

Codifica ed edizioni digitali: appunti da una conferenza per una riflessione diacronica.

"Essentially, all models are wrong, but some are useful"
George E. P. Box

La conferenza annuale della *Text Encoding Initiative (TEI)*¹, insieme alla relativa assemblea dei soci, in inglese *member's meeting*, termine a prima vista freddo ma che in realtà sta a indicare la variegata e interdisciplinare comunità scientifica che si raduna ormai da diversi anni attorno a questa associazione internazionale, si è svolta dal 2 al 5 ottobre 2013 a Roma², presso L'Università La Sapienza, ospitata nello specifico dal *DigiLab*³. L'evento è significativo per diversi motivi: principalmente per la valenza scientifica e culturale, dato il ruolo di primo piano che la *TEI* ha ormai, stabilmente e da diverso tempo, nel multiforme e sempre crescente panorama dell'Informatica Umanistica (o, se si preferisce il termine anglosassone, *Digital Humanities*, non privo sia di una connotazione sia di una denotazione altra rispetto a quello italiano), e può fornire lo spunto per qualche riflessione sul ruolo della codifica e il suo rapporto con l'edizione digitale, attraverso un'analisi diacronica di questi incontri.

Chi scrive ha avuto la possibilità di partecipare alla quasi totalità delle conferenze europee della *TEI*⁴, a partire dal 2001 a Pisa, la prima in Italia e che sancì la nascita del *Consortium*, l'assetto organizzativo raggiunto dopo 14 anni l'ormai storico incontro di Poughkeepsie⁵, da cui presero vita gli sforzi coordinati che di lì a poco produssero la prima versione⁶ delle *Guidelines for Electronic Text Encoding and Interchange*.

¹ <<http://www.tei-c.org>>

² <<http://digilab2.let.uniroma1.it/teiconf2013/>>

³ <<http://digilab.uniroma1.it/>>

⁴ Una regola non scritta all'interno della comunità prevede come le conferenze abbiano luogo alternativamente un anno in Nord America e il successivo in Europa, anche se, a partire dal 2001, hanno fatto eccezione le annate 2006-2007 con, rispettivamente, Victoria, Canada e Maryland, USA e 2010-2011, in Croazia e Germania. L'incontro europeo a cui non ho potuto partecipare è quindi proprio quello del 2010, a Zadar, in Croazia, essendo presente invece, a quello già citato a Pisa, al successivo a Nancy, poi Sofia, Londra, Würzburg, e infine Roma. Per un elenco completo vedi <<http://www.tei-c.org/Membership/Meetings/2011/>>. A conferma di questa regola ufficiale, l'incontro del 2014 si svolgerà a Chicago, ospitato dalla Northwestern University.

⁵ Vedi <<ftp://ftp.uic.edu/pub/tei/info/hist.html>> e <<http://www.tei-c.org/About/history.xml>>

⁶ *Proposal 1*, abbreviato *P1*, e da quel momento in poi ogni versione delle *Guidelines* ha mantenuto questa sigla iniziale, incrementando progressivamente la numerazione, fino ad arrivare all'attuale *P5*.

In quell'occasione la conferenza fu ospitata e organizzata a livello locale dall'Istituto di Linguistica Computazionale del CNR⁷, nella figura di Antonio Zampolli, all'epoca direttore dell'Istituto e presidente dell'*Association for Literary and Linguistic Computing (ALLC)*⁸, che dopo neppure due anni scomparve tragicamente, andando ad intitolare sia l'Istituto stesso sia, successivamente, un premio dell'*Alliance of Digital Humanities Organizations*⁹. Sempre legato a quel convegno del 2001 non si può non ricordare la figura di Giuseppe Gigliozi, esponente di primo piano della scuola romana d'Informatica Umanistica, che venne a mancare appena due settimane prima del convegno, non potendo quindi né partecipare alle elezioni del *Board*, cui era candidato, né presentare la relazione sul progetto Testi in Linea – TIL, che vedeva proprio l'utilizzo dello standard *TEI*, all'epoca declinato nella sintassi *SGML*¹⁰, come nucleo centrale di una biblioteca digitale disponibile sul Web di testi della letteratura italiana, congiuntamente al software di pubblicazione *DynaWeb*¹¹. La relazione fu presentata da uno dei suoi migliori allievi, Fabio Ciotti, che a distanza di dodici anni e di qualche centinaio di chilometri ha ricordato Gigliozi proprio nel discorso di apertura della *2013 Conference* ed è stato eletto nel *Council*, l'organo che si occupa delle questioni più specificatamente tecniche nello sviluppo dello standard¹².

La scarsa, se confrontata con quelle più recenti, dimensione quantitativa dell'incontro del 2001 era controbilanciata da un'atmosfera intima, di persone che si conoscevano e lavoravano insieme da diverso tempo¹³ su entrambe le sponde sia dell'Atlantico sia del Pacifico, in un mondo molto meno interconnesso di quello attuale, nonostante la forte familiarità dei partecipanti con le tecnologie telematiche. All'epoca il côté dell'Informatica Umanistica (in ambito anglosassone definita ancora come *Humanities Computing*) e della codifica testuale aveva già raggiunta una massa

⁷ <<http://www.ilc.cnr.it/>>

⁸ <<http://www.allc.org/>>

⁹ <<http://adho.org/awards/antonio-zampolli-prize>>

¹⁰ Ossia fino alla versione *P3*, in quanto già la *P4* segnò il passaggio a XML, anche se per utilizzare le parole di Sebastian Rathz durante la conferenza *TEI@Galway* nell'aprile del 2008, fu solo con la *P5* che le *TEI Guidelines* divennero un “fully-fledged citizen in the XML universe”, includendo, ad esempio, l'utilizzo dei *Namespaces*.

