

INDAGINE PER IDENTIFICARE I SOFTWARE OPEN SOURCE E I METADATI IN USO NELLE BIBLIOTECHE DIGITALI IN ITALIA

Report Risultati

Draft 1.0 9 settembre 2008

Anna Maria Tamaro
Damiana Luzzi

Indice

[\[nascondi\]](#)

- [1 Introduzione](#)
- [2 Cenni storici](#)
- [3 SOS e Free Software](#)
- [4 Diffusione del SOS in Italia](#)
- [5 Risultati dell'indagine](#)
- [6 Osservazioni e Conclusioni](#)

Introduzione

Nell'ambito progetto Accesso alle Biblioteche Digitali finanziato dalla Fondazione Rinascimento Digitale e coordinato da Anna Maria Tamaro, nel febbraio 2008 è stata avviata un'indagine per identificare i Software Open Source (d'ora in avanti citato SOS) e i Metadati in uso nelle biblioteche digitali in Italia. Lo scopo dell'indagine è quello di monitorare la situazione dell'uso di SOS, favorirne la conoscenza e diffonderne le migliori pratiche di utilizzo nell'ambito delle biblioteche digitali. Prima di esaminare i dati ottenuti dall'indagine si ritengono necessari una sintesi storica sulla genesi del fenomeno del SOS, la sua definizione e la diffusione di tale software in Italia.

Cenni storici

L'Open Source non è un fenomeno recente, va ricondotto agli albori dell'informatica, quando nei laboratori storici Xerox Park, IA lab del MIT, Bell Labs e Berkeley, il software e i relativi sorgenti venivano messi a disposizione della comunità. Il software era considerato un prodotto scientifico al pari della fisica e della matematica, pertanto era trattato come un esperimento scientifico. Se di un esperimento scientifico vengono distribuite le ipotesi, le tesi, il procedimento e i risultati, del software vengono allora distribuite le analisi dei requisiti e i codice sorgenti.

Quando il mondo commerciale vide nel software un prodotto da proteggere con licenze d'uso, sul quale esercitare il diritto di proprietà e da commercializzare in un mercato in rapida crescita, iniziò a produrre

software coperto da segreto industriale e a distribuirlo in formato eseguibile, ovvero senza accesso al codice sorgente, a pagamento. È in questo periodo che inizia la confusione maggiore tra SOS e Free Software, intendendo il primo, assimilato al secondo, come software gratuito e di scarsa qualità. Per circa un quindicennio l'attenzione degli utenti è stata orientata da campagne marketing verso prodotti commerciali.

Quando e cosa ha fatto invertire la rotta?

Negli anni Novanta la nascita di Linux ha avuto un impatto dirompente sulla comunità degli sviluppatori prima, sul mercato e gli utenti poi. Linux ha fatto da traino allo sviluppo del Free Software e/o SOS.

SOS e Free Software

La confusione terminologica tra Free Software e Open Source ci accompagna oltre la metà degli anni Novanta. Tale confusione è stata generata dal considerare sinonimi e quindi con la stessa valenza i termini free e open, anche se letteralmente open source significa a "sorgente aperta". La differenza è, al contrario, molto sottile e ha origine dal concetto comune di diffondere software libero e dall'ambivalenza del termine inglese free che significa sia libero che gratuito. Free Software è comunque l'idea che sta alla base anche di Open Source.

Il Free Software dal 1984 è patrocinato dalla Free Software Foundation fondata nello stesso anno da Richard Stallman. Il Free Software viene rilasciato con la General Purpose License (GPL), per evitare che possa essere cooptato dai produttori commerciali. I termini della GPL non permettono nessuna commercializzazione del software, ma permettono la possibilità di introito derivante dalla vendita del supporto fisico del programma o dalle varie forme di consulenza sul programma stesso. Il problema era garantire il diritto di libera distribuzione ai programmatori insieme al diritto di poter ottenere guadagno per gli investitori.

