

# Dig *Italia*

Anno VII, Numero 2 - **2012**

ISSN 1972-6201

Rivista del digitale nei beni culturali

ICCU-ROMA

# Comunicazione scientifica: la nuova frontiera

**Donatella Castelli - Paolo Manghi - Costantino Thanos**

ISTI - Consiglio Nazionale delle Ricerche

*Il nuovo paradigma della scienza, basato sull'utilizzo intensivo dei dati, sta provocando rivoluzionari cambiamenti non solo nei modi di condurre le attività di ricerca ma anche nel mondo della comunicazione scientifica, dove i dati stanno diventando "cittadini di prima classe". Si avvera così la visione pionieristica di Jim Gray di un mondo in cui dati scientifici e letteratura siano unificati. Il nuovo ruolo dei dati nella editoria scientifica è reso possibile dalla definizione di nuovi modelli di pubblicazione che concepiscono una pubblicazione come un "pacchetto di unità informative" che può includere testo, dati, immagini, video, audio e pacchetti software collegati fra loro mediante relazioni semantiche. I due pilastri della moderna comunicazione scientifica sono gli Archivi di Dati e le Biblioteche Digitali Scientifiche. Essendo stati ideati e costruiti per supportare fasi complementari del processo di ricerca e pubblicazione, la loro integrazione risulta difficoltosa. Questo articolo discute la necessità di una attività di ricerca avente l'obiettivo di rendere interoperabili tali Archivi e Biblioteche al fine di supportare con efficacia l'integrazione fra dati e pubblicazioni.*

## Il nuovo scenario della comunicazione scientifica

**M**oderni strumenti scientifici, quali telescopi, acceleratori, satelliti, supercomputer, reti di sensori, ecc., stanno producendo enormi quantità di dati scientifici. La disponibilità di questi enormi volumi di dati sta provocando una rivoluzione nel modo di pensare, organizzare e condurre la ricerca scientifica: stiamo entrando nell'era della Scienza dominata dai dati, chiamata anche "Data-Intensive Science". Questa Scienza avrà anche un considerevole impatto sul mondo dell'editoria scientifica, le cui pubblicazioni dovranno permettere di fare citazioni ai dati allo stesso modo con cui si fanno le citazioni bibliografiche. Infatti, sta emergendo la consapevolezza che la citazione dei dati costituisce una pratica fondamentale per riconoscere i dati come un risultato primario della attività scientifica piuttosto che un suo sottoprodotto. I dati sono destinati a diventare "cittadini di prima classe" della moderna comunicazione scientifica.

Si sta così avverando la visione di Jim Gray<sup>1</sup> secondo la quale è necessario creare un mondo nel quale dati e letteratura siano unificati e possano inter-operare fra loro.

<sup>1</sup> Jim Gray, *A transformed scientific method*, in: *The fourth paradigm: data intensive scientific discovery*, a cura di Tony Hey, Stewart Tansley e Kristine Tolle, Redmond, WA: Microsoft, 2009, p. xix-xxxiii.

Alcuni editori hanno già iniziato a produrre riviste scientifiche secondo modelli che permettono di integrare i dati con le pubblicazioni. Lo stesso problema dovrà essere affrontato dalle biblioteche digitali. Ovviamente ci sono barriere alla sostenibilità di questa pratica, dovute principalmente all'enorme volume dei dati e alla grande varietà dei loro formati. Ulteriori difficoltà riguardano la necessità di scegliere esattamente quali dati rendere disponibili inserendoli nella pubblicazione, oppure allegarli in file supplementari, oppure ancora collegarli mediante riferimenti (link).

### **I pilastri della moderna comunicazione scientifica**

I due pilastri della moderna comunicazione scientifica saranno gli Archivi di Dati Disciplinari (Discipline-specific Data Centers) e le Biblioteche Digitali Scientifiche (Research Digital Libraries). Gli archivi di dati svolgono la funzione di deposito dove i ricercatori depositano i dati da loro creati. Le biblioteche digitali svolgono la funzione di deposito e pubblicazione di documenti scientifici, sia nel campo accademico sia nel campo propriamente editoriale. Archivi di Dati e Biblioteche supportano fasi complementari del ciclo di vita dell'informazione scientifica. Infatti, gli archivi sono responsabili per l'archiviazione, gestione e conservazione dei dati, insieme alla creazione di metadati per la descrizione dei *data-set*. Le biblioteche, invece, si concentrano sulle fasi finali, quelle dell'acquisizione e gestione di documenti e della pubblicazione di documenti, attraverso un processo che frequentemente consiste in diversi passi (revisione, aggiornamento).