¹¹ Uno dei pochi strumenti, proprietari e dal non certo facile utilizzo, che all'epoca permettersero di sfruttare la potenza di *SGML* e quindi utilizzato in progetti di ricerca affini, come ad esempio il *Blake Archive* o il *Womens' Writers Project*. Il passaggio a XML, avvenuto proprio in quel periodo, ha favorito l'utilizzo di strumenti *open-source*, e di conseguenza una maggiore diffusione dello standard *TEI* con relativo allargamento della comunità.

¹² Laddove invece il *Board* ha finalità di tipo strategico/amministrativo.

¹³ Oltre al già citato Zampolli vanno menzionati C. M. Sperberg-McQueen, Lou Burnard, Allen Renear, Sebastian Rahtz, Syd Bauman, Laurent Romary e Chris Wittern, figure ancora estremamente attive e di rilievo nella comunità.

critica considerevole, e con uno sguardo retrospettivo quell'incontro pisano forse va più considerato come un punto di partenza, una sorta di riunione strategica in cui un informale nucleo operativo decise, oltre allo sviluppo delle linee guida per la codifica elettronica dei testi letterari e linguistici, le ormai famose *TEI Guidelines*, anche quelle programmatiche per la diffusione dello standard e l'ampliamento della comunità dei *textual encoders*.

A conferma di questa ipotesi, e della validità di quell'ipotetico quanto informale programma, già dal successivo incontro europeo, nel 2003 in Francia, a Nancy, la situazione era decisamente cambiata, presentando una platea decisamente più vasta, composta sia da studiosi affermati, esponenti di altre discipline interessati alle potenzialità della codifica digitale, sia da esponenti delle nuove generazioni, che con il tempo avrebbero assunto un ruolo sempre più stabile e significativo all'interno delle attività del consorzio.

L'incontro di Nancy segna anche la nascita degli *Special Interest Groups (SIG)*¹⁴, gruppi di lavoro incentrati su di un aspetto specifico della codifica testuale, e di conseguenza della testualità in generale¹⁵, o di aspetti derivati più legati alla dimensione digitale del testo. Per dovere di completezza va sottolineata la differenza tra i *SIG* e i *Workgroups*¹⁶, in quanto questi ultimi hanno come scopo principale quello di modificare o estendere le *Guidelines*, mentre nel caso dei primi non è detto.

Tra i vari *SIG*, il più rilevante ai fini del discorso che andremo a sviluppare è quello denominato originariamente *Presentation Tools*, e attualmente con un più generico ed onnicomprensivo *Tools*¹⁷. L'etichetta di *Presentation* infatti andò stretta sin dal primo incontro¹⁸ (ed è effettivamente riduttiva, in quanto semplifica e meccanizza i processi di elaborazione computazionale sottostanti, riducendo il tutto a una sorta di tipografia dinamica), anche se in quel particolare momento evidenziava

¹⁴ <<http://www.tei-c.org/Activities/SIG/>>

¹⁵ Come fatto notare da Patrick Sahle alla conferenza di Roma, durante un panel sul gruppo di lavoro del network europeo *NEDIMAH* <<http://www.nedimah.eu/>> incentrato proprio sulle edizioni elettroniche, in cui ha evidenziato sei diverse dimensioni della testualità, tra cui quella linguistica, quella visiva e quella bibliografica, e come ognuna di esse fosse appannaggio di uno specifico *SIG*.

¹⁶ <<http://www.tei-c.org/Activities/Workgroups/>>

¹⁷ <<http://wiki.tei-c.org/index.php/SIG:Tools>>

¹⁸ Così come si legge dal verbale informale del primo incontro, <http://wiki.tei-c.org/index.php/Minutes_of_the_TEI_Presentation_Tools_SIG_Meeting_in_Nancy_08_November_2003>, in cui viene citato il *SIG* complementare focalizzato sugli strumenti di *authoring*, per la creazione di testi elettronici, anch'esso confluito successivamente nel più generale *Tool SIG*, che così come riportato sulla prima pagina si occupa di strumenti per la codifica, conversione, linguaggi di trasformazione e presentazione, memorizzazione, interrogazione, pubblicazione e analisi. Il wiki del *SIG* al momento classifica 90 strumenti divisi in 13 categorie <<http://wiki.tei-c.org/index.php/Category:Tools>>.

un bisogno fortemente sentito all'interno della comunità, relativo alla condivisione di esperienze concrete e pratiche d'utilizzo per la pubblicazione sul web di testi *TEI*, evitando il più possibile la duplicazione degli sforzi e il continuo reinventare la ruota¹⁹.

Rispetto al panorama precedente, relativo agli strumenti disponibili per *SGML*, la diffusione sempre più pervasiva di programmi *open-source* in grado di manipolare documenti *XML* era sì un'opportunità enorme in quanto apriva nuovi scenari, ma come tutte le fasi di transizione era caratterizzata da un'instabilità confusiva. Va sottolineato come le discussioni dell'epoca fossero, per forza di cose, caratterizzate da una dimensione quasi esclusivamente pragmatica, e come le riflessioni teoriche su questo aspetto avrebbero iniziato a comparire solo una volta risolte le questioni pratiche più immediate²⁰, laddove invece il percorso evolutivo del mark up sembra aver bilanciato in maniera più equilibrata questi due aspetti.

