

***Uno standard per la digitalizzazione dei documenti medievali con XML. Cronaca di un Workshop internazionale: Monaco 5-6 aprile 2004\****

**Edizioni elettroniche.**

Internet sembra proprio presentarsi come il medium che rende facilmente accessibile alla scienza la documentazione. Ciò non dipende soltanto dall'ottima infrastruttura tecnica delle istituzioni accademiche, che agevola particolarmente l'accesso a Internet, ma anche dall'interesse peculiare che nei suoi confronti manifestano discipline caratterizzate da forti interconnessioni su scala internazionale: come, appunto, la Diplomatica. Internet per la sua disponibilità su scala mondiale è un appropriato mezzo di comunicazione per questa disciplina; inoltre i precoci sforzi, compiuti per altro proprio dai diplomatici, mirati a un corretto utilizzo delle tecnologie informatiche nella ricerca mostrano che il computer, proprio nelle discipline documentarie – così ricche di tradizione – gioca un ruolo specifico non soltanto come mezzo di comunicazione, ma come vero e proprio strumento di lavoro scientifico. Sul piano teorico, di applicazioni informatiche in Diplomatica si sono occupati in modo particolare Karsten Uhde e Michele Ansani. Uhde mira a cogliere il rapporto di collaborazione fra archivi e scienza<sup>1</sup>. Ansani riflette sull'alto grado di formalizzazione dei testi diplomatici, ideali per la sperimentazione di modelli di codifica, suggerendo di partire dalla sistematizzazione termi-

\* Si presenta la traduzione italiana (a cura della redazione di «Scrineum») del contributo destinato a «Archiv für Diplomatik» 50 (2004). Si ringraziano gli editori per il consenso prestato alla duplice pubblicazione.

<sup>1</sup> K. UHDE, *Urkunden im Internet. Neue Präsentationsformen alter Archivalien*, in «Archiv für Diplomatik», 45 (1999) pp. 441-464 [trad. it. (a cura di Antonella Ghignoli), *Documenti in Internet. Forme di presentazione nuove d'antichi documenti d'archivio*, in «Scrineum. Saggi e materiali on line di scienze del documento e del libro medieval», Biblioteca: <http://dabc.unipv.it/scrineum/biblioteca/kuhde.htm>].

nologica e concettuale realizzata con il *Vocabulaire International de la Diplomatique*<sup>2</sup>. Le prime prese di posizione per l'impiego dell'informatica nel lavoro coi documenti risalgono a Heide Dienst, che ha saggiato le possibilità di analisi dei testi con l'ausilio dell'elaborazione elettronica dei dati, proprio come hanno fatto Annegret Fiebig e, in molti lavori, Michael Gervers e Michael Margolin e i loro collaboratori al DEEDS-Project<sup>3</sup>.

Dalla pratica di concreti progetti per la digitalizzazione di documenti traggono invece origine le riflessioni di Patrick Sahle, dei collaboratori del *Thesaurus Diplomaticus* belga, di Andrea Rapp, di Paul Bertrand, di Benoît Tock, di Stuart Jenks o di Martin Glessgen<sup>4</sup>. Una ricognizione medi-

<sup>2</sup> M. ANSANI, *Diplomatica (e diplomatisti) nell'arena digitale*, in «Archivio Storico Italiano», 158 (2000), pp. 349-379 (online: [http://www.storia.unifi.it/asi-dspt/ASI/Testi\\_online/ansani.htm](http://www.storia.unifi.it/asi-dspt/ASI/Testi_online/ansani.htm)). Una diversa, più ampia versione in «Scrineum. Saggi e materiali online di scienze del documento e del libro medievali», Biblioteca: <http://dobb.unipv.it/scrineum/biblioteca/ansani.htm>.

<sup>3</sup> Cfr. H. DIENST, *Traditionsbücher. Editionsprobleme, Inhaltsanalysen und EDV-Einsatz am Beispiel der Klosterneuburger Traditionen*, in *Datenetze für die Historischen Wissenschaften?* (1987), pp. 51-62. M. GERVERS, *The DEEDS Database of Mediaeval Charters: Design and Coding for the RDBMS Oracle 5*, in «History & Computing», 2/1 (1990), pp. 1-11; ID., *The DEEDS Project and the Development of a Computerized Methodology for Dating Undated English Private Charters of the Twelfth and Thirteenth Centuries* e R. FIALLOS, *An Overview of the Process of Dating Undated Medieval Charters. Latest Results and Future Developments*, in *Dating Undated Medieval Charters*, ed. M. GERVERS (2000), pp. 13-50; A. FIEBIG, *Urkundentext. Computergestützte Auswertung deutschsprachiger Urkunden der Kuenringer auf Basis der eXtensible Markup Language (XML)*, (Schriften zur südwestdeutschen Landeskunde Bd. 33, 2000).

