

# Osservazioni preliminari al Progetto SRI (versione 20.11.2017)

## 0. Premessa

Scopo di queste righe è mettere in evidenza ( con osservazioni necessariamente sintetiche e preliminari) alcuni punti di forza e alcuni punti critici del progetto SRI, nell'intento di offrire un primo contributo costruttivo e usabile nelle "Fasi di lavorazione del progetto SRI" (paragrafo 9.2), con la massima disponibilità ad approfondire.

## 1. Modelli di riferimento per l'integrazione

La *progettazione preliminare* pare non tenere conto del necessario riferimento alla letteratura e alle soluzioni correnti di *integrazione ragionata* di servizi di accesso: si fa infatti riferimento a **modelli di integrazione nati tra la fine degli anni 90 e i primi anni del 2000**, quando i modelli di *portale* avevano l'obiettivo di *fornire all'utente un'esperienza unitaria di accesso* a fonti di informazione gestite da una Istituzione o da una Azienda (modello *one-stop-shop*). Quell'obiettivo ha ovviamente ancora la sua validità; oggi occorre però **tenere conto di almeno due cambiamenti fondamentali**:

- il web non può essere considerato semplicemente come un servizio di trasporto tra un servizio di accesso centrale e l'utente finale: **non possiamo ignorare l'ecosistema web** che oggi viene offerto all'utente (i **servizi nostri "concorrenti"** o quelli **complementari** all'informazione che offriamo);
- gli **utenti dei servizi offerti** sul web oggi non sono solo gli esseri umani, ma sono **anche macchine in grado di rielaborare l'informazione e di offrire altri servizi**. Una volta prodotta e pubblicata, l'informazione è riusabile anche in altri servizi non più definibili e controllabili dal produttore (modello *one-shop-stop*).

I paragrafi che seguono propongono alcune raccomandazioni proprio a partire dalla considerazione che un "servizio nazionale" non può disinteressarsi oggi di *a*) come i suoi dati interagiranno con i dati già presenti sul web e di *b*) come le *macchine* (pensiamo ad esempio ai motori di ricerca) useranno quei dati.

## 2. Integrazione, ruolo degli identificatori *interni* e "sviluppi successivi"

Il progetto SRI mette correttamente l'accento sul ruolo degli *identificatori* quali dispositivi ineludibili ed efficaci per mettere in relazione informazioni provenienti da fonti differenti - Indice SBN, Manus, Edit16 e Internet Culturale - e che si riferiscono alla stessa entità. Tuttavia il "livello ontologico proprio dell'universo bibliografico" (p. 2) preso in considerazione è esclusivamente quello *interno* delle fonti appena ricordate. Nel capitolo 4 si parla di "**sviluppo successivo**" di una "**Interfaccia**

**RDF**” per “l’arricchimento semantico” anche con le “informazioni esposte dall’aggregatore *Cultura italia*” [sic].

È naturalmente del tutto comprensibile che vengano definite delle priorità (non tutto si può fare subito), purché però il punto di arrivo sia ben definito e quello che viene messo in opera oggi sia coerente anche con il punto di arrivo stesso. In questo senso, ci domandiamo perché non procedere subito *anche* alla pubblicazione sul web come *Linked data/RDF* dei dati gestiti centralmente dalla cooperazione SBN, a maggior ragione sapendo che “l’arricchimento semantico”- **ovvero la maggiore usabilità/interconnessione delle informazioni bibliografiche nel web dei dati oltre che nel web dei documenti** - sarà il frutto di un impegno graduale e progressivo che conviene mettere in opera prima possibile. Così come del resto il progetto SRI affronta “contemporaneamente” sia lo sviluppo del protocollo SBNMARC negli ambienti MOL e EDIT16 per la “riconciliazione dei dati” (che sarà sfruttata in fase di ricerca integrata), sia l’avvio della ricerca integrata (che progressivamente potrà contare anche sulla “riconciliazione”). In altre parole e in generale la disponibilità di una nuova funzionalità in una determinata applicazione non significa automaticamente la possibilità di disporre di nuove informazioni provenienti dai dati già archiviati con la versione precedente della stessa applicazione, ma nondimeno la nuova funzionalità può/deve essere implementata costituendo la precondizione per integrare nuove informazioni.

### **3. Integrazione, Linked data e ruolo degli identificatori (interni e esterni). Punti di partenza e punti di arrivo.**

Il punto di arrivo dovrebbe **comprendere anche** la pubblicazione su web come **Linked data/RDF** delle informazioni gestite centralmente dalla cooperazione **SBN** (senza concessioni alle mode e valutando attentamente costi/benefici). Ovviamente senza identificatori persistenti (creati attraverso il controllo di autorità) i **Linked data** pubblicati sarebbero un “miserò brodino” del tutto inutilizzabile. Non è sufficiente però proporre strategie per la condivisione fin dalla creazione degli identificatori *interni* ai servizi informativi centrali, **è necessario con sistematicità prendere in considerazione anche tutti i riferimenti esterni messi in opera nella fase del controllo di autorità**. Buone pratiche (ad es. qui e qui) *quali il riferimento tramite URI* a identificatori appartenenti a domini esterni andrebbero estese con sistematicità a partire dalla creazione dei record. In questo contesto occorre **dare risposte adeguate quando servizi di accesso all’informazione esterni a SBN fanno riferimento a identificatori SBN**: qui - ad esempio - di fronte a un forte interesse esterno, ci sembra che il dominio SBN non abbia reagito con un adeguato disegno complessivo.

