

# Diffusione dell'Open Source nei Siti Web dei Comuni Italiani

Giuseppe Destefanis, Roberto Tonelli, Giulio Concas, Michele Marchesi

DIEE, Dipartimento di Ingegneria Elettrica ed Elettronica  
Università degli studi di Cagliari  
giuseppe.destefanis@diee.unica.it



**Abstract.** Negli ultimi anni l'utilizzo dei Content Management System (CMS) Open Source nel Web sta aumentando grazie alla flessibilità e alla semplificazione delle operazioni di gestione garantita da tali sistemi. Tante Pubbliche Amministrazioni hanno adottato CMS Open Source per la costruzione dei propri siti web, riducendo i costi operativi e assicurando scalabilità, performance avanzate e garantendo la tutela della privacy agli utenti. Una opportunità per analizzare un interessante caso di studio è data dai comuni italiani. Il Governo Italiano ha imposto alle pubbliche amministrazioni comunali di dotarsi (a partire dal primo Gennaio 2011) di un sito web nel quale pubblicare tutti gli atti relativi all'amministrazione (delibere di Giunta, di Consiglio, ordinanze, bandi, ecc.). Il nostro studio punta a quantificare, attraverso uno studio empirico, la diffusione dei CMS Open Source nei siti web degli 8092 comuni italiani. I risultati ottenuti mostrano che il numero dei comuni italiani che usa CMS Open Source sta aumentando con un incremento di circa il 5% in un anno, passando dal 15.8% del 2011 al 21.2% del 2012.

## 1 Introduzione

Negli ultimi anni il numero e l'importanza delle applicazioni del software libero sono cresciuti oltre le più ottimistiche previsioni e hanno dato vita a numerosi modelli di business che hanno coinvolto le aree applicative più importanti. Anche molte istituzioni governative, nazionali e locali, hanno deciso di promuovere politiche per l'adozione di software Open Source. La scelta del software Open Source può rappresentare una soluzione per i numerosi problemi che si presentano nell'attuazione di una politica dell'ICT. La principale ragione è di ordine economico. Il software libero consente notevoli risparmi sia nel capitolo degli investimenti, sia in quello dei costi di esercizio. Un Content Management System (CMS) è un sistema software progettato per facilitare la gestione di un sito web, consentendone l'amministrazione anche a chi non ha una conoscenza tecnico-informatica approfondita di programmazione orientata al Web. Esistono CMS progettati specificamente per la gestione di tipi particolari di contenuti (ad esempio negozi on line), e CMS generici utilizzati per la pubblicazione di differenti tipologie di contenuti (articoli, immagini, notizie). Tecnicamente, un

CMS è un'applicazione lato server, connessa a un database in cui vengono memorizzati i contenuti, e divisa principalmente in due parti:

- la sezione di amministrazione (back end), utile per gestire la produzione dei contenuti;
- la sezione applicativa (front end), utilizzata dagli utenti per visualizzare i contenuti e le applicazioni del sito.

I CMS possono essere sviluppati utilizzando diversi linguaggi di programmazione, come Python, PHP, Java; il tipo di linguaggio usato non condiziona le funzionalità del sistema. Il campo di utilizzo e di applicazione dei CMS è quello della gestione e sviluppo dei portali web, nel quale si rivelano uno strumento flessibile e multiutente utile per una gestione ottimale di contenuti testuali (news, articoli, liste di discussione) e multimediali (immagini, video, materiale scaricabile). E' possibile modificare in qualsiasi momento la struttura, il numero e l'organizzazione delle pagine. I CMS consentono di definire gruppi di utenti e diritti di accesso specifici in modo da consentire una distribuzione del lavoro tra più persone. E' possibile ad esempio definire una classe di utenti autorizzati all'inserimento delle notizie, riservando la scrittura degli articoli a un altro gruppo, e limitare tutti gli altri alla sola consultazione. Un CMS permette di costruire e gestire siti dinamici di grandi dimensioni senza scrivere una sola riga di HTML e senza conoscere linguaggi di programmazione lato server o tecniche di progettazione di database. L'aspetto delle pagine può essere personalizzato scegliendo un "template" grafico appositamente progettato per un particolare CMS. Molti sono i vantaggi derivanti dall'uso di un CMS Open Source da parte delle Pubbliche Amministrazioni:

- possibilità massima di personalizzazione;
- il codice aperto assicura trasparenza e rende possibile alla Pubblica Amministrazione verificare il reale funzionamento del CMS (analizzando il codice sorgente è possibile avere certezza del comportamento del sistema);
- inesistenza di "lock-in" e completa indipendenza dai fornitori;
- vantaggi derivanti dall'utilizzo di standard aperti (garanzia di accessibilità);
- massima compatibilità con la maggioranza dei browser web;
- aggiornamenti gratuiti e sempre disponibili;
- riduzione dei costi di sviluppo;

Diversi studi [2] sono stati condotti sulla diffusione dei tradizionali software Open Source, ma poco o niente è stato fatto per analizzare la diffusione dell'Open Source nel web. Questa analisi svolta sui comuni italiani ci permette di avere informazioni su un campione abbastanza grande di siti web e ci consente di capire qual'è, allo stato attuale, l'orientamento tecnologico delle pubbliche amministrazioni italiane.

La nostra analisi è stata ripetuta due volte: la prima volta nel Febbraio 2011 e la seconda nel Febbraio 2012, con lo scopo di comprendere il trend di diffusione dei CMS Open Source nel web. Lo studio dimostra che la diffusione dei CMS Open Source sta aumentando; un altro dato importante è dato dal fatto che

le Pubbliche Amministrazioni stanno iniziando a percepire le opportunità (in termini di qualità, sicurezza, prestazioni e risparmio) offerte da tali tecnologie.

## 2 Metodologia

Per poter analizzare la diffusione dei CMS Open Source nei siti web dei comuni italiani, abbiamo sviluppato uno strumento in Java che lavora in due passi. Nella prima fase vengono collezionati e memorizzati in un database tutti gli url dei siti web dei comuni. Gli indirizzi sono stati raccolti interrogando il sito web dell'ANCI [6] (il portale dell'Associazione Nazionale Comuni Italiani, contenente informazioni puntuali e abbastanza aggiornate sui comuni). Nella seconda fase lo strumento interroga il database nel quale sono stati memorizzati gli url dei siti web e inizia il processo di riconoscimento della tecnologia utilizzata per lo sviluppo del sito. I CMS Open Source sono tanti e differenti; per il nostro studio abbiamo deciso di considerare i più comuni e più conosciuti:

- Joomla! [9];
- Drupal [11];
- Wordpress [12];
- Open CMS [13];
- Mambo [14];
- XOOPS [15];
- Plone [16].

Per riconoscere in maniera automatica il tipo di CMS utilizzato, abbiamo provveduto a creare delle installazioni base dei CMS considerati nello studio (senza inserire contenuti personalizzati), in modo tale da avere punti di riferimento per poter confrontare i siti web. Lo strumento, nella seconda fase, analizza le home page dei siti web dei comuni italiani (una alla volta) depurandole (tramite l'utilizzo di espressioni regolari) dai contenuti testuali, e mantenendo soltanto codici HTML, PHP o Javascript. Il codice ottenuto viene confrontato con ciascuna installazione base dei diversi CMS, e viene calcolata la distanza di Levenshtein e di Hamming tra il codice del sito web comunale e il codice dell'installazione base dei CMS. Nella teoria dell'informazione, la distanza di Levenshtein (o distanza di edit), è una misura utile per calcolare le differenze tra due stringhe, e viene utilizzata per capire quanto due stringhe sono simili tra loro. La distanza di Levenshtein tra due stringhe è il numero minimo di cambiamenti elementari per convertire una stringa  $x$  in una stringa  $y$ . I cambiamenti elementari possono essere di tre tipi:

- cancellazione di un carattere;
- rimpiazzamento di un carattere con un altro;
- inserimento di un carattere;