<sup>4</sup> G. DECLERCQ, PH. DEMONTY, K. NAESSENS und G. TRIFIN, *L'informatisation de la 'Table chronologique' d'A. Wanters. Méthodologie du nouveau répertoire des documents diplomatiques belges antérieurs à 1200*, in «Bulletin de la Commission Royale d'Histoire», 153, (1987), pp. 223-302; A. RAPP, *Computergestützte Verfahren zur Erfassung, Katalogisierung, Bearbeitung und Edition mittelfränkischer Urkunden des 13. und 14. Jahrhunderts*, in *Maschinelle Verarbeitung altdentscher Texte*, hrsg. von ST. MOSER, P. STAHL, W. WEGSTEIN und N. R. WOLF (Tübingen 2001), pp. 247-261; ID., *Die elektronische Edition, Erschließung und Vernetzung des Trierer Korpus mittelfränkischer Urkunden des 14. Jahrhunderts*, in «Jahrbuch für Computerphilologie», 2, (2000), pp. 147-161; ID., *Die elektronische Edition, Erschließung und Vernetzung historischer Quellen und Grundlagenwerke. Das Trierer Korpus mittelfränkischer Urkunden*

tata di essi è stata effettuata nel 2003 da Patrick Sahle e Georg Vogeler<sup>5</sup>; progetti che vengono progressivamente documentati nella *Virtual Library Geschichtliche Hilfswissenschaften*<sup>6</sup>.

Nell'ambito dell'elaborazione informatica di testi destinati alla rete, XML (*eXtensible Markup Language*) si impone come standard. XML è uno dei linguaggi di marcatura dei testi. Di questi, il più noto è forse HTML, l'*Hypertext Markup Language*, nel quale sono codificate le pagine di Internet. Nei linguaggi di marcatura vengono fissate delle regole, da applicare nel trattamento elettronico del testo mediante descrizioni in lin-

*des 14. Jahrhunderts*, in *TUSTEP educa*, hrsg. von N. CASTRILLO BENITO und P. STAHL (2001), pp. 99-123; S. JENKS, *KISS (Keep It Simple, Stupid). Elektronische Quelleneditionen mit einfachsten Mitteln*, in *Quellen und Quellenedition im neuen Medienzeitalter*, hrsg. v. S. JENKS und F. SCHMIEDER, «Fundus; Forum für Geschichte und Ihre Quellen», 2 (2002), pp. 33-43 (online: [http://webdoc.sub.gwdg.de/edoc/p/fundus/html/heft\\_2html](http://webdoc.sub.gwdg.de/edoc/p/fundus/html/heft_2html)); P. SAHLE u. TH. SCHAASAN, *Das Hansische Urkundenbuch in der digitalen Welt. Vom Druckwerk zum offenen Quellenrepertorium*, in «Hansische Geschichtsblätter», 118, (2000), pp. 133-155 (v. anche ulteriore documentazione all'indirizzo: <http://www.hki.uni-koeln.de/projekte/hub/inhalt.htm>); P. SAHLE, *Vom editorischen Fachwissen zur digitalen Edition. Der Editionsprozess zwischen Quellenbeschreibung und Benutzeroberfläche*, in «Fundus; Forum für Geschichte und Ihre Quellen», 2, (2000), pp. 75-102 (online: [http://webdoc.sub.gwdg.de/edoc/p/fundus/html/heft\\_2html](http://webdoc.sub.gwdg.de/edoc/p/fundus/html/heft_2html)); M.-D. GLESSGEN, *Editorische, lexikologische und graphematische Erschließung altfranzösischer Urkundentexte mit Hilfe von TUSTEP. Stand der Arbeiten*, in *Drittes Trierer Urkundensprachekolloquium (20.-22. Juni 2001)*, hrsg. v. K. GÄRTNER u. G. HOLTUS, in stampa. Cfr. anche i cinque contributi sull'analisi di documenti mediante banche-dati nel volume *Resourcing Sources*, ed. K. S. B. Keats-Rohan, (2002): *La base de données 'Cartulaires' de la section de diplomatique de l'Institut de recherche et d'histoire des textes (Orléans) et l'entreprise du Répertoire des Cartulaires Français* (P. BERTRAND, pp. 145-152); *La base de données des chartes originales antérieures à 1121 conservées en France* (B.-M. TOCK, pp. 153-163); *Identifying irregularities and establishing chronology in medieval charters* (M. GERVERS, pp. 164-178); *Dating charters using textual evidence* (A. SPENCER, pp. 179-194); *Changing legal terminology in dated private documents in England in the twelfth and thirteenth centuries: a case study - quitclaims* (Á. JUHASZ-ORMSBY, p. 195 ss.).

