

# Sul perché, anche nel mondo dei Linked Data, non possiamo rinunciare al concetto di documento

di Alberto Salarelli

## Dal *Web of documents* al *Web of data*

Ormai da diversi anni Tim Berners-Lee, l'inventore del WWW, formulando l'auspicio che la sua creazione debba evolvere dal Web dei documenti al Web dei dati, ha offerto alla platea degli utilizzatori della Rete la visione di un nuovo scenario globale per il trattamento e la condivisione dell'informazione<sup>1</sup>. Questa visione, basata su alcune regole di base enunciate dallo stesso Berners-Lee nel 2006 e che vengono ormai considerate come le tavole della legge dei Linked Data, ovvero i 'Linked Data principles'<sup>2</sup>, ha riscosso in tempi recenti un notevole successo, non solo sul piano della speculazione teorica ma anche su quello dell'implementazione di servizi, come attestato dalla messa a disposizione di quantità sempre più ingenti di archivi di dati strutturati secondo le norme summenzionate, e il parallelo sviluppo di strumenti in grado di operare su questi dati allo scopo di trarne informazioni potenzialmente utili.

In che cosa consiste questo cambio di paradigma? In estrema sintesi, se il Web dei documenti è quello costituito da risorse HTML reciprocamente collegate tramite link 'non tipizzati' (*untyped*), e cioè link la cui unica funzione consiste, per l'appunto, nel rimandare da un documento all'altro del sistema senza esprimere alcuna valutazione in merito alla natura concettuale del collegamento stesso, il Web dei dati assume come punto di partenza l'adozione di uno schema (RDF) che impone una destrutturazione del documento individuandone, dal punto di vista concettuale, la natura, la tipologia e gli attributi degli elementi che lo costituiscono e dei legami che connettono questi elementi fra loro. L'idea di correlare i dati reciprocamente tramite legami che siano in grado di esprimere il significato di tali correlazioni in una forma interpretabile da un elaboratore, non è di certo una novità: in ambito informatico il model-

ALBERTO SALARELLI, Università degli studi di Parma, Dipartimento di lettere, arti, storia e società, piazzale della Pace 7/A, 43121 Parma, e-mail [alberto.salarelli@unipr.it](mailto:alberto.salarelli@unipr.it).

Ultima consultazione siti web: 8 settembre 2014.

1 Si vedano, ad esempio, le dichiarazioni rilasciate in un'intervista a Jason Rubin, *Tim Berners-Lee on the future of his invention*, «ON Magazine», (2009), n. 4, <<http://www.emc.com/leadership/features/berners-lee.htm>>. Fra i primi documenti attorno allo sviluppo del Web dei dati cfr. Tim Berners-Lee, *Semantic Web road map*, 1998, <<http://www.w3.org/DesignIssues/Semantic.html>>.

2 Tim Berners-Lee, *Linked Data*, 27 luglio 2006, <<http://www.w3.org/DesignIssues/LinkedData.html>>.



lo di rappresentazione concettuale dei dati denominato ER (Entità-Relazione) fu sviluppato negli anni Settanta del secolo passato proprio allo scopo di connettere, all'interno di un database, i vari elementi in modo significativo. Non a caso lo stesso Tim Berners-Lee, in una delle sue prime riflessioni attorno al futuro del Web semantico, parlava dei vantaggi – grazie all'utilizzo dei linguaggi di schema e di inferenza – di far sembrare tutti i dati al mondo come un unico enorme database<sup>3</sup>. Quindi, se l'idea di correlare i dati in modo significativo è vecchia, la novità dei Linked Data consiste nelle forme concrete di implementazione con cui oggi si va ad operare su collezioni di dati che sempre più massicciamente vengono resi disponibili in formato aperto. Ci riferiamo a schemi di descrizione delle risorse (RDF – Resource Description Framework), identificatori delle medesime (URI – Uniform Resource Identifier), linguaggi di rappresentazione delle entità (OWL – Ontology Web Language) che condividono un principio base, quello dell'interoperabilità, ovvero la condizione necessaria per far sì che i differenti sistemi siano in grado di essere trasparenti l'uno all'altro non solo per potersi scambiare dati, ma per offrire la possibilità a soggetti terzi di creare nuove aggregazioni informative basate sui contenuti dei differenti dataset.

I vantaggi di questo tipo di approccio sono evidenti: i Linked Data, infatti, rappresentano una delle forme di utilizzo più interessanti e proficue dei documenti digitali dal momento che offrono la possibilità di correlare reciprocamente i contenuti recuperati da differenti archivi realizzando, in tal modo, strumenti informativi del tutto nuovi che al momento è difficile, e forse inutile, tentare di denominare con terminologie mutuata dal passato e questo perché, come ha scritto George Steiner, «il periodo in cui viviamo è un tempo di transizione, più rapido, più difficile da 'leggere' di qualunque altra epoca precedente»<sup>4</sup>. Tuttavia, dal momento che ogni medaglia ha il suo rovescio, ci pare opportuno considerare come queste nuove forme di aggregazione di dati siano fondate su un modello di strutturazione logica degli elementi costitutivi dei documenti originari finalizzato alla decostruzione e successiva ri-significazione dei medesimi: per usare una metafora, i mattoni (dati) che costituiscono l'edificio di partenza (documento originario), vengono enucleati da esso (decostruzione) al fine di un reimpiego (ri-significazione) in una nuovo artefatto (RDF graph).

Perché questa operazione deve essere osservata con occhio critico? Perché al di là del ruolo dell'edificio originario, viene da chiedersi qual è la *ratio* architettonica che soggiace alla ricomposizione in forme nuove di elementi già utilizzati in altri contesti: in altre parole, è necessario considerare come la perdita del vincolo originario che teneva uniti i dati nel documento di partenza, nel quale si esplicitavano gli obiettivi della sua costituzione, le caratteristiche editoriali e la responsabilità autoriale, possa rappresentare un fattore di rischio o, perlomeno, di debolezza nei confronti della stabilità semantica dei nuovi costrutti. Si tratta, insomma, di considerare quali conseguenze, logiche e pratiche, si configurino nel momento in cui «the record disintegrates as the data becomes distributed»<sup>5</sup>. È in gioco, insomma, una partita estremamente delicata tra due esigenze non necessariamente contrapposte ma, di certo, difficili da conciliare: la qualità dei documenti in quanto aggregazioni di dati basate su principi e regole condivisi all'interno di una specifica comunità di utilizzatori e la loro messa

3 Id., *L'architettura del nuovo Web*. Milano: Feltrinelli, 2001, p. 162.

4 George Steiner, *I libri hanno bisogno di noi*. Milano: Garzanti, 2013, p. 76.

5 Martin Malmsten, *Cataloguing in the open: the disintegration and distribution of the record*, «JLIS.it», 4 (2013), n. 1, p. 421, <<http://leo.cineca.it/index.php/jlis/article/view/5512>>, DOI: 10.4403/jlis.it-5512.