La distanza di Levenshtein ha le seguenti proprietà:

- se le lunghezze delle due stringhe sono uguali, la distanza di Levenshtein non supera la distanza di Hamming;
- per due stringhe di lunghezze differenti il limite superiore è la distanza di Hamming aumentata della differenza delle lunghezze;
- è almeno la differenza fra le lunghezze delle due stringhe;
- non supera la lunghezza della stringa più lunga;
- è zero se e solo se le due stringhe sono identiche;

La distanza di Hamming [7] tra due stringhe di lunghezza identica è il numero di posizioni nelle quali i simboli corrispondenti sono diversi. La distanza di Hamming misura il numero di sostituzioni necessarie per convertire una stringa nell'altra. Il peso di Hamming di una stringa è la sua distanza di Hamming dalla stringa nulla (stringa costituita solo da zeri) della stessa lunghezza. Per una fissata lunghezza  $n$  la distanza di Hamming è una metrica sullo spazio vettoriale delle stringhe aventi quella lunghezza, poichè soddisfa le condizioni di non negatività, identità di due elementi aventi distanza nulla, simmetria, e si può dimostrare mediante induzione completa che essa soddisfa anche la disuguaglianza triangolare.

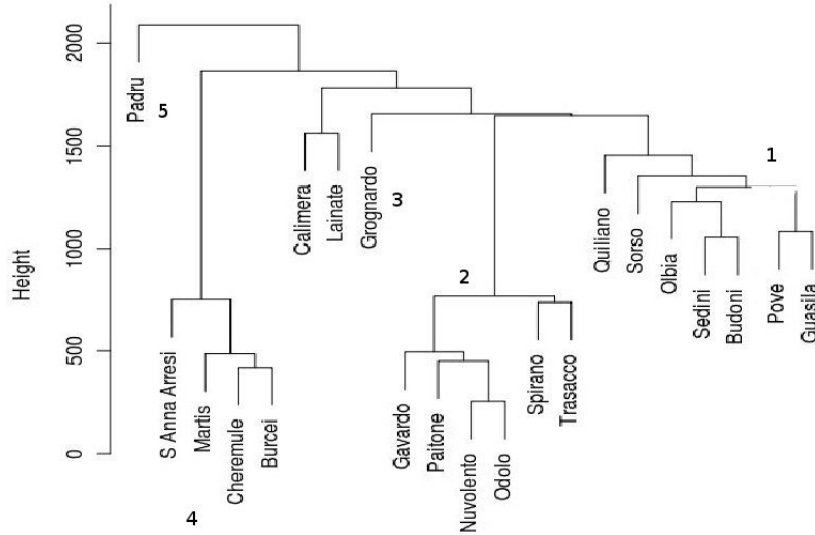
Successivamente, per migliorare la qualità del riconoscimento del tipo di CMS utilizzato, viene fatta una ricerca di parole chiave (come Joomla!, Drupal, ecc.) nel codice sorgente della pagina web, e in caso di presenza di queste parole viene diminuito il valore di distanza con le installazioni base dei sistemi. Come terzo passaggio, viene effettuato un clustering gerarchico tra tutti i siti e le installazioni base dei CMS in modo tale da raggrupparli in insiemi simili per evidenziare le caratteristiche comuni.

Come esempio riportiamo (in Fig. 1) un dendrogramma ottenuto confrontando i siti web di 21 comuni, i quali sono rappresentativi della situazione generale. Il dendrogramma è stato creato per dimostrare il funzionamento dell'algoritmo utilizzato. Dalla figura si nota che è possibile individuare 5 cluster. Il cluster 1 è quello dei siti creati con Joomla!, i cluster 2 e 3 sono formati da siti sviluppati utilizzando Drupal (la distanza tra i due cluster è dovuta al fatto che sono state utilizzate due versioni differenti di Drupal, e questo dimostra che l'algoritmo è in grado di riconoscere e classificare anche differenti versioni di uno stesso CMS). Il cluster 4 è riferito ai siti web creati da una stessa azienda, e si tratta di siti web che non utilizzano CMS Open Source, mentre il cluster 5 (formato da un solo elemento) riguarda un sito web che differisce tanto dagli altri. Dopo un controllo manuale, abbiamo verificato che si tratta di un sito web statico, abbastanza vecchio e che non fa uso di nessun tipo di CMS (semplici pagine HTML).