<sup>5</sup> P. SAHLE, G. VOGELER, *Urkundenforschung und Urkundenedition im digitalen Zeitalter*, in *Geschichte und Neue Medien. Kongress .hist2003*, Berlin April 2003, in stampa.

<sup>6</sup> Cfr. <http://www.vl-ghw.uni-muenchen.de/diplomatik.html>.

guaggio naturale, tali da poter essere riconosciute come metatesto dal computer: con il loro ausilio la macchina può assegnare a quel medesimo testo una struttura. In questo modo possono essere recuperate informazioni sul medesimo testo, che saranno orientate totalmente sulla struttura logica di esso e non verranno riprodotte grazie al sistema tipografico di una pagina di libro. La veste grafica che queste informazioni possono assumere è liberamente plasmabile e può in questo modo rappresentare diverse parti della struttura del testo. Questa struttura viene fissata in una cosiddetta *Document Type Definition* (DTD).

L'impiego dei linguaggi di marcatura è di fondamentale importanza per la Diplomatica, non soltanto a causa della sua crescente diffusione in Internet. Diversamente dalla rigidità, dal punto di vista tecnico, di una base di dati, i linguaggi di marcatura prendono le mosse da un testo continuo, le cui sequenze sono parte di una struttura gerarchica: precisamente come una classica edizione di documenti è un 'testo', che veicola una serie di documenti di natura ripetitiva in sé ma mai nell'identico modo, e precisamente come i documenti stessi sono testi, il cui 'testo' è formato da una serie mobile di parti formulari.

Il Workshop che qui presenteremo ha raccolto dunque quei ricercatori di tutto il mondo che hanno sperimentato il trattamento digitale di documenti mediante XML. La maggior parte di essi ha presentato i propri progetti con un'apposita relazione, gli altri hanno partecipato alla discussione: all'incontro era dunque rappresentata la larghissima maggioranza dei progetti che si conoscono in quest'ambito.

### **Relazioni ufficiali.**

La seduta è stata aperta dal discorso di saluto del prof. Walter Koch, che ha sottolineato l'importanza dell'informatica applicata alle discipline storiche, come nuova scienza ausiliaria. Ha introdotto la parte ufficiale del Workshop la relazione del Dr. Karsten Uhde (Marburgo) con «Considerazioni preliminari per una DTD per la descrizione di documenti secondo il punto di vista degli archivisti». Uhde fa presente come la stragrande maggioranza dei documenti degli archivi non sia stata ancora co-

nosciuta e utilizzata dalla ricerca. L'archivista prende le mosse dal documento come oggetto fisico e nel descriverlo produce informazioni che possono rappresentare una base di costruzione per un'edizione critica. Uhde ha studiato i due più importanti programmi di marcatura fondati su XML, ovvero MIDOSA e VERA, per saggiarne l'idoneità a marcare i documenti. È giunto alla conclusione che il software VERA, finora soltanto di rado applicato, disegna al meglio il trattamento dei documenti e consente di accedervi in modo molto approfondito. Ha proposto che nell'ambito delle strategie di inventariazione archivistica debba venir concordato uno standard comune, nel quale possano esser contenute le strutture XML per gli inventari, e che tale standard debba esser adatto a trattare la grande quantità dei documenti tardo medievali e dell'età moderna.

Il prof. Michele Ansani (Pavia) ha presentato il poderoso Codice Diplomatico digitale della Lombardia medievale<sup>7</sup>, per il quale è stata sviluppata un'apposita propria DTD. Ha poi sottolineato la necessità di partire dal già citato *Vocabulaire international de la Diplomatie*, qualora si volesse dar vita a un modello formale condiviso per l'edizione digitale di *corpora* documentari.