Successivamente si passa a Sofia nel 2005, un incontro caratterizzato da una dimensione più intima, causa anche la posizione non centralissima e poi a Londra nel 2008, su cui vale la pena di spendere qualche riflessione. La *TEI Conference* del 2008 era successiva a quella del ventennale (prendendo sempre Poughkeepsie come punto di riferimento) svoltasi all'University of Maryland, che nel 2009 ospiterà il *DH*, sigla che è al tempo stesso un'evidente abbreviazione di *Digital Humanities*, e che indica il convegno internazionale per eccellenza dedicato a questa disciplina²¹. Quasi fosse un ipotetico passaggio di testimone il *DH2010* si svolse proprio Londra, sempre al King's College, la stessa istituzione ospite del meeting *TEI* di due anni prima. Questa singolare coincidenza offre il destro per sottolineare il rapporto tra questi due appuntamenti scientifici²², il *TEI Members Meeting* e la *DH Conference* entrambi in

¹⁹ F. MESCHINI, *Free Yourself from C.C.M.S. (Commercial Content Management Systems)*, ALLC/ACH 2004, Göteborg, Sweden, 2004.

²⁰ Basta confrontare il diverso impianto di questi due articoli, *Storage, Retrieval, and Rendering* di Sebastian Rahtz <http://www.tei-c.org/About/Archive_new/ETE/Preview/rahtz.xml>, pubblicato in quello stesso periodo, e *TEI CHI and the Tools Paradox: Developing a Publishing Framework for Digital Editions* di Sebastian Pape et al. del 2012 <<http://jtei.revues.org/432>> in cui all'aspetto più prettamente tecnologico viene affiancata una riflessione metodologica e comparativa sullo sviluppo di strumenti per la pubblicazione di testi elettronici.

²¹ Il *DH* nasce ufficialmente nel 2006 a Parigi, prendendo il posto di quella che era la 'joint conference' delle due principali associazioni internazionali d'informatica umanistica, rispettivamente l'*Association for Computers and the Humanities (ACH)* <<http://www.ach.org>> e la già citata *ALLC*, recentemente divenuta *European Association for Digital Humanities (EADH)* e quindi denominata *ALLC/ACH Conference* sin dalla prima edizione nell'estate del 1989 a Toronto.

²² Per completezza vanno citate sia *Balisage* <<http://www.balisage.net/>> sia l'*International Conference on Electronic Publishing (ElPub)*, entrambe focalizzate su argomenti di livello più generale, rispettivamente il mark up e l'editoria elettronica. Anche se evidentemente collegate, non esiste però, in particolare con la seconda, lo stesso livello d'interazione presente tra il *TEI MM* e il *DH*.

un chiaro rapporto di interazione e sovrapposizione²³. Ora, va da sé come la conferenza ufficiale dell'informatica umanistica offra uno spettro maggiore rispetto alla sola codifica testuale, abbracciando ogni possibile approccio tecnologico, metodologico o teorico dell'incontro tra le discipline computazionali e le scienze umane. In ogni caso la codifica testuale è sempre o la base e il punto di partenza o perlomeno una componente rilevante di numerosi progetti di ricerca ed applicazioni. Basti pensare alle edizioni critiche digitali, in cui al testo elettronico viene sovrainposto uno strumento di pubblicazione, una struttura algoritmica in grado di dargli vita e renderlo fruibile, creando così un sistema informativo dotato di diverse funzionalità, in teoria potenzialmente infinite, non di rado eterogenee tra edizione ed edizione anche a partire da uno stesso testo; funzionalità che hanno lo scopo in un primo momento di replicare, e successivamente di estendere le sofisticate caratteristiche delle edizioni a stampa cui la lunga, nobile e variegata tradizione filologica ci ha abituato²⁴.

L'aver evidenziato il concetto di edizione digitale come processo al tempo stesso complementare, di evoluzione e rottura rispetto alla fase della codifica è affatto voluto. Una volta raggiunta una certa stabilità per ciò che riguarda gli aspetti relativi alla codifica uno spostamento verso il livello successivo è sia fisiologico sia auspicabile. Non essendo però, le pratiche dell'edizione digitale l'interesse scientifico primario della TEI, ma una sorta di interesse, per quanto importante, comunque derivato, fino ad ora le relazioni incentrate sulle edizioni si sono concentrate sulla presentazione di casi singoli, logicamente sempre con un'attenzione particolare verso gli aspetti relativi alla codifica, e raramente su un'analisi globale che mediasse tra le numerose singolarità specifiche ad ogni caso. Questo non senza le dovute eccezioni. Durante il già citato incontro del ventennale Fotis Jannidis ricorda come «But we all know that *there is more to an edition than to encode a text*. In designing an edition, we also create ways to approach the text and to interact with it, in short visualization and functionality, i.e. algorithms.»²⁵. Oppure a Würzburg, nel 2011, ben due relazioni

²³ O se si preferisce *overlapping*, utilizzando così un termine familiare che indica un concetto estremamente affascinante per chi si occupa di codifica testuale.

²⁴ Tra i vari esponenti di rilievo della comunità dei critici testuali, che si sono occupati del cambiamento paradigmatico provocato dall'utilizzo del medium digitale come veicolo dei risultati dell'indagine filologica, vanno citati in particolare Peter Shillingsburg, Jerome McGann, Peter Robinson, Paul Eggert e Hans Walter Gabler.

²⁵ F. JANNIDIS, *Tei in a Crystal Ball*, «Literary and Linguistic Computing», 24, nr. 3 (2009), pp. 253-265.

erano incentrate su questo aspetto, presentate da rispettivamente da Marjorie Burghart²⁶ ed Elena Pierazzo²⁷, entrambe sottolineando, da un punto di vista l'una più funzionale, l'altra più legata alla dimensione tecnologica, le difficoltà sia concettuali sia pratiche relative alla creazione di un'edizione digitale²⁸.