Attualmente, Archivi di Dati e Biblioteche sono poco integrati, anzi esiste fra loro una vera e propria dicotomia. Le tecnologie, le politiche e le "buone pratiche", che ciascuna istituzione ha adottato, non sono state concepite per rendere interoperabili questi due ambienti. È deludente che entrambe queste organizzazioni siano così poco e male integrate e che nessuna di loro faccia alcun tentativo di adottare gli aspetti di forza dell'altra. Come conseguenza, questa dicotomia ha ostacolato una piena integrazione fra dati e pubblicazioni. Tutto ciò ha dato origine ad un effetto pernicioso per il fatto che malgrado i dati costituiscano un ingrediente molto importante della ricerca, essi non sono trattati come componente primaria della comunicazione scientifica.

La situazione che abbiamo descritto è in conflitto con le esigenze della scienza moderna, che richiede che l'intero ciclo della produzione e diffusione della conoscenza (che consiste nella acquisizione di dati greggi/archiviazione/produzione di metadati/produzione di dati di secondo livello (elaborati)/pubblicazione dei risultati della ricerca) sia supportato in modo integrato. Oggi gli enti pubblici finanziatori della ricerca richiedono che i dati prodotti dalle attività scientifiche siano resi accessibili per permettere la loro condivisione e ri-utilizzo, ed infine il confronto fra risultati della ricerca, possibilmente diversi, ottenuti dagli stessi dati (Principio di "Open Science - Open Data").

L'integrazione fra dati e pubblicazioni può inoltre produrre significativi vantaggi. In particolare le pubblicazioni possono favorire una maggiore reperibilità dei dati, una loro migliore interpretabilità e permettere di assegnare i giusti crediti ai produttori dei dati. A loro volta i dati possono contribuire ad una più profonda comprensione di una pubblicazione scientifica.

È, quindi, necessario che Archivi di Dati Disciplinari (Discipline-specific Data Centers) e Biblioteche Digitali Scientifiche (Research Digital Libraries) siano resi interoperabili per poter supportare con efficacia l'integrazione dei dati con le pubblicazioni.

### **La piramide della pubblicazione dei dati**

In una recente indagine sulle pratiche di integrazione fra dati e pubblicazioni in atto nella editoria scientifica<sup>2</sup> sono stati individuati tre livelli principali di integrazione, illustrati mediante una piramide. La relazione esclusiva fra dati e pubblicazioni diminuisce scendendo dal vertice della piramide verso la sua base [ vedi Fig. 1].

Al primo livello, i dati sono contenuti all'interno della pubblicazione. Si tratta del modello di pubblicazione tradizionale dove l'editore assume tutta la responsabilità della pubblicazione dell'articolo scientifico e dei dati in esso inseriti. Questo modello di pubblicazione rende i dati facilmente reperibili, ma la loro ri-utilizzazione è resa difficoltosa quando risulta impossibile trovare questi dati in modo indipendente dall'articolo. Tale modello, inoltre, non è adatto quando il volume dei dati è tale da non poter rispettare il formato di pubblicazione tradizionale.

Al secondo livello, i dati risiedono in file supplementari che si aggiungono alla pubblicazione. In questo modello, l'autore ha la possibilità di aggiungere alla propria pubblicazione file contenenti dati che per volume (grandi quantità di dati) oppure per formato (immagini, animazione) non sono conformi al formato tradizionale dell'articolo. Una limitazione di questo modello riguarda il fatto che, in presenza di grandi volumi di dati, gli editori possono imporre delle limitazioni al tipo di contenuto da includere nei file di supplemento.

