Integrazione modulare di servizi informativi su web con software Open Source

Guglielmo Cresci {Guglielmo.Cresci@isti.cnr.it}
CNR - ISTI - Area della ricerca CNR, via G. Moruzzi 1, 56124 PISA, Italy

Diana Lari {Diana.Lari@isti.cnr.it}
CNR - ISTI - Area della ricerca CNR, via G. Moruzzi 1, 56124 PISA, Italia

Marco Righi {Marco.Righi@isti.cnr.it;righi@di.unipi.it;marco.righi@lipn.univ-paris13.fr}
CNR - ISTI - Area della ricerca CNR, via G. Moruzzi 1, 56124 PISA, Italia
Università di Pisa - Dipartimento di Informatica - Largo B. Pontecorvo, 3 56127 Pisa, Italia
Laboratoire d'Informatique de l'Université Paris Nord - Institut Galilée - Avenue J.B. Clément 93430 Villetaneuse - Francia

Abstract

1 Sommario

Questo documento presenta il risultato dello sviluppo di un ambiente per la gestione e dei servizi informativi di un Istituto CNR multi sede (IAMC: Istituto per l'Ambiente Marino Costiero) integrato nella Intranet del portale dell'Istituto. Il sistema è destinato a gestire i servizi gestionali in modo modulare e flessibile ed è progettato per essere esteso sia sviluppando indipendentemente singoli moduli in base alle necessità, sia replicando parti della struttura presso altre sedi di IAMC.
Uno degli aspetti chiave dell'implementazione risiede nell'architettura della Intranet che realizza l'integrazione tra i diversi moduli.
Ognuno di questi può essere sviluppato con una tecnologia diversa da quella degli altri pur presentandosi all'utente con un'interfaccia grafica omogenea come se fossero una unica suite.
La Intranet implementa il meccanismo della Single Sign-On (SSO) concentrando in un unico repository gli elementi che caratterizzano i singoli utenti, tra cui il ruolo che identifica i diritti di accesso alle varie applicazioni.
L'implementazione utilizza soltanto software freeware ed Open Source, in particolare:

- CentoOS 5.5 [6][7][8] come sistema operativo,
- Apache Tomcat 5.27 [9][10] come application server,
- Apache HTTP Server 2 [6][11][12][13] come web server,
- OpenLDAP 2 [14][15] come server LDAP,
- Plone 3 [16][17][18] come CMS,
- Shibboleth 2 [19][20] come sistema di identity management.

Le applicazioni finora realizzate utilizzano i linguaggi di programmazione JAVA [21], PHP [22], Javascript [23] e Python [24].
È stato scelto di utilizzare Shibboleth per lo sviluppo del sistema di identity management con il fine di seguire le indicazioni date dal progetto IDEM, patrocinato dal GARR [4].

2 Le applicazioni ed il loro obiettivo

Il portale pubblico è stato progettato per essere accessibile ed in grado di gestire contenuti multimediali e geografici. Ai sensi della legge italiana la gestione dei contenuti, per essere accessibile, deve tutelare i diritti di ogni individuo, infatti il primo articolo delle raccomandazioni relative all'accessibilità cita “La Repubblica riconosce e tutela il diritto di ogni persona ad accedere a tutte le fonti di informazione e ai relativi servizi, ivi compresi quelli che si articolano attraverso gli strumenti informatici e telematici. È tutelato e garantito, in particolare, il diritto di accesso ai servizi informatici e telematici della pubblica amministrazione e ai servizi di pubblica utilità da parte delle persone disabili, in ottemperanza al principio di uguaglianza ai sensi dell'articolo 3 della Costituzione.”[1][2]
Per la scelta del CMS atto a contenere i dati pubblici sono state prese in considerazioni varie soluzioni basate su CMS freeware, Open Source aderenti ai criteri di accessibilità. Nella scelta del CMS sono state molto utili le indicazioni e gli strumenti presenti su cmsmatrix [3] che recensisce moltissimi ambienti di questo tipo evidenziandone in modo chiaro e schematico pregi e difetti.
La disponibilità dei plug-in è stata un ulteriore elemento di valutazione utilizzato per l’individuazione del CMS. Era infatti richiesto uno strumento potente e flessibile per la pubblicazione e indicizzazione di vari tipi di documenti: da semplici pagine web, a documenti più complessi come PDF, immagini fotografiche e filmati (funzionando da video streaming server). Era anche richiesto il supporto multilingue incluso l’arabo.
Infine il CMS Plone è in grado di integrarsi completamente con Shibboleth e di gestire utenti e gruppi condivisi con le altre applicazioni residenti su un server LDAP.
Nell'architettura realizzata Plone è l'unica interfaccia per la pubblicazione delle informazioni: i servizi della Intranet hanno invece la funzione di gestire il flusso informativo interno alla struttura.
A fronte delle notevoli funzionalità, lo strumento Plone presenta una serie di limiti e difficoltà per lo sviluppatore, quali la complessità interna, in parte motivata dalla molteplicità di soggetti che contribuiscono al suo sviluppo, e soprattutto la scarsa documentazione. Queste caratteristiche hanno richiesto un impiego di risorse notevole per lo sviluppo.
La mancanza di documentazione deve essere un invito a riflettere al progetto PloneGov [5] e sull'impegno che in futuro dovranno mettere gli sviluppatori di Plone per produrre/mantenere documentazione.
L’architettura del portale web prevede una sezione pubblica gestita attraverso Plone e una sezione protetta, denominata intranet, ben distinta dall’altra, non gestita da un CMS e integrata con le applicazioni del sistema informativo. Strettamente integrata con la intranet è l’applicazione di gestione degli utenti che