Questa digressione è funzionale per descrivere il cambiamento, percepito o reale, o molto più probabilmente una combinazione di entrambi questi aspetti, cui mi trovai di fronte a Londra nel 2008, complice anche un'assenza di tre anni e la partecipazione a diverse conferenze, prima ALLC/ACH poi DH²⁹, e avendo quindi maturato una maggiore profondità di campo. La *TEI Conference* del 2008³⁰ ai miei occhi presentava sia uno scarto quantitativo, relativo all'ampio numero di persone presenti, dovuto alla centralità del luogo e all'importanza dell'istituzione ospite, sia, soprattutto, qualitativo, riguardo la tipologia e il contenuto delle presentazioni, che, come anticipato precedentemente, andavano a inglobare e a collegare al tema principale elementi originariamente alieni sviluppatasi parallelamente nel vasto ecosistema della testualità digitale³¹, delineando così strutture sempre più complesse e stratificate. Ad esempio lo schema concettuale FRBR, nato in ambito biblioteconomico, o l'ontologia CIDOC-CRM, relativa al patrimonio culturale museale, oppure applicazioni in cui a essere di primo piano è l'aspetto linguistico-computazionale rispetto a quello di modellazione e rappresentazione del testo.

²⁶ M. BURGHART, *Digital Editions as the Myth of Sisyphus*, TEI MM: Philology in the Digital Age, Würzburg 10-16 Ottobre 2011, <http://www.zde.uni-wuerzburg.de/tei_mm_2011/abstracts/abstracts_micropapers/#c249141>

²⁷ E. PIERAZZO, *The Role of Technology in Scholarly Editing*, TEI MM: Philology in the Digital Age. Würzburg 10-16 Ottobre 2011, <http://www.zde.uni-wuerzburg.de/tei_mm_2011/abstracts/abstracts_papers/#c249169>. Slide e post di approfondimento disponibili a <<http://epierazzo.blogspot.it/2011/10/role-of-technology-in-scholarly-editing.html>>.

²⁸ In un'altra sede il suggerimento era di affrontare questa stessa questione partendo, ed eventualmente adattando, gli studi già effettuati in un settore collegato, quello delle biblioteche digitali, in cui il punto di partenza è sulle modalità di disseminazione degli oggetti digitali, e non sulla codifica. Vedi F. MESCHINI, *Mercury Ain't What He Used to Be, but Was He Ever? Or Do Electronic Scholarly Editions Have a Mercurial Attitude?*, *The Marriage of Mercury and Philology: Problems and Outcomes in Digital Philology*. E-Science Institute, Edinburgh, Scotland, 2008.

²⁹ Nello specifico l'ALLC/ACH 2004 in Svezia, a Göteborg, l'anno successivo l'ALLC/ACH 2005 a Victoria, B.C., Canada, il DH 2006 a Parigi e quello del 2008 a Oulu in Finlandia.

³⁰ <<http://www.cch.kcl.ac.uk/cocoon/tei2008/index.html>>

³¹ Un'ottima introduzione alla vastità e alla complessità di questo ecosistema è la *Metadata Map* di Jenn Reilly dove vengono classificati 105 standard diversi di metadati <<http://www.dlib.indiana.edu/~jenrile/metadatamap/>> tramite un'efficace rappresentazione visuale. Tale complessità, così come viene accennato brevemente dalla stessa Reilly, esplode esponenzialmente se si pensa a tutte le possibili combinazioni, seppure parziali, dei suddetti standard in una o più ipotetiche applicazioni, che inseguano, così come uno dei punti fondamentali del programma hilbertiano, il traguardo della completezza.

Premessa di questo ragionamento è il livello di stabilità, una soglia minima di completezza e, laddove possibile coerenza, raggiunto dalla codifica testuale, che permetta quindi di concentrarsi sugli aspetti successivi. Questo però non sta certo a significare che le riflessioni teoriche sul ruolo del mark up, estremamente fiorenti tra la fine degli anni '80 e l'inizio degli anni '90, abbiano esaurito il loro ruolo, così come le innovazioni più prettamente tecniche. Non a caso una delle relazioni plenarie del convegno di Londra era affidata a Dino Buzzetti³², le cui posizioni critiche sul mark up, e in particolare sul mark up di tipo *embedded*, in cui testo e marcatori costituiscono una sequenza ininterrotta di caratteri³³, sono ben note e tra le più citate all'interno della comunità³⁴. E, sempre al convegno di Londra, prese il via il gruppo di lavoro incentrato sulla critica genetica, che pone problemi e requisiti di rappresentazione totalmente diversi da quelli della critica testuale di tipo Lachmaniano³⁵.

Il meeting di Würzburg del 2011, nonostante la diversa, e meno centrale, localizzazione geografica, conferma i princîpi generali già delineati per quello di Londra. Degno di menzione è l'evento satellite di *Interedition*³⁶, una *COST Action* svoltasi nel quadriennio 2008-2012, incentrata sull'interoperabilità delle edizioni digitali³⁷. La presenza di eventi di questo tipo sottolinea il ruolo di aggregazione e di

³² D. BUZZETTI, *The expediency of markup*, TEI: Supporting Cultural Heritage Research, Londra 6-8 Novembre 2008, <<http://www.cch.kcl.ac.uk/cocoon/tei2008/keynote/buzzetti.html>>.