### 3 Risultati

Il tool da noi sviluppato analizza tutti i siti web degli 8092 comuni italiani e verifica se sono stati utilizzati CMS Open Source, scelti tra i seguenti:

Fig. 1. Cluster Dendrogram



- Joomla!;
- Drupal;
- Wordpress;
- Open CMS;
- Mambo;
- XOOPS;
- Plone.

Abbiamo condotto due analisi: una nel Febbraio 2011 e una nel Febbraio 2012. I risultati sono mostrati in tabella 1.

Essi mostrano una crescita dei CMS Open Source nelle pubbliche amministrazioni comunali, con una penetrazione dei CMS open source pari al 21.2% nel 2012, in aumento del 5% se si confronta il valore del 2011 (15.8% è il valore registrato nel 2011). La colonna "% crescita" indica la crescita percentuale di ogni CMS Open Source sul totale dei siti web analizzati. Le ultime due colonne invece mostrano il peso relativo di ogni CMS analizzato nella categoria dei CMS Open Source. La categoria "Minors" contiene la somma dei siti che utilizzano PhpBB [17] e Japs [18] (con una bassa percentuale di utilizzo).

Joomla! risulta essere il CMS che contribuisce maggiormente a guidare la crescita dei CMS Open Source, seguito subito da Drupal, Wordpress e Open CMS. Il dato negativo mostrato da Mambo è dovuto al fatto che si tratta di una distribuzione obsoleta (sostituita da Joomla!) e per la quale è consigliata

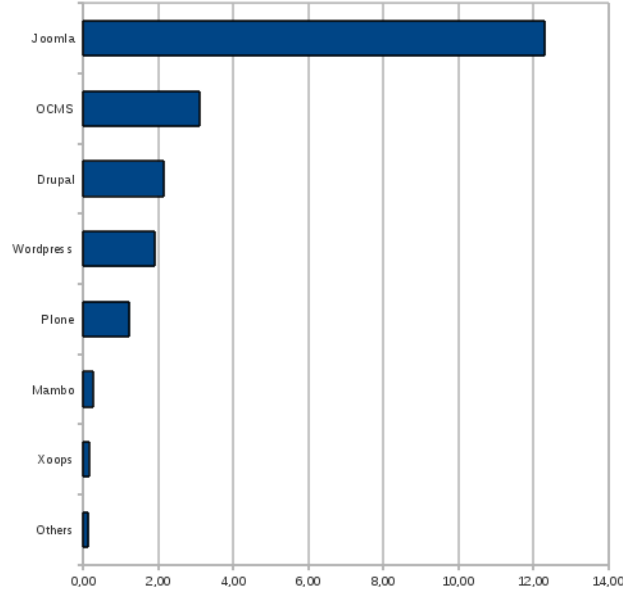
**Table 1.** Risultati

	% 2011	% 2012	% crescita	% 2011	% 2012
Drupal	1,36	2,15	0,8	8,6	10,1
Joomla!	8,77	12,32	3,5	55,4	57,9
Mambo	0,47	0,25	-0,2	3,0	1,2
OCMS	2,58	3,10	0,5	16,3	14,6
Plone	1,03	1,23	0,2	6,5	5,8
Wordpress	1,36	1,91	0,5	8,6	9,0
Xoops	0,20	0,15	0,05	1,2	0,7
Minors	0,06	0,14	0,1	0,4	0,7
totale	15,8	21,2	5,4	100	100
Altri	84,17	78,8			

la migrazione verso Joomla!. Nell'insieme dei CMS Open Source analizzati c'è una forte crescita di Joomla! che era già predominante nel 2011, e una netta crescita di Drupal (tasso di crescita del 15%). Al contrario, Open CMS, il quale ancora cresce e ha contribuito alla crescita di utilizzo dei CMS Open Source nel 2012, ha diminuito la sua rappresentatività all'interno dell'insieme dei CMS Open Source analizzati. Ciò significa che:

- nella competizione tra i CMS Open Source e il resto delle soluzioni usate, i primi stanno emergendo;
- nella competizione tra i CMS Open Source analizzati, Joomla! si conferma forte, mentre Open CMS sta perdendo terreno rispetto ai suoi concorrenti; Drupal guadagna notevolmente, e Wordpress aumenta leggermente (in maniera simile a Joomla!).

Il motivo della predominanza di Joomla! va attribuito al fatto che si tratta di un sistema molto conosciuto e completo, per il quale ci sono molti plug-in utilizzabili per qualsiasi tipo di funzionalità richiesta per il sito web [10]. Si tratta di un CMS estremamente semplice da usare da parte di utenti poco esperti, è completamente compatibile con tutti i tipi di browser web ed è compatibile con tutti i requisiti di accessibilità. Essendo uno dei CMS più utilizzati è soggetto ad esternalità di rete: ci sono molti tutorial, manuali, forum, e molti casi d'uso ben spiegati. Le motivazioni che spiegano la crescita di Drupal sono simili a quelle di Joomla!, e nell'ultimo periodo Drupal ha goduto di diversa pubblicità da parte di importanti enti governativi: basti pensare alla pubblicità fatta dalla Casa Bianca Americana [8]. La rapida crescita di Wordpress va probabilmente attribuita alla semplicità delle procedure di installazione del sistema. Si tratta di un'operazione che non richiede competenze tecniche (a differenza di Joomla! e Drupal, ad esempio) e sta crescendo la diffusione di template grafici e di plug-in studiati appositamente per Wordpress. La decrescita di Open CMS va probabilmente attribuita al tipo di linguaggio di sviluppo utilizzato: è scritto

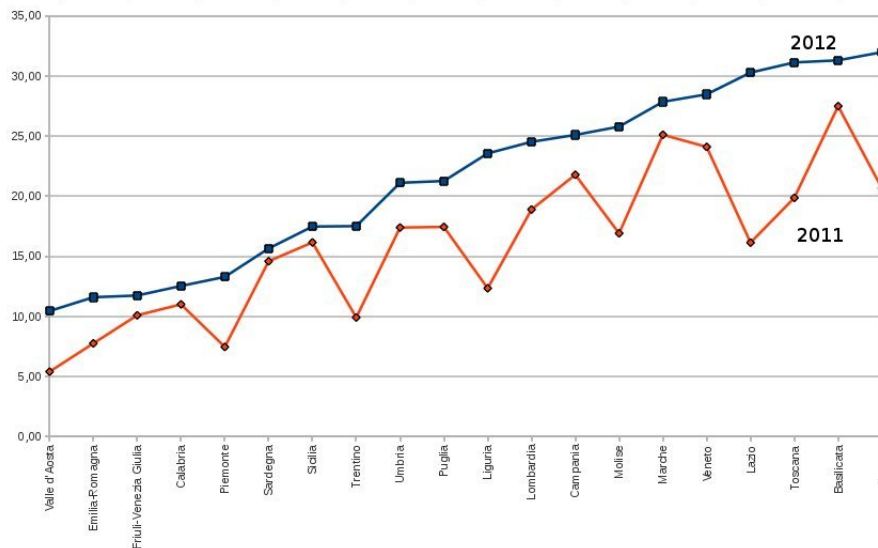
**Fig. 2.** Percentuale dei CMS Open Source nel 2012

in Java e ciò comporta una richiesta di forti conoscenze tecniche da parte degli utenti finali. Il grafico in Fig. 3 mostra la percentuale di presenza di CMS Open Source per regione.

