



UNIONE EUROPEA  
Fondo Sociale Europeo  
Fondo Europeo di Sviluppo Regionale



UNIVERSITÀ  
DEL SALENTO

Procedura aperta per la progettazione, sviluppo, rilascio, consegna, installazione dei  
seguenti sistemi software

LOTTO 1

Biodiversity and Ecosystems – LifeWatch INTERoperable DATA Portal (BE-LINDA)

LOTTO 2

BE-Biodiversity and Ecosystems - HUB (BE-HUB)

CAPITOLATO TECNICO

## 1 – Premessa

### 1.1 - Contesto Operativo

LifeWatch è un'infrastruttura europea di eScience (e-Infrastructure) per la ricerca su biodiversità e ecosistemi, istituita dall'UE il 17/03/2017, come Consorzio Europeo di Infrastruttura di Ricerca (LifeWatch-ERIC). Fa parte dei landmark dell'area Ambiente presenti nella roadmap di ESFRI, ed occupa una nicchia ben definita tra le infrastrutture del settore, unica infrastruttura di eScience distribuita a concentrarsi sullo studio della biodiversità e degli ecosistemi. In quanto tale, LifeWatch-ERIC costruisce i suoi ambienti e laboratori virtuali di ricerca (VRE) con una combinazione di tecnologia digitale, dati e risorse computazionali, e comunicazione a supporto del lavoro e della ricerca collaborativi. Pertanto, questi costituiscono gli strumenti e gli apparati di ricerca che LifeWatch-ERIC sta implementando e rendendo operativi sul web per i suoi utenti. Sostenendo reperibilità, accessibilità, interoperabilità e riutilizzo dei dati già raccolti, LifeWatch-ERIC ha un impatto positivo sia sull'efficienza complessiva del finanziamento della ricerca, sia sulla scoperta di settori della conoscenza con informazioni di base carenti e per i quali è necessaria una nuova e più intensa raccolta di dati. Come ESFRI distribuita, LifeWatch-ERIC ha un hub centrale, distribuito tra Spagna, Italia e Paesi Bassi, e Nodi Tematici Nazionali. Ospitando il Service Centre, che, come componente dell'hub centrale, è l'unico punto di accesso per gli utenti dell'infrastruttura, l'Italia ha un ruolo di primo piano in LifeWatch-ERIC, cui contribuisce con il Nodo Tematico Nazionale LifeWatch-ITA.

L'ampliamento e l'approfondimento delle conoscenze attuali sull'organizzazione della biodiversità e sulla salute degli ecosistemi sono essenziali per affrontare le principali sfide ambientali e sociali, quali: conservazione di biodiversità ed ecosistemi, sfruttamento delle risorse della biosfera, riscaldamento e cambiamenti globali, sviluppo sostenibile. La perdita di biodiversità già supera i limiti planetari di sostenibilità (Rockstrom et al., 2009; Nature 461) e di integrità biologica (Steffen et al., 2015; Science 347), l'appropriazione umana della produttività primaria della biosfera è molto elevata (Haberl et al., 2014; Ann. Rev. Environmental Resources 39; Pritchard et al., 2018; Ecological Economics 146), la perdita di habitat è intensa (Watson et al., 2016; Conservation Letters 9), il global warming e il degrado ecosistemico danno flebili segnali di miglioramento (Ripple et al., 2017; BioScience 67). Negli ultimi decenni, il contributo teorico e sperimentale per approfondire la conoscenza dei meccanismi di organizzazione della biodiversità (Chase, 2011, in The theory of Ecology, Scheiner & Willig Eds.; Kendall, 2015, Ecology 96), dello stato della biosfera (Keith et al., 2013, PLOSone 8; Pereira et al., 2013, Science 339; Bojinski et al, 2014, BAMS 95), delle scale (O'Neill, 2001, Ecology 82; Murcia et al., 2014, TREE 29; Heffernan, 2014, Front Ecol Environ 12) e dei servizi (Costanza et al., 2014, Global Environmental Change 26) ecosistemici è in aumento ed è stato proposto (Purves et al., 2013, Nature 493) un approccio meccanicistico, basato sulla teoria metabolica, per creare un modello della biodiversità della biosfera. Inoltre, progetti finanziati dall'UE, reti di osservatori europei, sistemi di osservazione della Terra, compresi sensori in situ e satellitari, infrastrutture di ricerca, organizzazioni e iniziative su scala mondiale producono dati sulla biodiversità e sugli ecosistemi ad una velocità e frequenza senza precedenti, con un potenziale d'uso per la conservazione di hotspot di particolare valore (Hoffman et al., 2018, Rapporti scientifici 8). Tuttavia, le tecnologie digitali macchina-macchina per la gestione e l'analisi dei dati non sono



UNIONE EUROPEA  
Fondo Sociale Europeo  
Fondo Europeo di Sviluppo Regionale



UNIVERSITÀ  
DEL SALENTO

sufficientemente avanzate da consentire uno sfruttamento pienamente redditizio di big data ed approfondire la nostra conoscenza della biodiversità e l'organizzazione e conservazione degli ecosistemi.