a disposizione in forma aperta, disaggregata e riutilizzabile da tutti. Allorché i dati erano conservati in archivi (oggi si usa dire 'silos') separati, inevitabilmente l'organizzazione responsabile esercitava su di essi un controllo assoluto, controllo che se da un lato impediva la loro messa in comunione, dall'altro era garanzia di coerenza nella definizione formale del sistema e nel conseguente trattamento documentario. Oggi, in un mondo sempre più interconnesso, qualsiasi arroccamento su posizioni di chiusura risulta difficilmente accettabile: come ha chiaramente affermato Caren Coyle riferendosi agli archivi di dati bibliografici, «We can no longer view the goal of our data creation to be a library catalog that looks much like the catalog we have today. And we can no longer view our catalog as a destination that is separate from the open web»<sup>6</sup>. Di fatto, però, se l'obiettivo da raggiungere è quello della condivisione dei propri patrimoni informativi, ci si chiede se da esso non derivi anche il rischio di una perdita più o meno ampia di verifica sui medesimi: al di là di ciò che i cittadini del mondo potranno fare con i dati che un'organizzazione renderà liberamente disponibili (e cioè come essi saranno riutilizzati) il vero punto chiave è come questa organizzazione dovrà cambiare mentalità e procedure per operare in senso aperto. Per rimanere in campo bibliografico, la transizione dal formato MARC al formato BIBFRAME sta sollevando esattamente questo tipo di contrasto tra i fautori di posizioni *top down*, più votati all'obiettivo del controllo, e quelli votati al *bottom up*, per i quali flessibilità e interoperabilità sono mantra irrinunciabili<sup>7</sup>.

Del resto, come ha spiegato Jacques Derrida, il vero potenziale ermeneutico dell'azione decostruttiva consiste nel saper fornire un punto di vista differente sulle cose del mondo, essa perciò serve, innanzitutto, a fornire una carica di energia al pensiero e a rimettere in moto situazioni sclerotizzate<sup>8</sup> e quindi, sotto questo aspetto, ogni confronto di posizioni, anche aspro, è un buon segno. Ciò non significa, a mio modo di vedere, che la decostruzione debba essere intesa come un annichimento dell'esistente per la fondazione di un mondo del tutto nuovo: un approccio radicale, questo, da rammentare le illusioni di Leopardi sulle «magnifiche sorti e progressive». Significa, invece, che si debba procedere con accortezza lungo la strada dell'apertura vagliando i rischi che lungo di essa si frappongono, rischi riconducibili in generale alle dinamiche tipiche del trattamento dei dati aperti e che consistono nella difficoltà di ricavare delle correlazioni significative tra dati provenienti da fonti diverse, soprattutto se si voglia porre l'accento sul senso della significanza nel contesto di una specifica correlazione fra entità concepite in modi e in tempi differenti per differenti sistemi di trattamento e archiviazione<sup>9</sup>. Questo problema si è

6 Karen Coyle, *Library linked data: an evolution*, «JLIS.it», 4 (2013), n. 1, p. 60, <<http://leo.cineca.it/index.php/jlis/article/view/5443>>, DOI: 10.4403/jlis.it-5443.

7 Ne dà conto Kim Tallerås, *From many records to one graph: heterogeneity conflicts in the linked data restructuring cycle*, «Information Research», 18 (2013), n. 3, <<http://InformationR.net/ir/18-3/colis/paperC18.html>>.

8 Jacques Derrida, *Lettera a un amico giapponese*, «Rivista di estetica», 25 (1984), n. 17, p. 5-10.

9 «Un modo di affrontare la questione ontologica consiste nel costruire un modello che definisca le entità, gli attributi e le relazioni. La creazione di modelli come mezzo per esemplificare la teoria riflette una tendenza secondo la quale si concettualizza una base dati per mezzo di un processo di formalizzazione astratta. La concettualizzazione è soggetta ad alcune limitazioni, la principale delle quali è lo scopo che devono assolvere le basi dati da rappresentare come modello» (Elaine Svenonius, *Il fondamento intellettuale dell'organizzazione dell'informazione*. Firenze: Le lettere, 2008, p. 46).

presentato in tutta la sua evidenza nel caso di Google Flu Trends, un servizio messo a disposizione qualche tempo fa da Google per monitorare, attraverso l'analisi delle ricerche effettuate dagli utenti, il diffondersi in tempo reale delle sindromi influenzali, cercando altresì di anticipare le forme di diffusione dell'epidemia. Google Flu, dopo una partenza promettente, ha presentato diverse falle e questo perché, com'è stato osservato<sup>10</sup>, gli errori di predizione del modello sono imputabili a diversi fattori critici che, in un'ipotesi più generale, possono applicarsi all'intero ambito dei dati aperti, fattori che comprendono le eterogeneità di creazione, in termini di modalità e di obiettivi, dei diversi dataset<sup>11</sup>.

Le modalità di trattamento dei Linked Data sono state sviluppate proprio per superare queste limitazioni caratteristiche dei 'semplici' dati aperti, e in particolare delle loro massive aggregazioni e utilizzazioni (Big Data). Nondimeno è opportuno osservare come anche attorno ad essi non manchino voci fortemente critiche in relazione alla possibilità che i Linked Data siano sufficientemente attrezzati per rispondere agli auspici del Web semantico.

Per intendere in modo compiuto la valenza e le potenzialità dei Linked Data occorre considerare come essi non sostituiscano il Web dei documenti, ma si appoggino su di esso come un nuovo strato (*layer*) costituito da specifiche interconnessioni semantiche. Questo aspetto emerge in maniera indiscutibile dalle parole di Christian Bizer, Tom Heath e Tim Berners-Lee laddove affermano che: «the Web of Data can therefore be seen as an additional layer that is tightly interwoven with the classic document Web»<sup>12</sup>. Ciò significa che lo 'strato' dei Linked Data non ha senso in una dimensione indipendente dal 'sostrato' documentario sottostante e cioè, banalmente, che non è sufficiente ridurre un documento in triple per esplicitarne l'intero potenziale comunicativo, dal momento che il modello su cui si fondano i Linked Data ancora soffre di limiti strutturali che ne minano l'effettiva efficacia in termini d'uso: tale modello, al momento, non è in grado di esprimere in maniera convincente la complessità dei legami che uniscono in maniera reciproca i dati provenienti da archivi differenti a motivo della incoerenza espressiva dei valori semantici di un dataset rispetto a un altro<sup>13</sup>. Ma non basta: un documento, infatti, non è costituito solo dai dati e dai legami fra loro intercorrenti ma anche da quelle caratteristiche progettuali che, come accennato sopra, tengono questi dati uniti in un insieme sensato e coerente; si tratta di una dimensione, questa, che i Linked Data non contempiano, e che quindi necessita di ulteriori strumenti per poter essere esplicitata ai fini,

**10** David Lazer, *The parable of Google Flu: traps in Big Data analysis*, «Science», 343 (2014), n. 6176, p. 1203-1205.

**11** Gary Marcus; Ernest Davis, *Eight (no, nine!) problems with Big Data*, «The New York Times», 6 April 2014, <[http://www.nytimes.com/2014/04/07/opinion/eight-no-nine-problems-with-big-data.html?\\_r=1](http://www.nytimes.com/2014/04/07/opinion/eight-no-nine-problems-with-big-data.html?_r=1)>; Kaiser Fung, *Google Flu Trends' failure shows Good Data > Big Data*, «HBR Blog Network», 25 March 2014, <<http://blogs.hbr.org/2014/03/google-flu-trends-failure-shows-good-data-big-data/>>.

**12** Christian Bizer; Tom Heath Tim Berners-Lee, *Linked Data - The story so far*, «International Journal on Semantic Web and Information Systems», 5 (2009), n. 3, p. 4, DOI: 10.4018/jswis.2009081901.

