26/05/2017

09:30 - 15:00 | Sala Conferenze
Rettorato Università del Salento

LifeWatch-ERIC

L'infrastruttura di ricerca Europea sulla biodiversita' per scienza e societa'

Conferenza Scientifica

LifeWatch ITALY | UNIVERSITÀ DEL SALENTO | BPP Banca Popolare Pugliese

DATA PORTAL

OVERVIEW DEL SISTEMA

Relatore Daniele Viva, Links Management & Technology S.p.A.

LINKS – CHI SIAMO

Links è uno dei principali fornitori italiani di servizi di consulenza, tecnologia, prodotti e outsourcing per Banche, Pubbliche Amministrazioni e Aziende.

MILANO – Via Messina n° 38 (Torre B)

ROMA – Via XX Settembre n° 4

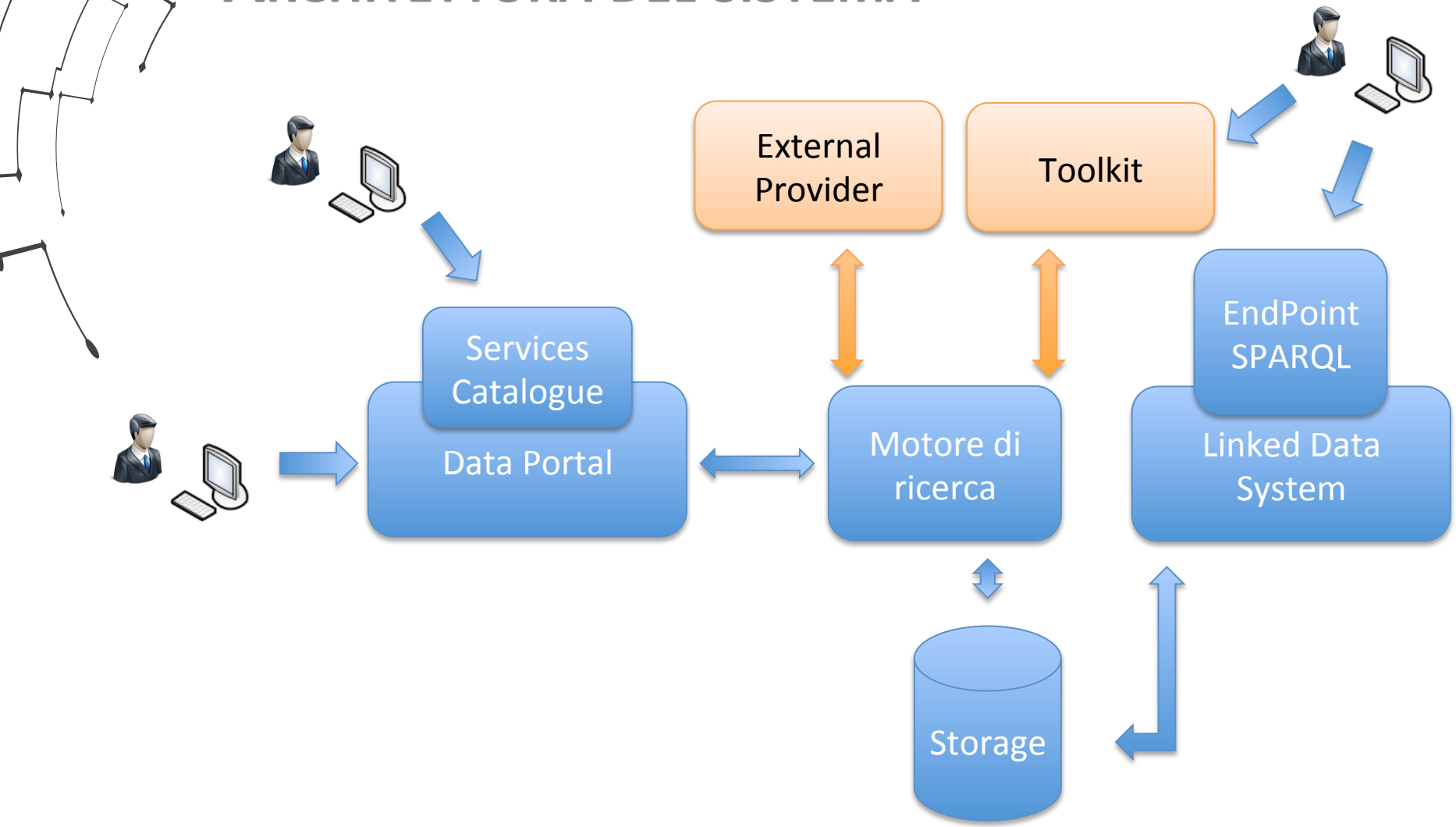
BARI – Via Dante Alighieri n° 25

LECCE – Via Rocco Scotellaro n° 55

TIRANA – Via Andon Zako Çajupi n° 6

