

# Infrastrutture Digitali Aperte per la Comunicazione Scientifica

**Costantino Thanos**

*Istituto di Scienza e Tecnologie dell'Informazione (ISTI) del Consiglio Nazionale delle Ricerche (CNR)*

*Le caratteristiche della scienza moderna richiedono la creazione di un record scientifico interconnesso, aperto e online. Per realizzare questa visione è necessario sviluppare Infrastrutture Digitali Aperte per la Comunicazione Scientifica (IDACS). L'articolo descrive l'evoluzione del record scientifico e la funzionalità delle future IDACS in grado di supportare nuove prassi di lettura, apprendimento e reperimento dell'informazione scientifica.*

## 1. Introduzione

**S**ta emergendo un nuovo paradigma scientifico caratterizzato da: (a) un calcolo intensivo di dati scientifici reso possibile dalla disponibilità pubblica di enormi volumi di dati (data intensive science); (b) un ricorso a diverse discipline scientifiche per risolvere problemi complessi (*multidisciplinary science*); (c) un accesso aperto ai risultati scientifici prodotti dalla comunità scientifica (*open science*); (d) un crescendo di collaborazioni scientifiche su scala globale rese possibili dalla interconnessione delle varie componenti dell'ecosistema scientifico (*globalism of science*); (e) un crescente utilizzo dei risultati ottenuti dai recenti progressi della tecnologia dell'informazione, delle infrastrutture di calcolo e di Internet a supporto delle investigazioni scientifiche (*science 2.0*).

Questo nuovo paradigma scientifico accompagnato da un processo rivoluzionario di digitalizzazione dell'informazione scientifica sta creando enorme pressione per cambiamenti radicali nelle prassi seguite dagli studiosi. Esso, inoltre, sta inducendo domande di radicale cambiamento e creando nuove aspettative. Infatti, oggi una delle più grandi sfide che gli studiosi devono affrontare è quella di fare il miglior uso possibile della crescente quantità di informazione scientifica su scala globale. Gli studiosi hanno bisogno di trovare la più esauriente e accreditata informazione che riguarda un certo argomento; trovare l'introduzione a un argomento creata da un esperto; condurre un'analisi di prospettiva della letteratura scientifica, per esempio, quali sono stati gli argomenti per confutare un certo articolo; condurre un'analisi di provenienza, per esempio alla ricerca delle origini di un certa idea, ecc. C'è inoltre un diffuso desiderio di poter navigare in uno spazio di dati scientifici allo scopo di scoprire ed estrarre utile conoscenza contenuta in questo spazio.

Una pressione per radicali cambiamenti nelle prassi degli studiosi è causata anche dai rivoluzionari cambiamenti in corso nella comunicazione scientifica:

- i dati scientifici stanno diventando cittadini di prima classe della comunicazione scientifica e come tali devono essere integrati con le pubblicazioni scientifiche allo scopo di supportare ripetibilità, riproducibilità e rianalisi.
- i dati scientifici e le pubblicazioni devono essere in grado di attraversare confini disciplinari e quindi per poter mantenere il contesto interpretativo devono essere rinforzate semanticamente, per esempio collegando termini testuali ad ontologie, terminologie, e vocabolari.
- la pubblicazione della letteratura scientifica avviene sempre più online. La forma digitale dell'articolo permette la definizione di nuovi modelli basati su tecniche di modularità che permettono di superare la tradizionale forma lineare dell'articolo.
- tecnologie avanzate di connessione (linking) permettono l'interconnessione fra insiemi di dati e moduli di articoli in molti e vari modi. Questa possibilità permette la creazione di vari percorsi d'interesse per studiosi e ricercatori.

Lo sviluppo di infrastrutture digitali aperte per la comunicazione scientifica sarà decisivo per rendere fattibili i radicali cambiamenti richiesti.

## 2. Il Record scientifico moderno

Con il termine "record scientifico" s'intende l'insieme di giornali scientifici, letteratura grigia, presentazioni alle conferenze, collezioni di dati scientifici e ogni altra evidenza di supporto ai risultati pubblicati. Il record scientifico moderno è arricchito anche da comunicazioni meno formali, rese possibili dai recenti progressi nella scienza dell'informazione e Internet, come per esempio, blog posting, tweet, presentazioni video, ecc.

Questa documentazione è distribuita fra biblioteche scientifiche, archivi istituzionali, archivi di case editrici, e centri di dati disciplinari (*discipline-specific data centers*). Tuttavia è poco interconnessa e questo costituisce un grande ostacolo a un diretto e pieno coinvolgimento degli studiosi.

I due principali costituenti del record scientifico moderno sono la collezione di dati (dataset) e l'articolo.

### 2.1 Collezione di Dati Scientifici

Non esiste una singola definizione formale del concetto "collezione di dati" (dataset). Informalmente, s'intende una collezione di dati pubblicata e mantenuta da un singolo fornitore, che tratta un certo argomento e che viene originata da un esperimento/osservazione/processo. Nell'ambito dei "Linked Data", per dataset si intende un insieme di triple RDF pubblicate e mantenute da un singolo fornitore.

Vari tipi di relazione possono esistere fra i dati di una collezione:

- Relazione circostanziale: i dati di una collezione possono essere legati dal tempo, dal luogo, dallo strumento, da un'osservazione che li ha prodotti.
- Relazione sintattica: i dati di una collezione, tipicamente, hanno la stessa struttura sintattica.
- Relazione semantica: i dati di una collezione, tipicamente, riguardano lo stesso soggetto.

Una collezione di dati, una volta archiviata, si registra assegnandole un identificatore (DOI). L'identificatore è un nome unico nello spazio dei nomi di un ambiente scientifico interconnesso e supporta un sistema per l'identificazione permanente di collezioni di dati. Gli identificatori devono: (a) identificare inequivocabilmente la collezione; (b) essere globali; (c) essere associati con un servizio di risoluzione che prende come input il DOI di una collezione e ne genera il suo indirizzo. In aggiunta, una collezione deve essere accompagnata da "metadati".