La relazione del dr. Georg Vogeler (Monaco) ha descritto l'utilità di una elaborazione elettronica dei *Traditionsbücher* bavaresi, finora solo parzialmente pubblicati a stampa. Accanto a cattive edizioni del XVIII secolo alcuni di questi libri risultano ancora inediti. Pertanto egli ha espresso l'esigenza che lo standard da definire sia tale da permettere tanto l'elaborazione digitale di edizioni a stampa quanto la realizzazione di edizioni digitali *ex novo*. Le nuove edizioni elettroniche, se concepite come 'edizioni ibride', dovrebbero potersi inserire con una loro versione a stampa all'interno del *corpus* di fonti edite tradizionalmente, esattamente come le stesse edizioni a stampa, una volta elaborate in una forma digitale, possono acquisire le funzionalità di una edizione digitale.

<sup>7</sup> Cfr. <http://cdlm.unipv.it>.

Il prof. Michael Gervers insieme a Michael Margolin (Toronto) ha presentato non soltanto i positivi risultati metodologici dell'elaborazione elettronica per la datazione dei documenti nell'ambito del DEEDS-Project<sup>8</sup>, ma anche uno schema per la codifica XML dei documenti, che mediante un "Virtual Markup" (definito anche *stand-off markup*) può descrivere diversi livelli testuali che altrimenti, in uno schema XML strettamente gerarchizzato, non sarebbe possibile rappresentare, perché si sovrapporrebbero gli uni agli altri.

Il contributo di Andreas Kuczera (Magonza) ha esposto le ragioni per le quali i *Regesta Imperii*<sup>9</sup> hanno deciso di ricorrere a XML sia per il trattamento e la conversione digitale dei volumi già pubblicati a stampa sia di quelli ancora da pubblicare. Egli ha sottolineato l'esigenza per i *Regesta Imperii* di seguire degli standard estensibili, e ha formulato la necessità di implementare del software che possa rendere più agevole la marcatura dei testi ai collaboratori dei *Regesta Imperii*, inesperti dal punto di vista tecnico-informatico.

I progetti norvegesi, presentati da Christian-Emil Ore (Oslo), erano stati avviati già nel 1988, vantando perciò una notevole esperienza nell'ambito dell'elaborazione elettronica dei dati. Ore ha perciò potuto dar conto di diversi cambiamenti tecnologici avvenuti, che partendo dal *Diplomatarium Norvegicum*<sup>10</sup> hanno sviluppato un eccellente strumento, utile in particolare per la ricerca linguistica e lessicografica. La raccolta di dati elettronici è parte del progetto MeNoTA (*Medieval Nordic Text Archive*)<sup>11</sup>, avviato nel 2001 in Danimarca, Islanda, Norvegia e Svezia; esso consiste in materiale retrodigitalizzato del *Diplomatarium Norvegicum*<sup>12</sup>, dei

<sup>8</sup> Cfr. <http://www.utoronto.ca/deeds/research/research.html>.

<sup>9</sup> L'offerta on line dei *Regesta Imperii* si trova al seguente indirizzo: [http://regesta-imperii.uni-giessen.de/index\\_ri.htm](http://regesta-imperii.uni-giessen.de/index_ri.htm).

<sup>10</sup> Cfr. [http://www.dokpro.uio.no/dipl\\_norv/diplom\\_field\\_eng.html](http://www.dokpro.uio.no/dipl_norv/diplom_field_eng.html).

<sup>11</sup> Cfr. <http://www.menota.org>.

<sup>12</sup> *Diplomatarium Norvegicum. Oldbreve til Kundskab om Norges indre og ydre Forbolde, Sprog, Slaegter, Saeder, Longvining og Rettergang i Middelalderen*, cur. CHR. C. A. LANGE u. C. R. UNGER et al. (1847-).

*Regesta Norvegica*<sup>13</sup>, dei ‘nomi di corte’ norvegesi<sup>14</sup> e di documenti norvegesi redatti esclusivamente in forma digitale. Sono salvati in file compatibili con lo standard della *Text Encoding Initiative* (TEI) e sulla base dei quali possono essere impostate diverse analisi e presentazioni.

