Case Study

-

Portale del Garante per la Protezione dei Dati Personali

11 Giugno 2012

Il Portale del Garante per la protezione dei dati personali è ancora in fase sperimentale, non è aperto alla pubblica consultazione



Garante Privacy

- Il **Garante per la Protezione dei Dati Personali** è un'autorità amministrativa indipendente istituita dalla legge n. 675 del 31 dicembre 1996 (cosiddetta legge sulla privacy), per assicurare la tutela dei diritti e delle libertà fondamentali e il rispetto della dignità nel trattamento dei dati personali
- Etcware si è aggiudicata la gara per lo sviluppo e la manutenzione del sito web ufficiale del Garante per la protezione dei dati personali
- Esigenze da soddisfare:
  - Rinnovamento dell'architettura tecnologica (alta affidabilità)
  - Nuovo layout grafico caratterizzato da maggiore flessibilità
  - Massima attenzione all'aderenza ai principi dell'*information retrieval* e alla visibilità delle relazioni logico-giuridiche tra i provvedimenti
  - Realizzazione di un motore di ricerca che sfrutti a pieno gli schemi di classificazione e i principi del Web Semantico
  - Aderenza ai principi di accessibilità e usabilità per garantire la possibilità di accesso anche ad utenti diversamente abili



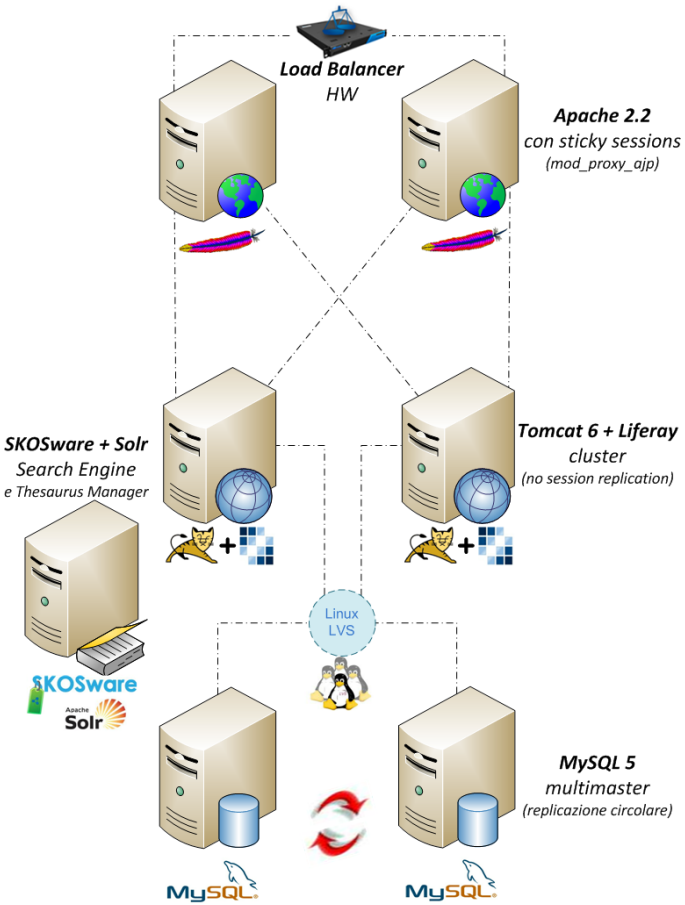
Attività e requisiti

Attività:

- Progettazione logico-fisica della soluzione
- Installazione, configurazione e adattamento di un Content Management System (CMS) stabile, maturo e affidabile (**Liferay**)
- Interfaccia Web di navigazione multilingua e con predisposizione per Homepage multiple con caratteristiche di elevata accessibilità
- Motore di ricerca avanzato (*faceted searches*) con caratteristiche di classificazione semantica dei contenuti
- Integrazione delle funzionalità mancanti nel CMS:
  - geolocalizzazione dei contenuti
  - georeferenziazione dei contenuti
  - integrazione con Apache Solr come default *search engine*
  - integrazione di un motore di classificazione semantico (argomenti come *facet*)
- Produzione del nuovo layout grafico (HomePage multiple intercambiabili, layout multipli, ecc.) in conformità alla Corporate Identity dell'Amministrazione
- Gestione del pregresso (importazione della base dati dei provvedimenti completi di metadati ed allegati binari all'interno del CMS e mantenimento degli URI)
- Supporto alle attività di testing , collaudo, configurazione e avviamento in produzione
- Servizi di manutenzione ordinaria,correttiva ed evolutiva e assistenza sistemistica specialistica
- Il Capitolato Tecnico contiene l'espressione di **105 (!!!)** requisiti di conformità da validare al collaudo