### **4. Quello che possiamo fare a breve termine (anche senza modificare le applicazioni che producono i dati)**

Le modalità di pubblicazione come **Linked data** possono adottare soluzioni differenti (anche complementari). Quello che non può mancare è oggi una **strategia complessiva**. Con investimenti del tutto modesti *a breve termine* si può usare **schema.org** (che in ogni caso è un’ontologia con la quale occorre fare i conti se si decide che **anche i motori di ricerca** - si noti il plurale - **possono essere utenti**

**dei servizi e dei dati che pubblichiamo su web**). In realtà è interessante notare che anche nell'attuale opac dell'Indice SBN ci sono già dei tentativi di applicazione di un'ontologia (Opengraph proposta originariamente da **Facebook**). Qui e qui ad esempio come *i motori di ricerca* oggi possono interpretare i dati pubblicati da un concorrente con l'ontologia schema.org .

Sempre *a breve termine* (e anche senza modifiche alle modalità di immissione/strutturazione dei dati) sarebbe possibile prendere in conto delle funzionalità di **raggruppamento logico dei record ispirate a FRBR** già operative in strumenti consolidati come EDS, Primo, Worldcat discovery, in strumenti open source come versioni di Vufind, ma anche in sperimentazioni italiane come Scoprirete FRBR.

Anche qui si tratta di prendere decisioni di tipo pragmatico nell'interesse dei nostri utenti (decisioni che potranno poi essere riviste e che naturalmente risulteranno più efficienti se - vedi il paragrafo successivo - si introdurranno cambiamenti anche al modo di produrre i dati bibliografici).

Si suggerisce inoltre - in questo contesto dove si sono appena richiamati i più importanti c.d. *discovery tool* che "per mestiere" aggregano fonti eterogenee - di valutare i pro e i contro di una possibile alternativa: sviluppare un *aggregatore ad hoc* - scelta della progettazione SRI - o riutilizzo di un aggregatore già consolidato. E' possibile che questa valutazione sia stata già fatta, ma ci sembrerebbe opportuno che fosse illustrata nel documento.

## **5. Quello che possiamo fare a medio termine (fare i conti anche con i nuovi modi di produzione)**

Naturalmente nel proporre un piano di evoluzione *a medio termine* sarebbe consigliabile **fare i conti anche con modi nuovi di produrre servizi informativi (non solo bibliografici)** direttamente come Linked data/RDF senza attendere *l'ontologia unica e definitiva*: tra le direzioni da prendere in conto - anche per prospettare in maniera ragionata alternative - vi sono sicuramente Folio e Wikibase). Può essere **rischioso vedere l'attuale architettura SBN sub specie aeternitatis** senza confrontarsi con i punti di forza e i punti critici.

Qualche precisazione:

- **ontologia**: come è stato osservato, con l'affermarsi del Web semantico (o web dei dati) e dei *Linked data* l'espressione **ontologia ha sostituito** - anche nel mondo dell'informazione bibliografica - l'espressione **schema di metadati** (o insieme strutturato di metadati o formato). Si potrebbe anche dire che ad esempio UNIMARC è una *ontologia di fatto*: ovvero una ontologia che *non* si presenta in forma esplicita e che non si appoggia sulla grammatica standard per lo scambio di dati su web (RDF). **Quando si parla di "arricchimento semantico" occorre precisare se si tratta semplicemente di un cambio di "sintassi"** (i "vecchi dati" vengono pubblicati grazie all'ospitalità di nuove strutture) **oppure se si tratta di un vero e proprio cambio di ontologia** (servirebbe ad esempio questa precisazione quando a p. 28 il progetto SRI parla di una possibile ontologia di riferimento della "Interfaccia RDF" ovvero del modello "IFLA LRM ontologizzato");
- **Folio**: si tratta come è noto di una recente iniziativa internazionale aperta nata dalla collaborazione tra biblioteche, sviluppatori e aziende che operano nel

settore con un piano di lavoro molto ambizioso. Viene citata qui perché ha preso in conto con sistematicità anche l'obiettivo di proporre nuovi modi di produrre l'informazione bibliografica. In ogni caso **Folio** è sicuramente un punto di riferimento per un progetto di sviluppo, di integrazione di applicazioni informatiche in campo biblioteconomico.