A questo livello di integrazione la conservazione dei dati è responsabilità dell'editore. Al terzo livello, i dati risiedono in archivi disciplinari (discipline-specific data center) pubblici, e sono legati alla pubblicazione mediante collegamenti (link) bi-direzionali. In questo modello, nella pubblicazione si trova inserita una citazione e un collegamento (link) ai dati, i quali però risiedono in un archivio separato. La responsabilità per la conservazione dei dati è dell'amministratore dell'archivio dei dati. L'accesso ai dati, il loro ri-uso e la combinazione di essi con altri dati può avvenire direttamente attraverso i servizi dell'archivio, indipendentemente dalla rivi-

<sup>2</sup> Susan Reilly - Wouter Schallier - Sabine Schrimpf - Eefke Smit - Max Wilkinson, *Report on integration of data and publications*, Opportunities for Data Exchange (ODE), 2011, p. 36-61, <<http://www.alliancepermanentaccess.org/index.php/community/current-projects/ode/outputs/#Integration>>.

sta in cui i dati sono citati. Questo modello tuttavia dipende molto dalla esistenza di meccanismi bi-direzionali persistenti di collegamento (citation/link). All'ultimo livello, i dati risiedono in apparecchiature di memoria di uso personale, non pubblicati e generalmente non reperibili.



Figura 1 . La piramide della pubblicazione dei dati

### Nuovi modelli di pubblicazione

Per rispondere alle nuove esigenze della comunicazione scientifica è stato necessario ampliare il concetto di “pubblicazione” introducendo nuovi modelli che considerano una pubblicazione come un insieme di “unità informative” consistenti in testo, dati, immagini, e animazione, e collegate fra loro mediante relazioni semantiche. Questi modelli rendono possibile l’integrazione fra dati e pubblicazioni al secondo e terzo livello della piramide della pubblicazione. Nella letteratura, sono state proposte quattro principali classi di modelli di pubblicazione: pubblicazioni di oggetti composti, pubblicazioni strutturate, pubblicazioni “live”, e pubblicazioni orientate al riuso sperimentale, tutte descritte nel seguito.

**Pubblicazioni di oggetti compositi (*compound objects*).** Queste pubblicazioni sono oggetti nati digitali (*born-digital*) che connettono mediante relazioni semantiche oggetti/metadati per creare un nuovo oggetto composto con i propri identificatore e metadati. Esempi significativi di tale tipo di pubblicazione sono la *enhanced publication* e il *modular article*:

- *Enhanced Publication (Pubblicazione potenziata)*<sup>3</sup>. Per “pubblicazione potenziata” si intende una pubblicazione dotata anche di dati a supporto dei risultati da essa riportati. Questi dati possono anche includere informazioni di tipo contestuale, oppure di provenienza dei dati. Tutti questi componenti sono esplicitamente collegati all’interno di un singolo pacchetto che costituisce, quindi, un oggetto digitale composito, contenente una varietà di componenti eterogenei collegati fra di loro, eventualmente distribuiti in diverse ubicazioni.
- *Modular Article (Articolo modulare)*<sup>4</sup>. Secondo questo modello, dati, immagini, suoni, simulazione e video fanno parte dell’ambiente di pubblicazione circostante il testo. Tutti questi componenti sono legati insieme e costituiscono un articolo modulare, essendo un modulo definito come una unità informativa concettuale auto-contenuta. Ciascuna unità informativa dovrebbe essere ben definita e dotata di un insieme di metadati ciascuno dei quali ne descrive un aspetto. Lo sviluppo di un articolo come aggregazione di moduli indipendenti ma interagenti fra loro è resa possibile dalla tecnologia Internet.

**Pubblicazioni strutturate.** Per pubblicazione strutturata si intende un oggetto informativo testuale strutturato in ben definite sotto-parti, le quali possono includere sezioni, paragrafi, figure, tabelle, oppure riferimenti, mediante il web, a risorse esterne e applicazioni interattive. Ricerche su questo tipo di pubblicazione sono state iniziate una decade fa, per esempio con l’“OpenDLib data model”<sup>5</sup>, ma sono state riproposte recentemente come modelli di supporto alle pubblicazioni Web 2.0, vedi il progetto “Article of the Future” di Elsevier<sup>6</sup>.

<sup>3</sup> Saskia Woutersen-Windhouwer - Renze Brandsma - Peter Verhaar - Arjan Hogenaar - Maarten Hoogerwerf - Paul Doorenbosch - Eugène Durr - Jens Ludwig - Birgit Schmidt - Barbara Sierman, *Enhanced publications*, a cura di M. Verwooy-Gerritsen, Amsterdam: Amsterdam University Press, 2009.