Nel 1997 la nascita della Open Source Initiative (OSI), organizzazione non a scopo di lucro guidata da Eric Raymond, rimette a posto le carte e evita nel contempo la frattura nella comunità hacker tra pro e contro il Free Software. La OSI è nata per un'esigenza reale dovuta al lancio da parte di Netscape del progetto Mozilla e della distribuzione dei sorgenti del browser Netscape Navigator. Dovevano essere coperti gli interessi commerciali della libera impresa e allo stesso tempo assicurare il libero utilizzo del codice agli sviluppatori. La OSI ha registrato il marchio OSI Certified che garantisce agli sviluppatori la possibilità legale del riutilizzo del software e della distribuzione del codice. Contemporaneamente ha rilasciato la Open Source Definition che detta sia un insieme di regole che la tipologia di licenze che accompagnano il SOS devono rispettare per dotarsi del marchio OSI Certified. La Open Source Definition non è una licenza software e non ha valore legale, ma dà una specifica delle caratteristiche ammesse in una licenza software perché possa essere riferita come Open Source. Tutti e 10 i termini della Open Source Definition devono essere applicati al programma originale come pure alle versioni derivate da esso.

Di seguito si riporta la traduzione in italiano della Open Source Definition:

Introduzione

Open Source non significa solo accesso al codice sorgente. I termini di distribuzione di un programma Open Source devono essere consoni ai criteri seguenti:

1. Libera redistribuzione

La licenza non può limitare alcuno dal vendere o donare il software che ne è oggetto, come componente di una distribuzione aggregata, contenente programmi di varia origine. La licenza non può richiedere diritti o altri pagamenti a fronte di tali vendite.

2. Codice sorgente

Il programma deve includere il codice sorgente e ne deve essere permessa la distribuzione sia come codice sorgente che in forma compilata. Laddove alcune forme di un prodotto non siano distribuite con il relativo codice sorgente, deve essere chiaramente indicato il modo per ottenerlo, ad un costo non superiore ad una ragionevole spesa di distribuzione, preferibilmente scaricandolo gratuitamente da Internet. Per codice sorgente si intende la forma in cui un programmatore preferirebbe modificare il

programma. Codice sorgente deliberatamente reso illeggibile non risponde ai requisiti. Forme intermedie come l'output di un preprocessore o compilatore non rispondono ai requisiti.

3. Prodotti derivati

La licenza deve permettere modifiche e prodotti derivati, e deve permetterne la distribuzione sotto le stesse condizioni della licenza del software originale.

4. Integrità del codice sorgente originale

La licenza può impedire la distribuzione del codice sorgente in forma modificata, a patto che venga consentita la distribuzione dell'originale accompagnato da "patch", ovvero file che permettono di applicare modifiche automatiche al codice sorgente in fase di compilazione. La licenza deve esplicitamente permettere la distribuzione del software prodotto con un codice sorgente modificato. La licenza può richiedere che i prodotti derivati portino un nome o una versione diversa dal software originale.

5. Discriminazione contro persone o gruppi

La licenza non deve discriminare alcuna persona o gruppo di persone.

6. Discriminazione per campo d'applicazione

La licenza non deve impedire di far uso del programma in un ambito specifico. Ad esempio non si può impedire l'uso del programma in ambito commerciale o nell'ambito della ricerca genetica. 7.

Distribuzione della licenza

I diritti allegati a un programma devono essere applicabili a tutti coloro a cui il programma è redistribuito, senza che sia necessaria l'emissione di ulteriori licenze.

8. Specificità ad un prodotto

I diritti allegati al programma non devono dipendere dall'essere il programma parte di una particolare distribuzione di software. Se il programma è estratto da quella distribuzione e usato o redistribuito secondo i termini della licenza del programma, tutti coloro che ricevano il programma dovranno avere gli stessi diritti che sono garantiti nel caso della distribuzione originale. 9. Vincoli su altro software

La licenza non deve porre restrizioni su altro software distribuito insieme al software licenziato. Per esempio, la licenza non deve richiedere che tutti gli altri programmi distribuiti sugli stessi supporti siano software open source.

10. Neutralità rispetto alle tecnologie

La licenza non deve contenere clausole che dipendano o si basino su particolari tecnologie o tipi di interfacce.

