

HUMBOLDT: armonizzazione dei dati a supporto delle Infrastrutture di Dati Territoriali

Pasquale DI DONATO (*), Mauro SALVEMINI (*), Laura BERARDI (*)

(*) Università Sapienza di Roma – Dipartimento CAVEA - LABSITA, piazza Borghese 9 00186 Roma, +39 0649918830, +39 0649918873, { *pasquale.didonato* } { *mauro.salvemini* } { *laura.berardi* } @uniroma1.it

Riassunto

La direttiva INSPIRE è stata pubblicata in Gazzetta Ufficiale dell'Unione Europea il 25 Aprile 2007 ed è entrata in vigore il 15 Maggio 2007. La direttiva ha l'obiettivo di rendere disponibile un'informazione geografica rilevante, armonizzata e di qualità per la formulazione, attuazione, monitoraggio e valutazione di politiche comunitarie. Il termine armonizzazione appare in molta della documentazione relativa ad INSPIRE, anche se non ne viene fornita una definizione.

Il progetto HUMBOLDT, finanziato nell'ambito del 6° Programma Quadro di Ricerca Europeo, ha lo scopo di analizzare in dettaglio gli aspetti relativi all'armonizzazione di dati e servizi territoriali, al fine di fornire, laddove possibile, procedure di automatizzazione del processo di armonizzazione.

Abstract

The INSPIRE Directive was published in the official Journal of the European Union on the 25th April 2007, and entered into force on the 15th of May 2007. The directive intends to trigger the creation of an European Spatial Data Infrastructure that delivers to the users relevant, harmonized and quality geographic information for the formulation, implementation and evaluation of European Union policies. The term "harmonisation" appears often in most of the INSPIRE related documentation, unless a clear definition is not given.

The EU HUMBOLDT, funded under the 6° Framework Programme of the European Union, intends to analyse in details the issues related to spatial data and services harmonisation with the aim to provide a framework for the automation of the harmonisation processes.

Introduzione

La direttiva INSPIRE¹ è stata pubblicata in Gazzetta Ufficiale dell'Unione Europea il 25 Aprile 2007 ed è entrata in vigore il 15 Maggio 2007. INSPIRE (INfrastructure for SPatial InfoRmation in Europe) costituisce il quadro giuridico di riferimento per la creazione di un'Infrastruttura Europea di Dati Territoriali che ha l'obiettivo di rendere disponibile un'informazione geografica rilevante, armonizzata e di qualità per la formulazione, attuazione, monitoraggio e valutazione di politiche comunitarie a vari livelli e che hanno un impatto sul territorio, attraverso la messa a disposizione di servizi basati sull'informazione geografica stessa. L'implementazione di INSPIRE seguirà un approccio per fasi, che tenderà in prima istanza a sbloccare il potenziale delle infrastrutture e dei dati territoriali esistenti, per poi proseguire gradualmente verso l'armonizzazione di dati e servizi ed eventualmente integrare gli stessi in un'unica e coerente Infrastruttura di Dati Territoriali Europea.

INSPIRE sarà supportata dai risultati di una serie di progetti finanziati nell'ambito del 6° Programma Quadro di Ricerca Europeo (6PQ)²: HUMBOLDT è uno di questi progetti.

¹ <http://eur-lex.europa.eu/JOHtml.do?uri=OJ:L:2007:108:SOM:IT:HTML>

² <http://cordis.europa.eu/fp6/dc/index.cfm?fuseaction=UserSite.FP6HomePage>

Il progetto HUMBOLDT³ è stato finanziato nell'area tematica "Space" del 6°PQ per una durata di quattro anni, a partire dall'Ottobre del 2006: ha 27 partner provenienti da 14 Paesi dell'Unione e risorse per circa 13.5 milioni di euro e 110 anni/uomo.

Scopo principale di HUMBOLDT è la creazione di un *framework* a supporto del processo di armonizzazione di dati ed integrazione di servizi territoriali (il sottotitolo del progetto è "Development of a Framework for Data Harmonisation and Service Integration").