consente la creazione / cancellazione / modifica di un utente e di tutti i suoi attributi: da nome e password per l'accesso al sistema, al ruolo che identifica i diritti per l'accesso alle applicazioni del sistema informativo e al portale.
Questo programma si interfaccia direttamente con un server LDAP utilizzato come archivio per la gestione di tutte le informazioni degli utenti. Al fine di centralizzare tutte le informazioni relative agli utenti in un unico repository è stato modificato il normale funzionamento di Plone bypassando il suo sistema della gestione degli utenti: l'autenticazione avviene direttamente sfruttando Shibboleth mentre il riconoscimento delle caratteristiche di un utente avviene leggendo direttamente i dati in LDAP.
Poiché Plone e la intranet hanno una diversa classificazione dei ruoli degli utenti, LDAP viene utilizzato come unificatore: esso contiene le informazioni per distinguere il gruppo di afferenza di un utente sia quando accede al portale pubblico sia quando accede alla Intranet.
L'applicazione di gestione degli utenti, con il fine di garantire un adeguato livello di sicurezza, ogni qual volta che sono modificati i diritti di accesso o le generalità di un utente, avvisa l'utente tramite una e-mail e l'operazione viene registrata in un apposito giornale. Il giornale è visibile da Web per mezzo dell'applicazione sviluppata; ogni utente vede le informazioni relative alla propria configurazione. L'amministratore vede tutte le operazioni e tutte le caratteristiche di un utente ad eccezione della password.
Poiché il server LDAP è soggetto ad operazioni concorrenti e non atomiche, si è reso necessario gestire in modo esplicito la concorrenza. La gestione della concorrenza avviene attualmente sfruttando i meccanismi di accesso in mutua esclusione forniti dal sistema operativo CentOS 5.5.
In previsione di un possibile sviluppo di applicazioni distribuite per la gestione degli utenti (su più calcolatori) è prevista l'implementazione di un monitor con il fine di serializzare le richieste.
Il servizio di Single Sign-On è stato realizzato per permettere con una unica azione di login l'accesso a tutte le aree riservate (la gestione dei diritti in ogni area è delegata alle singole applicazioni).
Lo studio delle soluzioni di Single Sign-On presenti sul mercato ci ha condotto ad utilizzare Shibboleth per i motivi di seguito riportati:

- è uno strumento flessibile capace di gestire reti federate, permette quindi una elevata modularità nello sviluppo di più sistemi
- è freeware e Open Source
- è conforme alle specifiche dell'infrastruttura di Autenticazione e Autorizzazione federata della rete Garr
- è diffuso in molti contesti pubblici e privati

Inoltre Plone presenta opportuni plug-in capaci interfacciarvisi con il fine di autenticare gli utenti sfruttando una connessione crittata.