Una prassi emergente nel mondo scientifico riguarda la pubblicazione di collezioni di dati. Per pubblicazione di una collezione di dati s'intende il processo che permette alla comunità scientifica di scoprire, comprendere e fare asserzioni circa la attendibilità dei dati contenuti nella collezione. In aggiunta, essa permette di assegnare un credito accademico ai creatori dei dati.

## 2.2 L'Articolo scientifico moderno

In un mondo scientifico sempre più interconnesso e online la struttura, funzionalità e presentazione dell'articolo scientifico sono destinate a cambiare radicalmente. L'articolo diventerà una finestra attraverso la quale gli studiosi e i ricercatori potranno comprendere un risultato scientifico, riprodurlo o anche estenderlo. Un altro modo di considerare l'articolo è come un'interfaccia mediante la quale autori e lettori interagiscono.

L'articolo scientifico del futuro sarà caratterizzato da:

- disaggregazione che permette di superare la sua forma lineare e presentarlo come una rete di moduli interconnessi in modo significativo mediante relazioni. In sostanza, emergerà un nuovo modello modulare composto da due principali componenti: moduli e relazioni. I moduli sono unità informative autonome contenenti informazioni di vario tipo, per esempio informazioni che riguardano gli aspetti strutturali dell'articolo, oppure l'argomentazione scientifica, per esempio ipotesi, dimostrazioni, risultati ottenuti, ecc. Tali moduli potranno essere localizzati, acceduti e consultati separatamente. Fra i moduli potranno essere stabilite vari tipi di relazione: di causalità, di similitudine, di contraddizione, ecc.
- mutabilità che permette all'autore di aggiornarlo in qualsiasi momento, di farne una revisione cambiandone il contenuto, di espanderlo aggiungendo annotazioni,

commenti, collegamenti ipertestuali, di includere informazione non statica, per esempio, animazione, e suono.

– inglobamento di software che permette di calcolare una formula, oppure di visualizzare i risultati mentre l'articolo viene letto, ecc.

Diversi modelli per descrivere un articolo scientifico sono apparsi nella letteratura. Il nome utilizzato per indicare tutti questi nuovi modelli è "pubblicazione arricchita" (*enhanced publication*). Per pubblicazione arricchita s'intende una combinazione dinamica e indetificabile di oggetti, risultante dall'associare a una pubblicazione elettronica dati scientifici incorporati o remoti, materiale suppletivo (per esempio, contenuto visivo, video streaming, ecc.), metadati, ecc. La pubblicazione arricchita è un concetto ombrella che abbraccia molti e diversi modelli di articolo scientifico.

Una buona prassi sempre più seguita, l'assegnazione di identificatori, DOI agli articoli scientifici, permette di stabilire collegamenti attivi (online) fra articoli citanti e citati. Identificatori possono essere assegnati anche ai singoli moduli di un articolo permettendo così di stabilire connessioni fra moduli di articoli. È inoltre importante la possibilità di arricchire semanticamente articoli e moduli con dei metadati appropriati.

### 3. Record scientifico connesso

Le esigenze della scienza moderna richiedono la creazione di record scientifici disciplinari connessi, capaci di supportare efficacemente l'indagine scientifica. In effetti, i ricercatori e gli scienziati hanno la necessità di poter muoversi da ipotesi a dimostrazioni, da un articolo all'altro, da una collezione dati all'altra, e da un articolo a una collezione dati. Essi devono essere in grado di poter scoprire percorsi e connessioni potenzialmente utili fra i vari costituenti di uno spazio informativo disciplinare. Da un punto di vista concettuale, un record scientifico connesso implica che i suoi singoli costituenti, e cioè collezioni di dati, articoli, moduli di articoli, ecc. costituiscano i nodi di una rete dai quali può essere acceduto dai ricercatori in qualsiasi momento e da qualsiasi luogo.

#### 3.1 Spazi di dati disciplinari connessi

Nuovi strumenti scientifici caratterizzati da un'elevatissima capacità di produrre dati come telescopi, satelliti, acceleratori, super-computer, reti di sensori e programmi di simulazione stanno generando enormi volumi di dati. La disponibilità di questi volumi di dati sta rivoluzionando il modo con cui si conduce la ricerca scientifica e fa emergere un nuovo paradigma scientifico basato sul calcolo intensivo di dati. Una sfida difficile che frequentemente devono affrontare i ricercatori riguarda l'esigenza di condurre ricerche che si basano su varie e diverse collezioni

di dati correlate senza però possedere le necessarie metodiche per la gestione dello spazio di dati scientifici.

È necessario, quindi, sviluppare meccanismi e approcci che permettano di connettere collezioni di dati prodotti da diversi gruppi di ricercatori rendendole condivisibili. Connettere una collezione di dati si riferisce alla capacità di connetterla ad altre collezioni esterne e anche di essere, a sua volta, connessa da altre collezioni. La connessione di collezioni di dati permette la loro condivisione su scala globale. Il processo che rende possibile la connessione dei dati è noto come “pubblicazione dei dati”. Una generalizzazione del concetto di connessione dei dati conduce alla creazione di spazi di dati disciplinari o interdisciplinari connessi. Uno spazio di dati connesso permette ai ricercatori e ai motori di ricerca di esplorarlo seguendo i collegamenti fra le varie collezioni di dati.

La creazione di uno spazio connesso richiede l’abilità di descrivere formalmente e significativamente le collezioni che ne devono far parte. I metadati costituiscono l’informazione descrittiva. Egualmente importante è descrivere/modellare formalmente le relazioni (link) fra le collezioni. Ad esempio, è importante poter modellare la relazione “la collezione B è un’astrazione temporale/spaziale della collezione A”, oppure la relazione “le collezioni A e B sono state create indipendentemente l’una dall’altra ma entrambe riflettono la stessa attività sperimentale”, ecc.

Poter sfruttare a pieno il potenziale degli spazi dati connessi richiede di rendere comprensibile la semantica dei dati contenuti nelle varie ed eterogenee collezioni facenti parte dello spazio connesso. All’uopo si devono collegare alle collezioni specifiche ontologie disciplinari.

Un’iniziativa che realizza il concetto di spazi dati connessi utilizzando tecnologie del Web semantico è la ben nota “Linked Data”.