www.linksmt.it | info@linksmt.it

ARCHITETTURA DEL SISTEMA



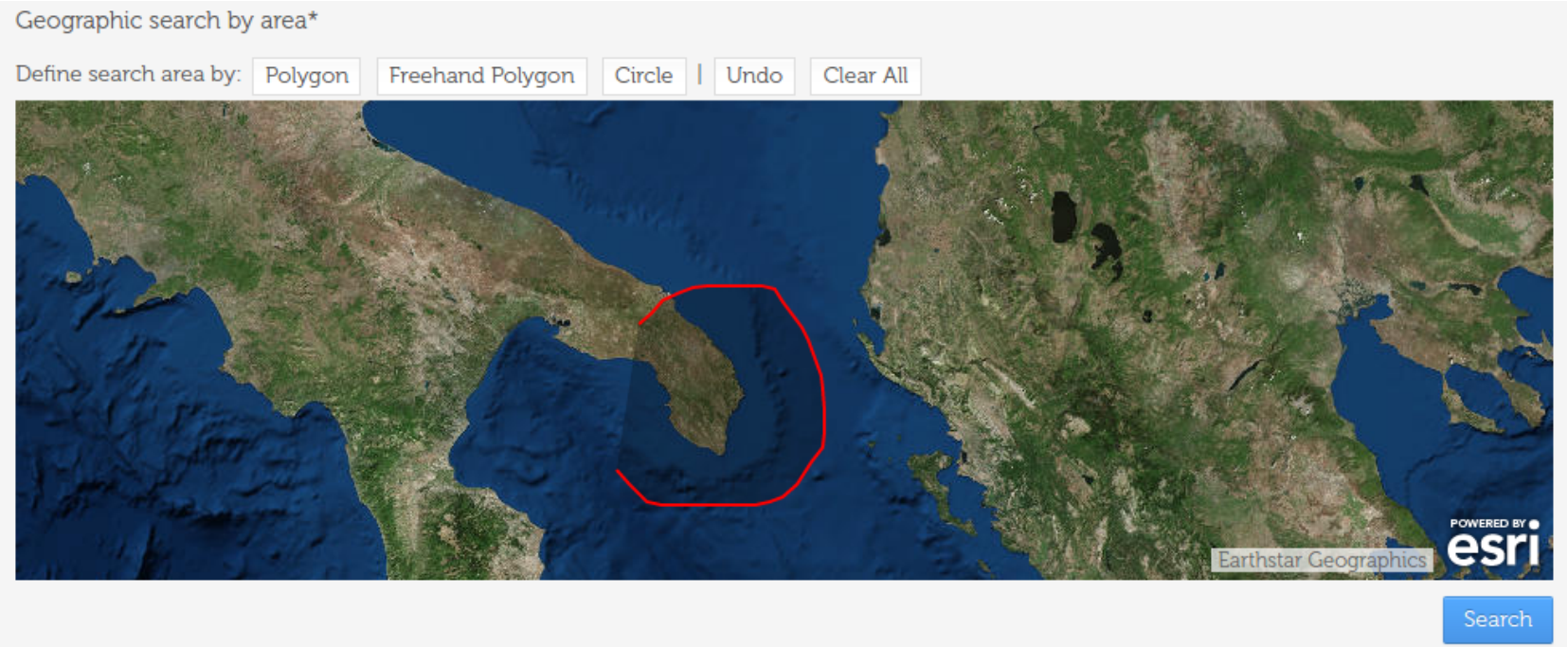
ARCHITETTURA DEL SISTEMA

Il sistema è basato su soluzioni **Open Source**:

- Content Management System: **Liferay Community Edition** si tratta di un Enterprice Portal scritto in Java;
- DBMS: **MongoDB** (non relazionale) si tratta di un document store con alto indice di scalabilità e flessibilità. E' classificato come un database di tipo NoSQL;
- Search Engine: **ElasticSearch** si tratta di un server di ricerca basato su Apache Lucene. Tutte le sue funzionalità sono esposte tramite interfaccia RESTful;
- Linked Data: **Virtuoso OpenLink** si tratta di un Universal Server caratterizzato da un Graph Data Management;