Diffusione del SOS in Italia

Negli ultimi anni la diffusione del fenomeno Open Source in Italia è in aumento, come segnalato dal I° Rapporto sull'Open Source in Italia che, anche se datato 2004, ha messo fin da allora in luce alcune tendenze:

- la crescita e la diffusione del SOS in Italia si è realizzata per moto spontaneo e in pratica quasi esclusivamente solo su base locale;
- forte e crescente coinvolgimento alle grandi community di sviluppo e di numerosi singoli individui che hanno accumulato nel tempo approfondite competenze.

I dati sulle competenze nell'ambito dello sviluppo di SOS è rafforzata dall'indagine della Berlecon Research di Berlino che pone l'Italia al quarto posto al mondo con quasi l'8% degli sviluppatori. Dal 2004 ad oggi tali tendenze sono state confermate e la diffusione ha avuto una ulteriore crescita che si evince anche dalla presenza in Italia di reti attive in ambito locale di sviluppatori SOS specialisti o semplici appassionati che si riuniscono in associazioni come Linux User Group, Free Software User Group. Entrambi gli User Group si occupano, fra le altre attività, di promuovere la conoscenza all'utilizzo di SOS e Free Software.

Una forte impulso allo sviluppo del SOS è arrivato agli inizi del 2000 dalla Comunità Europea con il finanziamento e gestione di progetti sul SOS. La Comunità Europea partiva da un solido background poiché le Pubbliche Amministrazioni europee, agli inizi degli anni Novanta, sono state fra le prime utilizzatrici di SOS .

Un ulteriore input all'uso di SOS in Italia giunge dal 2002, con la messa in atto di iniziative, normative, disegni di legge e leggi sia nazionali che regionali volte a promuoverne e incentivarne l'applicazione. Per un elenco delle più significative si rinvia alla bozza di studio sullo Stato dell'Open Source in Italia, redatta dalla Commissione per il software "open source" nella Pubblica Amministrazione istituita dal Ministro Nicolais in data 16 maggio 2007. Qui si segnalano:

- la costituzione all'interno del Centro nazionale per l'Informatica nella Pubblica amministrazione (CNIPA) dell'Osservatorio Open Source per la rilevazione continua sull'uso di FLOSS presso le PA italiane.
- la realizzazione all'interno del CNIPA, nel maggio 2007, dell'Ambiente di Sviluppo Cooperativo (ASC) che consente alle PA e ad attori di mercato provenienti dal mondo della ricerca di collaborare allo sviluppo di applicazioni Open Source per le esigenze del settore pubblico.
- la firma del protocollo d'intesa (gennaio 2008) fra Regione Toscana, Provincia di Pisa, Università e Polo tecnologico di Navacchio per la creazione di un centro regionale - e in futuro nazionale - dedicato allo studio delle tecnologie e allo sviluppo del software libero. Il progetto rientra in un più ampio quadro di iniziative intraprese dalla Regione Toscana per portare l'open source negli uffici pubblici e nelle amministrazioni, in sinergia con le imprese.

Risultati dell'indagine

Come descritto, l'attenzione e la diffusione crescente rivolta al SOS in Italia hanno portato gli esperti del progetto Accesso alle biblioteche digitali a valutare in maniera positiva il lancio di una prima indagine di carattere esplorativo per verificare e monitorare l'uso che viene fatto del SOS nelle biblioteche digitali in Italia.

L'indagine è stata suddivisa in tre parti per organizzare e circoscrivere al meglio l'oggetto da indagare e anche per agevolare i tempi di compilazione di ogni singola indagine. Le domande di ciascuna indagine sono state elaborate dagli esperti che partecipano al progetto Accesso alle biblioteche digitali.

1. La prima indagine ha proposto domande volte ad indagare l'opinione degli operatori del settore in relazione all'uso di SOS nelle biblioteche digitali.
2. La seconda indagine ha approfondito e monitorato la situazione dell'uso di SOS.
3. La terza indagine ha raccolto informazioni sull'utilizzo di metadati.