- **Wikibase**: si tratta di un **modello di successo consolidato a livello mondiale** di produzione e di pubblicazione dei dati con la grammatica del web (RDF). Come è noto Wikibase è l'infrastruttura tecnologica del servizio Wikidata (la centralizzazione di *tutti i dati* presenti nelle *pagine* delle varie versioni linguistiche di Wikipedia e degli altri servizi Wiki, per esempio Wikisource: **oltre 1 miliardo di triple RDF gestite e realmente disponibili nel web dei dati**). L'infrastruttura Wikibase è open source e liberamente disponibile e non è legata ad una determinata visione ontologica dei dati (si tratta di uno strumento che facilita sia la collaborazione nella creazione di dati strutturati, sia lo sviluppo condiviso di ontologie). Per una ipotesi di uso possibile di wikibase in questo contesto si rinvia alla pagina GLAM alla voce Wikibib. (Tra parentesi **il mondo Wiki\* è l'unico che in questo momento può costruire alternative di peso al fenomeno crescente della concentrazione di strumenti informatici e informativi in mano ai soliti big**).

## **6. Integrazione, risorse digitali, decisioni di natura tecnica e posizioni condivise nella cooperazione SBN**

È sicuramente importante dedicare il 6. capitolo a *SBNTECA: integrazione tra aggregatore digitale e catalogo collettivo* con la proposta di **“sviluppare e/o riutilizzare un sistema di teca digitale funzionalmente autonomo usabile in linea teorica anche con altri LMS di Polo”**. È altrettanto importante l'osservazione contenuta nel paragrafo 5.2.2 dove, con l'obiettivo di **“ampliare considerevolmente il volume di risorse digitali attualmente indicizzate da Internet Culturale (ribattezzato Indice unico del digitale)”**, si ipotizzano per il polo nuove modalità per “informare l'Indice SBN” sull'offerta di oggetti digitali.

**Ci si potrebbe però chiedere quando si parla di *punto di accesso unico* se la cooperazione SBN:**

- **ha proposte condivise solo per gli attuali poli SBN** (sia per l'aggregazione dell'informazione bibliografica, sia per quella che punta alle risorse digitalizzate) e si disinteressa di quello che accade *in partibus infidelium*; le tecnologie che abbiamo oggi a disposizione ci possono aiutare - pragmaticamente - anche sul versante “apertura di SBN”. Perché ad esempio aggregare solo MOL, Edit16 e IC e non altri sistemi che non hanno un gestionale SBN? L'aggregazione può essere fatta in molti modi e in molti contesti: ad esempio se il nostro punto di arrivo comprende anche la pubblicazione dei dati come Linked data/RDF può essere possibile offrire a esperienze come SHARE CAT adeguati livelli di integrazione;
- **ha una posizione condivisa sulle offerte che altri aggregatori** (per es. Worldcat) oggi offrono in Italia anche agli stessi poli SBN;

- **ha una posizione condivisa sui progetti di digitalizzazione di massa** (per es. Google Books è ancora in corso e sta per arrivare a 1 milione di documenti digitalizzati);
- **ha una posizione condivisa sull'ecosistema del digitale (soprattutto del digitale nativo)** dove fornire l'indirizzo di una risorsa in rete è cosa ben diversa dal garantire accesso nel tempo (il deposito legale come servizio fondamentale in questo ecosistema; accanto a servizi specifici come il c.d. accesso perpetuo).

Non si può chiedere a una "progettazione preliminare" la risposta a queste domande, ma **decisioni di natura tecnica non possono sostituire** di fatto gli indirizzi da condividere a livello di cooperazione SBN nella **definizione di livelli di adesione, di ruoli e responsabilità e di politiche per l'ampliamento della cooperazione tra le biblioteche italiane.**

## 7. Servizi nazionali e servizi locali

A p. 7 il progetto SRI parlando del modello gestionale SBN sintetizza la distribuzione di compiti tra Indice e poli: "tale modello gestionale prevede che l'informazione sia creata dalla periferia e immediatamente condivisa attraverso servizi resi disponibili dal sistema centrale di indice. La descrizione bibliografica, la gestione del documento fisico e l'erogazione dei servizi sono compiti di cui si fanno carico gli LMS di Polo". In realtà negli ultimi anni in varie sedi e occasioni si è sempre convenuto che occorre recuperare e valorizzare il ruolo di SBN come erogatore di servizi e che i servizi centrali sono (almeno) due, Indice e ILL (nel progetto il sistema/servizio ILL-SBN viene citato solo nel paragrafo 2.5 a pag. 10). In ogni caso le integrazioni con ILL-SBN, Nilde, PagoPa, SPID etc. dovrebbero costituire una parte essenziale anche in una progettazione preliminare. Inoltre anche prendendo spunto dalle indicazioni di Tommaso Giordano nella sua lectio "Lost in transition", si potrebbe ricordare la molteplicità di servizi non solo bibliografici che si potrebbero realizzare a partire da un knowledge base nazionale di dati bibliografici

## 8. Sviluppo di software e investimenti pubblici

Richiamandoci a quanto detto sopra al paragrafo 5 soprattutto quando si parla di investimento pubblico nel software sarebbe importante fare riferimento a modelli internazionali (per esempio sia Folio che Vufind sono su GitHub) e/o almeno a raccomandazioni nazionali (riuso Agid).

(02/01/2018)