

English

Cultural Objects in Digital Resources: Imagining the Text

Italian

Oggetti culturali in risorse digitali: immaginando il testo

Arianna Ciula (arianna.ciula@kcl.ac.uk)
King's College London (London, United Kingdom)

PAPER

[English]

This paper proposes an evaluation of humanities computing resources based on their ability to represent texts. The discussion is based on the analysis of several projects – to the development of several of which the author has directly collaborated – different in nature, but all related to the representation of textual material connected to its visual or graphic ‘counterpart’.

It is well known that humanities scholars use images for their research in various formats and combinations. In particular, copies, illustrations, surrogates are used to observe and interpret the original cultural artefacts otherwise not directly available and/or not legitimately transformable. When historical texts – in the broad sense as understood by CLiP – are taken into account, the connection between the textual content and its physical object is particularly explicit. This is valid for philology or palaeography as well as for archaeology and art history. However, in the former case the text of interest generally refers to a support perceived as a two dimensional image. It is an extreme case of symbiosis between the text as a string of print/digital characters – already a first level of interpretation and mark-up in itself – and its extant materiality as a physical artefact (codex, roll, early print edition etc.) recalled by a two dimensional representation either analogical or digital.

Although humanities computing projects can be very different in methodologies and aims, within the digital resources created for supporting academic research it is possible to identify some categories relevant to the focus of this paper:

1. Digital resources which give priority to texts
 - a. encoded texts
 - b. visualisation of the textual encoding
2. Digital resources which give priority to images
 - a. low resolution images of texts
 - b. high resolution images of texts
 - c. virtual restoration of images of texts
 - d. annotation of images of texts
3. Digital resources which combine images and texts equally
 - a. juxtaposition of independent sources (textual and graphic)
 - b. cross-references and connection of linked sources (textual and graphic)

This categorisation is not exhaustive nor exclusive – indeed, different subcategories can vary a lot semantically and can co-occur in the same resource –, but it is appropriate for the evaluation which the paper aims to accomplish. Although point 1.a is not relevant here, it has been included so as to contextualise point 1.b. The visualisation of the encoding of a text is usually hidden from the main interface of whatever

encoded textual material a digital resource provides. When the encoding is shown, a visualisation of the interpretative layers beneath the textual sequence is made available. What is visible is not the correspondent folio of a manuscript transcription for instance, but the components that the encoder or the editor has considered as structural parts of the text. This visualisation cannot be considered a graphic correspondent. It is rather a visual representation of some selected elements of which the user/scholar may already possess a mental image or, even more interestingly, a new and unusual visual representation of the text generated by the encoding process and open to further research.

Similarly, point 2.a is not relevant to the issues this paper deals with. However, a consistent amount of digital resources offer just low resolution images for a combination of reasons that can range from financial restrictions to copyright permissions, from the aim of producing a cataloguing resource rather than an imaging resource to the lack of consideration for image-based analysis. Whatever the reason, the result is the same: the image of the text is subsidiary and can only be explored to a minimal extent. Points 2.b, c and d apply to all those cases in which the images are recognised as fundamental interfaces to the text they contain. Provided that the browsing resources are good enough, detailed observations of the material aspects of the image-pages can be carried out (given point 2.b). The editorial effort related to the images themselves is minimal, but the results can be very appealing for image-based research. Points 2.c and d require instead careful planning of imaging pre-processing and post-processing respectively, with no guarantee of the output results. When the operations of image processing and image mark-up are successful, the developers of the resource add their level of analysis to the graphical objects and make available a layer of interpretations which adds value to the resource by casting, at the same time, the representation of the primary source as such.

Point 3 represents the balanced combination of the previous categories and implies several complications both in terms of editorial process and fruition tools. Indeed, if so far the analysis has not mentioned the dynamic aspects of a digital resource, the active role of the user/reader as editor/developer is emerging more and more in relation to this type of resources (especially point 3.b).

To conclude, the objectives of the paper are to analyse the digital resources of historical textual materials under a perspective where:

- the encoded text, when made visible in all its stratification, is considered itself a visual counterpart of the "normal" text;
- the representation of the physical materiality of the text – often plural because witnessed in many occurrences, especially in the case of classic and medieval traditions – is evaluated in different grades on the basis of the implications for image-based research.

[Italian]

Questo articolo propone una valutazione delle risorse di informatica umanistica sulla base della capacità delle stesse di rappresentare testi. La discussione si basa sull'analisi di vari progetti – allo sviluppo di alcuni dei quali l'autore ha partecipato direttamente – di natura diversa, ma tutti correlati alla rappresentazione di materiale testuale connesso alla sua "controparte" visiva o grafica.