³³ Seppure direttamente mappabile in una struttura dati isomorfa, come ad esempio quella ad albero utilizzata dal modello DOM <<http://www.w3.org/TR/REC-DOM-Level-1/introduction.html>>.

³⁴ D. BUZZETTI, *Rappresentazione digitale e modello del testo*, «Il ruolo del modello nella scienza e nel sapere», Roma, 1998. Sintassi e modelli alternativi rispetto a quello di XML sono da tempo presenti nel panorama della codifica testuale, essendo in molti casi il risultato di anni di ricerca e riflessioni sul noto problema delle *overlapping hierarchies*, la sovrapposizione delle gerarchie che non può essere gestita nativamente sia in SGML sia in XML. Tra i più rilevanti il *MarkUp Languages for Complex Documents* di C. M. Sperberg-McQueen e Claus Huitfeldt <<http://mlcd.blackmesatech.com/>>, LMNL di Wendell Piez <<http://xml.coverpages.org/LMNL-Abstract.html>> e il *Freestyle Markup Language* di Denis Pondorf e Andreas Witt <<http://www.freestyle-markup.org/>>. Un modello incentrato sul problema delle varianti testuali è quello di Desmond Schmidt con il *Multi-Version Documents* <<http://multiversiondocs.blogspot.it/>>. Per maggiori informazioni e approfondimenti vedi la pagina del SIG sull'*Overlap* <<http://wiki.tei-c.org/index.php/SIG:Overlap>>.

³⁵ Gruppo di lavoro che ha prodotto il capitolo 'aggiunto' delle *Guidelines* denominato *An Encoding Model for Genetic Editions* <<http://www.tei-c.org/Activities/Council/Working/tcw19.html>>. Un bell'esempio concreto di questa estensione è il *Proust Prototype* realizzato dalla già citata Elena Pierazzo <http://research.cch.kcl.ac.uk/proust_prototype/about.html>; questo prototipo dimostra quanto le tecnologie innestate su una codifica di base, come i fogli di trasformazione XSLT, gli script Javascript o le immagini vettoriali in SVG, per una mappatura topografica tra livello iconografico e livello testuale, facciano integralmente parte dell'assetto dell'edizione, al pari della codifica, elemento sì fondativo ma non esclusivamente essenziale.

³⁶ <<http://www.interedition.eu/>>

³⁷ Uno dei principali risultati di *Interedition* è il programma *CollateX* <<http://collatex.net/>>, al tempo stesso aggiornamento e salto evolutivo dello storico *Collate* di Peter Robinson, lo strumento di collazione automatica utilizzato per la realizzazione di numerose edizioni scientifiche. Collegato alla necessità di una modellazione formale delle edizioni digitali, l'ultimo incontro ufficiale di *Interedition*

'ombrello scientifico' che la *TEI* ha ormai assunto nei confronti di aspetti sia complementari sia, in certi casi, alternativi alla principale e originaria visione scientifica veicolata, allineando così sempre di più i suoi convegni alla pluralità di prospettive che è da sempre una delle caratteristiche principali dell'informatica umanistica. Oltre al tradizionale *Bootcamp*, un campo di lavoro in cui programmatori e umanisti discutono, progettano e sviluppano fianco a fianco strumenti digitali, in questo caso incentrati sulla trascrizione e sull'annotazione³⁸, l'evento d'*Interedition* includeva un *TEI & Interoperability Think Tank*³⁹, una giornata di studio incentrata proprio sul rapporto tra il modello di codifica proposto dalla TEI e il concetto d'interoperabilità. La giornata era strutturata con una serie di relazioni a invito, seguite da una discussione aperta⁴⁰, che non hanno fatto altro che confermare come il concetto di interoperabilità, o perlomeno di un'interoperabilità che si posizioni ad un livello superiore di quella sintattica resa possibile da *XML*, sia tanto discusso e desiderato quanto sfuggente ed estremamente difficile da definire in maniera univoca.

Arriviamo così all'incontro da cui eravamo partiti, quello conclusosi da poco a Roma, che, come ormai dovrebbe essere chiaro, va inquadrato in una prospettiva diacronica cui sia appartiene in maniera inscindibile sia ne compone un ulteriore tassello, influenzato dai precedenti e che andrà a influenzare i successivi. Sempre chi scrive ha una visione privilegiata sui processi che hanno portato all'organizzazione di questo convegno a Roma, in quanto membro del consiglio dell'*Associazione per l'Informatica Umanistica e la Cultura Digitale (AIUCD)*⁴¹, a cui arrivò negli ultimi mesi del 2012 la proposta da parte del *TEI Consortium* dell'Italia come possibile sede del meeting, poco prima del precedente, in Texas, in quanto per tradizione, durante l'ultimo giorno viene annunciato dove si terrà l'incontro successivo. La proposta di riportare la *TEI Conference* in Italia dopo dodici anni incontrò immediatamente l'approvazione unanime del consiglio, facendo subito partire contemporaneamente una riflessione su quello che sarebbe dovuto essere il tema portante del convegno.

era incentrato proprio sulla definizione di un *Domain Specific Language (DSL)*

<<http://www.interedition.eu/?p=414>> relativo a questo ambito.