Un aspetto interessante dello studio (che potrà essere esteso a qualsiasi tipo di pubblica amministrazione) è legato al Total Cost of Ownership [1] dei siti Web. L'uso di un CMS open source comporta un risparmio che può essere stimato approssimativamente in circa 50.000 euro per il semplice sito web di un comune di 200.000 abitanti. A questi costi va aggiunta la ricapitalizzazione delle competenze del personale e dei costi di sviluppo annuali (che possono essere stimati in circa il 30% del valore riportato in precedenza). Il relativo valore aggiunto è legato al fatto che le competenze acquisite da parte del personale possono essere reinvestite per la gestione di altri portali web che esistono nella rete comunale (come portali per promuovere il turismo, biblioteche, circoscrizioni, ecc.).

#### 4 Minacce alla validità

Le minacce alla validità dello studio da noi condotto sono relative alla generalizzazione delle conclusioni. Abbiamo considerato un insieme limitato di CMS Open Source e abbiamo studiato soltanto il caso dei comuni italiani (non considerando altre istituzioni pubbliche come scuola, università ecc.) e questo può ledere la generalità della trattazione.

**Fig. 3.** Percentuale di diffusione dei CMS Open Source per regione

## 5 Conclusioni

Gli obiettivi di questa ricerca sono stati quelli di analizzare la diffusione dei CMS Open Source nei comuni italiani. Abbiamo sviluppato uno strumento Java che lavora in due fasi. Durante la prima fase vengono memorizzati in un database gli URL dei siti web di tutte le amministrazioni comunali italiane (8092 comuni), recuperati dal sito web dell'ANCI [6]. Durante la seconda fase viene interrogato il database creato nella prima fase e si esegue l'elaborazione del contenuto del sito web. Infine, si applica un clustering gerarchico tra tutti i siti e le installazioni di base del CMS, al fine di raggrupparli in insiemi simili e in modo da evidenziarne le caratteristiche comuni. Abbiamo condotto due analisi, una nel mese di Febbraio 2011 e una nel mese di Febbraio 2012 e i risultati mostrano che il numero di comuni italiani che utilizzano CMS Open Source sono in aumento: il 21,2% delle amministrazioni comunali utilizza CMS Open Source, come osservato nel mese di febbraio 2012. I dati mostrano che dopo un anno la diffusione dei CMS Open Source è aumentata del 5%, e si prevede una crescita anche per l'immediato futuro.

## References

1. Shaikh, Maha and Cornford, Tony (2011) Total cost of ownership of open source software: a report for the UK Cabinet Office supported by OpenForum Europe. UK Cabinet Office, London, UK.



2. Rishab Aiyer Ghosh: Study on the Economic impact of open source software on innovation and the competitiveness of the Information and Communication Technologies (ICT) sector in the EU, UNU-MERIT, (2006).
3. Are government internet portals evolving towards more interaction, participation, and collaboration? Revisiting the rhetoric of e-government among municipalities. Rodrigo Sandoval-Almazan, J. Ramon Gil-Garcia. 27 November 2011
4. Global Trends in Municipal E-Government: An Online Assessment of Worldwide Municipal Web Portals. Marc Holzer and Aroon Manoharan.
5. A Content Analysis of Selected Government Web Sites: a Case Study of Nepal. Jitendra Parajuli, High Level Commission for Information Technology, Kathmandu, Nepal
6. <http://www.anci.it>
7. [www.ippari.unict.it/infapp/](http://www.ippari.unict.it/infapp/)
8. <http://www.whitehouse.gov/blog/2011/02/11/whitehousegov-releases-second-set-open-source-code>
9. [www.joomla.org](http://www.joomla.org)
10. <http://extensions.joomla.org/>
11. [www.drupal.org](http://www.drupal.org)
12. [www.wordpress.org](http://www.wordpress.org)
13. [www.opencms.org](http://www.opencms.org)
14. [www.mambo-foundation.org](http://www.mambo-foundation.org)
15. [www.xoops.org](http://www.xoops.org)
16. [www.plone.org](http://www.plone.org)
17. [www.phpbb.com](http://www.phpbb.com)
18. [www.japsportal.org](http://www.japsportal.org)