Il progetto è iniziato con una dettagliata analisi dello stato dell'arte in termini di metodi e strumenti disponibili, sulla cui base sono state identificate le varie fasi necessarie per l'armonizzazione di dati territoriali ed i software utili per un eventuale utilizzo ed integrazione nel *framework* finale. Il progetto, inoltre, è strutturato intorno ad otto scenari applicativi che hanno lo scopo di aiutare nella definizione delle necessità utente (in termini di armonizzazione di dati ed integrazione di servizi) e costituiscono allo stesso tempo un "ambiente di test" di quanto viene man mano sviluppato.

L'immagine seguente mostra le relazioni tra il *framework* HUMBOLDT e gli attori (interni ed esterni al progetto) che a vario titolo sono interessati a procedure di armonizzazione (HUMBOLDT WP2, 2006).

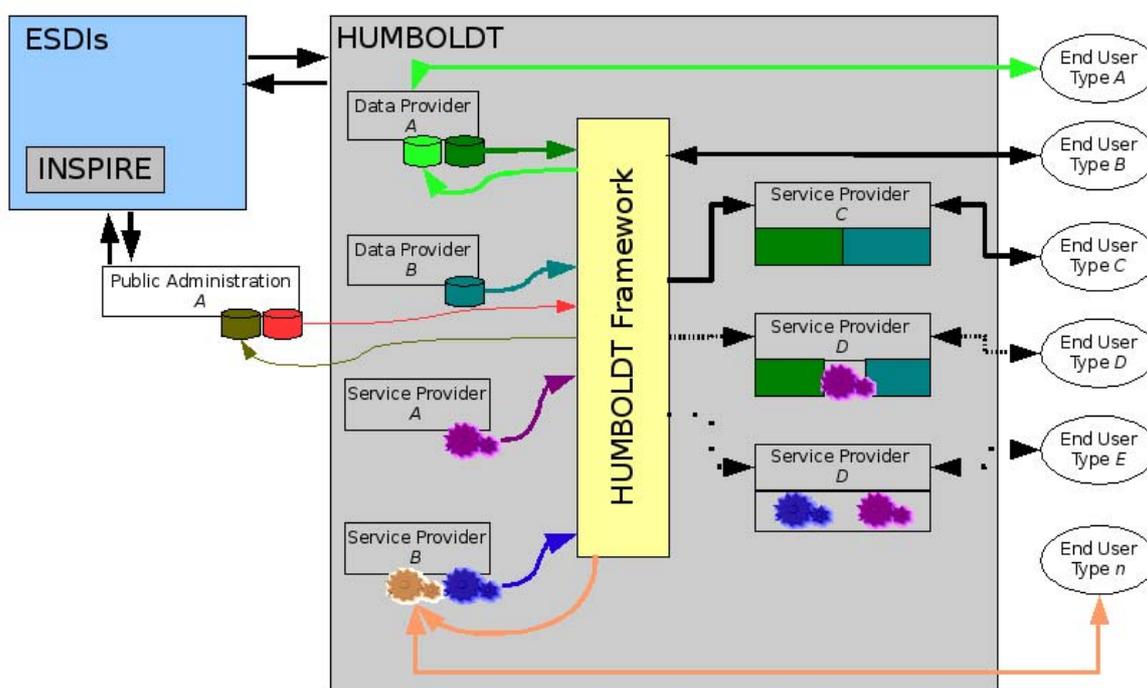


Figura 1 – Relazioni del framework HUMBOLDT

Lo schema prende in considerazione gli attori interni ed i casi d'uso definiti con gli scenari: allo stesso tempo presenta alcune possibili relazioni con attori non-HUMBOLDT:

- **Caso 1 – Fornitore di dati A:** un fornitore di dati può essere interessato ad utilizzare il *framework* allo scopo di rendere armonizzati i propri dati. Il set di dati armonizzati può essere fornito ad un utente finale (End User A) direttamente o tramite un fornitore di servizi HUMBOLDT (Service Provider C > End User C);
- **Caso 2 – Fornitore di servizi B:** caso 1 applicato a servizi invece che a dati;