### 3.2 Spazi di letteratura scientifica connessi

Nel paragrafo 2.2 abbiamo descritto la natura reticolare del futuro articolo scientifico. Disaggregare l’articolo scientifico in vari moduli semanticamente connessi permette di organizzare l’informazione scientifica come una moltitudine di moduli interconnessi. L’organizzazione reticolare permette, per esempio, di rispondere a domande quali: qual è la dimostrazione a sostegno di questa tesi; quali sono i fondamenti concettuali di questa tesi?; chi ha contestato questa tesi e con quali argomenti?; qual’ è l’origine di una certa tesi?; ci sono inconsistenze in questa scuola di pensiero? ecc.

Nella linguistica computazionale è stata studiata estensivamente la struttura della dissertazione scientifica come pure la relazione fra semantica della dissertazione e disaggregazione. È stato evidenziato che la disaggregazione di un articolo scientifico in moduli non deve basarsi esclusivamente sulla semantica dei moduli ma deve anche considerare la natura retorica delle relazioni fra moduli.

In sostanza, disaggregare un articolo scientifico in diversi moduli è una difficile

operazione concettuale. In letteratura sono stati proposti vari modelli di relazione fra moduli, per esempio, prova/confuta, supporta/contraddice, suggerisce/suggerisce di non, ecc.

Una generalizzazione della struttura reticolare del moderno articolo scientifico conduce alla creazione di spazi di letteratura scientifica disciplinare/ interdisciplinare connessi. Di conseguenza, un contributo scientifico diventa un nodo, rigorosamente connesso, di uno spazio di letteratura scientifica.

Creare uno spazio di letteratura scientifica connesso richiede la capacità di poter descrivere formalmente e significativamente i vari moduli degli articoli come pure le relazioni fra essi. È, quindi, necessario definire modelli di metadati, ricchi dal punto di vista semantico, per descrivere i moduli e le relazioni fra loro.

In aggiunta, è importante il concetto di “identificatore di modulo” (DOI) come un meccanismo che permetta di riferire un modulo di articolo.

### 3.3 Interconnessione fra spazi di dati e spazi di letteratura

Connettere la pubblicazione scientifica con i sottostanti dati che supportano i risultati da essa riportati sta diventando un’esigenza importante della scienza moderna. Infatti, l’interconnessione fra pubblicazione e dati permette ai ricercatori, mentre leggono un articolo, di accedere ai sottostanti dati e poter rifare un’analisi, riprodurre o verificare i risultati riportati nell’articolo; oppure mentre esaminano una collezione di dati ritornare indietro e trovare tutta la letteratura che riguarda quella collezione. Le case editrici cominciano ad accogliere l’opportunità di integrare dati con articoli scientifici nonostante diversi problemi di natura tecnologica, per esempio l’enorme volume dei dati e la varietà dei loro formati creano grandi ostacoli alla sostenibilità di questa pratica.

Si possono realizzare parecchi livelli d’integrazione fra dati e pubblicazioni:

- Una stretta integrazione si realizza quando i dati sono contenuti all’interno dell’articolo. Questo livello d’integrazione non è praticabile quando il volume dei dati da incorporare all’interno dell’articolo è grande.
- Un livello d’integrazione meno stretta si realizza quando i dati risiedono in archivi (file) supplementari associati all’articolo. Con questo livello d’integrazione nascono alcuni problemi che riguardano principalmente la conservazione dei file supplementari e la difficoltà di accedere ai dati indipendentemente dall’articolo.
- Infine, un livello d’integrazione debole si realizza quando i dati risiedono in Depositi di dati istituzionali oppure in Centri di dati (Data Center) disciplinari. In questo caso, la connessione articolo-dati avviene mediante collegamenti (link) bi-direzionali. La conservazione dei dati è responsabilità dell’amministratore del Centro/Deposito dati. Questo livello d’integrazione rende i dati più facilmente rintracciabili e permette di accedere e di riutilizzare i dati separatamente dall’articolo. Realizzare questo livello d’integrazione richiede l’esistenza di meccanismi permanenti i quali consentono una citazione bi-direzionale. È evidente che nell’era dei

“Big Data” soltanto questo livello d’integrazione è praticabile.

La possibilità di connettere pubblicazioni con collezioni di dati permette di connettere spazi di letteratura scientifica con spazi di dati scientifici creando, così, un record scientifico connesso. Un “record scientifico connesso” accrescerà la velocità dell’informazione scientifica e contribuirà a incrementare la produttività scientifica.

### 3.4 Potenzamento semantico del record scientifico

Nell’ambito di una scienza multidisciplinare, la comunicazione scientifica deve attraversare confini disciplinari e quindi necessita di un potenziamento semantico per render il testo e i dati intelligibili ad una vasta audience composta da specialisti in diverse discipline scientifiche. Questa necessità ha motivato l’emergere della “pubblicazione semantica” (*semantic publishing*). Per pubblicazione semantica s’intende il potenziamento del significato di un articolo scientifico online permettendo una automatica disambiguazione e una definizione semantica di termini specialistici. Questo può realizzarsi connettendo direttamente i termini a ontologie disciplinari e/o terminologie standardizzate, oppure connettendosi direttamente a tutti i riferimenti contenuti nell’articolo. In un mondo scientifico multidisciplinare e interconnesso, quando l’informazione scientifica attraversa confini semantici, è necessario mantenere il suo ambito interpretativo. Questo si realizza allineando le ontologie disciplinari che supportano l’accesso a collezioni di dati distribuite.

Contrassegnare (mark-up) semanticamente il testo è una tecnologia che facilita la sua comprensione. Vengono attualmente sviluppati strumenti avanzati per l’estrazione automatica d’istanze testuali e di elaborazione del linguaggio naturale che permettono di riconoscere istanze testuali e di connetterle automaticamente a ontologie specifiche di un certo settore scientifico. Un ulteriore potenziamento semantico del testo si può ottenere connettendolo sapientemente a siti Web, o ad archivi di presentazioni e commenti fatti da terzi.

Potenziare semanticamente una pubblicazione richiede di arricchirla con appropriati metadati che si prestano a un’elaborazione e analisi automatica. Il potenziamento semantico accresce il valore di una pubblicazione perché contribuisce ad accrescerne l’intelligibilità e facilitare l’estrazione della conoscenza in essa contenuta.