Le tre indagini sono state redatte nella forma di questionari a risposta chiusa. I questionari sono stati resi disponibili on line utilizzando il SOS LimeSurvey installato in un sottodominio della Fondazione Rinascimento Digitale.

Il lancio dei questionari è avvenuto il 20 febbraio e sono rimasti aperti fino al 30 giugno. I metodi utilizzati per diffondere i questionari sono stati:

- annuncio in home page nel sito della Fondazione Rinascimento Digitale;
- invio di mail con invito alla diffusione agli esperti che partecipano al progetto e che, in larga parte, operano a vario titolo nel settore delle biblioteche;
- selezione di indirizzi dalla mailing list della Fondazione Rinascimento Digitale;
- invio alla mailing list AIB.

1. Risultati dell'indagine sull'uso di SOS nelle biblioteche digitali in Italia

- Totale dei questionari compilati: 19
 - Biblioteche pubbliche: 15,
 - Biblioteche private: 4.

Il 78,9% sono concordi nel ritenere che la filosofia del SOS sia adeguata alla missione della biblioteca digitale, il 15,8 concordano in parte con tale affermazione, mentre il 5,3% è in totale disaccordo.

Il 57,9% concordano in parte e il 5,25% è concorde nell'affermare che il SOS può essere una soluzione più economica per le applicazioni tecnologiche della biblioteca digitale, il 31,6% ha espresso un parere discorde, il 5,25% è senza opinione.

Il 31,6% concorda in parte sul fatto che il SOS fornisca una soluzione flessibile alla personalizzazione a seconda della necessità della biblioteca digitale, il 21% non ha espresso opinioni, il 26,4% è in disaccordo e il 21% è in forte disaccordo.

Tutti sono concordi (il 47,3% solo in parte) nel dichiarare che il SOS include una scarsa documentazione. Tutti sono concordi (il 42% solo in parte e 5,25% è senza opinione) nell'affermare che il SOS è user-friendly.

Il 43% è in disaccordo nell'affermare che il SOS fornisce meno funzionalità di quello "close source" commerciale, il 5,25% è senza opinione, mentre il 21% concorda in parte e il 30,75 % concorda.

2. Risultati dell'indagine sull'utilizzo di SOS nelle biblioteche digitali in Italia

- Totale dei questionari compilati: 10
 - Biblioteche pubbliche: 7,
 - Biblioteche private: 3.

Tutti hanno dichiarato di utilizzare SOS

I SOS utilizzati sono:

- DSpace: 5
- Scout Portal Toolkit: 2
- Firefox: 3
- Open Office: 3
- Fedora: 1
- UIUC OAI Metadata Harvesting Project: 1
- MySQL Community server: 1
- MySQL: 1
- Apache server: 2
- Protégé ontology editor: 1
- JSP Wiki: 1
- GNU E-prints: 2
- MediaWiki: 1
- Linux: 1
- Gyp: 1
- SendMail: 1
- OJS: 1
- Bind: 1
- Tomcat: 1
- Plone: 1
- DjVu Libre: 1
- Clamav: 1
- Mailman: 1
- CODEX[ml]: 1
- BD2 viper: 1
- lucene: 1
- longwell: 1
- moddle: 1
- xml exist open cms: 1
- museo&web: 1
- typo3: 1
- IIPImage: 1
- COA: 1

In merito a questa risposta si segnala che due istituzioni non si sono attenute alle indicazioni di compilazione segnalando più di un SOS all'interno di un singolo questionario, pertanto non è stato possibile indicare i dati in percentuale. Questo ha influenzato anche alcune delle risposte seguenti.

Suddividendoli per categoria, dominio e funzione del SOS emerge sia dalla lista precedente che nella tabella di seguito che la maggioranza (7) sono Repository.