Il progetto di digitalizzazione del *Corpus der altdutschen Originalurkunden* di Friedrich Wilhelm<sup>15</sup> (CAO), esposto dal dr. Andrea Rapp (Gottinga) e da Andreas Gniffke (Treviri), presenta un’ulteriore e particolare esigenza. La versione digitale dovrebbe permettere di realizzare una vera e propria riproduzione dei testi. Il progetto persegue però, oltre a questo, anche lo scopo di elaborare materiale per i suoi molteplici contesti di impiego. I relatori hanno esposto in relazione a ciò i particolari problemi legati agli indici, contenenti informazioni che il lettore facilmente deduce dal contenuto, ma che in una struttura formalizzata non si riescono a rappresentare. Nonostante questi ostacoli i relatori hanno comunque con chiarezza sostenuto che l’edizione digitale del CAO migliora in maniera vistosa l’accesso al patrimonio di informazioni grazie al suo sistema differenziato di indici dei luoghi e delle persone.

Gautier Poupeau (Parigi) ha illustrato i progetti di edizioni documentarie digitali avviate presso l’*École des Chartes*<sup>16</sup>, che si avvalgono totalmente dello standard TEI. Egli ha mostrato – con impressionante efficacia – quanto grande sia il lavoro da fare specialmente con gli elementi dello standard per la trascrizione di testi manoscritti. Anche la creazione di un indice sarebbe possibile senza problemi e potrebbe venire trasportata in rete. Egli dunque bene interpreta la posizione di chi sostiene che

<sup>13</sup> *Regesta Norvegica*, cur. SVERRE BAGGE et al. (1898-).

<sup>14</sup> *Norske gaardnavne. Oplysninger samlede til brug ved matrikelens revision*, cur. O. RYGH et al. (1897-1936).

<sup>15</sup> *Corpus der altdutschen Originalurkunden bis zum Jahr 1300*, bearb. v. F. WILHELM u.a. (1932 - 1980).

La descrizione online del progetto si trova all’indirizzo: <http://www.corpus.uni-trier.de>.

<sup>16</sup> Cfr. <http://elec.enc.sorbonne.fr>.

sia possibile produrre una completa edizione documentaria ricorrendo agli elementi del TEI, se ‘generosamente’ e opportunamente interpretati.

Il contributo di Patrick Sahle e di Bernhard Assman (Colonia) ha trattato innanzitutto della strategia per digitalizzare le edizioni documentarie a stampa. Si è descritta la procedura che trasforma la rappresentazione di una pagina di libro in un testo logicamente accessibile, vale a dire da una immagine a un *fulltext* digitale marcato in profondità. Poiché i testi stampati sono strutturati fra loro in modi molto differenti, Sahle conclude con la proposta che debba essere stabilito e formalizzato uno standard comune e abbastanza ‘morbido’, tale da poter ricomprendere le varietà che caratterizzano le edizioni documentarie a stampa. Successive marcature più profonde, secondo regole più stringenti, non sarebbero con ciò necessariamente escluse. Bernhard Assman ha presentato, da ultimo, un sistema compiuto di termini canonici per le edizioni digitali di documenti e le loro varie forme di rappresentazione, che potrebbe venire realizzato sul server con l’ausilio dell’*Extensible Stylesheet Language for Transformation* (XSLT).

Durante la discussione che ha fatto seguito alle relazioni, il dr. Mark Mersiowsky (Monaco) ha brevemente riferito intorno al progetto dei *Monumenta Germaniae Historica* di sostituire le attuali versioni elettroniche, gli *e-MGH* (pubblicati presso Brépols), con forme più aggiornate sotto il nome di *d-MGH*; la prima ad esser rielaborata dovrebbe esser proprio la serie dei *Diplomata*.

### **Discussione non aperta al pubblico.**

Benché i partecipanti si siano denominati “Working Group” soltanto alla fine della discussione, il clima di questa era già sin dall’inizio così animato da spirito di cooperazione che vorrei parlare di “Working Group” subito, qui di seguito.

I contributi hanno presentato possibilità fra loro molto differenti per l’accesso ai documenti: perciò la discussione si è occupata innanzitutto di ricondurre queste diverse prospettive a una posizione concorde. Le prospettive principali che si sono profilate sono: quella innanzitutto degli ar-

chivisti, che mira al ‘documento’ come a un oggetto fisico, marcato come parte di un inventario. Il testo del documento, invece, occupa per gli editori e gli utenti di edizioni a stampa il primo piano: esso costituisce la seconda prospettiva, per la quale l’oggetto fisico è soltanto uno dei tanti portatori del testo insieme, per esempio, a tutta una tradizione in copia. La terza prospettiva prende le mosse dal contenuto del documento: la dimensione fisica e testuale serve solo come argomentazione per stabilire la veridicità del contenuto. Le ultime due prospettive sono proprie dell’editore. Una unità fisica ancora diversa contemplanò poi i bibliotecari, nel momento in cui vanno a digitalizzare i documenti come componenti testuali di un volume di edizioni a stampa.