E' ampiamente noto il fatto che gli studiosi di discipline umanistiche usino nel loro processo di ricerca immagini di vario formato in varie combinazioni. In particolare, copie, illustrazioni, sostituti di varia tipologia sono ritenuti utili nell'osservare e interpretare gli artefatti culturali originali, altrimenti non direttamente accessibili e/o legittimamente modificabili. Nel caso in cui si prendano in considerazione testi – intesi nel senso ampio che CLiP considera – di natura storica, la connessione tra il contenuto testuale e l'oggetto fisico è particolarmente esplicita. Ciò vale per discipline come la filologia e la paleografia, così come per l'archeologia e la storia dell'arte. Tuttavia, nel primo caso il testo d'interesse ha generalmente come referente un supporto percepito in forma di immagine bidimensionale. E' questo un caso estremo di simbiosi tra il testo in qualità di stringa di caratteri a stampa/digitali – già un primo livello di interpretazione e codifica in se stesso – e la materialità esistente come artefatto fisico (codice, rotolo, edizioni a stampa antica etc.) richiamata da una rappresentazione bidimensionale che sia analogica o digitale.

Anche se i progetti di informatica umanistica possono variare di molto in metodologie e obiettivi, all'interno

delle risorse digitali create per supportare la ricerca accademica è possibile identificare delle categorie rilevanti al tema del presente articolo:

1. Risorse digitali che danno priorità ai testi
 - a. testi codificati
 - b. visualizzazione della codifica testuale
2. Risorse digitali che danno priorità alle immagini
 - a. immagini di testi a bassa risoluzione
 - b. immagini di testi ad alta risoluzione
 - c. restuaro virtuale di immagini di testi
 - d. annotazione di immagini di testi
3. Risorse digitali che combinano equamente immagini e testi
 - a. giustapposizione di risorse indipendenti (testuali and grafiche)
 - b. riferimenti e connessione tra risorse collegate (testuali and grafiche)

La categorizzazione di sopra non ambisce ad essere né esaustiva, né esclusiva – sotto categorie possono variare di molto tra loro in termini semantici e possono co-occorrere nella stessa risorsa –; tuttavia essa sembra appropriata per la valutazione che il presente articolo intende delineare. Nonostante il punto 1.a non sia qui rilevante, esso è stato incluso così da contestualizzare il punto 1.b. La visualizzazione della codifica di un testo è solitamente nascosta a vantaggio dell'interfaccia principale di qualsivoglia testo codificato messo a disposizione da una risorsa digitale. Allorquando la codifica è resa evidente, si rende disponibile una visualizzazione dei livelli interpretativi al di là della sequenza testuale. Ciò che è visibile non è, per esempio, la carta corrispondente alla trascrizione di un manoscritto, bensì i componenti che chi fa la codifica o l'editore ritengono parti strutturali del testo. Questa visualizzazione non può essere considerata un corrispondente grafico; si tratta piuttosto di una rappresentazione visuale di alcuni elementi selezionati dei quali l'utente/studioso potrebbe possedere già un'immagine mentale o, meglio ancora, di una nuova e inusuale rappresentazione del testo scaturita dal processo di codifica e aperta ad ulteriore ricerca.

Allo stesso modo, il punto 2.a non è rilevante relativamente agli argomenti di questo articolo. Tuttavia, un numero consistente di risorse digitali offre soltanto immagini a bassa risoluzione per una serie di ragioni che variano dalle restrizioni finanziarie ai permessi di copyright, dall'intenzione di produrre una risorsa catalografica piuttosto che prettamente grafica alla mancanza di considerazione dell'analisi basata sulle immagini. Qualsiasi sia la motivazione, il risultato rimane lo stesso: l'immagine del testo risulta essere sussidiaria e si presta di fatto ad una esplorazione minima. I punti 2.b, c e d sono validi per tutti quei casi nei quali alle immagini è riconosciuto il ruolo di interfaccia fondamentale al testo che contengono. Se gli strumenti per la navigazione in rete lo permettono, osservazioni dettagliate degli aspetti materiali possono essere effettuate sulle pagine immagine (premessò il punto 2.b). Il lavoro editoriale relativo alle immagini è minimo, ma i risultati possono risultare alquanto appetibili per la ricerca basata sulle immagini. I punti 2.c e d richiedono invece una pianificazione accurata, rispettivamente, del *pre-processing* and *post-processing* delle immagini, senza garanzia sui risultati finali. Allorquando le operazioni di elaborazione dell'immagine e di codifica della stessa hanno successo, coloro che implementano la risorsa aggiungono il loro livello di analisi agli oggetti grafici e rendono disponibile uno strato interpretativo che attribuisce valore aggiunto alla risorsa, nascondendo allo stesso tempo la rappresentazione della fonte primaria in quanto tale.