³⁸ <http://www.interedition.eu/wiki/index.php/Würzburg102011_CFP>

³⁹ <http://www.interedition.eu/wiki/index.php/WürzburgTEI102011_CFP>

⁴⁰ Le relazioni su invito erano di Wendell Piez

<<http://www.interedition.eu/wiki/index.php/WendellPiezTEI2011>>, Doug Reside

<<http://www.interedition.eu/wiki/index.php/DougReside2011>> e Sebastian Rahtz

<<http://www.interedition.eu/wiki/index.php/SebastianRahtzTEI2011>>. Successivamente ognuno dei partecipanti all'incontro ha potuto pubblicare le sue riflessioni sulla pagina wiki resa disponibile

<http://www.interedition.eu/wiki/index.php/WürzburgTEI102011_Participants>.

⁴¹ <<http://www.umanisticadigitale.it/>>

Essendo il consiglio un microcosmo, si spera sufficientemente rappresentativo, del panorama italiano delle pratiche dell'informatica umanistica, le posizioni riguardo la TEI sono sì variegate, ma in ogni caso sempre estremamente consapevoli dei pro e contro di questo standard, senza dimenticare che il ruolo del Presidente dell'associazione era, ed è tuttora, assolto da Dino Buzzetti, e già è stata sottolineata la sua importante funzione di voce critica e consapevole nei confronti della testualità digitale e della codifica.

The Linked TEI, il titolo scelto per l'edizione del 2013, aveva l'intenzione quindi di unire *the best of both worlds*: da un lato la tradizione della codifica testuale, sviluppata a partire dai *Poughkeepsie Principles*⁴², a sua volta erede e paladina del modello del mark up descrittivo incarnato da SGML per una rappresentazione efficace dell'informazione documento-centrica⁴³; dall'altro l'attuale e innovativo paradigma dei *LinkedData*, l'attesa e concreta diffusione su larga scala della visione del *Semantic Web* di Tim Berners-Lee⁴⁴, con alle spalle quindi la modellazione logica implicita delle basi di dati relazionali⁴⁵ e quella esplicita delle ontologie⁴⁶ e delle *Description Logics*⁴⁷ per la rappresentazione della conoscenza nel campo dell'Intelligenza Artificiale, e quindi orientato principalmente verso l'informazione data-centrica.

L'interazione tra queste due metà del cielo non costituiva certo una novità, in quanto nella sua storia, ormai più che venticinquennale, la TEI aveva già avuto modo di confrontarsi con questa tema, che veniva però ufficialmente proposto per la prima volta per un intero convegno. Tra i diversi esempi di questa interazione troviamo: la mappatura dei documenti TEI in sistemi di *Knowledge Representation* o basi di dati *Object Oriented*⁴⁸; l'utilizzo delle *Topic Maps*⁴⁹, per l'organizzazione di archivi

⁴² <<http://www.tei-c.org/Vault/ED/edp01.htm>>

⁴³ J. H. COOMBS, A. RENEAR, S. J. DE ROSE, *Markup Systems and the Future of Scholarly Text Processing*, «Communications of the ACM», 30, nr. 11, 1987: pp. 933 - 947.

⁴⁴ T. BERNERS-LEE, *Weaving the Web*, San Francisco: Harper Collins, 1999.

⁴⁵ E. F. CODD, *A Relational Model of Data for Large Shared Data Banks*, «Communications of the ACM» 13, nr. 6, 1970, pp. 377-387; P. P. CHEN, *The Entity-Relationship Model - toward a Unified View of Data*, «ACM Transactions on Database Systems», 1, nr. 1, 1976: pp. 9-36.

⁴⁶ T. R. GRUBER, *Toward principles for the design of ontologies used for knowledge sharing?*, «International Journal of Human-Computer Studies», 43, nr. 5-6, 1995, pp. 907-928.

⁴⁷ F. BAADER ET AL. (a cura di), *The Description Logic Handbook: Theory, Implementation and Applications*, Cambridge University Press, Cambridge, 2004.

⁴⁸ N. IDE, T. MCGRAW, C. WELTY, *Representing TEI documents in the CLASSIC knowledge representation system*, «Proceedings of the Tenth workshop of the Text Encoding Initiative», 1997; G. F. SIMONS, *Using Architectural Forms to Map TEI Data into an Object-Oriented Database*, «Computers and the Humanities», 33, 1999, pp. 85-101.

⁴⁹ Un modello alternativo al *Semantic Web* per la rappresentazione della conoscenza. Grazie a una maggiore vicinanza al linguaggio naturale, soprattutto se paragonate ai formalismi logici di *RDF/RDFS/OWL*, seppure a costo di una maggiore verbosità, le *Topic Maps* hanno conosciuto una certa diffusione verso la metà degli anni 2000, in particolare nei paesi dell'area scandinava, e sono

elettronici testuali, per la modellazione dei contenuti e delle tematiche veicolate da un testo letterario o addirittura per la classificazione degli strumenti utilizzati per la pubblicazione dinamica dei testi TEI, in base sia alle funzionalità offerte sia alle tecnologie sottostanti⁵⁰; la rappresentazione dei linguaggi di marcatura tramite linguaggi logici o più recentemente il modello del *Semantic Web*⁵¹; il rapporto tra le entità definite da *FRBR* e la natura di un documento XML⁵² e il collegamento tra le descrizioni di oggetti fisici nei testi letterari e la loro controparte reale o eventuali rappresentazioni digitali⁵³.