Queste divergenze vengono ad assumere un particolare peso una volta inserite nell’ambiente tecnologico comune che si è scelto. Come abbiamo ricordato, XML è un linguaggio di marcatura strettamente gerarchico. Le divergenze d’impostazione cominciano però già a partire dall’‘elemento radice’, vale a dire dal punto in cui si snoda la gerarchia. Come Karsten Uhde ha fatto notare, se per esempio il ‘fondo archivistico’ per l’archivista è un’unità logica che comprende diversi documenti, in un’edizione esso invece compare come parte dei dati sulla tradizione di un singolo documento. Il Working Group ha espresso l’opinione che sia possibile far dialogare queste divergenti prospettive mediante due strategie. Ci si è potuti in primo luogo trovare d’accordo sul fatto che gli standard TEI costituiscano una ragionevole cornice per tutte le impostazioni, che potranno adottare strutture proprie elaborate di volta in volta. In secondo luogo, il Working Group ha ritenuto che ci siano ampie zone di intersezione fra le diverse impostazioni, che dovrebbero poter essere rappresentate mediante una semantica comune.

I partecipanti hanno ritenuto gli standard proposti dalla TEI<sup>17</sup> quelli che meglio si adattano a elaborare i documenti con XML. La DTD della TEI esiste ora nella versione 4 (TEI P4). Una nuova versione però è già

<sup>17</sup> Cfr. <http://www.tei-c.org>.

in fase di elaborazione<sup>18</sup>. Essa tenterà di andare a coprire ulteriori ambiti in relazione coi testi manoscritti, con testi multimediali, coi prodotti delle arti grafiche, così come per le procedure di marcatura dei testi in unità elettroniche separate (*standoff annotation*) e per l'introduzione di XML *namespaces*, vale a dire di strutture testuali parallele. Per quanto concerne il lavoro coi documenti, di grande importanza è soprattutto il primo ambito di applicazione. Le proposte della MASTER *initiative*<sup>19</sup>, che si occupa della descrizione elettronica dei manoscritti, dovrebbe così divenire una parte della TEI-DTD. Mentre non sembrano così centrali per il lavoro coi documenti le implementazioni nei campi dei multimedia e delle arti grafiche, gli ambiti della *standoff annotation* invece e degli XML *namespaces* vengono incontro a un ulteriore problema posto dal trattamento digitale dei testi documentari. Come l'impostazione nel DEEDS-Project ha dimostrato, la *standoff annotation* può rappresentare una soluzione per le strutture che potrebbero sovrapporsi. Durante la discussione, inoltre, anche gli XML *namespaces* sono apparsi come una possibilità per salvare strutture di testo parallele. Sono un buon esempio di tipo di strutture di testo parallele le *litterae elongatae* e il protocollo di un documento: strutture, vale a dire, che sono l'una grafica, l'altra linguistica e che possono, ma non devono necessariamente, riferirsi l'una all'altra. Che TEI-P5 abbia molto riguardo per gli interessi specifici del lavoro storico, lo dimostra anche la recente discussione tenuta dalla *mailinglist* della TEI intorno al problema dell'integrazione delle informazioni prosopografiche in TEI-L<sup>20</sup>.

In particolare è lo Header della TEI che dovrebbe venire assunto dallo standard per i documenti. Con i suoi elementi *file description* (<fileDesc>), *encoding description* (<encodingDesc>), *text profile* (<profileDesc>) e *revision history* (<revisionDesc>), esso serve a documentare le metainformazioni del file, ovvero informazioni sul titolo, sull'editore, sul-

<sup>18</sup> Cfr. <http://www.tei-c.org/P5/>.

<sup>19</sup> Cfr. <http://www.tei-c.org.uk/Master/Reference/>.