- Comunicazione: audio e video conferenza: 1
- Comunicazione: connettività: 1
- Comunicazione: e-mail: 2
- Comunicazione: instant messenger: 1
- Comunicazione: open publishing: 2
- E-Learning: 1
- Repository: 7
- Strumento di produttività: antivirus: 2
- Strumento di produttività: collaborazione/GroupWare: 2
- Strumento di produttività: content management system (CMS): 4
- Strumento di produttività: elaborazione audio: 1
- Strumento di produttività: elaborazione delle immagini: 2
- Strumento di produttività: forum: 1
- Strumento di produttività: Grafica ed elaborazione digitale: 1
- Strumento di produttività: HTML browser/editor: 4
- Strumento di produttività: knowledge management: 3
- Strumento di produttività: library automation: 2
- Strumento di produttività: mailing list: 3
- Strumento di produttività: office suite/word processor: 4
- Strumento di produttività: ontologia: 2
- Strumento di produttività: optical character recognition (OCR)/Trascrizione di testi: 1
- Strumento di produttività: player multimediale: 1
- Strumento di produttività: project management: 1
- Library 2.0: blog: 1
- Library 2.0: opac 2.0: 1
- Library 2.0: wiki: 2
- Server web: 2
- Sistema operativo: 3
- ALTRO:: 3
 - Catalogo: 1
 - Raccolta e pubblicazione metadati: 1
 - DBMS: 1

Se organizzati per tipologia di licenza emerge che la maggioranza (5) ha licenza GNU GPL, 4 hanno licenza BSD License, 2 GNU LGPL, 1 University of Illinois/NCSA (National Center for Supercomputing Applications) Open Source License, 1 Apache License, Version 2.0, 1 Mozilla Firefox EULA e 1 MPL - Mozilla Public License.

Fra le problematiche di utilizzo sono state rilevate: Dspace:

- scarsa documentazione,
- personalizzazione dell'interfaccia limitata,
- usabilità che può essere migliorata,
- difficoltà segnalata dagli utenti per l'iscrizione.

UIUC OAI Metadata Harvesting Project:

- approccio tecnico impegnativo.

E-Prints:

- difficoltà nella risoluzione di problematiche tecniche.

Il 20% partecipa alla comunità di sviluppo del SOS.

Il 70% ha apportato modifiche al SOS (Software modificati: DSpace, Eprints, UIUC OAI Metadata Harvesting Project, JSP Wiki). Di quest il 71,4% ha affidato le modifiche al personale interno, mentre il 28,6% ha delegato ad un'azienda esterna.

L'azienda, in entrambi i casi, ha la gestione completa del SOS e si occupa della manutenzione e/o aggiornamento. In un solo caso ha svolto anche attività di formazione.

3. Risultati dell'indagine sull'uso di Metadati nelle biblioteche digitali in Italia

- Totale dei questionari compilati: 13
 - Biblioteche pubbliche: 10
 - Biblioteche private: 3

Il set o sistema di metadati utilizzato sono:

- Dublin Core: 12
- MAG: 7
- MARC: 4
- Meta Tag HTML/XHTML: 4
- METS: 2
- CEDARS: 1
- MOA2: 1
- EAD: 1
- PREMIS: 1

Il linguaggio di codifica utilizzato maggiormente è XML (12), meno lo è XHTML (2). Non viene utilizzato SGML.

Il tipo di identificatore più utilizzato è Open URL (5), seguito da HANDLE (3) , DOI (2), PURL (2), in altro sono stati segnalati: URL (2), Identificatori interni (1), ID proprio del software (1).

Nessuno utilizza CIDOC CRM.

L'84,6% utilizzano OAI-PMH.

Tutti (100%) utilizzano set o sistema di metadati base, il 46,1% utilizza un set o sistema di metadati esteso, il 23% utilizza un set o sistema di metadati personalizzato.

Fa coloro che utilizzano un set o sistema di metadati personalizzato è stata data una sola risposta su quale sia il metadato interessato e i relativi qualificatori oggetto della personalizzazione: subject = settori scientifico disciplinari MIUR.

Il 15,4% ha utilizzato mappature/crosswalk con un altro set o sistema di metadati, fra questi c'è stata una sola risposta ad indicare quale set o sistemi di metadati siano stati oggetto della mappatura/crosswalk: MAG-METS mappatura in forma sperimentale.

Osservazioni e Conclusioni

Le osservazioni e conclusioni saranno frutto della discussione sui dati raccolti con il gruppo di esperti del progetto Accesso alle biblioteche digitali.