**13** Cfr. Prateek Jain [et al.], *Linked Data is merely more data*. In: *Linked Data meets Artificial Intelligence, Palo Alto, California, 22-24 March 2010*, edited by Dan Brickley [et al.]. Menlo Park, California: AAAI Press, 2010, p. 82-86. Sulle carenze nella descrizione concettuale dei differenti dataset si vedano anche le considerazioni di Vuk Miličić, *Problems of Linked Data (1/4): Identity*, «Bew Citnames», 26 July 2011, <<http://milicicvuk.com/blog/2011/07/26/problems-of-linked-data-14-identity/>>.

per esempio, della possibilità di interpretare e riutilizzare in modo corretto i dati collegati a un determinato esperimento scientifico<sup>14</sup>.

Riassumendo: il modello attuale dei Linked Data presenta il forte rischio di non poter esprimere le proprie reali potenzialità; occorre quindi agire per poterlo migliorare perché se le limitazioni che lo caratterizzano rimarranno inalterate, tale modello «will merely be more data that suffers from the same kinds of problems which plague the Web of Documents, and hence the vision of the Semantic Web will fall short»<sup>15</sup>.

### Provenienza e granularità: due sfide per i Linked Data

Quali sono, dunque, gli elementi critici dei Linked Data su cui si potrebbe agire in vista di un miglioramento della loro funzionalità? Per poter rispondere a questa domanda bisogna, a mio parere, partire da un interrogativo di respiro più ampio. Vale a dire che, nelle considerazioni seguenti, vorrei anzitutto porre l'attenzione sul come i Linked Data possano inserirsi all'interno di una teoria del documento nella quale si possano distinguere diversi gradi di autorevolezza e stabilità dei documenti stessi, sulla base delle loro specifiche caratteristiche formali.

Com'è noto, la domanda «che cos'è un documento?» ricorre ormai da diversi decenni in quella parte di letteratura scientifica relativa alle discipline del libro, dell'archivio e dell'informazione, legata a studiosi interessati, fra le altre cose, anche alla definizione della natura epistemologica dei propri oggetti di studio<sup>16</sup>. Si tratta, a mio avviso, di una domanda cruciale dal momento che le società complesse come la nostra si fondano sui documenti, e che quindi la definizione di cosa essi siano rappresenta una chiave ermeneutica di primaria importanza per comprenderne struttura e funzionamento.

È pur vero che, come ha rilevato Bernd Frohmann<sup>17</sup>, la pretesa di fornire una risposta esaustiva a tale interrogativo potrebbe condurci verso una teoresi talmente astratta da non risultare di alcuna utilità per le pratiche concrete di trattamento dei documenti stessi: ora, a parte il fatto che tale considerazione può rappresentare una via di fuga fin troppo comoda dal problema, se accettiamo di prescindere dal piano ontologico della questione non ci rimane che ordinare le diverse tipologie di documenti in base a un approccio pragmatico, creando cioè delle categorie che si basano unicamente sui modi d'uso dei documenti stessi. Il che, com'è facilmente intuibile, può forse essere sufficiente per un'organizzazione personale della propria dimensione documentaria mentre risulta del tutto

14 Sean Bechhofer [et. al.], *Why linked data is not enough for scientists*, «Future Generation Computer Systems», 29 (2013), n. 2, p. 602, DOI: 10.1016/j.future.2011.08.004.

15 P. Jain [et al.], *Linked Data is merely more data* cit., p. 82.

16 Il riferimento imprescindibile è, ovviamente, l'articolo di Michael Buckland, *What is a "document"?*, «Journal of the American Society for Information Science», 48 (1997), n. 9, p. 804-809, DOI: 10.1002/(SICI)1097-4571(199709)48:9<804::AID-AS15>3.0.CO;2-V. Di notevole interesse anche le considerazioni sul concetto di 'documento' in E. Svenonius, *Il fondamento intellettuale dell'organizzazione dell'informazione* cit., *passim*. Una sintesi del dibattito internazionale attorno al concetto di documento si trova in David Bawden; Lyn Robinson, *Introduction to Information Science*. London: Facet Publishing, 2012, p. 75-78. In ambito italiano hanno discusso sul tema Claudio Gnoli, Paola Rescigno, Riccardo Ridi e Alberto Salarelli su «Bibliotime», 15 (2012), n. 1, <<http://www.aib.it/aib/sezioni/emr/bibtime/num-xv-1/index.html>>.

17 Bernd Frohmann, *Revisiting "what is a document?"*, «Journal of Documentation», 65 (2009), n. 2, p. 291-303, DOI: 10.1108/00220410910937624.

inadeguato sul piano sociale, laddove una collettività deve necessariamente riconoscere a particolari categorie di documenti specifiche valenze, per esempio in ordine al loro valore probatorio. Valenze che il documento acquisisce, in primo luogo, sulla base degli elementi formali che sottendono alla sua redazione e, in secondo luogo, in relazione alle soluzioni tecniche con le quali esso viene registrato su un supporto materiale<sup>18</sup>. Se questi due piani del discorso sono essenziali per intendere il concetto di documento a tutto tondo, in riferimento agli obiettivi di questo intervento è certamente il primo quello che deve essere tenuto in somma considerazione. Infatti, al di là del 'come' il documento assuma forme e caratteristiche differenti a seconda dei modi della sua registrazione, qui si vuole in specifico sottolineare la logica tramite la quale i dati che lo compongono vengono ad essere organizzati per le finalità specifiche con cui esso è stato elaborato<sup>19</sup>.

Quindi, la prima fondamentale categorizzazione riguarda i documenti che supportano informazioni 'consapevolmente' registrate dai rispettivi autori per essere trasmesse nel tempo, e documenti che emergono come portatori di informazioni solo in relazione alle capacità dei soggetti interpretanti di saperne trarre informazioni utili. Tale suddivisione che, per inciso, ha utilizzato Alberto Salarelli ne *La biblioteca digitale*<sup>20</sup> mutuandola da Jacques Le Goff e Henri Moniot<sup>21</sup>, è stata considerata di recente più o meno negli stessi termini da Riccardo Ridi<sup>22</sup> e da Maurizio Ferraris il quale, a ben vedere, si spinge più in là proponendo una scala di intensità che consenta di distinguere tra documenti forti, documenti deboli, iscrizioni, registrazioni e tracce. Secondo Ferraris il documento forte consiste nell'iscrizione di un atto, mentre quello debole nell'attestazione di un fatto:

ontologicamente, tra il documento in senso forte e il documento in senso debole intercorre una differenza rilevante, giacché il primo è per l'appunto un *atto*, il secondo una *prova*, che può eventualmente venire adoperata in un atto, ma che non necessariamente (anzi, quasi mai) lo è<sup>23</sup>.

**18** La migliore definizione di documento, in grado di riassumere il rapporto imprescindibile tra il dato e il suo supporto, rimane quella di Alfredo Serrai: «i documenti sono oggetti che portano segni» (Alfredo Serrai, *Biblioteconomia come scienza: introduzione ai problemi e alla metodologia*. Firenze: Olschki, 1973, p. 21).