## 4. Innovazioni nelle prassi di indagine scientifica

Stiamo entrando in una nuova era caratterizzata dalla disponibilità di enormi collezioni di articoli scientifici. Si stima che almeno 114 milioni di documenti scientifici sono accessibili sul Web. Di questi, si stima che almeno 27 milioni sono ad accesso libero “open access”. Inoltre, come già accennato, enormi volumi di dati scientifici si producono dalle varie attività di ricerca. Questa esplosione d’informazione scientifica da un lato rende difficile il compito della ricerca delle informazioni di cui necessita

uno studioso e dall'altro lato estende in modo significativo gli obiettivi della ricerca. Noi pensiamo che l'avanzamento tecnologico nei campi della semantica, del modellamento dell'informazione, dei motori di ricerca e della connessione delle informazioni (linking) contribuirà all'emergere di nuove prassi da parte degli studiosi permettendo loro di affrontare con successo le sfide dovute al diluvio informativo. Di seguito descriviamo alcune delle attese innovazioni nelle prassi degli studiosi che le nuove tecnologie renderanno possibili.

#### 4.1 Nuove prassi nel reperimento dell'informazione scientifica (*Information seeking*)

La disponibilità di enormi volumi d'informazione scientifica è destinata a produrre un cambiamento nel tradizionale metodo scientifico: da una ricerca guidata da ipotesi (*hypothesis-driven*) a una ricerca guidata da connessioni e correlazioni che si trovano fra i diversi tipi di risorse informative. Scoprire informazione scientifica sconosciuta, potenzialmente utile, richiede la scoperta di percorsi all'interno di un record scientifico connesso. Per percorso s'intende un cammino attraverso un certo numero di moduli di articolo e collezioni di dati, significativamente interconnessi da relazioni materializzate da collegamenti (link).

Le relazioni possono essere espresse da:

- equazioni matematiche quando connettono campi numerici di due collezioni di dati;
- connessioni logiche fra moduli di articoli e (sotto)collezioni di dati; e
- connessioni semantiche/retoriche fra moduli di articoli.

Un percorso forma una catena causale, e il processo di scoperta può prendere forme complesse. Il percorso è specificato in un linguaggio di alto livello e costituisce l'input ad un motore di ricerca.

Data l'enorme dimensione del record scientifico moderno, un motore di ricerca può assistere meglio uno studioso nel trovare percorsi interessanti se è in grado di comprendere l'intensione dello studioso dietro ad una richiesta di percorso. Permettere a un motore di ricerca di comprendere l'intento di uno studioso richiede di definire precisamente la semantica della rappresentazione dell'intensione dello studioso; e di specificare precisamente i confini semantici del dominio dell'intensione.

La rappresentazione dell'intento è inclusa nella specificazione del percorso inviata a un motore di ricerca.

##### 4.1.1 Un nuovo paradigma di reperimento dell'informazione: esplorazione

Nell'era caratterizzata da un diluvio di informazioni scientifiche, le tradizionali tecnologie di reperimento dell'informazione (*information retrieval*) e elaborazione delle interrogazioni (*query processing*) non sono adeguate.

Stanno emergendo nuovi paradigmi di reperimento delle informazioni che permetteranno di:

- navigare all'interno di un record scientifico connesso seguendo percorsi idonei;
- esplorare il record scientifico connesso alla ricerca di percorsi di interesse;
- muoversi rapidamente all'interno del record scientifico e di individuare informazioni rilevanti in movimento.

L'esplorazione di un grande spazio informativo, supportato da adeguati sistemi software, è un paradigma emergente di reperimento dell'informazione ed è estremamente utile agli studiosi che cercano percorsi interessanti senza però conoscere a priori l'oggetto della ricerca.

L'esplorazione può essere condotta in due modi:

- “Interrogare Navigando” (*Navigational Querying*): in questo stile di navigazione, l'esplorazione si conduce avendo per obiettivo il raggiungimento di un preciso nodo dello spazio informativo connesso; la parte critica riguarda il processo di selezione del “nodo successivo” nella navigazione. Per migliorare l'efficacia di questo processo è importante essere a conoscenza della struttura dello spazio informativo.
- “Scorrere Navigando” (*Navigational Browsing*): in questo stile di navigazione si visita un certo numero di nodi in modo casuale nella speranza di trovare qualcosa d'interessante; non si è in grado di formulare una precisa richiesta ma solo di riconoscere informazione rilevante.

Lo stile “*scorrere*” si distingue dallo stile “*interrogare*” per l'assenza di un preciso obiettivo nella mente dello studioso. La distinzione, quindi, fra questi due modi di esplorazione è determinata dallo stato cognitivo dello studioso.

#### 4.1.2 Mappe tematiche (*Topic maps*)

Una tecnologia che aiuta gli studiosi a trovare utili informazioni in uno spazio di informazioni scientifiche connesso è quella relativa alle Mappe tematiche. Una Mappa tematica è un modo di rappresentare conoscenza connessa in termini di “temi, associazioni e occorrenze”:

- i temi sono una rappresentazione di un concetto che può essere capita ed elaborata da un computer. Essi possono rappresentare qualsiasi concetto; per esempio nell'ambito scientifico un tema può essere un articolo, una collezione dati, un autore, un esperimento, un algoritmo, ecc.
- le associazioni rappresentano relazioni fra temi: per esempio, un articolo che suggerisce una tesi può essere connesso ad un altro articolo che supporta questa tesi; un algoritmo di estrazione dati (*data mining*) può essere connesso alla collezione dati sulla quale esso opera, ecc.
- le occorrenze rappresentano risorse informative rilevanti per un tema.

Temi, associazioni e occorrenze sono tipizzati; i vari tipi sono definiti dal creatore della Mappa tematica. Le definizioni dei tipi permessi costituiscono l'ontologia della Mappa tematica. Ciascun tema che partecipa a una associazione svolge un ruolo che viene definito dal tipo dell'associazione.

Le Mappe tematiche permettono di sviluppare un punto di vista logico svelando connessioni e aiutando gli studiosi, per esempio, a scoprire la discendenza di un'idea, oppure come idee individuali possono contribuire a formare concetti/idee più complesse.