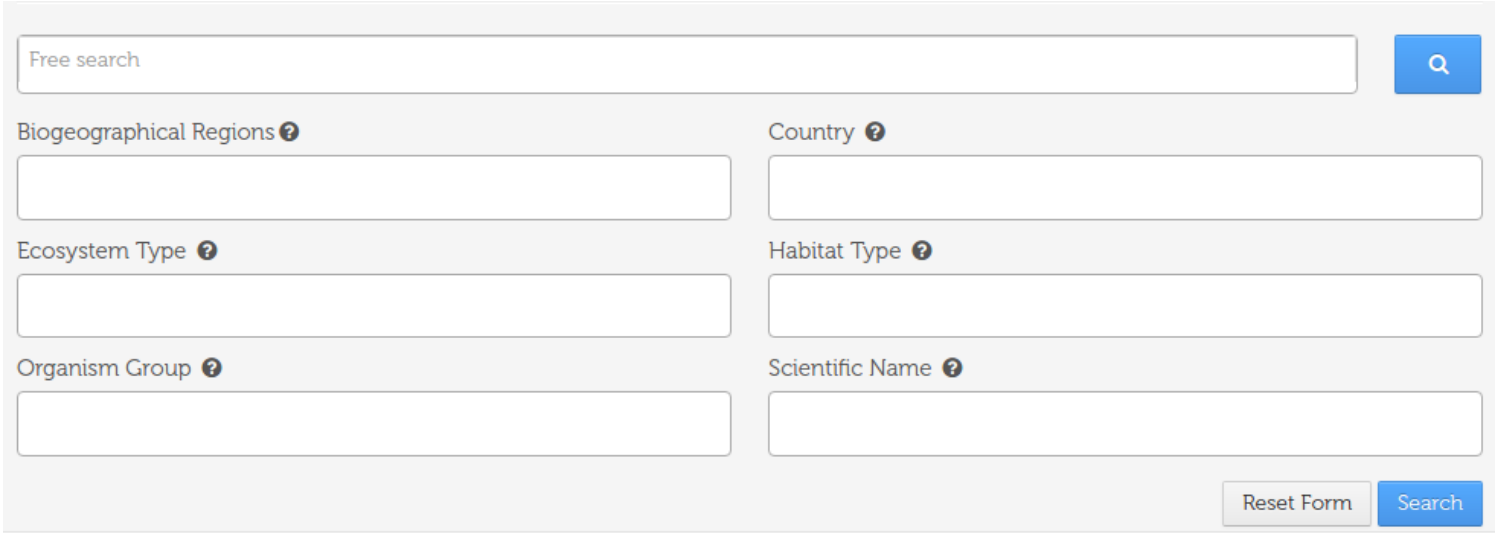
DATA PORTAL: RICERCA DATASET

- Ricerca Dataset:
  - Area **geografica** sui dataset del dataportal (Disegnando un poligono; Disegno a mano libera; Disegnano un cerchio)



DATA PORTAL: RICERCA DATASET

- Ricerca Dataset:
  - **Fulltext** sui dataset di dataportal, toolkit e provider esterni;
  - **Filtri** per particolari filed (biogeographical regions, country, ecosystem, habitat, organism, scentific name) sui dataset di dataportal e toolkit;