<sup>4</sup> Joost G. Kircz, *New practices for electronic publishing: new forms of the scientific paper*, «Learned Publishing», 15 (2002), p. 27-32.

<sup>5</sup> Donatella Castelli - Pasquale Pagano, *OpenDLib: A digital library service system*, in: *Proceedings of the 6<sup>th</sup> European conference on research and advanced technology for digital libraries (ECDL'02), Rome, Italy, September 2002*; a cura di Maristella Agosti e Costantino Thanos. Berlin: Springer, 2002, p. 292-307.

<sup>6</sup> Elsevier, “Article of the Future”, <<http://www.articleofthefuture.com/>>.

**Pubblicazioni “Live”.** Questo tipo di pubblicazioni è emerso recentemente nel contesto di infrastrutture per dati scientifici. Si tratta di pubblicazioni testuali (tipicamente rapporti di ricerca) che includono descrizioni di dati, tabelle, istogrammi e statistiche basate su dati aggiornati, che vengono generati al momento di accesso (detti perciò *live*) dalla sottostante infrastruttura. Una pubblicazione *live* può, quindi, essere generata in un dato istante per descrivere lo stato di un scenario applicativo in quell’istante. Esempi di tale tipo di pubblicazioni si possono trovare nelle infrastrutture D4Science e iMarine<sup>7</sup>.

**Pubblicazioni orientate al riuso sperimentale.** Queste pubblicazioni sono simili alle pubblicazioni di oggetti compositi, ma generalmente contengono anche unità informative finalizzate a rendere riutilizzabile il loro contenuto. Esempi di tali tipi di pubblicazioni sono il “pacchetto” di informazioni scientifiche e l’“oggetto di ricerca” :

- *Scientific Information Package (Pacchetto di informazioni scientifiche)*<sup>8</sup>. Questo tipo di pubblicazione incapsula dati greggi, algoritmi, software, testo e metadati associati, fornendo strumenti che permettono ai ricercatori di specificare le singole componenti all’interno del pacchetto: funzioni matematiche, specifiche software, e documenti testuali. Le diverse unità informative devono essere specificate e possono essere inserite sia come riferimenti ad un unico identificatore oppure incorporate come sequenza di bits. Il pacchetto viene rappresentato come un pacchetto PDF.
- *Research Object (Oggetto di ricerca)*<sup>9</sup>. Un oggetto di ricerca è un oggetto composto che esibisce le seguenti proprietà: ripetibilità, riproducibilità, riusabilità, affidabilità, ri-eseguibilità. L’idea che sostiene questo modello di pubblicazione è quella di sostituire i modelli tradizionali di pubblicazione con altri in grado di rappresentare oggetti digitali condivisibili e riutilizzabili.

Infine, sono stati proposti e accettati dalle comunità scientifiche diversi standard per esportare tali pubblicazioni, indipendentemente dal modello di pubblicazione sottostante. Fra le diverse proposte vale la pena di menzionare l’*OAI-ORE*<sup>10</sup>.

<sup>7</sup> Leonardo Candela - Donatella Castelli - Pasquale Pagano - Manuele Simi, *From heterogeneous information spaces to virtual documents*. In: *Digital libraries: implementing strategies and sharing experiences: 8<sup>th</sup> International conference on Asian digital libraries, ICADL 2005, Bangkok, Thailand, 12-15 December, 2005*. Proceedings, a cura di Edward A. Fox, Erich J. Neuhold, Pimrumpai Premssmit, Vilas Wuwongse, p. 11- 22.

<sup>8</sup> Jane Hunter, *Scientific models: a user-oriented approach to the integration of scientific data and digital libraries*, in: *VALA 2006, Melbourne, February 2006*, <<http://www.vala.org.au/vala2006-proceedings/vala2006-session-2-hunter>>.

<sup>9</sup> Sean Bechhofer - David De Roure - Matthew Gamble - Carole Goble - Iain Buchan, *Research objects: towards exchange and reuse of digital knowledge*, in: *Proceedings. The future of the web for collaborative science (FWCS 2010)*, Raleigh, NC, USA, <<http://www.w3.org/wiki/HCLS/WWW2010/Workshop>>.