Architettura



- L'Amministrazione ha messo a disposizione 7 Server RHEL 6.1 su architettura virtualizzata VMWare con almeno 3 GB di RAM ciascuno (sugli Application Server la memoria è stata aumentata a più di 5 GB)

La configurazione logico-fisica prevede:

- 1 Load Balancer hardware (centralizzato)
- 2 Web Server Apache 2.2 configurati con sticky sessions e mod_proxy_ajp per il bilanciamento degli AS
- 2 Application Server Tomcat 6 con Liferay Portal Server in configurazione clusterizzata senza replicazione delle sessioni (per motivi di performance)
- 2 Database MySQL 5 in configurazione multimaster a replicazione circolare (IP virtuale con Linux LVS)
- 1 Application Server JBoss 6 che fornisce:
  - il servizio di *search engine* integrato (Apache Solr)
  - il servizio di gestione di classificazione semantica (SKOSware)
  - altri strumenti di supporto:
    - monitoraggio dello stato dei Server
    - collezione delle statistiche di accesso al Portale (Piwik)
    - strumenti per la verifica della consistenza dei link all'interno delle pagine del Portale e per l'invio di report periodici via mail delle info collezionate



SKOSware

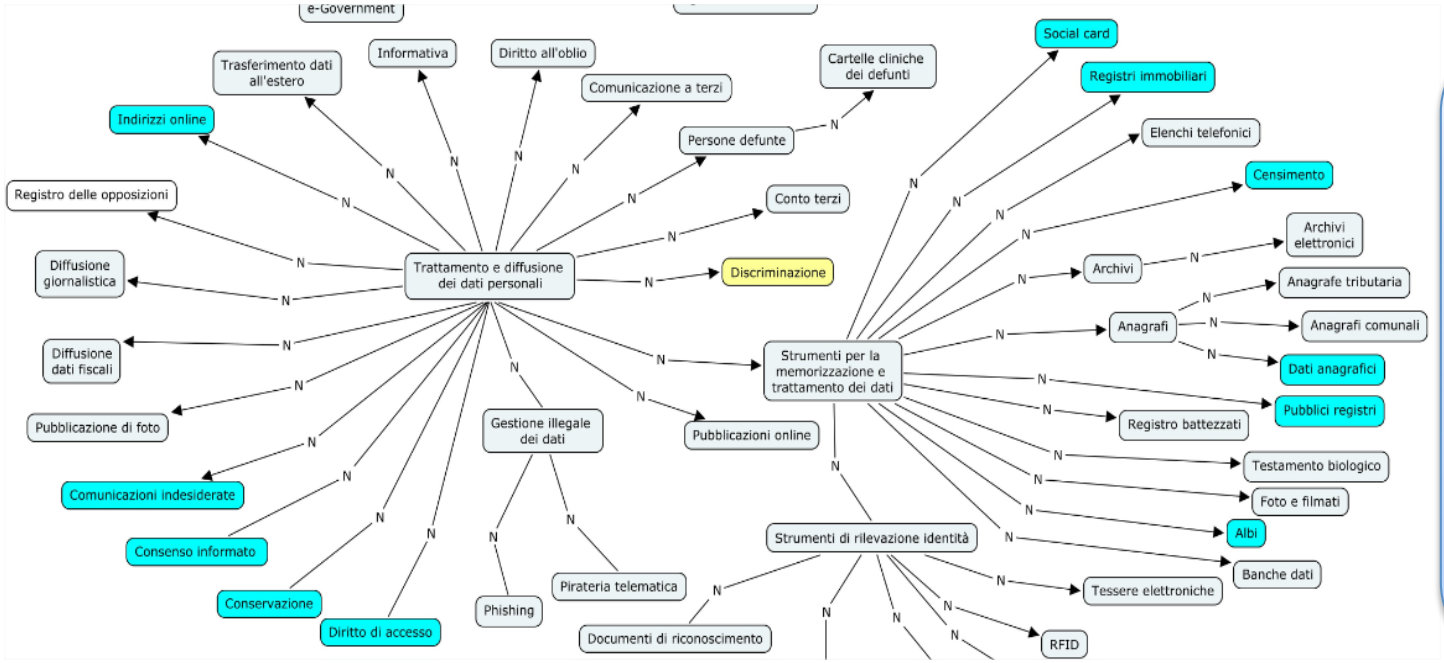
- SKOSware permette la completa gestione del tesauro del Garante, una sorta di grafo delle categorie concettuali pertinenti ai provvedimenti emanati e alle materie trattate dal Garante per la Protezione dei Dati Personali
  - E' possibile manipolare i "concetti", attribuendo loro delle definizioni univoche preferenziali e associare ad essi una serie di sinonimi (o note)
  - E' possibile collegare i concetti tra loro, indicando la relazione semantica che esiste tra di essi, come la generalizzazione (e il suo contrario) e la correlazione
- SKOSware **è stato integrato con Liferay** per consentire la classificazione semantica dei contenuti del Portale
- SKOSware **è stato integrato anche con il motore di ricerca Solr** in modo da consentire di restringere o estendere le ricerche effettuate attraverso *facet* gerarchiche collegate con il dizionario semantico dei concetti
- Grazie alle interfacce REST che il prodotto espone e a strumenti di *text analysis* è stato possibile realizzare uno strumento di catalogazione automatica di TUTTE le migliaia di provvedimenti emanati dal Garante in questi anni (manualmente avrebbe richiesto uno sforzo di mesi)