<sup>20</sup> Cfr. <http://listserv.brown.edu/archives/cgi-bin/wa?A1=ind0405&L=tei-1#8>.

la fonte dei dati, sulla modalità di codifica, sulle rielaborazioni, permette una descrizione del contenuto e una sua classificazione. Con ciò, da una parte, viene attestato il lavoro compiuto sui dati elettronici, mentre dall'altra il file viene ordinatamente inserito all'interno di un *corpus* più grande, come per esempio una biblioteca elettronica o un *corpus* ancora più ampio.

I partecipanti al Workshop non hanno avuto la possibilità di discutere se gli elementi di codifica testuale presenti nel TEI siano utilizzabili in un'edizione critica. Qui in effetti troviamo da un lato elementi utili alla descrizione di trascrizioni, come per esempio: abbreviazioni (<abbr> ed <expan>), correzioni (<add>, <corr>, <del>), danni (<damage>), cambi di mano (<handshift>) e così via; e dall'altro lato elementi utili per un apparato critico, come per esempio: elementi dell'apparato (<app>), lezioni (<rdg>) e testimoni (<wit>). Il gruppo di lavoro non ha poi ancora avuto la possibilità di discutere le modalità definite da TEI per la descrizione di luoghi, di persone e di date.

Il Working Group assumerebbe tuttavia le proposte del TEI per la descrizione della tradizione. In questo modo una lista (<witlist>) dei singoli testimoni (<witness>) entrerebbe a far parte del complesso di informazioni sulla tradizione. Il caso dei 'testimoni' induce a illustrare un problema generale, che ha dato luogo a discussione anche durante i lavori. L'utente impartisce degli ordini al computer impiegando, per fortuna, termini del linguaggio naturale. Questi termini, però, sono fra loro legati da regole: in altre parole essi devono essere univoci, con un solo significato. Il vantaggio di una tecnologia fondata sulla marcatura dei testi consiste nel fatto che le informazioni vengono inserite in testi che sono espressi in linguaggio naturale, ma che dal computer vengono compresi come informazioni aggiuntive al testo. Un 'elemento' del testo <abbr> è per uno studioso relativamente facile da interpretare come il segno di una abbreviazione. Per il computer esso è invece una catena inquivocabile di segni, sulla quale sono da applicare determinate regole, come per esempio la regola di rappresentare quella catena in un certo modo. Ora, la lingua contiene purtroppo delle ambiguità che in determinati contesti non emergono, ma che alla macchina, che non conosce quei contesti, procurano problemi. Così i partecipanti al Workshop hanno discusso se fosse

meglio adottare il concetto di ‘testimone’ per i ‘testimoni del testo’ o non piuttosto per i ‘testimoni del documento’. Hanno deciso di aderire all’uso linguistico di TEI, quindi di adottare *witness* per indicare i testimoni del testo e di proporre il termine latino *testis* per indicare i testimoni del documento. Sarebbe tuttavia da augurarsi che TEI, nell’ambito della trascrizione, passi al termine *textwitness*.

Durante la discussione sono emerse ulteriori debolezze dello standard, che per esempio non presuppone alcun riferimento di tipo archivistico e che riesce a rappresentare soltanto indirettamente la struttura formale di un testo documentario. Il Working Group considera TEI come una cornice, che mette a disposizione un comune *tag-set* (un gruppo di elementi di marcatura), che deve essere completato attraverso dei sottogruppi di elementi per specifici tipi di testo. Il Working Group si è perciò prefissato, fra gli altri, lo scopo di proporre a TEI degli elementi che possano costituire un sottogruppo per la marcatura di documenti medievali e della prima età moderna.

Il Working Group ritiene inoltre che quello proposto da W3C come modello di descrizione di strutture, “Schema”, sia più adatto del modello attualmente esistente (“Document Type Definition”). Il Working Group inoltre considera l’integrazione di *namespaces* nel progetto TEI ragionevole e di ausilio per risolvere gli specifici problemi della marcatura del testo nei documenti.

\*\*\*\*\*

Baricentro della discussione del Working Group non sono state però le questioni tecniche ricordate, ma il tentativo di sviluppare una semantica comune. La semantica elaborata dal Working Group verrà descritta dettagliatamente sul suo sito Web<sup>21</sup> e ulteriormente sviluppata. Verrà qui presentata solo brevemente.

<sup>21</sup> Attivo fra breve: <http://www.cei.lmu.de>.

