer-Verlag, pp. 522-525.  
URL=[http://link.springer.com/chapter/10.1007%2F978-3-642-15464-5\\_72](http://link.springer.com/chapter/10.1007%2F978-3-642-15464-5_72). [ultima visita 3.3.2014].
- Manmatha R., Feng S. (2006). *A Hierarchical, HMM-Based Automatic Evaluation of OCR Accuracy for a Digital Library of Books*. In *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries*, ACM, pp. 109-118.  
URL=<http://dl.acm.org/citation.cfm?doid=1141753.1141776>. [ultima visita 3.3.2014].
- Smyth H.W. (1920). *A Greek Grammar for Colleges*. American Book Company.