Il punto 3 rappresenta una combinazione equilibrata delle categorie precedenti e implica delle complessità sia in termini di processo editoriale che di strumenti di fruizione. Infatti, se finora l'analisi non ha fatto menzione degli aspetti dinamici di una risorsa digitale, il ruolo attivo dell'utente/lettore in qualità di editore/"implementatore" emerge proprio in relazione a questo tipo di risorse (in modo specifico per il punto 3.b).

In conclusione, gli obiettivi dell'articolo sono quelli di analizzare le risorse digitali di materiale testuale di valore storico in una prospettiva in cui:

- il testo codificato, se reso visibile in tutta la sua stratificazione, è considerato esso stesso una controparte visuale del testo "normale";

- la rappresentazione della materialità fisica del testo – spesso molteplice, poichè testimoniata in diverse occorrenze, specialmente nel caso di tradizioni classiche e medievali – è valutata in gradazioni diverse a seconda delle implicazioni per la ricerca basata sull'immagine.

References

- Archimedes Project*. <http://archimedes2.mpiwg-berlin.mpg.de/archimedes_templates/> (December 2005)
- Brown, Michael S. and W. Brent Seales. 2001. 'The Digital Atheneum: New Approaches for Preserving, Restoring and Analyzing Damaged Manuscripts'. *ACM*: 437-443.
- Bunke, Horst, and A. Lawrence Spitz, eds. 2006. 'Document Analysis Systems VII'. *Lecture Notes in Computer Science* 3872.
- Burnley, David and Alison Wiggins, eds. 2003. *The Auchinleck Manuscript*. <<http://www.nls.uk/auchinleck/>> (December 2005)
- Buzzetti, Dino. Forthcoming. 'Biblioteche digitali e oggetti digitali complessi: Esaustività e funzionalità nella conservazione'. In *Atti del Workshop ErpaNET Trusted Repositories for Preserving Cultural Heritage* (Roma, Accademia dei Lincei, 17-19 November 2003).
- Coombs, James, and Allen H. Reaner, Steven J. De Rose. 1987. 'Markup Systems and the Future of Scholarly Text Processing', *Communications of the ACM* 30: 933-47.
- Codices Electronici Ecclesiae Coloniensis (CEEC)*. <<http://www.ceec.uni-koeln.de/>>
- Dekhtyar, Alex, and Ionut E. Iacob, Jerzy W. Jaromczyk, et al. 2006. 'Multihierarchical XML Markup of Image-based Electronic Editions'. *International Journal on Digital Libraries (IJDL)* 6.1: 55-69. <<http://dblab.csr.uky.edu/~eiaco0/publications/IJDL2004.pdf>> (April 2006)
- Duggan, Hoyt N. gen. ed. *The Piers Plowman Electronic Archive*. <<http://jefferson.village.virginia.edu/seenet/piers/>> (December 2005)
- Guyotjeannin, Olivier, and Gautier Poupeau. *Le Cartulaire blanc de Saint-Denis*. <<http://elec.enc.sorbonne.fr/cartulaireblanc/>> (December 2005)
- Harvard Law School Library. 2003. *Bracton Online: Bracton: De Legibus Et Consuetudinibus Angliæ*. <<http://hisl.law.harvard.edu/bracton/>> (December 2005)
- Henry III Fine Rolls Project*. 2005-2008. <<http://www.frh3.org.uk/>> (April 2006)
- Institute for New Testament Textual Research of University of Münster. 2003-2005. *Digital Nestle-Aland Prototype*. <<http://nestlealand.uni-muenster.de/>> (December 2005)
- Irish Script on Screen*. <<http://www.isos.dias.ie/>> (December 2005)
- Kiernan, Kevin, ed. 2002-2006. *Electronic Boethius*. <<http://beowulf.engl.uky.edu/~kiernan/eBoethius/inlad.htm>> (April 2006)
- Kiernan, Kevin, Alex Dekhtyar, Jerzy W. Jaromczyk, et al. 2004. 'Edition Production Technology (EPT) and the ARCHway Project'. *DigiCULT Newsletter*. <<http://dblab.csr.uky.edu/~eiaco0/publications/DigiCULT04.html>> (December 2005)
- McGann, J. J. 2001. *Radiant Textuality: Literature after the world wide web*. Palgrave: New York.
- Online Chopin Variorum Edition Pilot Project*. 2003-2004. <<http://www.ocve.org.uk>> (December 2005)
- Scholarly Digital Editions*. <<http://www.sd-editions.com/>> (December 2005)
- Sperberg-McQueen, C. M. 1991. 'Text in the Electronic Age: Textual Study and Text Encoding with Examples From Medieval Text'. *Literary and Linguistic Computing* 6.1: 32-46.