SKOSware è un prodotto Etcware



Mappe concettuali

- Per la modellazione iniziale del tesauro del Garante, nell'ottica di offrire al personale dell'Amministrazione uno strumento visuale per la modellazione della propria **mappa concettuale**, è stato usato il prodotto IHMC Cmap Tools (IHMC = Institute for Human and Machine Cognition, www.ihmc.us)



- Seguendo un semplice standard di compilazione, è stato possibile importare direttamente la mappa concettuale all'interno di SKOSware

- SKOSware include infatti procedure automatiche in linguaggio Java per l'importazione di tesauri in formato SKOS da alcuni formati similari (ma non direttamente compatibili)



Integrazione con Solr

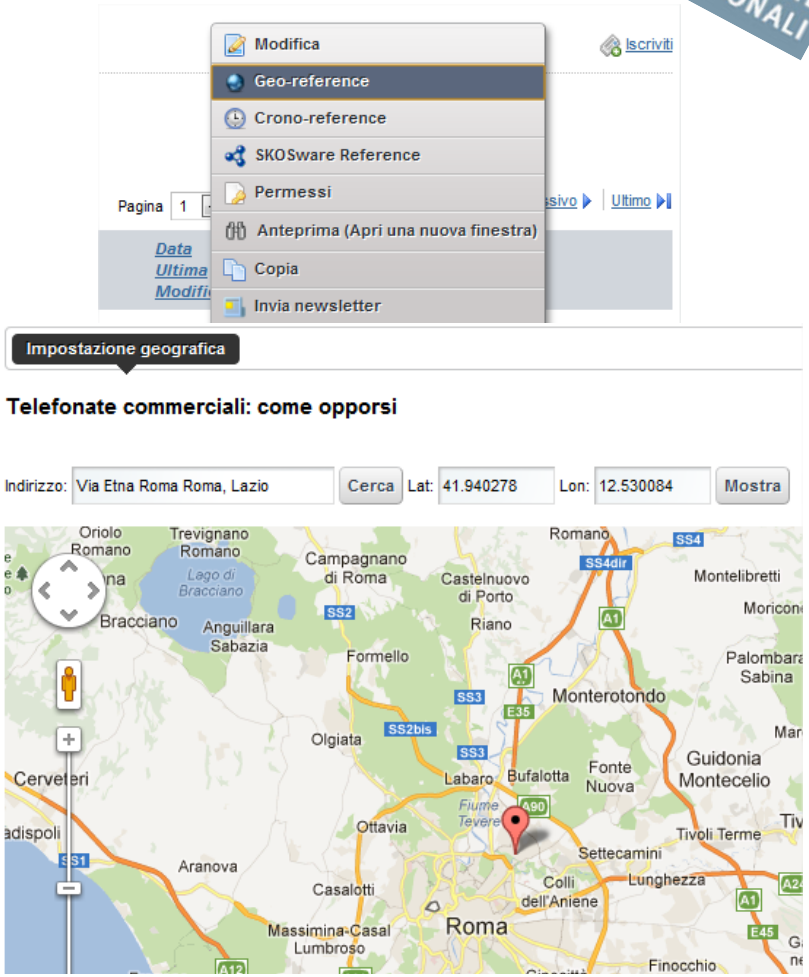
- **Apache Solr** (lucene.apache.org/solr) è un motore di ricerca con caratteristiche Enterprise (clusterizzazione, indici di ricerca multipli per basi documentali disgiunte, ecc.), che consente tipologie di ricerca molto avanzate (ricerche condizionali in AND e OR, ricerche di prossimità, ricerche per similarità, ricerche *fuzzy*, ecc.), e che rende accessibile l'intera gamma delle proprie funzionalità in modalità REST
- Lo schema di Solr è stato modificato per poter gestire i **metadati** associati ai provvedimenti del Garante (Docweb ID, titolo, abstract, corpo, autore, data di pubblicazione, ecc.)
- Sono state utilizzate classificazioni dei risultati delle ricerche basate su *facet* (tipologia, argomento e data di pubblicazione) per consentire successivi raffinamenti