³ <http://esdi-humboldt.eu>

- **Caso 3 – Fornitore di dati B:** un fornitore di dati può partecipare a HUMBOLDT mettendo a disposizione i propri dati, che vengono poi forniti ad un utente finale tramite un fornitore di servizi HUMBOLDT (Service Provider C);
- **Caso 4 – Fornitore di servizi A:** caso 3 applicato a servizi invece che a dati;
- **Caso 5 – Utente finale B:** è il caso di un utente finale particolare, come ad esempio un professionista, che può accedere direttamente al *framework* senza il tramite di un fornitore di servizi;
- **Caso 6 – Amministrazione Pubblica A:** una Pubblica Amministrazione può utilizzare il *framework* per armonizzare i propri dati e servizi al fine di renderli compatibili con le richieste di INSPIRE, ad esempio.

Armonizzazione

Il termine armonizzazione appare in molta della documentazione relativa ad INSPIRE, ma non ne viene fornita una definizione chiara. Nel testo della Direttiva appare l'indicazione implicita che armonizzazione è qualcosa di diverso e probabilmente migliore dell'interoperabilità, senza tuttavia precisarne le differenze.

Nell'ambito del progetto HUMBOLDT si è optato per la seguente definizione di armonizzazione: “(Creare) la possibilità di combinare dati da fonti eterogenee in un prodotto informativo non ambiguo, consistente ed integrato in una modalità trasparente e semplice per l'utente” (HUMBOLDT WP3, 2007). Da un punto di vista tecnico armonizzazione significa adattare dati, modelli dati e servizi a standard e regole definite in funzione delle necessità specifiche di un'infrastruttura di dati territoriali.

La domanda naturale a questo punto è: quali tipi di eterogeneità esistono nel dominio dell'informazione geografica? Diversi schemi di classificazione si ritrovano in letteratura: una delle classificazioni più utilizzate (Curtis e Muller, 2006) prevede tipologie di eterogeneità a livello di sintassi (differenze in termini di formato dati), di struttura (differenze in termini di schema) e di semantica (differenze nel significato dei termini in un contesto specifico). Sebbene utile, tale classificazione può dare adito a diverse interpretazioni ed incomprensioni, per cui nell'ambito di HUMBOLDT, in funzione anche delle analisi fatte sullo stato dell'arte e nell'ambito degli scenari, è stato deciso di meglio specificare e cercare di risolvere gli aspetti di eterogeneità dovuti a differenze di:

- Formato dati;
- Sistema di riferimento spaziale;
- Modello concettuale (struttura e vincoli);
- Nomenclatura, classificazione e tassonomia;
- Terminologia, thesauri, ontologie;
- Modello di metadati;
- Scala;
- Visualizzazione (legenda, classificazione, stile);
- Processing;
- Estensione (spaziale, tematica e temporale);
- Procedure di acquisizione dati.

Molti di questi elementi si ritrovano tra le componenti di armonizzazione definite nell'ambito di una collaborazione tra il progetto RISE (*Reference Information Specifications for Europe*)⁴ ed il *Data Specification Drafting Team* di INSPIRE.

Il progetto è attualmente impegnato sugli aspetti di armonizzazione a livello di modello dati. A tale scopo, HUMBOLDT utilizza l'approccio MDA (*Model Driver Approach*), che presenta due aspetti: creare un nuovo modello dati in funzione dell'analisi delle necessità e quindi trasformare i dati in maniera tale che essi risultino strutturati secondo il nuovo modello; generare il modello tecnico per l'implementazione fisica (ad esempio uno schema XML/GML) sulla base del nuovo modello concettuale (HUMBOLDT WP7, 2007).

Tali funzionalità saranno fornite dal *framework* tramite servizi web: saranno anche implementate procedure di *geoprocessing* a supporto di ulteriori necessità di armonizzazione a livello dati (es. generalizzazione, *edge matching*, riclassificazione).