<sup>10</sup> [www.openarchives.org/ore](http://www.openarchives.org/ore).

OAI-ORE definisce standard per la descrizione e scambio di aggregazioni di risorse del web. Queste aggregazioni, spesso chiamate oggetti digitali composti/compleksi, permettono di combinare risorse distribuite di vario tipo (testo, immagini, dati, video). L'obiettivo di queste aggregazioni è rendere disponibile il loro ricco contenuto a programmi applicativi di supporto ad attività di scrittura, deposito, scambio, visualizzazione, ri-uso e conservazione di risorse informative.

### Citazione di dati

La principale tecnologia per rendere fattibile l'integrazione fra dati e pubblicazioni è quella che permette la citazione dei dati. Per citazione dei dati si intende la pratica di fornire riferimenti ai dati nello stesso modo con cui i ricercatori forniscono riferimenti bibliografici. Attualmente, non esistono standard universali per la citazione dei dati. Le pratiche variano da settore a settore e da archivio da archivio. È stato proposto<sup>11</sup> uno standard composto di cinque elementi atti ad identificare un insieme di dati, o data-set:

- l'autore del data-set
- la data in cui è stato reso pubblico il data-set
- il titolo del data-set
- un Identificatore Globale Unico; si tratta di un breve nome o stringa di caratteri atti a garantire la sua unicità nel dominio dei nomi e ad identificare permanentemente il data set indipendentemente dalla sua locazione fisica (esempi: Life-Science Identifier – LSID, Digital Object Identifier – DOI, Uniform Resource Name – URN).
- una Impronta Numerica Universale (Universal Numeric Fingerprint – UNF); si tratta di stringa di numeri e caratteri, di lunghezza fissa, che riassume il contenuto del data set in modo tale che un cambiamento in qualche parte del data set produca un diverso UNF.

Nel 2009 quindici istituzioni provenienti da dieci paesi hanno costituito una importante organizzazione internazionale, DataCite<sup>12</sup>, con l'obiettivo di rendere i dati visibili ed accessibili e, in particolare, renderli citabili in modo armonizzato, interoperabile e persistente, tale da permettere: un più facile accesso a dati scientifici su Internet; accrescere l'accettazione dei dati quali legittimi e citabili contributi alla pubblicazione scientifica; supportare l'archiviazione di dati in modo da permettere la loro verifica e riutilizzo anche per studi futuri.

DataCite offre, inoltre, una serie di servizi a ricercatori, archivi di dati ed editori scientifici, aiutando i primi a trovare, identificare e citare data set scientifici, for-

<sup>11</sup> Micah Altman - Gary King, *A proposed standard for the scholarly citation of quantitative data*, «D-Lib Magazine».

<sup>12</sup> [www.datacite.org](http://www.datacite.org).



nendo ai secondi identificatori di data set permanenti unitamente a standard e workflow per la pubblicazione dei dati, e, infine, permettendo agli editori di collegare articoli scientifici con i sottostanti dati.

Il mezzo principale usato da dataCite per creare un facile accesso ai data set scientifici è assegnare loro un DOI (digital object identifier). Tuttavia DataCite adotta un approccio “aperto” al processo di assegnazione di identificatori, nel senso che prende in considerazione anche altri sistemi e servizi che permettano di raggiungere i suoi obiettivi. I 15 membri fondatori di DataCite hanno creato e mantengono una infrastruttura per registrare data sets e assegnare loro identificatori unici persistenti. Il successo maggiore conseguito da DataCite è quello di avere coinvolto in questo sforzo tutte le più importanti istituzioni interessate: archivi di dati, biblioteche ed editori. Infatti, DataCite opera nella convinzione che il progresso nella integrazione fra dati e pubblicazioni possa essere ottenuto soltanto combinando i punti di forza di tali istituzioni.

### Stato dell'arte

Sono in fase di sviluppo diversi progetti aventi l'obiettivo di indagare il problema dell'integrazione fra dati e pubblicazioni, individuare soluzioni ed implementarle. Di seguito, descriviamo alcuni di queste iniziative.