- Le ***facet* gerarchiche associate agli argomenti** trattati sono collegate direttamente alle categorie concettuali espresse nel tesauro SKOS del Garante





- Backend - geolocalizzazione

- Abbiamo eseguito <u>significative personalizzazioni di Liferay</u> anche nell'area di **backend amministrativo** per l'inserimento e la modifica dei contenuti
- Grazie a questi interventi è possibile impostare la **localizzazione geografica** di pertinenza per eventi, articoli e provvedimenti
- E' anche possibile effettuare ricerche in base ai nomi dei luoghi e quindi utilizzare meccanismi di "*reverse geocoding*"
- In questo modo è possibile mostrare la **mappa del luogo** associato al contenuto con **Google Maps**

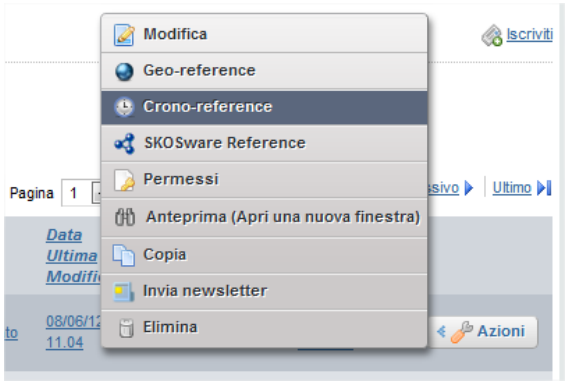




- Backend - cronoreferenziazione

- Sul backend del Portale del Garante è anche possibile impostare la **referenziazione temporale** o cronoreferenziazione di eventi, articoli e provvedimenti
- In questo modo è possibile mostrare la **collocazione temporale** della lista di documenti corrispondente ai risultati di una ricerca su una *timeline* interattiva scorrevole realizzata in Javascript

  In particolare per la visualizzazione è stato utilizzato il widget "Timeline" proveniente dal SIMILE Project (SIMILE = Semantic Interoperability of Metadata and Information in unLike Environments)