**19** In buona sostanza è nell'ontologia del documento che emerge, con specifica evidenza, la compresenza di elementi idiografici (espressioni di un'identità) e nomotetici (in conformità a un canone di riproducibilità formale), cfr. Maurizio Ferraris, *Sans papier: ontologia dell'attualità*. Roma: Castelvecchi, 2007, p. 93-94. Ciò non esclude che una formulazione generica come 'risorsa', utilizzata nei Principi internazionali di catalogazione, possa mirare al medesimo scopo ma, di certo, con diminuita efficacia. Il fatto stesso che, secondo ICP, 'risorsa' possa riferirsi a qualsiasi entità del gruppo 1 di FRBR ne rende improprio l'utilizzo a livello ontologico, potendo essa indicare tanto l'opera (concetto di per sé quanto mai ambiguo, cfr. Mauro Guerini, *I Principi internazionali di catalogazione (ICP)*. Milano: Editrice Bibliografica, 2012, p. 109-114), come la sua espressione formale e, infine, la sua registrazione su un determinato supporto materiale.

**20** Alberto Salarelli; Anna Maria Tammaro, *La biblioteca digitale*. Milano: Editrice Bibliografica, 2006, p. 19-23.

**21** Jacques Le Goff, *Documento/monumento*. In: *Enciclopedia*. Torino: Einaudi, 1978, vol. 5, p. 46; Henri Moniot, *La storia dei popoli senza storia*. In: *Fare Storia*, a cura di Jacques Le Goff, Pierre Nora. Torino: Einaudi, 1988, p. 76.

**22** Riccardo Ridi, *Il mondo dei documenti*. Roma-Bari: Laterza, 2010, p. 10-14.

**23** Maurizio Ferraris, *Documentalità*. Roma-Bari: Laterza, 2009, p. 300.

In altri termini, se al vertice della piramide documentale stanno i documenti dichiarativi (quelli riconosciuti a livello sociale, che definiscono norme, attribuiscono status, creano istituzioni), mano a mano che scendiamo verso la base ci imbattiamo prima in documenti concepiti come portatori di una precisa finalità informativa ma non tali da produrre oggetti sociali (come uno scambio epistolare tra due amici che si raccontano come va la vita), per poi rinvenire, alla base dello schema, quei documenti inintenzionali costituiti da mere tracce interpretabili come informazioni solo grazie alla perspicacia di un lettore specificamente interessato ad esse e opportunamente addestrato per poterle riconoscere. L'elemento chiave che consente di stabilire questa fenomenologia del documento consiste nel tipo di iscrizione di cui esso è portatore: più l'iscrizione sarà di livello qualitativamente elevato, vale a dire caratterizzata da spiccata nomoteticità (e cioè registrata in modo peculiare, secondo stilemi formali riconosciuti socialmente), più il documento potrà essere collocato in alto nel prospetto tassonomico. Fin qui Ferraris.

Ora, ciò che mi preme considerare in questa sede è l'ambito di applicazione dei Linked Data: mi interessa, in altri termini, capire su quali documenti può estendersi il loro strato semantico. In linea di principio tutti i documenti, intenzionali e non, possono essere triplizzati tuttavia, in considerazione del fatto che l'architettura di una tripla è costituito da un 'predicato' (e cioè da una affermazione che indica una proprietà del soggetto), diventa cruciale distinguere l'autorità di chi esprime tale giudizio di valore in relazione ai contenuti e alle caratteristiche formali che il documento già di per sé propone oltre, ovviamente, alla capacità di formulare in modo corretto tale espressione secondo le regole stabilite dalla grammatica dei Linked Data. Dico, insomma, che quegli elementi caratterizzanti la forza di un determinato documento – idiografici e nomotetici – non possono essere affatto tralasciati anche nella fase di creazione di quella rete di significati in grado di correlare i dati reciprocamente in modo sensato. Per capirci, quando Tim Berners-Lee ne *L'architettura del nuovo Web* provò a spiegare a un pubblico di non specialisti il concetto di Web semantico, fece l'esempio di un tizio intenzionato a vendere un'automobile gialla usata<sup>24</sup>. Ebbene, questo tizio posta in Rete il relativo annuncio, ma i motori di ricerca non ne comprendono gli elementi essenziali (marca, modello, colore, stato di manutenzione ecc.) perché il messaggio è stato composto in testo semplice. Invece, aggiunge Berners-Lee, se l'annuncio fosse stato composto tramite la compilazione di un modulo, i dati sarebbero stati leggibili e interpretabili automaticamente dai computer, consentendo di creare un sistema di informazioni omogenee e correlate. Il problema è: chi stabilisce l'individuazione degli elementi essenziali (le entità) e le reciproche relazioni? Chi, insomma, disegna il modulo? I responsabili di un portale di vendita online come Ebay? L'autoconcessionario che viene incaricato della pratica? L'ACI o la FIA? Il PRA? O, addirittura, il singolo venditore, cioè l'utente stesso, che supponiamo appassionato di informatica e, dunque, in grado di disegnarsi la propria personale ontologia?

La risposta di Berners-Lee è: non importa. Quello che conta è che esistano delle stanze di compensazione in grado di tradurre una categoria in un'altra rendendole reciprocamente trasparenti: i linguaggi inferenziali. Francamente mi pare una visione un po' ingenua o, perlomeno, semplicistica, che lascia in un cantone la responsabilità di chi disegna, per un determinato scopo, un'architettura dei dati. Affermare che, in relazione a questo aspetto, «non esiste nessuna autorità su alcunché»,

24 T. Berners-Lee, *L'architettura del nuovo Web* cit., p. 157.

ovvero statuire il cosiddetto *AAA principle*: «Anyone can say Anything about Any topic», significa esprimere un atto di fede pieno e assoluto nei confronti della capacità euristica dei computer connessi in rete e nell'intelligenza collettiva che scaturisce dalle menti dei loro utilizzatori<sup>25</sup>. Il che potrebbe anche starci quando il problema consiste nel comprare o vendere un'auto usata, un po' meno quando i dati da correlare riguardano il genoma umano o le varianti del titolo di un'opera letteraria o musicale. In egual misura il problema si ripropone in fase di applicazione di un determinato schema e cioè, supponendo di aver individuato l'ontologia più appropriata da applicare per triplizzare un documento, un'ontologia che ipotizziamo istituzionalmente certificata, chi può garantirci che essa verrà successivamente utilizzata in maniera opportuna e che, quindi, ogni entità e ogni predicato saranno l'effettiva espressione dei dati tratti dalla fonte documentaria?

In questi (ma anche in moltissimi altri) casi quel valore che Ridi evidenzia come 'terzietà' nel trattamento dell'informazione assume un'importanza fondamentale:

dovrebbe quindi risultare chiaro – anche se non sempre questo aspetto viene sufficientemente enfatizzato – che oltre alla dimensione tecnica dell'indicizzazione ne esiste anche una etica, che diverrà sempre più importante quanto più si allargherà la platea degli indicatori, che sempre più spesso non sono dei tecnici legati a un ordine professionale o a un codice deontologico, ma liberi cittadini guidati solo dalla propria coscienza e dai propri interessi personali<sup>26</sup>.

Insomma: l'autorità di chi individua le entità e ne stabilisce i relativi predicati e la competenza di chi applica gli schemi sono ingredienti essenziali per il buon funzionamento dei Linked Data proprio in virtù del fatto che esse contribuiscono a costruire quella 'fiducia' considerata dallo stesso Berners-Lee come un «prerequisito fondamentale di una società reticolare»<sup>27</sup>.