*Dryad.* Dryad<sup>13</sup> è un deposito internazionale di dati che supportano pubblicazioni (sottoposte a revisione paritaria) nel settore delle bioscienze applicate. Dryad funziona come deposito di tabelle, fogli (spreadsheets), file piatti e permette agli studiosi di convalidare risultati già pubblicati, esplorare nuove metodologie di analisi, eseguire studi sintetici. Le riviste, attraverso il loro sistema di gestione dei manoscritti, presentano al deposito Dryad i dati relativi agli articoli accettati per la pubblicazione attraverso una “procedura integrata”. Il sistema si interfaccia con Dryad ed invia una notifica che permette a Dryad di creare i record provvisori per registrare i dati relativi ai nuovi manoscritti. L'articolo pubblicato contiene un legame (link) ai dati contenuti nel deposito Dryad, e Dryad a sua volta lega i dati all'articolo.

*Dataverse Network.* Dataverse Network<sup>14</sup> è una applicazione *open source* sviluppata dal MIT-Harvard Data Center nell'ambito di una collaborazione con la Biblioteca dell'Università di Harvard e riguarda principalmente le scienze sociali. Essa permette di pubblicare, ricercare, condividere, estrarre e analizzare dati scientifici e citarli in modo agevole, standardizzato e persistente.

*Pangaea.* Pangaea<sup>15</sup> è un servizio informativo che supporta l'archiviazione, condivisione, e pubblicazione di dati geo-referenziali. Ogni dataset viene registrato e

<sup>13</sup> [www.datadryad.org](http://www.datadryad.org).

<sup>14</sup> [www.thedata.org](http://www.thedata.org).

<sup>15</sup> [www.pangaea.de](http://www.pangaea.de).

reso citabile attraverso un identificatore DOI (Digital Object Identifier).

I dati vengono archiviati come supplemento alle pubblicazioni o come collezioni di dati citabili. Le citazioni vengono rese disponibili attraverso il portale della “German National Library of Science and Technology”. Pangaea mira a collegare dati scientifici dell’ambiente marino ad articoli di riviste scientifiche. Il collegamento fra dati ed articoli deve essere bi-direzionale.

### **Le infrastrutture per la comunicazione scientifica del futuro**

Come già discusso, i due pilastri della moderna comunicazione scientifica saranno gli Archivi di Dati Disciplinari (Discipline-specific Data Centers) e le Biblioteche Digitali Scientifiche (Research Digital Libraries). È stata anche sottolineata la dicotomia esistente fra queste istituzioni, sebbene svolgano ruoli complementari nella comunicazione scientifica. È quindi necessario intraprendere una attività di ricerca per poter costruire un ponte fra questi due pilastri dando soluzione ai numerosi problemi di ordine tecnico, organizzativo, politico ed economico che già stanno mettendo in discussione le attuali pratiche della comunicazione scientifica. L’obiettivo è la futura infrastruttura per la comunicazione scientifica.

Il compito principale di questa infrastruttura deve essere quello di rendere interoperabili gli Archivi di Dati e le Biblioteche Digitali Scientifiche per poter supportare l’intero ciclo di vita della ricerca e pubblicazione dei dati senza soluzioni di continuità.

La futura infrastruttura per la comunicazione scientifica dovrebbe, quindi, permettere l’intrecciarsi dei dati scientifici con gli articoli scientifici in una complessa varietà di modi creando un più ricco collegamento bi-direzionale fra dati e pubblicazioni, dando così origine a nuovi tipi di pubblicazioni (pubblicazioni potenziate). Un articolo scientifico potrebbe trasformarsi in una finestra che permetta allo scienziato non solo di capire un risultato scientifico ma anche di riprodurlo o di estenderlo<sup>16</sup>.

La futura infrastruttura per la comunicazione scientifica dovrebbe anche avere un considerevole impatto sulle modalità di lettura (*reading practices*) in quanto permetterebbe agli scienziati di andare oltre l’articolo scientifico collegandosi molto più efficacemente con i sottostanti dati scientifici e di muoversi fra un articolo ed un altro oppure fra un articolo ed una collezione di dati con maggiore facilità, precisione, e flessibilità<sup>17</sup>.