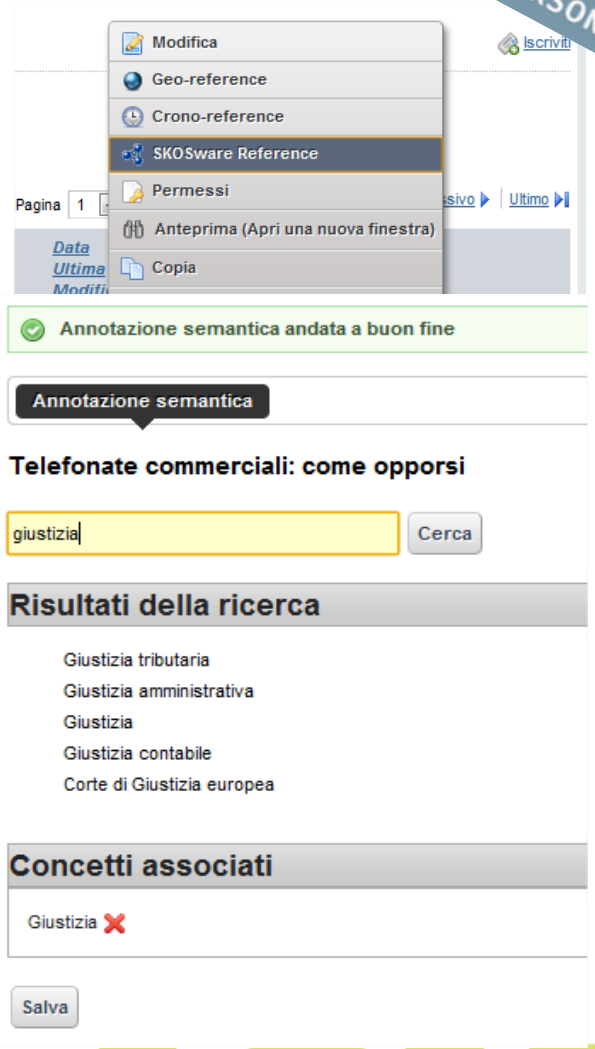






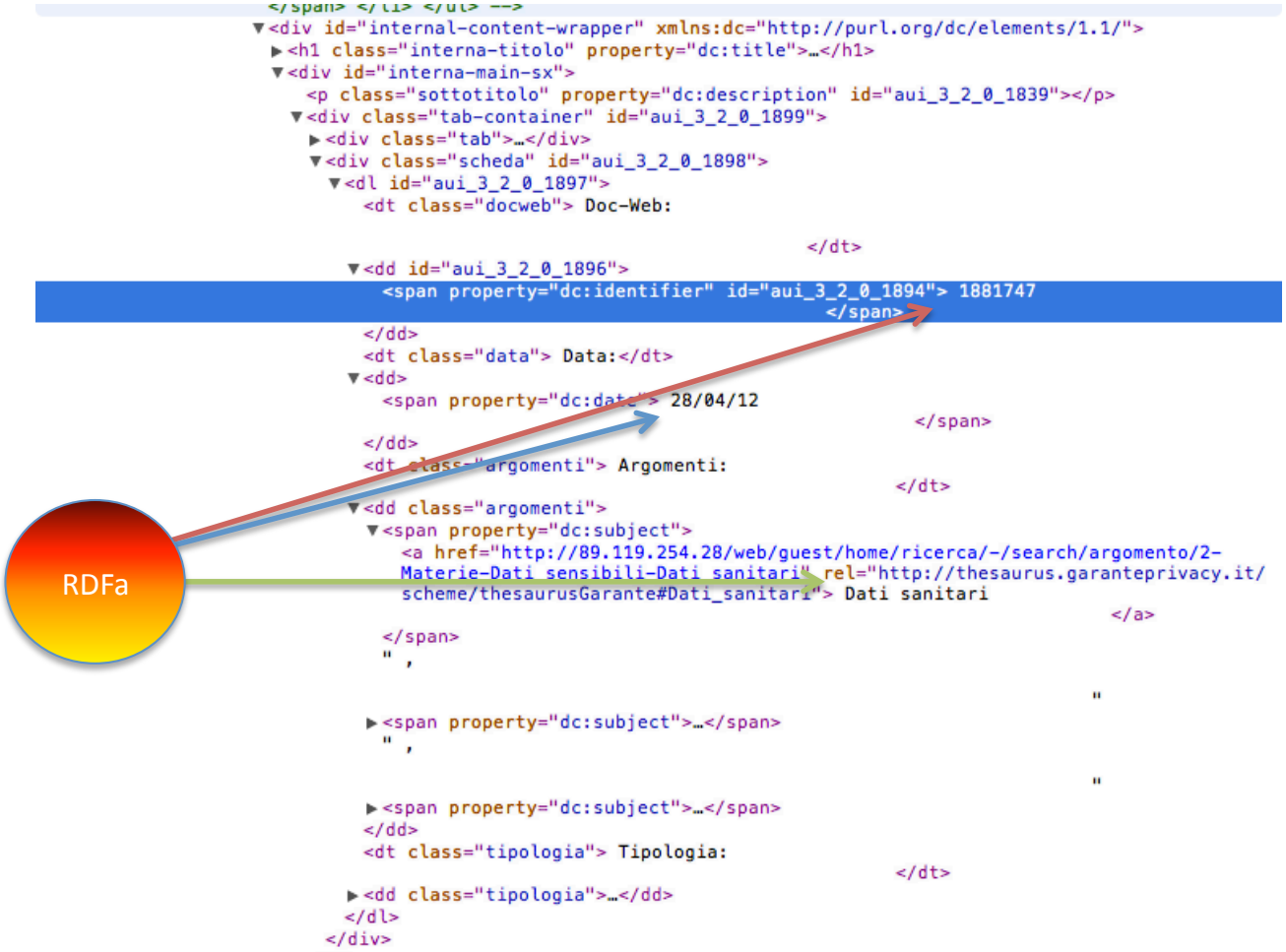
- Backend - tagging semantico

- Come accennato, **l'integrazione con SKOSware** e con la mappa semantica degli argomenti trattati dal Garante consente la classificazione ed il tagging dei contenuti del Portale
- Tramite un intuitivo meccanismo basato sul *drag and drop* è possibile associare ad un contenuto uno o più tag collegati ai lemmi ricercati sulla tassonomia gerarchica delle categorie
- Tale classificazione si riflette, in maniera immediata, all'interno del **motore di ricerca**
- La classificazione può essere effettuata in modo **automatico** e **massivo** su tutti i contenuti presenti
- Le meta-informazioni semantiche vengono utilizzate anche nella pubblicazione dei contenuti sul frontend (ottimizzandone la reperibilità e l'analisi semantica da parte dei motori di ricerca); si utilizzano a questo scopo tag **RDFa** integrati





Metadati RDFa



RDFa è un modo per associare etichette ai contenuti al fine di descrivere specifici tipi di informazioni, come la recensione di un ristorante, un evento, una persona o la scheda di un prodotto. Questi tipi di informazioni sono definiti **entità o elementi**. Ogni entità possiede differenti proprietà. Noi abbiamo utilizzato lo standard **Dublin Core** per descrivere l'identificativo univoco, la data e gli argomenti di cui tratta un articolo o un provvedimento del Garante.

In generale, RDFa utilizza semplici attributi nei tag XHTML (spesso o) per assegnare nomi brevi e descrittivi alle entità e alle loro proprietà.

DOMANDE???



Grazie per l'attenzione…

- Per qualsiasi ulteriore approfondimento, o per organizzare una demo delle nostre soluzioni, potete contattarci a questi recapiti:

Andrea Ciapetti
Etcware srl (www.etcware.it)

Via Etna 13, 00141 Roma

tel./fax +39 06 64495131
email a.ciapetti@etcware.it
cell. +39 320 6197534

Alessandra Donnini
Etcware srl (www.etcware.it)

Via Etna 13, 00141 Roma

tel./fax +39 06 64495131
email a.donnini@etcware.it
cell. +39 333 8914865