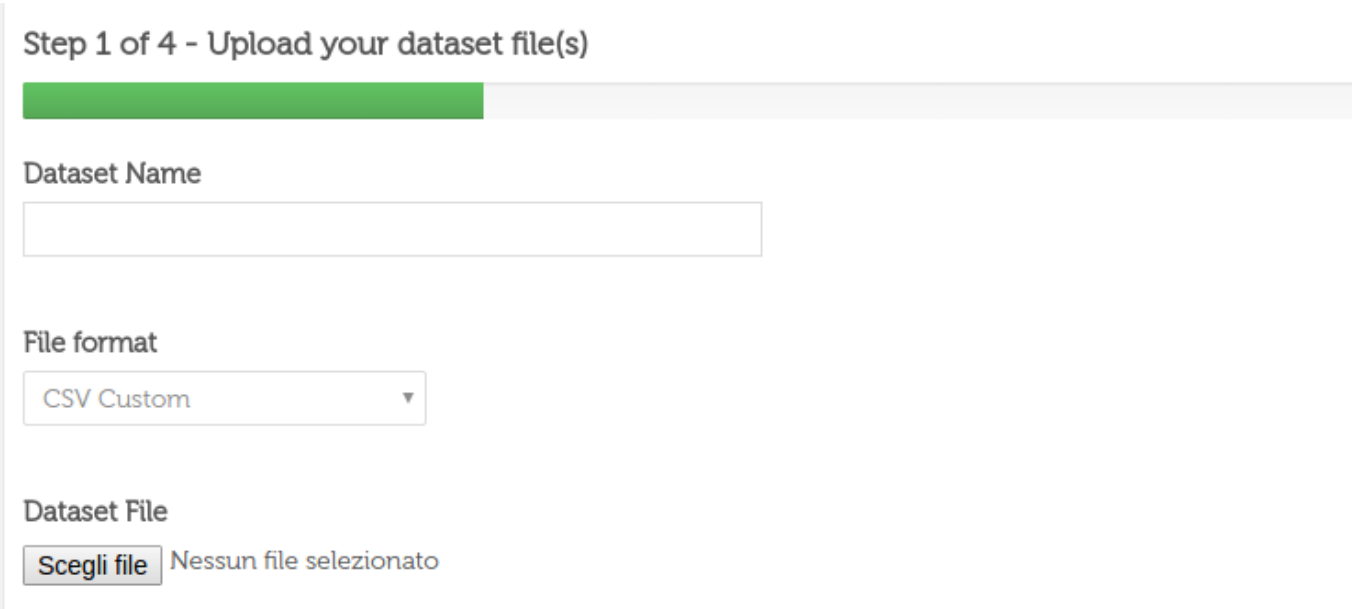
DATA PORTAL: RICERCA DATASET

- Video Ricerca Dataset

datasetSearch.mp4

DATA PORTAL: CARICAMENTO DATASET

- Caricamento Dataset (Darwin Core, ABCD, Csv custom)
- Il dataset viene **elaborato dal sistema** e viene trasformato secondo la struttura semantica definita in LifeWatch;
- L'elaborazione avviene tramite un **mapping semi-automatico**;



DATA PORTAL: CARICAMENTO DATASET

- Video Caricamento Dataset

submitYourFile_02.2.mp4

DATA PORTAL: TOOLKIT

- **Toolkit**: si tratta di una web application da installare presso il data center del centro di ricerca;
- **Goal**: condividere dataset di centri di ricerca;

Step da eseguire:

1. Download della webapp;
2. Collegamento ad un **DataSource** (SQL Server, MySql, Csv);
3. Generazione di un **dataset custom** a partire dal DataSource;
4. **Condivisione** sia di **metadati** e sia dell'intero **dataset**;

Risultato: I metadati ed il dataset sono messi a disposizione rispettivamente per la ricerca e la richiesta di inserimento nel sistema (previa approvazione da parte dell'amministratore);

DATA PORTAL: TOOLKIT

- Video Toolkit

AddDataset.mp4

DATA PORTAL: KNOWLEDGE BASE

- La knowledge base è una «fotografia» di un insieme di dataset a seguito di una ricerca;
- Ogni ricercatore può salvare fino ad un max di 10 KB;
- Il ricercatore può effettuare ulteriori ricerche avanzate sulle KB;
- Export dei dati in formato json o csv;

DATA PORTAL: ADMINISTRATION

- **Metadata Standard**: permette di definire e gestire lo standard dei metadati dei dataset LifeWatch;
- **Data Standard**: permette di gestire lo standard dei field dei dataset Lifewatch
- **Toolkit Registry**:
  - permette di accettare/rifiutare richieste di abilitazioni;
  - Verificare la raggiungibilità dei toolkit;
  - Gestire l'elenco dei toolkit abilitati;
- **External provider registry:** permette di gestire l'elenco dei provider esterni sui quali la ricerca semplice avrà visibilità;
- **Dataset request**: permette di gestire l'accesso o meno al dataset da parte del ricercatore;

DATA PORTAL: ADMINISTRATION

- **Service Catalogue Management**: permette di visualizzare e gestire il catalogo dei servizi messi a disposizione da Lifewatch (Rest Services, algoritmi, riferimenti a servizi esterni ecc.);
- **External Standard Management**: permette di creare un mapping tra uno standard esterno (es. ABCD) con i campi dello standard Lifewatch;
- **Lifewatch Data Validation**: permette di
  - verificare la correttezza dei dataset in fase di caricamento;
  - Correggere eventuali errori ed eventualmente inserire nel sistema nuove GNA;
- **Global Name Architecture (GNA):** permette di gestire tutte le specie conosciute dal sistema Lifewatch;

GRAZIE

daniele.viva@linksmt.it