Le Mappe tematiche possono offrire una panoramica di alto livello sul dominio di conoscenza contenuto in uno spazio informativo. Esse possono servire non soltanto come guida per localizzare risorse utili allo studioso ma anche come un mezzo che permette allo studioso di organizzare la sua conoscenza in modo strutturato, così da permettere ai non esperti di afferrare i concetti basilari prima di immergersi nelle risorse informative che forniscono più dettagli.

Le Mappe tematiche possono anche essere considerate come un strato di metadati semantico per indicizzare risorse informative eterogenee e distribuite.

Infine un altro modo di guardare alle Mappe tematiche è quello di considerarle come spazi virtuali dove studiosi e studenti possono esplorare quello che conoscono e quello che non conoscono.

Una Mappa tematica può essere creata sia da un autore umano che automaticamente. Essa rende l'informazione rintracciabile assegnando a ciascun concetto dello spazio informativo un identificatore e fornendo diversi percorsi di navigazione nello spazio informativo connesso.

In sostanza, le Mappe tematiche possono essere utilizzate per creare viste semantiche personalizzate di un record scientifico interconnesso che possono soddisfare esigenze di lettura, apprendimento e ricerca degli studiosi.

La standardizzazione delle Mappe tematiche avviene sotto l'ombrello del "ISO/IEC JTC1/SC34/WG3 Committee (ISO/IEC Joint Technical Committee 1, Subcommittee 34, Working Group 3 - Document description and processing languages - Information Association).

#### **4.2 Nuove prassi di lettura (*New reading practices*)**

La creazione di spazi connessi d'informazione scientifica, accompagnata da un crescente aumento di pubblicazioni scientifiche e dal conseguente tempo limitato da dedicare alla lettura, sta modificando le prassi di lettura, principalmente, in due direzioni:

"Lettura focalizzata": data l'abbondanza di pubblicazioni disponibili, gli studiosi evitano di leggere le pubblicazioni meno rilevanti e meno recenti, dedicandosi soltanto alla lettura delle parti rilevanti di un piccolo nucleo di articoli. Essi tendono, quindi, a restringere lo spazio della letteratura da scorrere (*tuned vision*). A questo scopo è usato un certo numero di indicatori di rilevanza di un articolo quali indi-

cizzazione e citazioni, estratti come surrogati di articoli completi, e “social networks” di colleghi come servizio di suggerimento personale.

“Lettura orizzontale/esplorazione”: un’altra forma di lettura consiste nel navigare nello spazio di letteratura connesso per trovare non uno specifico articolo, ma piuttosto esplorare parte dello spazio scrutando porzioni di vari articoli, cioè lettura orizzontale. In sostanza, la lettura orizzontale consiste nell’esplorare grandi quantità pubblicazioni.

“Lettura strategica”: lettura focalizzata e lettura orizzontale conducono ad un nuovo paradigma di lettura: “lettura strategica”, intesa come lettura dei moduli di un articolo in ordine di rilevanza piuttosto che in ordine narrativo.

#### 4.3 Nuove prassi di apprendimento (*New learning practices*)

Uno spazio d’informazione scientifica connesso ha il potenziale per accrescere la capacità di apprendimento degli studiosi in quanto supporta i loro processi cognitivi permettendo la creazione di associazioni fra concetti. Questo implica la capacità di creare significato/senso stabilendo percorsi, connessioni e relazioni. Uno spazio connesso consente a uno studioso di costruire percorsi di apprendimento che permettono di acquisire nuova conoscenza. Seguire percorsi di apprendimento pre-costruiti rende possibile: (a) l’esplorazione e confronto di idee; (b) identificare divergenze/disaccordi; (c) tenere traccia dei contributi di un ricercatore; (d) risalire all’origine di un’idea ecc.

In aggiunta, uno spazio d’informazione connesso facilita lo stabilire di connessioni significative fra i suoi diversi componenti: interpretazione, causalità, predizione, argomentazione (a sostegno/a contestazione), ecc.

### 5. Infrastrutture digitali aperte per la comunicazione scientifica

Si definisce *Infrastruttura digitale aperta per la comunicazione scientifica* un ambiente interconnesso e amministrato che supporta le attività condotte dagli studiosi. Questa infrastruttura rende possibile la creazione di un ecosistema aperto, composto di Biblioteche digitali, Archivi istituzionali, Archivi di case editrici, e Centri di dati disciplinari. Essa rende possibile l’interoperazione fra dati e letteratura creando un record scientifico aperto e distribuito e costituisce, quindi, un ambito favorevole per l’indagine scientifica, per l’analisi avanzata della letteratura, e per nuove prassi basate su tecnologie semantiche e Internet.

In particolare, queste Infrastrutture permettono:

– la creazione di un contesto che supporta l’interconnessione dell’informazione scientifica mediante:

- un insieme di servizi che permettono la creazione di spazi disciplinari connessi, sia di dati che di letteratura, e l’interconnessione di spazi dati con spazi di letteratura.

- un insieme di “connettori”, cioè moduli software, i quali, sfruttando la conoscenza codificata nei metadati di dati e di articoli, creano relazioni fra collezioni di dati, moduli di articolo e fra articoli e collezioni di dati. Diversi tipi di connettori creano diversi tipi di relazione, per esempio, relazioni di causalità, di similitudine, ecc.
  - registri di metadati per moduli di articoli.
  - metadati per connettori.
- la creazione di un contesto di mediazione mediante:
- un insieme di servizi che rendono i contenuti degli archivi di dati e di letteratura rintracciabili, intelligibili e utilizzabili.
  - un insieme di “mediatori”, cioè moduli software, i quali, sfruttando la conoscenza codificata nei metadati permettono la realizzazione di servizi di mediazione. I mediatori realizzano diversi tipi di mediazione, per esempio, mappatura, verifica di consistenza logica, corrispondenza di modelli, ecc.
  - dizionari di dati, ontologie e terminologie disciplinari.
- la creazione di un contesto di navigazione, il quale offre agli studiosi la possibilità di accedere ad una collezione di dati/moduli di articoli e da lì cominciare una navigazione seguendo i legami che connettono collezioni dati e moduli di articoli.
- la creazione di un contesto di Lettura/Apprendimento il quale supporta:
- la creazione e attivazione di un insieme “flussi di lavoro” (workflows) di lettura e di apprendimento. Un workflow consiste di un certo numero di compiti (attività); tali compiti possono includere, per esempio, analisi di testo, analisi di letteratura, analisi di provenienza, ripetibilità di un esperimento, ecc. L’istanza di un workflow di lettura/apprendimento consiste in un percorso che conduce a un insieme di articoli e/o collezioni dati che supportano una attività di lettura/apprendimento. In sostanza, tali percorsi costituiscono sentieri all’interno di un record scientifico connesso.
  - Supporta la creazione e mantenimento di profili di lettura/apprendimento allo scopo di poter creare “percorsi di lettura/apprendimento personalizzati”.
  - Supporta la creazione di mappe tematiche.