Da tutto questo ragionamento emergono un paio di punti critici che, a mio avviso, dovrebbero essere presi in considerazione in vista di un auspicabile miglioramento del sistema nel suo complesso.

Il primo di essi riguarda la dimensione sociale dei Linked Data: se per il concetto di documento (anche quello digitale, ovviamente) vale la considerazione che esso «va concepito, piuttosto che come qualcosa di dato una volta per tutte, e costituente una classe di oggetti stabile, come la reificazione di atti sociali i quali, a loro volta, mutano nella storia e nella geografia»<sup>28</sup>, bisognerebbe estendere le conseguenze di tale affermazione

**25** Maggior cautela emergeva in un documento sulla struttura formale di RDF nel quale si affermava che «To facilitate operation at Internet scale, RDF is an open-world framework that allows anyone to say anything about anything. In general, it is not assumed that all information about any topic is available. A consequence of this is that RDF cannot prevent anyone from making nonsensical or inconsistent assertions, and applications that build upon RDF must find ways to deal with conflicting sources of information» (*Resource Description Framework (RDF): Concepts and Abstract Data Model. W3C Working Draft 29 August 2002*, ed. Graham Klyne and Jeremy Carroll, <http://www.w3.org/TR/2002/WD-rdf-concepts-20020829/>). Un obiettivo dei Linked Data dovrebbe essere proprio quello di limitare le asserzioni illogiche o incoerenti e di mediare i conflitti fra differenti fonti informative. Obiettivo che, per poter essere perseguito, presuppone l'ammissione del fatto che tali situazioni possano in concreto verificarsi.

**26** R. Ridi, *Il mondo dei documenti* cit., p. 57.

**27** T. Berners-Lee, *L'architettura del nuovo Web* cit., p. 115.

anche alla nuova sovrastruttura semantica della Rete. Il fatto che i documenti siano strettamente collegati alle condizioni sociali della loro produzione e, quindi, ai mutamenti spazio-temporali che influiscono su di essi, comporta la necessità di esplicitare tutti gli elementi utili alla loro definizione identitaria. Il problema della provenienza dei dati – problema che può ormai vantare una letteratura scientifica di tutto rispetto<sup>29</sup> – verte, in sostanza, proprio su questo: sulla possibilità, cioè, di collegare la provenienza dei dati con un'identificazione di responsabilità in merito all'origine degli stessi che dia conto del dove, del quando e del perché essi sono stati creati: quali sono, in altri termini, le condizioni 'documentali' della loro origine. Questo consentirebbe, ad esempio, nella formulazione di una *query* o nella progettazione di un *mash-up*, di escludere le fonti di dati provenienti dai social network in generale (per esempio, nella definizione di un quadro eziologico) o, al contrario, di includervi solo quelle provenienti da alcuni specifici gruppi del mondo social (se, putacaso, sto valutando l'impatto di una campagna di marketing). La questione, a livello più astratto, può essere posta in questo modo: se vogliamo che i Linked Data rappresentino un salto di qualità realmente significativo nell'organizzazione della documentazione presente in Rete, abbiamo bisogno di specifici meta-dati in grado di identificare la provenienza, la coerenza e l'età dei dati stessi:

The openness of the Web means that once the data and links are made available on the Web, these different copies of statements about the same set of entities - which might be in conflict and of varied quality - become completely interconnected and intertwined. Finding data about a specific entity may result in multiple URIs identifying this entity and linking to Linked Data objects from different sources. Which of these links should be followed? Which of the Linked Data objects provides more trustworthy or more up-to-date information about the entity? To answer these questions we need not only data about the entity but also information about how the data became available. Hence, we require information about the provenance of Linked Data<sup>30</sup>.

28 M. Ferraris, *Documentalità* cit., p. 281.

29 Fra i primi lavori sul tema si veda: Andreas Harth; Axel Polleres; Stefan Decker, *Towards a social provenance model for the Web*, «The Health Well», 2007, Workshop on Principles of Provenance (PrOPr), <<http://hdl.handle.net/10379/527>>. Altri contributi di interesse, con proposte di diversa natura per l'implementazione di soluzioni in grado di rendere possibile la tracciabilità di origine dei dati, sono i seguenti: Olaf Hartig; Jun Zhao, *Publishing and consuming provenance metadata on the Web of Linked Data*. In: *Provenance and annotation of data and processes. Second International Provenance and Annotation Workshop, IPAW 2008, Salt Lake City, UT, USA, 17-18 June 2008*, edited by Juliana Freire, David Koop. Berlin-Heidelberg: Springer, 2010, p. 78-90; Harry Halpin, *Provenance: the missing component of the Semantic Web for privacy and trust*. In: *Proceedings of the ESWC2009 Workshop on Trust and Privacy on the Social and Semantic Web (SPOT2009), Heraklion, Greece, 1 June 2009*, <<http://ceur-ws.org/Vol-447/paper9.pdf>>; Simon Schenk; Renata Dividino; Steffen Staab, *Reasoning with provenance, trust and all that other Meta Knowledge in OWL*. In: *Proceedings of the First International Workshop on the role of Semantic Web in Provenance Management (SWPM 2009), Washington DC, USA, 25 October 2009*, <[http://www.researchgate.net/publication/221466060\\_Reasoning\\_With\\_Provenance\\_Trust\\_and\\_all\\_that\\_other\\_Meta\\_Knowledge\\_in\\_OWL](http://www.researchgate.net/publication/221466060_Reasoning_With_Provenance_Trust_and_all_that_other_Meta_Knowledge_in_OWL)>; Andre Freitas [et al.], *W3P: building an OPM based provenance model for the Web*, «Future Generation Computer Systems», 27 (2011), n. 6, p. 766-774, DOI:10.1016/j.future.2010.10.010; Maurizio Lunghi; Chiara Cirinnà; Emanuele Bellini, *Trust and persistence for internet resources*, «JLIS.it», 4 (2013), n. 1, p. 375-390, <<http://leo.cineca.it/index.php/jlis/article/view/5494>>, DOI: 10.4403/jlis.it-5494.

30 O. Hartig; J. Zhao, *Publishing and consuming provenance metadata on the Web of Linked Data* cit., p. 79.

Il secondo aspetto critico che va sottolineato concerne il livello massimo di granularità a cui può essere ridotto il documento allorché venga destrutturato in entità distinte senza che per questo motivo si metta a repentaglio il suo portato semantico. Se infatti, come abbiamo già accennato, a livello generale l'estrazione di un dato strutturato da un documento segna un punto di crisi in relazione alle condizioni ideali della sua piena comprensione, visto che viene a perdersi il contesto nel quale il dato stesso era originariamente stato trattato, nondimeno anche a un livello ermeneutico più basso c'è il rischio che, quando il dato risulti eccessivamente frammentato, si verifichino perdite di informazione tali non dico da renderlo inutilizzabile, ma di certo inutile al fine di un effettivo miglioramento delle funzioni semantiche del Web. Poniamo il caso di un grafo RDF relativo a Giuseppe Verdi composto da due distinte triple 'nome:Giuseppe' e 'cognome:Verdi'. Questa struttura rende possibile l'utilizzo delle rispettive triple per creare nuove aggregazioni che condividono o quel nome o quel cognome, ma si sarà perso il legame utile per individuare il Cigno di Busseto. Questo esempio, che potrebbe sembrare banale, è il medesimo utilizzato nel 2005 da un gruppo di informatici delle università di Stanford e del Maryland per introdurre il concetto di 'molecola RDF', un'entità intermedia tra il documento originario e le singole triple allo scopo di evitare esiziali perdite di informazione<sup>31</sup>. D'altra parte non è affatto scontato che, anche aggregando opportunamente le molecole invece delle singole triple, si riesca a ricompattare il totale portato informativo del documento originario<sup>32</sup>, anzi è praticamente certo che nella decostruzione qualcosa vada perduto: è questo l'inevitabile dazio da pagare a fronte dei vantaggi di fruire di un livello informativo più strutturato e più flessibile come quello dei Linked Data?