Architettura del Framework e prototipo

Il prototipo, così come l'intero *framework*, si basa sull'approccio alla progettazione di sistemi informativi noto come *Reference Model of Open Distributed Processing* (RM-ODP). RM-ODP è uno standard internazionale [ISO/IEC 10746]⁵ che definisce specifiche astratte per l'implementazione di sistemi operanti in un ambiente distribuito ed aperto: si basa sui concetti di federazione, interoperabilità, trasparenza, etc. ereditati dal tecnologie tipo SOA (Service Oriented Architecture), CORBA e RPC (Remote Procedure Call). ISO RM-ODP, inoltre, costituisce la base concettuale della serie di standard ISO 19100⁶, nonché del modello di riferimento⁷ dell'Open Geospatial Consortium.

Lo scopo del prototipo è dimostrare l'applicabilità di quanto previsto dal progetto e di provarne l'usabilità da parte degli utenti: allo stesso tempo si intende testare le tecnologie di base che faranno presumibilmente parte del *framework*. In particolare verrà implementata e testata un'architettura di base di un ambiente *client/server* distribuito dove l'informazione geografica è fornita da diversi servizi web, cui accedono utenti finali che lavorano con *client* differenti ma in un'ottica di condivisione di interessi operativi.

Le funzionalità del prototipo si basano sul seguente scenario applicativo (HUMBOLDT WP5, 2007):

- Un utente richiede dei dati da uno o più servizi WMS a lui noti, opera alcune operazioni GIS di base e salva in contesto della sua attività;
- Un secondo utente legge il contesto salvato dal primo utente, opera anch'egli alcune operazioni GIS di base, eventualmente aggiunge dati da altri servizi WMS, e salva in contesto della sua attività;

nel prototipo si assume che il modello concettuale (*information model*) entro il quale operano i due utenti sia diverso e che i due utenti siano di madrelingua diversa.

Sulla base dello scenario appena descritto, il prototipo ha lo scopo di:

- Testare la cooperazione tra utenti che parlano lingue diverse;
- Testare le possibilità di armonizzazione di set di dati provenienti da Paesi diversi, nello specifico:

⁴ www.eu-rise.org

⁵ [http://standards.iso.org/ittf/PubliclyAvailableStandards/c020696_ISO_IEC_10746-1_1998\(E\).zip](http://standards.iso.org/ittf/PubliclyAvailableStandards/c020696_ISO_IEC_10746-1_1998(E).zip)

⁶ <http://www.isotc211.org>

⁷ http://portal.opengeospatial.org/files/?artifact_id=3836

- Testare l'armonizzazione di dati provenienti da due WMS diversi che forniscono dati di pertinenza di Paesi diversi;
 - Testare l'utilizzo di un servizio di *portrayal* per l'armonizzazione dello stile di presentazione delle informazioni;
 - Testare le possibilità di armonizzazione di servizi WMS che forniscono dati a scala diversa e in sistemi di riferimento diversi;
- Definire un motore di ricerca di *web map context*;
 - Testare strumenti e modalità di gestione e condivisione di modelli dati;
 - Testare strumenti di editing di modelli dati;
 - Testare modalità e tecnologie per la ricerca di modelli dati con la possibilità di una eventuale procedura automatica di confronto dei vari modelli.

La figura seguente mostra la struttura logica del prototipo tramite un diagramma delle componenti:

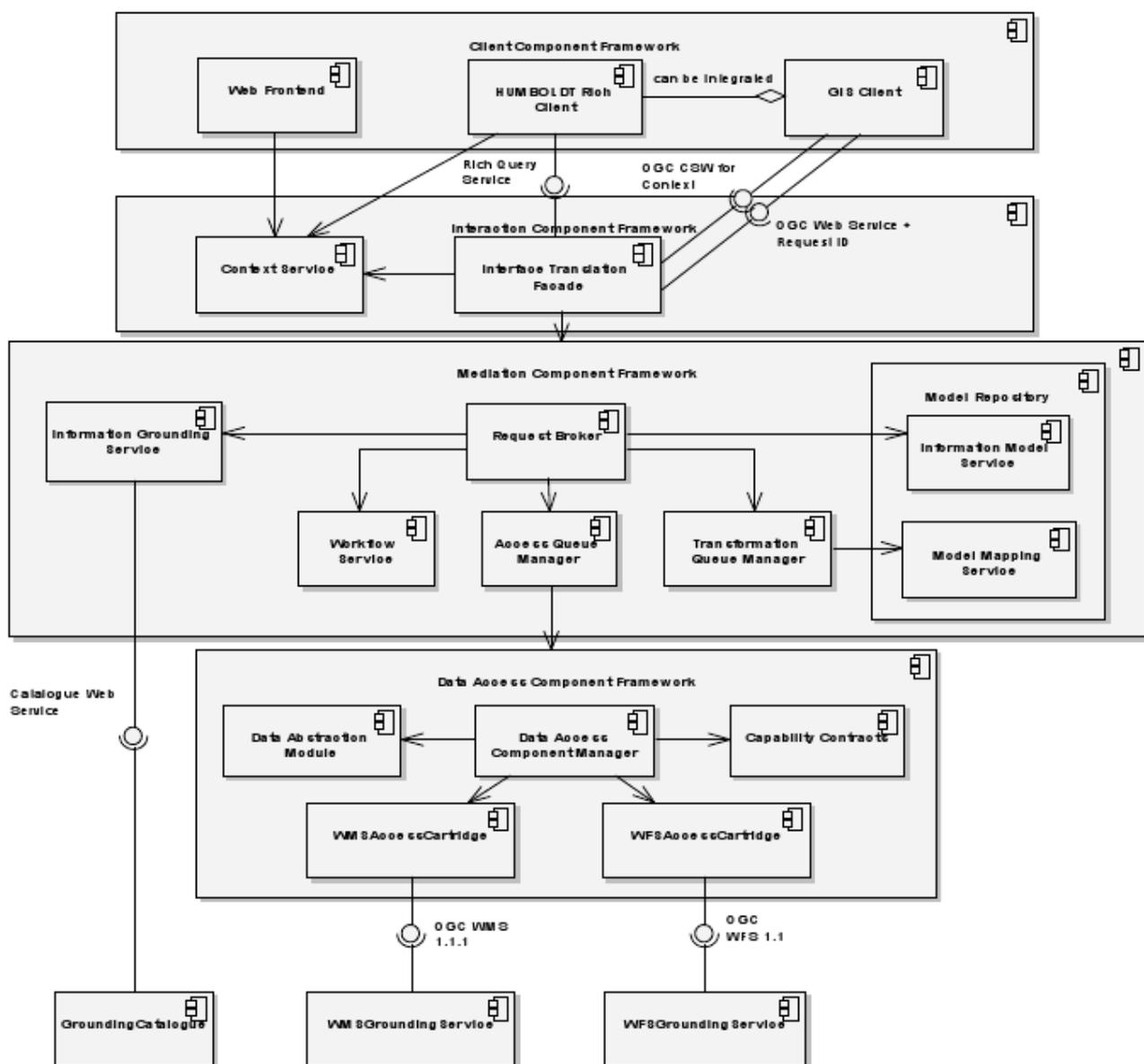


Figura 2 – Diagramma delle componenti del prototipo

Riferimenti

HUMBOLDT WP2 (2007), *A2.1-D1 Process Analysis*

HUMBOLDT WP3 (2007), *A3.5-D1 State of the Art in Data Harmonisation and Data Management*

Curtis E, Muller H. (2006), “Schema Translation in Practice”, *Snowflake Software White Paper*

HUMBOLDT WP7 (2007), *A4.4-D1 Data Harmonisation Processes*

HUMBOLDT WP5 (2007), *A5.1-D1 Framework specification Prototype*