Sempre riguardo questa interazione vorrei citare due rilevanti contributi scientifici, in cui, come descritto dagli stessi titoli (*MarkUp Reconsidered* e *From Data Representation to Data Model*⁵⁴), il mark up viene ripensato e il suo status di rappresentazione di dati è confrontato con quello di un modello di dati, analizzando quindi la diversa natura dei linguaggi di marcatura e quella dei database relazionali. Da un lato abbiamo delle strutture dati estremamente flessibili⁵⁵, mentre dall'altro un modello dati formalmente definito, laddove la differenza sostanziale è la presenza di un insieme di operazioni che operano su una determinata struttura; nel caso dei database relazionali si tratta di quelle stabilite a partire dalle operazioni insiemistiche e logiche e definite tramite l'algebra relazionale, o il computazionalmente equivalente calcolo relazionale⁵⁶. Oltre a equiparare la potenzialità espressiva di entrambe le

ancora utilizzate in diverse contesti. Vedi F. MESCHINI, *Le mappe topiche: come imparai a non preoccuparmi e ad amare i metadati*, «Bollettino AIB», 45, nr. 1, 2005, pp. 59-72.

⁵⁰ C. TUOHY, *Topic Maps and TEI - Using Topic Maps as a Tool for Presenting TEI Documents*, TEI Day in Kyoto, 2006; J. WALSH, *TM4DH - Topic Maps for Digital Humanities: Examples and Tool Kits*, ACH/ALLC 2005 Joint International Conference, 2005; F. MESCHINI, *Tms - Tei Management Systems*, E. PIPP (a cura di), «Zugang Zum Fachwissen - Odok '05», Graz-Feldkirch, Wolfgang Neugebauer Verlag, 2007, pp. 67-81.

⁵¹ D. DUBIN, A. RENEAR, C. M. SPERBERG-McQUEEN, C. HUITFELDT, *A Logic Programming Environment for Document Semantics and Inference*, «Literary and Linguistic Computing», 18, nr. 1, 2003, pp. 39-47; A. DI IORIO, S. PERONI, F. VITALI, *A Semantic Web approach to everyday overlapping markup*, «Journal of the American Society for Information Science and Technology», 62, nr. 9, pp. 1696-1716.

⁵² A. RENEAR, C. PHILLIPPE, P. LAWTON, D. DUBIN, *An XML document corresponds to which FRBR Group 1 entity?*, B. T. USDIN (a cura di), «Proceedings of Extreme Markup Languages 2003», Montreal, Quebec, Agosto 2003.

⁵³ C. E. ORE, Ø. EIDE, *Tei and Cultural Heritage Ontologies: Exchange of Information?*, «Literary and Linguistic Computing», 24, nr. 2, 2009.

⁵⁴ D. RAYMOND, F. TOMPA, D. WOOD, *Markup Reconsidered*, «First International Workshop on Principles of Document Processing», Washington, DC, 1992; D. RAYMOND, F. TOMPA, D. WOOD, *From data representation to data model: Meta-semantic issues in the evolution of SGML*, «Computer Standards & Interfaces», 18, 1996, pp. 25-36.

⁵⁵ La sintassi di XML nonostante sia isomorfa, come scritto precedentemente, a un albero gerarchico, è in grado di simulare qualsiasi struttura dati esistente: array, liste collegate, matrici, grafi. Questo grazie al meccanismo degli attributi di *id*, in grado di identificare in maniera univoca un elemento, e dei puntatori a tali identificatori.

⁵⁶ Nei tipi di dati astratti, come pile, code, grafi, o anche le tabelle del modello relazionale, non è necessario sapere quale sia la struttura dati sottostante o gli algoritmi utilizzati per implementare una

tecnologie⁵⁷, e tralasciando quindi quelle che sono le ricadute nelle implementazioni concrete a livello di convenienza ed efficienza, in questi articoli viene sottolineata l'importanza dell'*out-of-line* mark up, conosciuto anche come *stand-off* markup, un approccio data-centrico all'informazione documento-centrica, in cui testo e meta-informazioni sono separate, e tramite un meccanismo di indirizzamento è possibile applicare a uno stesso strato testuale più livelli di codifica. Va da sé come questo approccio, e il concetto stesso sottostante che al tempo stesso estrapola ed esplicita il rapporto tra testo e codifica, si sposi efficacemente al modello delle basi di dati⁵⁸ e a quello della formalizzazione logica⁵⁹. Lo *stand-off* markup ha già acquisito una certa importanza, così come di sicuro succederà anche per le riflessioni teoriche, e i relativi impieghi concreti, sui modelli di dati per le edizioni elettroniche, ed è giusto quindi ricordare l'importante contributo di questi due articoli.

A questo punto dovrebbe essere chiaro come la conferenza di Roma da cui siamo partiti è stata sia lo spunto per una descrizione del percorso effettuato dal *TEI Consortium* sia per delle riflessioni sul rapporto tra codifica ed edizione digitale. Sarebbe però ingeneroso concludere senza dedicare perlomeno qualche parola a *The Linked TEI*, partendo innanzitutto dai due *keynotes*, Allen Renear e Marie-Luce Demonet⁶⁰ il primo sul già citato rapporto tra codifica del testo ed ontologie, nell'ottica della natura multi livello degli oggetti digitali, la seconda su una rete di

determinata operazione, come l'inserimento o l'eliminazione di un elemento, quanto l'aderenza a dei requisiti generali in termini di parametri in ingresso e risultati in uscita. In certo senso anche l'edizione digitale è paragonabile a un tipo di dato astratto, che rende disponibili all'utente delle funzioni implementate a partire da un testo la cui struttura è stata esplicitata tramite il mark up, e sulla cui base vengono implementate delle istruzioni formali. Vale la pena ricordare come anche un foglio di stile XSLT, una delle tecnologie più utilizzate per la pubblicazione di testi TEI, sia basato su di un linguaggio *Turing-complete*. La riflessione teorica sulle edizioni digitali dovrebbe quindi orientarsi verso la definizione di un insieme minimo di funzioni che definiscano cosa sia o no un'edizione di questo tipo e soprattutto a quale livello di granularità questa definizione sia possibile, data la vasta eterogeneità dei casi possibili.