Quello che si può osservare è come, da qualche tempo a questa parte, attorno a questi punti critici<sup>33</sup> si stiano moltiplicando proposte per poterli affrontare in modo da aumentare l'affidabilità complessiva dei Linked Data, proposte che – e torno al nucleo del mio ragionamento – riprendono le istanze tipiche di una cultura del documento, istanze evidentemente ineliminabili anche nella dimensione del Web semantico. Ciò vale sia nella direzione di un collegamento complessivamente più articolato tra i dati e i documenti originari da cui essi sono tratti, sia sul versante dei nuovi documenti risultanti da agglomerazioni di questi dati reciprocamente correlati: in entrambi i casi, evidentemente, ciò che si cerca di fare è contenere la 'liquidità' dell'informazione in strutture che possano presentare quegli elementi di stabilità caratteristici (anche se non esclusivi) della tradizione bibliografica. In mancanza di questi elementi i Linked Data rischiano di rivelarsi uno strato informativo paragonabile a un semilavorato, collocato su una dimensione documentale forte senza però la prospettiva di poterne costruire una altrettanto solida, seppur basata su presupposti indiscutibilmente innovativi.

**31** Li Ding [et al], *Tracking RDF graph provenance using RDF molecules*, report TR-CS-05-06, Computer Science and Electrical Engineering, University of Maryland, Baltimore County, 30 April 2005, <ftp://www.ksl.stanford.edu/local/pub/KSL\_Reports/KSL-05-06.pdf>.

**32** Vuk Miličić, *Problems of Linked Data (2/4): Concept*, «Bew Citnames», 28 July 2011, <http://milicivuk.com/blog/2011/07/28/problems-of-linked-data-24-concept/>.

**33** In realtà i problemi relativi all'utilizzo dei Linked Data non si limitano ai due punti evidenziati ma comprendono anche questioni relative alla formalizzazione dei dati, alla loro obsolescenza, alla privacy e all'usabilità. Per una rassegna dettagliata si veda: Jorge Luis Morato [et. al.], *Evaluation of semantic retrieval systems on the semantic web*, «Library Hi Tech», 31 (2013), n. 4, p. 638-656, DOI 10.1108/LHT-03-2013-0026.

### L'imprescindibile concetto di documento

La tendenza a privilegiare la dimensione connettiva di un sistema informativo rispetto agli elementi da connettere non è sbocciata dal nulla con l'avvento dei Linked Data. Già negli anni Settanta, infatti, un eminente scienziato come Lewis M. Branscomb affermava che i documenti sono solo «sottoprodotti occasionali dell'accesso all'informazione e non la sua principale incarnazione»<sup>34</sup>. Erano gli anni – lo ricordavamo all'inizio – in cui si iniziavano a sviluppare i database relazionali, il che ci porta a osservare un fenomeno ricorrente nella storia delle tecniche di trattamento dei dati: ogni qualvolta si ipotizza (ma soprattutto si implementa) un gestionale che, in modo più o meno esplicito, si ricolleggi al Memex di Vannevar Bush, si celebrano i funerali del documento o, perlomeno, la sua ricollocazione in subordine, con la contestuale apoteosi di un concetto più brillante, scattante e flessibile: quello di informazione. Così è accaduto con l'avvento dei sistemi ER, poi con l'invenzione del Web, di nuovo con la sua evoluzione in senso partecipativo, infine con i Linked Data. In questi frangenti, insieme al documento, gli *information technologists* tendono a considerare sotto una luce negativa anche le istituzioni tradizionali che sovrintendono alla sua organizzazione e conservazione, in primo luogo biblioteche e archivi, le quali vengono accusate di essere rigide, ingessate nella pervicace applicazione di tassonomie calate dall'alto, dedite – per dirla con Petrucci – ad una «'pietas' dello scritto, con le sue liturgie e i suoi sacerdoti»<sup>35</sup>, in altre parole incapaci di far emergere il potenziale informativo contenuto nei materiali oggetto del loro trattamento e, come tali, destinate a perire con l'avanzare dei tempi nuovi<sup>36</sup>. Questo potenziale, al contrario, si ritiene possa essere scatenato da un approccio emergente dal basso (definito come folksonomia o come intelligenza collettiva, a seconda dei casi) e caratterizzato da ordinamenti prometeici e multidimensionali volta per volta concordati tra i diversi soggetti partecipanti all'atto comunicazionale. Come scriveva, nel 2001, Elena Esposito, il modello di organizzazione del sapere proposto dalla biblioteca è obsoleto perché «orientato all'ideale di una conoscenza unitaria e convergente, mentre si impone piuttosto un approccio performativo, rivolto alla capacità di mantenere delle connessioni piuttosto che dei contenuti»<sup>37</sup>. Pochi anni dopo si sarebbe sostenuto che «le biblioteche digitali sono conversazioni»<sup>38</sup> e, infine, sarebbe arrivato David Weinberger a spiegarci che il mondo è bello perché «everything is miscellaneous»<sup>39</sup>.

34 Lewis M. Branscomb, *Information: the ultimate frontier*, «Science», 203 (1979), n. 4376, p. 143-147.

35 Armando Petrucci, *Prima lezione di paleografia*. Roma-Bari: Laterza, 2002, p. 123.

36 La bibliografia sul tema della fine delle biblioteche e degli archivi è sterminata. Per limitarci alle biblioteche si potrebbe partire dalla considerazione contenuta in un celebre articolo di Licklider («It seems reasonable to envision, for a time 10 or 15 years hence, a 'thinking center' that will incorporate the functions of present-day libraries», Joseph Carl Robnett Licklider, *Man-Computer Symbiosis*, «IRE Transactions on Human Factors in Electronics», HFE-1 (1960), p. 4-11, <<http://groups.csail.mit.edu/medg/people/psz/Licklider.html>>), per poi passare a James Thompson, *The end of libraries*, «The Electronic Library», 1 (1983), n. 4, p. 245-255, DOI: 10.1108/ebo44603. Per una visione ad ampio spettro sul dibattito recente attorno al tema dell'obsolescenza delle biblioteche rimandiamo a Michael A. Keller; Victoria A. Reich; Andrew C. Herkovic, *What is a library anymore, anyway?*, «First Monday», 8 (2003), n. 5, <<http://firstmonday.org/ojs/index.php/fm/article/view/1053/973>>, DOI: 10.5210/fm.v8i5.1053.

37 Elena Esposito, *La memoria sociale*. Roma-Bari: Laterza, 2001, p. 224.

38 *Manifesto per le biblioteche digitali*, AIB – Gruppo di studio sulle biblioteche digitali, 2005, <<http://www.aib.it/aib/cg/gbdigdo5a.htm3>>.