## 6. Conclusioni

Lo sviluppo della prossima generazione d’infrastrutture digitali aperte per la comunicazione scientifica avrà un impatto importante:

- sulla accelerazione della transizione verso un sistema di comunicazione scientifica che faciliti la creazione e la disseminazione, di forme nuove e senza precedenti di ricerca scientifica, e il loro sostegno, utilizzando il potenziale innovativo delle tecnologie digitali;

- sul processo di maturazione di nuovi modelli di editoria in grado di supportare le prassi di promozione che ricompensino gli sforzi di pubblicazione;
- sul rendere la conoscenza scientifica liberamente accessibile;
- sul rendere i risultati della ricerca riproducibili e trasparenti ed esplicitando aspetti della conoscenza nascosti;
- sul superamento della distinzione fra le due culture nel mondo scientifico contemporaneo, cioè la cultura dei dati e la cultura della narrazione; e
- sul rendere possibili macro-analisi delle dinamiche della ricerca di interesse alle università, case editrici, agenzie di finanziamento della ricerca, ecc.

---

**Ringraziamenti:** *Questo articolo è stato ispirato da un certo numero di articoli elencati nei Riferimenti Bibliografici. Questi articoli hanno affrontato alcuni dei temi discussi in questo articolo. L'autore è riconoscente agli autori di questi articoli.*

*The characteristics of modern science entail the creation of a linked scholarly record that is online and open. Instrumental in making this vision happen is the development of the next generation of Open Cyber-Scholarly Communication Infrastructures (OCCIs), i.e., enablers of an open, evolvable, and extensible scholarly ecosystem.*

*The paper delineates the evolving scenario of the modern scholarly record and describes the functionality of the future OCCIs as well as the radical changes in the scholarly practices including new reading, learning and information seeking practices enabled by OCCIs.*

L'ultima consultazione dei siti Web è avvenuta nel mese di dicembre 2016.

## RIFERIMENTI BIBLIOGRAFICI

- Alexander - Cyganiak  
Hausenblas - Zhao      Keit Alexander, Richard Cyganiak, Michael Hausenblas, Jun Zhao, "Describing Linked datasets" in Linked Data Workshop at 2009-04-20, Madrid, Spain.
- Altman - King  
2007      Micah Altman and Gary King, "A Proposed Standard for the Scholarly Citation of Quantitative Data" in D-Lib Magazine 13 (March/April 2007), n.3/4.
- Bardi - Manghi  
2014      A. Bardi, P. Manghi, "Enhanced Publications: Data Models and Information Systems" in LIBER Quarterly, 23(2014), n.4, p. 240-273.  
Bechhofer - De Roure
- Bechhofer - De Roure  
Gamble - Goble  
Buchan 2010      Sean Bechhofer, David De Roure, Matthew Gamble, Carole Goble, Iain Buchan, "Research Objects: Towards Exchange and Reuse of Digital Knowledge". In: Proc. of *The Future of the Web for Collaborative Science (FWCS 2010)*, April 26, 2010 Raleigh, NC, USA.
- Belhajjame - Corcho  
Garijo - Zhao -  
Missier et al.  
2012      K. Belhajjame, O. Corcho, D. Garijo, J. Zhao, P. Missier, et al, "Workflow-Centric Research Objects: First Class Citizens in Scholarly Discourse" in proceedings of the ESWC2012 Workshop on the Future of Scholarly Communication in the Semantic Web (SePublica2012), Heraklion, Greece, May 2012.
- Bizer - Heath  
Berners - Lee  
2009      Christian Bizer, Tom Heath, Tim Berners-Lee, "Linked Data – The Story So Far" in International Journal on Semantic Web & Information Systems, 5 (2009) 3, p. 1-22.
- Bizer  
2011      Christian Bizer, "Evolving the Web into a Global Data Space" in: Advances in Databases: 28<sup>th</sup> British National Conference on Databases (BNCOD28), Manchester, UK, July 12-14, 2011: Revised selected papers, Eds. Alvaro A.A. Fernandes, Alasdair J.G.Gray, Khalid Belhajjame, Springer, 2011.
- Bizer  
2013      Christian Bizer, "Interlinking Scientific Data on a Global Scale", in Data Science Journal, Vol. 12, (2013), GRDI6-GRDI12 [DOI, pdf].
- Borgman  
2008      Christine Borgman, "Data, Disciplines, and Scholarly Publishing", in Learned Publishing Vol.21, No. 1, January 2008 p. 29-38, [DOI, pdf].
- Bourne  
2005      P. Bourne, "Will a Biological database Be Different from a Biological Journal?", in PLoS Computational Biology, Vol.1, Issue 3, 2005
- Bourne - Clark - Dale  
de Waard - Herman  
Hovy - Shotton 2011      P. Bourne, T. Clark, R. Dale, A. de Waard, I. Herman, E. Hovy, D. Shotton, (Eds) "Improving The Future of Research Communications and e-Scholarship", in: Dagstuhl Manifesto, Vol.1, Issue 1, 2011.