Il singolo documento dovrebbe essere contrassegnato come *document*. I documenti possono poi essere descritti sulla base di due grossi blocchi: le metainformazioni (registro, descrizione formale, commentario) e le informazioni sul testo (edizione critica, elementi linguistici, struttura della pagina, elementi del testo rilevanti dal punto di vista del contenuto).

Nell'ambito delle metainformazioni il Working Group propone di descrivere il documento con l'ausilio di un *abstract*, di un complesso di informazioni in merito alla tradizione (lista dei testimoni del testo, <witnessList>) e di una *diplomaticAnalysis*. L'*abstract* rende possibile una sintesi del contenuto.

Come s'è detto, con l'ausilio dell'elemento *witnessList* assunto dal TEI possono essere inclusi i testimoni materiali del testo. Sarebbero nello stesso contesto da descrivere anche le caratteristiche materiali dei documenti originali. Questa <physDesc> (*physical description*) dovrebbe descrivere, con l'ausilio degli elementi previsti da MASTER, le caratteristiche materiali di un oggetto documentario. La determinazione degli elementi per la descrizione di un sigillo dovrebbe esser compito degli specialisti di sfragistica. All'interno della lista dei testimoni del testo dovrebbero venir accolti anche elementi come le annotazioni e segnature apposte sul documento dal destinatario o comunque in archivio. Analogamente potrebbero venirvi concepiti elementi della validazione del documento e che si riferiscono allo scrittore.

Questi elementi stanno in una zona a confine con l'analisi diplomaticistica, cosicché il Working Group fa presente che è questione innanzitutto di elaborare un comune vocabolario per le descrizioni degli elementi, che non deve essere ancora strutturato a questo livello. In senso proprio la *diplomaticAnalysis* contiene informazioni per lo studio sia diplomatico sia storico del documento, vale a dire il rinvio a più antiche edizioni a stampa e a registi, così come a monografie diplomaticistiche, quindi a un testo di critica documentaria come è per esempio buona parte delle note di presentazione di una edizione MGH.

Il testo vero e proprio e le informazioni su di esso potrebbero venire riuniti sotto il comune denominatore *tenor*. A questo apparterebbero gli elementi che solo un originale contiene, gli elementi della struttura linguistica (il formulario del documento) e del contenuto. Il Working Group

ha proposto per esempio di concepire le parti ‘esterne’ del testo come una classe *nota*, che al livello di attributi sarebbero poi da marcare più precisamente come annotazioni di registrazione o di tasse, come ordini di documentazione, come annotazioni di *vidit* etc.

La struttura linguistica del documento verrebbe descritta adottando i termini del formulario, prioritariamente raccogliendo come classe *formula* la componente del formulario definita a livello di contenuto, la cui funzione come clausola, formula di *non obstante*, eccezione e così via dovrebbe poi venire marcata a livello di attributi. La ricognizione, il rigo di *subscriptio* etc. dovrebbero venire descritti, secondo l’opinione del Working Group, come forme della *subscriptio*.

Il contenuto del documento dovrebbe diventare accessibile mediante gli elementi desunti dal TEI come *name*, *date*, *place*, *quote*, e in questo il Working Group intende certo utilizzare l’*index* di elementi generali ma vi vedrebbe volentieri aggiunti un elemento *res* e un elemento *persona* per insiemi di informazioni relative all’oggetto (non sempre materiale) dell’azione documentata e alle persone, con le diverse funzioni, che compaiono all’interno del testo documentario.

L’elenco delle denominazioni degli elementi non è ancora stato perfezionato. Lo stato delle scelte in merito verrà documentato sul sito Web del Working Group.

## Risultato

Questi i risultati del seminario, i suoi punti fermi. Anzitutto, i documenti sono pubblicabili con XML; il potenziale e la necessità di integrazione, da una parte, sono molto alti e, dall’altra, le strutture linguistiche e formali sono così simili, da alimentare la speranza di poter sviluppare un linguaggio formale comune, con il quale testi documentari digitali, in contesti d’impiego anche differenti, possano venire marcati. Con questa semantica comune può essere almeno creata una zona di intersezione, sulla quale definire di volta in volta i modelli di edizione locali. Si può sperare che uno schema concordato a livello internazionale si diffonda in modo tale che, grazie alla comune forma di pubblicazione digitale, siano

prossime anche le riflessioni per lo sviluppo di un motore di ricerca per i fondi documentari internazionali, che riunirebbe una parte importante della comune eredità del Medioevo europeo.

GEORG VOGELER