39 David Weinberger, *Elogio del disordine*. Milano: Rizzoli, 2010.

Ora, in linea di massima, si può senz'altro convenire sul fatto che la definizione del reticolo semantico in grado di collegare fra loro i vari documenti rappresenti la più squisita dimensione euristica dell'attività di mediazione informativa: del resto è su questo assunto che si statuisce ogni teoria bibliografica. Tuttavia, ciò che si rischia di perdere di vista assumendo per buone le posizioni innovative (presunte tali, invece) di chi esalta unicamente la dimensione sintattica del sapere, è il ruolo essenziale costituito dalla conformazione dei territori che si vorrebbero connettere con i ponti della comunicazione. Ritengo che, quando trent'anni fa, nell'ultimo capitolo de *La bibliografia: storia di una tradizione*, Luigi Balsamo affermava «l'insufficienza, per la domanda di informazione attuale, della dimensione bibliografica»<sup>40</sup>, non intendesse di certo gettare via, insieme a un approccio 'librocentrico' ritenuto evidentemente obsoleto rispetto ai problemi della mediazione informativa, anche il concetto di documento. Eppure, poco tempo dopo, due informatici statunitensi avrebbero sostenuto che

We must now create not only content, but behavior and relationships. That means new tools and new techniques.

The very way we use language will change. Composing an information element that will be dynamically combined with others to create an individualized just-in-time instruction for one person doing one task, is very different from writing a traditional manual. The document is dead. Everything we knew is wrong. It is time to begin<sup>41</sup>.

Perché, chiediamoci, la nefasta profezia sulla morte del documento non si è avverata? Perché, al contrario, in questo lasso di tempo si sono cercate con insistenza soluzioni per garantire, anche nel contesto digitale, quelle condizioni di validazione autoriale, di datazione, di integrità ereditate dal mondo dei documenti analogici? Insomma: perché «pronosticato come velocità, rumore, potenza, prevalenza della immagine e della parola, il futuro si realizza come silenzio, scrittura archivio»<sup>42</sup>?

La risposta è che la costruzione della realtà sociale si fonda sui documenti (in particolar modo quelli 'forti', dichiarativi), e cioè su quei particolari oggetti sociali che trasformano la volatilità dell'informazione in aggregazioni di dati sufficientemente stabili da poterci, per l'appunto, costruire sopra qualcosa<sup>43</sup>. Quando affermo 'sufficientemente stabili', beninteso, non intendo documenti graniticamente costituiti una volta per tutte: come ha osservato Robert Darnton, «abbiamo a che fare non con documenti fissi e immutabili, ma con testi plurali e mutevoli»<sup>44</sup> e, tuttavia, è indubbio che quel 'physical embodiment' (per usare la terminologia di FRBR) costituito dall'azione di registrazione del dato su un supporto fisico, così come non può essere considerata sempre alla stessa stregua sul piano della prassi, allo stesso modo

<sup>40</sup> Luigi Balsamo, *La bibliografia: storia di una tradizione*. Firenze: Sansoni, 1984, p. 143.

<sup>41</sup> Mark Baker; Carol Miksik, *The document is dead*. In: *Proceedings of IPCC '96. Communication on the Fast Track, Saratoga Springs, NY, 18-20 September 1996*. New York: IEEE, p. 61, DOI: 10.1109/IPCC.1996.552581.

<sup>42</sup> M. Ferraris, *Documentalità* cit., p. 203.

<sup>43</sup> «L'informazione nuda, come il sorriso del gatto del Cheshire, non può essere manipolata o recuperata. Solo quando è circostanziata, nel senso che è confezionata in un documento, elettronico o di altro tipo, diventa accessibile, e solo allora può essere oggetto di una descrizione bibliografica» (E. Sveinonius, *Il fondamento intellettuale dell'organizzazione dell'informazione* cit., p. 151).

<sup>44</sup> Robert Darnton, *Il futuro del libro*. Milano: Adelphi, 2011, p. 51.

non produce documenti considerabili con la medesima valenza per la vita sociale. Essa, almeno fino a oggi, impone l'esistenza (e quindi la definizione) di precise forme di autorialità, di certificazione e di integrità per determinate – e fondamentali – categorie documentarie<sup>45</sup>.

Questa stabilità è frutto di regole, di apparati e di autorità che hanno il preciso onere di creare quelle condizioni minime di convivenza senza le quali nessun consenso sociale, ma in particolare il nostro, caratterizzato da una complessità senza pari, potrebbe reggersi in piedi.

Provate a immaginare un mondo, come quello ipotizzato da Weinberger, basato sulla continua riorganizzazione di frammenti di documenti (le foglioline del grande albero) da parte degli utenti<sup>46</sup>. Un mondo dove ogni principio di autorità è tendenzialmente escluso perché ognuno (per dote naturale?) è in grado di misurare la certezza e la serietà delle informazioni di cui entra in possesso<sup>47</sup>. Un mondo di informazioni continuamente re-interpretabili in quanto totalmente decontestualizzate<sup>48</sup>. Ebbene, se le cose mutassero di segno in tal fatta, il nostro mondo attuale si affloscerebbe di colpo su sé stesso: oggetti sociali quali 'norma di legge', 'atto di compravendita', 'protocollo sanitario', 'articolo scientifico' non avrebbero più alcuna specifica caratteristica e funzione rispetto a un coacervo informativo senza riferimenti di autorità o di provenienza<sup>49</sup>.

Potremmo sopravvivere? Certo che sì, lascio tuttavia all'immaginazione del lettore le condizioni di tale sopravvivenza: le regole del contratto sociale sarebbero da riscrivere daccapo. Se anche un autore come Steven Johnson, in genere così fiducioso nella rivoluzione democratica apportata dalle reti sociali, è disposto ad ammettere che in taluni casi il processo decisionale *bottom-up* non possa rappresentare la soluzione migliore rispetto a una decisione di segno opposto<sup>50</sup>, forse dovremmo ricavarne qualche motivo di apprensione nei confronti dei rischi di un mondo popolato in prevalenza da documenti 'deboli'.

45 «The stability and identifiability of the object world in which human activities are usually embedded are key to the forms of experience such a world sustains. The instrumentation of means–ends sequences, the attribution of cause–effect relationships and sense–making in general are essentially supported by the stability of the tools and objects on which actors draw upon. In this respect, the malleable and transfigurable character of digital objects undermines basic facts of human experience and may end up constructing a less accountable environment» (Jannis Kallinikos; Aleksi Aaltonen; Attila Marton, *A theory of digital objects*, «First Monday», 15 (2010), n. 6, <<http://firstmonday.org/ojs/index.php/fm/article/view/3033/2564>>, DOI: 10.5210/fm.v15i6.3033).

46 D. Weinberger, *Elogio del disordine* cit., p. 334.

47 *Ivi*, p. 317.

48 *Ivi*, p. 235.

49 E, su Weinberger, mi fermo qui. Per una critica più puntuale e dettagliata nei confronti di un approccio che, portato alle sue estreme conseguenze, «rischia di condurre a decisioni molto pericolose sul fronte educativo e culturale», rimando a Riccardo Ridi, *Apocalittici e integrati del web: internet ci rende stupidi o intelligenti?*, «AIB Studi», 53 (2013), n. 1, p. 135-142, <<http://aibstudi.aib.it/article/view/8783>>, DOI: 10.2426/aibstudi-8783.