3 Considerazioni architetturali e possibili sviluppi

L'architettura logica del sistema è flessibile e modulare e si presta ad essere coniugata in un ampio spettro di soluzioni architetturali: da quelle monolitiche, ospitate su un singolo server, a situazioni più articolate a livello logico e fisico. Sono state svolte approfondite indagini sulle possibili architetture al fine di studiare l'estensione di quanto realizzato per la sede di Capo Granitola alle altre 6 sedi dell'Istituto prevedendo modelli organizzativi diversi con sedi autonome nella gestione del proprio sistema Hw / Sw ed altre che condividono uno stesso ambiente logico e fisico.
In pratica l'architettura implementata consente di operare, con marginali interventi di personalizzazione, su uno o più sistemi con dati (ad esempio l'archivio degli utenti) allocati su un unico archivio (nel caso dell'archivio degli utenti un server LDAP) o frazionato su più archivi che eseguono su altrettanti sistemi elaborativi (sempre per l'archivio degli utenti tanti server LDAP, ognuno dei quali contiene una parte dell'archivio, ed è gestito in autonomia).
Considerazioni analoghe a quelle riportate per l'applicativo della gestione degli utenti, si applicano ad altri componenti del sistema: in particolare si applicano alla gestione dei contenuti. Questi sono memorizzati nel web container Zope e gestiti per mezzo del CMS Plone. L'utilizzo di Plone permette di avere per ogni singolo utente un ambiente di lavoro in grado di offrire editing testuale e condividere documenti testuali, immagini e filmati. Queste funzioni sono naturalmente fruibili dall'utente indipendentemente dalla propria locazione geografica. Inoltre l'utilizzo di un sistema di astrazione come Zope permette di svincolare l'applicazione usata dall'utente dai dati dando così la possibilità di deallocarli e riallocarli liberamente, garantendo le caratteristiche flessibilità, scalabilità e efficienza che oggi sono rese necessarie per gestire l'evoluzione dei sistemi informatici e non solo.
Lo scenario che si configura è sostanzialmente coerente con il modello di elaborazione generalmente riferito come "Cloud computing" e, in particolare, con la tipologia nota come "Software as a Service".

4 Conclusioni

Il sistema realizzato sfrutta tecnologie aggiornate e innovative in grado di integrare ambienti eterogenei in accordo con standard di accessibilità e multilinguismo.
Il controllo degli accessi tramite Single Sign-On permette l'integrazione tra le varie applicazioni e si propone anche per future estensioni, ad esempio, per accogliere applicazioni che attualmente sono gestite indipendentemente in altri ambienti elaborativi.
L'architettura, oltre a rispecchiare elevati standard di sicurezza, è modulare e permette di scalare il sistema su più macchine con varie possibili articolazioni (ad es. un portale per macchina o più portali su una stessa macchina).
Gli aspetti di sicurezza rispecchiano elevati standard di qualità facendo sì che tutte le informazioni sensibili viaggino su connessioni crittate.

Una alpha-release del portale alla data della redazione di questo contributo è presente sul server di sviluppo di ISTI al seguente indirizzo:

- http://ictserver.isti.cnr.it/IAMC

Nei prossimi mesi il portale sarà accessibile sul sistema di esercizio alla URL:

- http://www.iamc.cnr.it/

5 Bibliografia

[1] Legge Stanca - 9 gennaio 2004, n. 4
[2] www.pubbliaccesso.gov.it/normative/legge_20040109_n4.htm (attivo in data 20 maggio 2010)
[3] www.cmsmatrix.org
[4] https://www.idem.garr.it/
[5] www.plonegov.it
[6] http://a2.pluto.it/
[7] www.centos.org
[8] Peter Membrey, Tim Verhoeven, and Ralph Angenendt; The Definitive Guide to CentOS Paperback; 2009
[9] http://tomcat.apache.org/
[10] Jason Brittain, Ian F. Darwin ; Tomcat: The Definitive Guide; Paperback; 2007
[11] http://httpd.apache.org/
[12] Ben Laurie; Apache: The Definitive Guide (3rd Edition); Paperback; 2002
[13] Ivan Ristic; Apache Security; Paperback; 2005
[14] http://www.openldap.org/
[15] Gerald Carter ; LDAP System Administration; Paperback; 2003
[16] http://plone.org/
[17] Redomino, Andy McKay; The Definitive Guide to Plone (2nd Edition); Paperback; 2009
[18] Sam Knox, Jon Stahl, Martin Aspeli, and David Convent; Practical Plone 3: A Beginner's Guide to Building Powerful Websites; Paperback; 2009
[19] http://shibboleth.internet2.edu/
[20] http://www.garr.it/eventiGARR/idem09/
[21] http://www.java.com
[22] http://php.net/
[23] http://www.w3schools.com/js/default.asp
[24] http://www.python.org/