Per poter costruire la futura infrastruttura per la comunicazione scientifica devono essere risolti diversi problemi di carattere tecnologico, legale e politico che ostaco-

<sup>16</sup> Clifford Lynch, *Jim Gray’s fourth paradigm and the construction of the scientific record*, in: *The fourth paradigm*, cit.

<sup>17</sup> Allen H. Renear – Carole L. Palmer, *Strategic reading, ontologies, and the future of scientific publishing*, «Science», 325 (2009), n. 5942, p. 828-832.

lano l'interoperabilità fra Archivi di Dati e Biblioteche Digitali. Si potranno così creare le condizioni per una permanente simbiosi fra questi due pilastri.

Attualmente sono in corso alcune importanti iniziative aventi come obiettivo di portare al pieno compimento i principi di Accesso Aperto (Open Access) e di Dati Aperti (Open Data). Questo obiettivo viene perseguito attraverso lo sviluppo di infrastrutture informatiche in grado di supportare l'intero processo di ricerca e pubblicazione.

Una iniziativa, che si muove in questa direzione, è il progetto OpenAIREplus (Open Access Infrastructure for Research in Europe)<sup>18</sup> finanziato dalla Commissione Europea. Questo progetto si prefigge di realizzare e rendere operativa una infrastruttura Europea di supporto alla comunicazione scientifica moderna. Questa infrastruttura sarà in grado di collezionare e collegare contenuti residenti in Archivi digitali di pubblicazioni scientifiche (sia che applichino il principio di accesso aperto che non), in Archivi digitali di dati scientifici (sia che applichino il principio di dati aperti che non), unitamente al contenuto di sistemi CRIS (Current Research Information Systems, adibiti alla gestione di informazioni relative al finanziamento e risultati della ricerca da parte dei governi nazionali e della Comunità Europea). In particolare, OpenAIREplus fornisce servizi per: (1) popolare e curare la qualità di questo "grafo" di metadati, il quale abbraccia tutte le discipline scientifiche e tutti i paesi dell'Unione Europea; (2) arricchire questo "grafo" con nuove relazioni semantiche tra pubblicazioni e dati scientifici mediante strumenti d'inferenza automatica; (3) fornire l'accesso aperto al "grafo" via interfacce utente (portali web) e interfacce software (API); (4) sperimentare sistemi per la gestione di "pubblicazioni potenziate" (*enhanced publications*); e (5) fornire metriche e calcolare statistiche per poter misurare l'impatto dell'accesso aperto sulla ricerca pubblica.

L'infrastruttura OpenAIREplus è abilitata dal software D-NET<sup>19</sup>. D-NET realizza un framework di servizi concepito per facilitare la realizzazione di workflow di manipolazione dati mediante la combinazione e orchestrazione di servizi per la gestione dati. D-NET mette a disposizione una vasta gamma di servizi per la collezione, trasformazione, memorizzazione e indicizzazione di dati, ma è progettato per consentire l'integrazione di nuovi servizi e quindi l'estensione del framework a soddisfare sempre nuovi requisiti funzionali.

<sup>18</sup> <http://www.openaire.eu/>.

<sup>19</sup> Paolo Manghi - Marko Mikulicic - Leonardo Candela - Donatella Castelli - Pasquale Pagano, *Realizing and maintaining aggregative digital library systems: D-NET software toolkit and oAlster system*, «D-Lib Magazine», 16 (2010), n.3/4.

*The new Science paradigm, based on data-intensive computing, leads to a new data-centric way of carrying out research activities. The pioneer vision of Jim Gray on the need of unifying all scientific data with the related literature is emerging as a reality and revolutionizing scientific publishing as a whole. In fact, data is becoming a first class citizen of modern scientific communication. This new role of data in the scientific communication is made feasible by the definition of new formal publication models, where a publication is intended as a set of "information units", including text, datasets, images, videos, sound recordings, mathematical models, workflows, presentation materials, and software packages meaningfully connected by relationships. The two pillars of modern scientific communication are discipline-specific Data Centers and Research Digital Libraries. Being realized to implement complementary phases of scientific research and publication process, these are poorly integrated with one another. We envision that research efforts must be undertaken in order to bridge the gap between published ideas and the underlying data, and make the relationship between data and publication more dynamic.*

Per tutti i siti web, l'ultima consultazione è avvenuta nel mese di dicembre 2012.