50 Si vedano le dichiarazioni rilasciate nell'intervista a firma di Marco Magrini, *Il network dei pari ci salverà*, «Nòva», supplemento de «Il Sole 24 Ore», 12 maggio 2013, p. 12.

### Osservazioni conclusive

Che l'invenzione e lo sviluppo delle tecnologie digitali abbiano moltiplicato la complessità del mondo dei documenti è un fatto difficilmente contestabile, ed è per questo motivo che i Linked Data rappresentano una possibilità di enorme portata per rendere più efficace la fruizione di informazioni potenzialmente utili. Abbiamo però una necessità, che consiste nel fare emergere, all'interno di questa innovativa dimensione connettiva, le differenti caratteristiche che costituiscono il panorama documentario sottostante, sul quale essa si appoggia. Come ci ricorda Birger Hjørland:

A document has a history, one or more authors or producers, a connection to other documents, and so on. All this is very well known and understood in many areas in the humanities, where there are disciplines such as the history of literature, criticism of documents (including films), and source criticism in history, but often less well understood in technological fields<sup>51</sup>.

La storia, l'autorialità, la provenienza di un documento sono elementi fondamentali per definire la qualità di una fonte e, quindi, per una valutazione di merito in relazione all'affidabilità dell'informazione recuperata. Per questo motivo il fatto che, fino a tempi relativamente recenti, la riflessione attorno ai Linked Data e il loro sviluppo sia stata portata avanti all'interno di un'accolita di esperti provenienti in gran parte dal mondo delle tecnologie informatiche<sup>52</sup>, non ha certamente giovato non solo a una loro più ampia diffusione, ma anche a una presa di coscienza dei loro aspetti critici sia sul piano teorico sia su quello applicativo.

Oggi, fortunatamente, si avvertono segni di cambiamento. Come si è provato a delineare in questo scritto, si stanno moltiplicando le istanze relative a un miglioramento dell'affidabilità delle prestazioni dei Linked Data, un miglioramento che necessariamente passa anche attraverso un rapporto più articolato e complesso con le fonti dei dati. Ed è su questo punto che, a mio avviso, le biblioteche hanno un ruolo fondamentale da giocare, da un lato perché sono fra le poche istituzioni che interpretano con vocazione primigenia l'idea di rendere aperti all'uso libero e gratuito della collettività i propri archivi e le notizie in essi contenute<sup>53</sup>, dall'altro perché la cultura della standardizzazione descrittiva e dell'interoperabilità fra i sistemi è da tempo entrata a far parte in modo sostanziale delle procedure biblioteconomiche di organizzazione e gestione dei sistemi catalografici. Le biblioteche, a differenza della maggior parte di coloro che inseriscono documenti in rete, «hanno sempre

**51** Birger Hjørland, *Documents, memory institutions and Information Science*, «Journal of Documentation», 56 (2000), n. 1, p. 35, DOI: 10.1108/EUM000000007107.

**52** «Particularly when compared to web 2.0 applications, linked data can seem rather inaccessible; anyone can create a Twitter account or promote user tagging or even contribute to mashups, but the world of linked data, for the moment, remains firmly in the hands of the experts» (Gillian Byrne; Lisa Goddard, *The strongest link: libraries and Linked Data*, «D-Lib Magazine», 16 (2010), n. 11-12, <<http://www.dlib.org/dlib/november10/byrne/11byrne.html>>, DOI: 10.1045/november2010-byrne).

**53** Un obbligo, quello della messa a disposizione della collettività dei dati aperti, che – secondo Giovanna De Minico – deriva direttamente dall'art. 97 della Costituzione, «in quanto se il dovere di trasparenza impone all'amministrazione la visibilità dei suoi percorsi decisionali, essa prescriverà anche l'esibizione dei risultati dell'agire pubblico: i dati» (Giovanna De Minico, *Nella Pa l'obbligo di open data deriva dalla Costituzione*, «Il Sole 24 Ore», 1 giugno 2014, p. 15).

prodotto dati di qualità in record bibliografici e di autorità fortemente strutturati»<sup>54</sup>, quindi esse hanno sempre ragionato, a livello logico, su ciò che ci si aspetta dal Web semantico. Quello che è mancato fino a ora, semmai, è stato un approccio politico, prima ancora che tecnico, in grado di favorire la messa a disposizione dei propri dati disponibili in una forma utilizzabile da altri soggetti fuori dall'ambito bibliotecario. Adesso, finalmente, si può cambiare, tuttavia questo cambiamento non deve essere subito ma accompagnato verso una migliore definizione, affinché quelle 'invarianti concettuali' della disciplina bibliografica, così definite da Maurizio Vivarelli<sup>55</sup>, possano contribuire alla creazione di legami realmente significativi tra i documenti, a servizio delle capacità euristiche e delle riflessioni critiche dei lettori.

Articolo proposto il 10 giugno 2014 e accettato il 5 settembre 2014.

---

**ABSTRACT** AIB studi, vol. 54 n. 2/3 (maggio/dicembre 2014), p. 279-293. DOI 10.2426/aibstudi-10128.

---

ALBERTO SALARELLI, Università degli studi di Parma, Dipartimento di lettere, arti, storia e società, piazzale della Pace 7/A, 43121 Parma, e-mail alberto.salarelli@unipr.it.

#### **Sul perché, anche nel mondo dei Linked Data, non possiamo rinunciare al concetto di documento**

I Linked Data rappresentano una delle forme di utilizzo più interessanti e proficue dei dati aperti, dal momento che offrono la possibilità di correlare i contenuti di differenti archivi digitali realizzando strumenti informativi estremamente innovativi. Tuttavia comportano anche alcuni problemi, fra i quali spiccano la provenienza dei dati e la loro granularità. Per far fronte a questi problemi è importante non abbandonare il concetto di documento, tanto più cruciale proprio nel momento in cui la logica della decostruzione sottesa ai Linked Data sembra porlo in secondo piano. Solo mantenendo in vita il concetto di documento, infatti, è possibile contenere la "liquidità" dell'informazione in strutture che possiedono quegli elementi di stabilità caratteristici (anche se non esclusivi) della tradizione bibliografica, e che sono di fondamentale importanza per la costruzione della realtà sociale.

#### **Why we can't give up the concept of document in the era of Linked Data**

Linked Data represent one of the most interesting and effective uses of open data, allowing the connection of data hosted by multiple digital archives and thus producing innovative information systems. On the other hand, they involve a few problems such as data sources and coarse data: that's why the more the deconstructive logic implied by Linked Data moves the concept of document to the background, the more we can't give it up, because only by preserving the concept of document we can contain the information "liquidity" into stable structures that are essential for our lives and society.

<sup>54</sup> Mauro Guerrini, *Classificazioni del sapere: web semantico, linked data e ontologie. Il ruolo rinnovato delle biblioteche nella trasmissione della conoscenza registrata*. In: *Noetica vs informatica: le nuove strutture della conoscenza scientifica, Atti del convegno di studi internazionale, Roma, 19-20 novembre 2013*, in corso di stampa.

<sup>55</sup> Maurizio Vivarelli, *Le dimensioni della bibliografia*. Roma: Carocci, 2013, p. 15.