⁵⁷ Non va inoltre dimenticato che linguaggi di marcatura e linguaggi di programmazione hanno lo stesso livello espressivo, appartenendo entrambi alle grammatiche libere dal contesto, dette anche di tipo 2, della gerarchia di Chomsky. Altrimenti il caso di XSLT, un linguaggio di programmazione basato proprio sulla sintassi XML, non sarebbe possibile. L'ipotetica, e spesso percepita, maggiore complessità dei linguaggi di programmazione, soprattutto nella realizzazione delle edizioni digitali, è probabilmente dovuta al fatto questi linguaggi devono gestire l'informazione contenuta nel mark up e di conseguenza a questa complessità di partenza va aggiunta quella del linguaggio stesso e delle operazioni che deve compiere.

⁵⁸ Y. LIU, J. SMITH, *A Relational Database Model for Text Encoding*, B. Nelson (a cura di), «Reassembling the Disassembled Book», CHWP, 2008, <http://projects.chass.utoronto.ca/chwp/CHC2007/Liu_Smith/Liu_Smith.htm>

⁵⁹ N. GAVIN, *Core Range Algebra: Toward a Formal Model of Markup*, «Proceedings of Extreme Markup Languages 2002», <<http://conferences.idealliance.org/extreme/html/2002/Nicol01/EML2002Nicol01.html>>

⁶⁰ Il libro degli abstract è liberamente scaricabile in diversi formati: PDF, Epub e TEI/XML <<http://digilab2.let.uniroma1.it/teiconf2013/abstracts/>>.

biblioteche digitali distribuite, le *Bibliothèques Virtuelles Humanistes (BVH)*⁶¹, caratterizzata da un'architettura in grado di permettere diversi tipi di analisi dei testi (letteraria, filologica e linguistica) e sulla creazione di una navigazione a faccette basata proprio sui metadati semantici. Appare evidente l'equilibrio simmetrico tra le due relazioni: da un lato una riflessione teorica supportata da esempi concreti, dall'altro la descrizione di un progetto concreto in cui le scelte architettoniche sono basate su solide basi teoriche, e, a rifletterci, questo dinamico rapporto dialettico sembra essere un po' la costante dell'Informatica Umanistica.

Tra il *keynote* di apertura di Renear e quello di chiusura della Demonet, i quasi duecento partecipanti hanno potuto assistere a più di cinquanta tra relazioni singole, sessioni e poster in cui la stessa codifica *TEI* o i suoi impieghi sono stati sia presentati sia analizzati sotto diversi punti di vista: le già citate edizioni critiche; l'utilizzo nei progetti di digitalizzazione bibliotecaria insieme ad altri standard di metadati; modellazione semantica; analisi linguistiche e statistiche; utilizzi nella critica letteraria; strumenti per la creazione, l'analisi e la pubblicazione dei testi digitali; l'insegnamento nei curricula accademici; la conservazione degli oggetti digitali e altro ancora. Il tutto senza dimenticare le riflessioni filosofiche sul concetto di opera, documento e testo e il ruolo della codifica digitale in questo già complicato rapporto. Naturalmente non si ha la pretesa di voler fare un elenco completo, che in ogni caso non renderebbe giustizia all'evento.

Vanno poi aggiunti i workshop e i tutorial⁶² svoltisi nei due giorni precedenti il convegno vero e proprio, andando così a riempire una settimana intera. Anche in questo caso i temi erano sia direttamente collegati, come *EpiDoc*⁶³ e *Clarin*⁶⁴, sia di più ampio respiro, sugli strumenti di annotazione semantica⁶⁵ e sull'iniziativa *AthenaPlus*⁶⁶, per la conversione e l'aggregazione di metadati eterogenei, sia per tipologia sia per provenienza, nel contesto di *Europeana*⁶⁷. In particolare questi due ultimi eventi erano organizzati rispettivamente dal *CINECA* e dall'*ICCU*, portando quindi un'importante testimonianza delle attività di ricerca e dei progetti in corso di queste due realtà nazionali in un contesto internazionale.

⁶¹ <<http://www.bvh.univ-tours.fr/>>

⁶² <<http://digilab2.let.uniroma1.it/teiconf2013/program/workshop/>>;
<<http://digilab2.let.uniroma1.it/teiconf2013/program/tutorials/>>

⁶³ <<http://sourceforge.net/p/epidoc/wiki/Home/>>

⁶⁴ <<http://www.clarin.eu/>>

⁶⁵ <http://demo.cineca.it/index_en.html>

⁶⁶ <<http://www.athenaplus.eu/>>

⁶⁷ <<http://www.europeana.eu/>>

Concludendo, a poco più di venticinque anni dalla sua nascita la *TEI* è un tassello fondamentale nel panorama delle *Digital Humanities*, sia a livello qualitativo sia quantitativo; come detto da un relatore durante una delle sessioni, è ormai impossibile prescindere in qualsiasi progetto di ricerca dotato di una forte componente testuale, non fosse altro per evidenziare in che modo sia possibile superare le limitazioni dello standard, spesso aggirate ad un livello superiore, quello dell'applicazione, e non quello della codifica.

L'invito per tutta la comunità è quindi, in questa fase in cui il traguardo sembra essere stato tagliato da tempo e la quasi totalità delle resistenze vinte, di non perdere di vista quello spirito pionieristico e di sperimentazione che ha caratterizzato per così tanti anni la *Text Encoding Initiative*.