- Buckingham Shun  
Simon - Li - Sereno  
Mancini  
2007
- S. Buckingham Shun, J. Simon, V. UrenG. Li, B. Sereno, C. Mancini, *"Modeling Naturalistic Argumentation in Research Literatures: Representation and Interaction Design Issues"*, in International Journal of Intelligent Systems, 22(1), 2007
- Burke - Hammond  
Young 1996
- R. Burke, K. Hammond, B. Young, *"Knowledge-Based Navigation of Complex Information Spaces"* in AAAI-96 Proceedings, 1996
- Castelli - Manghi  
Thanos  
2013
- D. Castelli, P. Manghi, C. Thanos, *"A Vision Towards Scientific Communication Infrastructures"* in Int. J. on Digital Libraries, Vol. 13 No. 3-4, Sept. 2013
- Clark - Ciccarese  
Goble  
2013
- T. Clark, P. Ciccarese, C. Goble, *"Micropublications: A Semantic Model for Claims, Evidence, Arguments and Annotations in Biomedical Communications"* submitted to J. Biomed Semantics, Sept. 2013
- De Vocht - Coppens  
Verborgh - Vander  
Sande - Mannens  
Van de Walle 2013
- L. De Vocht, S. Coppens, R. Verborgh, M. Vander Sande, E. Mannens, and R. Van de Walle, *"Discovering meaningful Connections between Resources in the Web of Data"* in LDOW2013 May 14, 2013, Rio de Janeiro, Brazil
- de Waard - Kircz  
2008
- A. de Waard, J. Kircz, *"Modeling Scientific Research Articles – Shifting Perspectives and Persistent Issues"* in Proc. ELPUB 2008 Conference on Electronic Publishing, Toronto, Canada, June 2008
- de Waard  
Buckingham - Carusi  
Park - Samwald  
2009
- A. de Waard, S. Buckingham, A. Carusi, J. Park, M. Samwald, Sor, *"Hypotheses, Evidence and relationships: The Hyper Approach for Representing Scientific Knowledge Claims"*, in Proc. 8<sup>th</sup> International Semantic Web Conference, Workshop on Semantic Web Applications in Scientific Discourse. Lecture Notes in Computer Science, Springer Verlag, Berlin, 2009
- de Waard  
2010
- A. de Waard, *"From Proteins to Fairytales: Directions in Semantic Publishing"* in IEEE Intelligent Systems, 2010
- Dillon - Richardson  
McKnight
- A. Dillon, J. Richardson, C. McKnight, *"Navigation in Hypertext: A Critical Review of the Concept"* in D. Diaper, D. Gilmore, G. Cockton and B. Shackel (eds.) Human Interaction – INTERACT'90. North Holland: Amsterdam, 587-592
- Evans  
2008
- J. Evans, *"Electronic Publication and the Narrowing of Science and Scholarship"*, in Science Vol. 321, 2008
- Fink - Fericola -  
Chandran - Parastatidis  
Wade Naim - Quinn  
2010
- L. Fink, P. Fericola, R. Chandran, S. Parastatidis, A. Wade, O. Naim, G. Quinn, *"Word add-in for Ontology Recognition: Semantic Enrichment of Scientific Literature"* in Bioinformatics 2010, 11: 103

- Franklin - Halevy  
Maier 2005 M. Franklin, A. Halevy, D. Maier *"From Databases to Dataspaces: A New Abstraction for Information Management"* in SIGMOD Rec. 2005
- Ginsparg 2009 P. Ginsparg, *"Text in a Data-centric World"* in The Fourth Paradigm: Data Intensive Scientific Discovery, Microsoft, Redmond, 2009.
- Goble - De Roure C. Goble, D. De Roure, *"The Impact of Workflows on Data-centric Research"* in [X]
- Gray - Szalay - Thakar  
Stoughton - van de  
Berg 2002 J. Gray, A. Szalay, A. Thakar, C. Stoughton, J. van de Berg, *"Online Scientific data: Curation, Publication and Archiving"*, Technical Report MSR-TR- 2002-74. Redmond, WA: Microsoft Research, 2002.
- Gruber T. Gruber, *"Towards Principles for the Design of Ontologies Used for Knowledge Sharing"* in *"Formal Ontology in Conceptual Analysis and Knowledge Representation"*, Technical Report KSL 93-04, Knowledge Systems Laboratory, Stanford University
- Halevy - Franklin  
Maier 2006 A. Halevy, M. Franklin, D. Maier, *"Principles of Dataspace Systems"* in PODS'06, June 26-28, 2006, Chicago, Illinois, USA
- Harmsze F. Harmsze, *"A Modular Structure for Scientific Articles in an Electronic Environment"*. Ph.D. thesis, University of Amsterdam
- Herman Clark - Hovy  
de Waard 2012 I. Herman, T. Clark, E. Hovy, A. de Waard, Report on the *"Future of Research Communications"* Workshop, Dagstuhl, 2012
- Hey - Tansley, S., &  
Tolle, K. 2009 Hey, T., Tansley, S., & Tolle, K. (Eds) (2009) *"The Fourth Paradigm: Data Intensive Scientific Discovery"*, Redmond, WA: Microsoft Research
- Hu - Wang  
Lochovsky - Sun  
Chen 2009 J. Hu, G. Wang, F. Lochovsky, J-T Sun, and Z. Chen, *"Understanding User's Query Intent with Wikipedia"* in WWW 2009, April 20-24, 2009, Madrid, Spain
- Hunter 2006 J. Hunter, *"Scientific Models – A user –oriented Approach to the Integration of Scientific Data and Digital Libraries"*, VALA 2006, Melbourne, February, 2006.
- Idreos 2013 S. Idreos, *"Big Data Exploration"* in Big Data Computing, Taylor and Francis, 2013
- Johnsen 2010 L. Johnsen, *"Topic Maps"* in Journal of Information Architecture, Vol.2 Issue1, Spring 2010
- Kavuluru - Thomas  
Sheth - Chan - Wang R. Kavuluru, C. Thomas, A. Sheth, V. Chan, W. Wang, and A. Smith, *"An Up-to-date Knowledge-Based Literature Search and Exploration"*

- Smith 2012 Framework for Focused Bioscience Domains" in IHI 2012 – 2<sup>nd</sup> ACM SIGHIT International Health Informatics Symposium, January 28-30, 2012
- Khabsa - Giles 2014 M. Khabsa, C.L. Giles, *"The number of Scholarly Documents on the Web"*, PLoS -2014
- Kircz - Harmsze 2000 J. Kircz, F. Harmsze, *"Modular Scenarios in the Electronic Age"*, in: Conferentie Informatiewetenschap 2000
- Lynch 2009 C. Lynch, *"Jim Gray's Fourth Paradigm and the Construction of the Scientific Record"* in *The Fourth Paradigm: Data Intensive Scientific Discovery*, Microsoft, Redmond, 2009.
- Mackenzie Owen 2005 J.S. Mackenzie Owen, *"The Scientific Article in the Age of Digitization"*, Ph.D. thesis, University of Amsterdam, 2005
- McPherson 2010 T. McPherson, *"Scaling Vectors: Thoughts on the Future of Scholarly Communication"* in *Journal of Electronic Publishing*, Vol.13, Issue 2, 2010.
- Nicholas Huntington - Jamali Rowlands 2007 D. Nicholas, P. Huntington, H. Jamali, I. Rowlands, T. Dobrowoski, *"Characterizing and Evaluating Information Seeking Behavior in a Digital Environment: Spotlight on the "Bouncer"* in *Information Processing and Management* 43 (2007)
- Paskin 2005 N. Paskin, *"Digital Object Identifier for Scientific Data"* *Data Science Journal*, Vol.4, 18 March 2005
- Phelps - Wilensky 1996 T. Phelps, R. Wilensky, *"Toward Active, Extensible, Networked Documents: Multivalent Architecture and Applications"* in *Proc. of ACM Digital Libraries '96/Bethesda, Maryland, March 20-23, 1996*
- Poggi - Lembo Calvanese - de Giacomo - Lenzerini Rosati 2008 A. Poggi, D. Lembo, D. Calvanese, G. de Giacomo, M. Lenzerini, R. Rosati, *"Linking Data to Ontologies"* in *Journal on Data Semantics*, Springer-Verlag, 2008
- Porter - Souther 1999 B. Porter, A. Souther, *"Knowledge-based Information Retrieval"* in *AAAI Technical Report FS-99-02*, 1999
- Renear - Palmer 2009 A. Renear, C. Palmer, *"Strategic Reading, Ontologies, and the Future of Scientific Publishing"*, in *Science* Vol. 325, 2009
- Renear - Sacchi Wickett 2010 A. Renear, S. Sacchi, K. Wickett, *"Definitions of Dataset in the Scientific and Technical Literature"* in *ASIST 2010*, October 22-27, 2010, Pittsburgh, PA, USA

- Seringhaus - Gerstein 2007 T. Seringhaus, M. Gerstein, "*Publishing Perishing? Towards Tomorrow's Information*", in BMC Bioinformatics 2007, 8:17
- Shotton 2009 D. Shotton, "*Semantic Publishing: The Coming Revolution in Scientific Journal Publishing*", in Learned Publishing Vol. 22 No. 2, April 2009
- Taylor -. Gannon - Shields (eds) I. Taylor, D. Gannon, M. Shields (eds) "*Workflows for e-Science*", Springer, ISBN 978-1-84628-519-6
- Tenopir - King 2008 C. Tenopir, D. King, "*Electronic Journals and Changes in Scholarly Article Seeking and Reading Patterns*" in D-Lib Magazine 2008, Vol. 14 No. 11/12
- Thanos 2014 C. Thanos, "*The Future of Digital Scholarship*" in Procedia Computer Science Vol. 38, pages 22-27, 2014, Elsevier
- Thanos 2016 C. Thanos, "*A Vision for Open Cyber-Scholarly Infrastructures*", Publications 2016
- White 2013 Colin White, "*Data Exploration and Discovery: A new Approach to Analytics*" in BI Research, October 2013
- Woutersen Windhouwer Brandsma - Verhaar Hogenaar - Hoogerwerf Doorenbosch - Durr Ludwig - Schmidt Sierman 2009 S. Woutersen-Windhouwer, R. Brandsma, P. Verhaar, A. Hogenaar, M. Hoogerwerf, P. Doorenbosch, E. Durr, J. Ludwig, B. Schmidt, B. Sierman, "*Enhanced Publications*", edited by M. Vernooij-Gerritsen, SURF Foundation, Amsterdam University Press, 2009

## SITOGRAFIA

- Buckingham Shun S. Buckingham Shun, "Net-Centric Scholarly Discourse?" in:  
<http://slidesha.re/qvoqoU>
- Decker S. Decker, "From Linked Data to Networked Knowledge" in:  
[http://videlectures.net/eswc2013\\_decker\\_networked\\_knowledge/](http://videlectures.net/eswc2013_decker_networked_knowledge/)
- Lagoze - Van de Sompel C. Lagoze, H. Van de Sompel, "The OAI Protocol for Object Reuse and Exchange", in: <http://www.openarchives.org/ore>
- Microsoft Microsoft, "Patterns & Practices" in: <https://msdn.microsoft.com/en-us/library/ff646997.aspx>
- Microsoft Microsoft, "An Introduction to Topic Maps" in:  
[https://msdn.microsoft.com/en-us/library/aa480048\(d=printer\)](https://msdn.microsoft.com/en-us/library/aa480048(d=printer))
- Nicholas Huntington - Jamali Rowlands - Dobrowoski D. Nicholas, P. Huntington, H. Jamali, I. Rowlands, T. Dobrowoski, "Viewing and Reading Behavior in a Virtual Environment" in:  
<http://www.emeraldinsight.com/loi/ap>
- Simon - Miklos Nejd - Sintek Salvachua 2003 B. Simon, Z. Miklos, W. Nejd, M. Sintek, J. Salvachua "Smart Space for Learning: A Mediation Infrastructure for Learning Services", in:  
<http://www.conference.org/www2003/cdrom/papers/alternate/P460/HTML/460-gamper.html>
- Thearling K. Thearling, "An Introduction to Data Mining" in:  
[http://www.thearling.com/dmintro/dmintro\\_2.htm](http://www.thearling.com/dmintro/dmintro_2.htm)
- Vieu L. Vieu "On the Semantics of Discourse Relations" in:  
<http://www.irit.fr/publis/LILAC/V-DRSemantics-CID11.pdf>
- Waterworth - Chignell J. Waterworth, M. Chignell, "A Model for Information Exploration" in:  
<http://www8.informatik.umu.se/~jwworth/infomodel.pdf>