## 2 – Oggetto della fornitura

All'interno dell'Infrastruttura LifeWatch su descritta, sono oggetto della presente fornitura la realizzazione e messa in opera dei beni e servizi necessari al rafforzamento del Service Centre e del Data Portal di LifeWatch-ITA descritti nel progetto LifeWatchPLUS (come da Allegato 2 del DD R.902 del 9/5/2019) corrispondenti alle seguenti voci:

Codice univoco del bene:

- **Biodiversity and Ecosystems – LifeWatch INTERoperable DATA Portal (BE-LINDA)**
- **BE-Biodiversity and Ecosystems - HUB (BE-HUB).**

In particolare, si richiede:

- **la progettazione esecutiva ai fini di definire in dettaglio i requisiti del sistema da realizzare;**
- **la realizzazione delle piattaforme applicative;**
- **la loro interconnessione/integrazione;**
- **la produzione di adeguata documentazione tecnica di progetto e manuali d'uso;**
- **i servizi di assistenza, formazione e messa in esercizio.**

Tutti i componenti dei servizi dovranno essere dimensionati al fine del raggiungimento degli obiettivi di progetto e secondo stime numeriche e valutazioni effettuate dalle Ditte aggiudicatrici affinché l'intera attuazione del progetto stesso ed il suo mantenimento funzionale per l'intero periodo contrattuale venga effettuato senza alcun onere aggiuntivo a carico dell'Amministrazione e senza necessità di acquisizione di ulteriori componenti, servizi di qualsiasi genere o sottoscrizione di contratti di fornitura e servizi con altri soggetti.

Gli stessi beni sono inclusi nel Programma Biennale di Acquisti 2019/2020 dell'Università del Salento e corrispondono alle voci identificate dai seguenti "Codice Unico Intervento (CUI)" che nella tabella seguente sono messi in relazione con i corrispondenti codici univoci dei beni:

- **F84001850589201900005 PIR01\_00028\_360879 BE-LINDA**
- **F84001850589201900006 PIR01\_00028\_360440 BE-HUB**

La fornitura si compone dei seguenti lotti:

- **Lotto n.1: Bene BE-LINDA**
- **Lotto n.2: Bene BE-HUB**

### 2.1 – Proprietà Intellettuale

L'Università del Salento acquista la proprietà intellettuale dell'ambiente integrato.

### 3 – Piano di Progetto

La durata complessiva del progetto è di mesi 8 e comunque dovrà completarsi entro il 31.07.2022. Il progetto prevedrà il rilascio dei seguenti moduli software:

- Per il LOTTO 1 bene BE-LINDA:
  - **servizi di Catalogo,**
  - **software per la produzione di mappature tra schemi/profilo di MetaDati,**
  - **ingestione e persistenza dei metadati,**
  - **miglioramento delle funzioni di ricerca avanzata,**
  - **attivazione di servizi web per l'accesso e il controllo delle funzionalità dei moduli in momenti diversi.**
  
- Per il LOTTO 2 bene BE-HUB:
  - **ingestione e persistenza dei dati,**
  - **Harvester,**
  - **Exposer per Data Provider**
  - **Global Species Databases (GSDs)**
  - **Global Names Architecture (GNA)**
  - **interfaccia BE-LINDA**

#### 3.1 – Pianificazione del progetto

Le attività necessarie al raggiungimento degli obiettivi previsti nell'oggetto della fornitura si articolano nelle seguenti 4 fasi per entrambi i Lotti.

##### 3.1.1 – Definizione dell'architettura del sistema

L'obiettivo della fase risiede nell'elaborazione delle specifiche tecniche e funzionali dell'architettura dell'ambiente, derivandole dai bisogni espressi dall'utenza, dalle soluzioni tecnologiche disponibili sul mercato e dalle soluzioni tecnologiche risultanti dallo stato dell'arte della ricerca nelle tecnologie del software rispettando le linee guida sia architetture che relative al disegno dell'Information Architecture specificate nella sezione 6.

Le principali attività da sviluppare riguardano:

- l'analisi delle soluzioni tecnologiche disponibili sul mercato e delle soluzioni tecnologiche risultanti dallo stato dell'arte della ricerca;
- la definizione dei requisiti utente;
- la definizione delle specifiche funzionali di ogni modulo della piattaforma software;
- la predisposizione delle specifiche architetture. Verranno elencate le varie componenti del sistema da collaudare unitamente alla strategia di collaudo ed in varie parti specifiche (una per componente) con le singole attività, tecniche, strumenti di collaudo.

Alla fine della fase saranno conseguiti i seguenti risultati:

- rapporto sui requisiti utente;



UNIONE EUROPEA  
Fondo Sociale Europeo  
Fondo Europeo di Sviluppo Regionale



UNIVERSITÀ  
DEL SALENTO

- specifiche funzionali di ogni singolo modulo delle piattaforme software BE-LINDA e BE-HUB;
- specifiche tecniche e architetture del sistema BE-LINDA e BE-HUB.

### 3.1.2 – Sviluppo dell'applicazione

L'obiettivo della fase consiste nella completa realizzazione, sperimentazione, integrazione e validazione delle piattaforme software BE-LINDA e BE-HUB, secondo le specifiche e le modalità definite nella fase precedente.

In questa fase saranno condotte le seguenti attività:

- setup dell'infrastruttura tecnologica di sviluppo applicativo e di comunicazione;
- progettazione di dettaglio e realizzazione dei servizi di catalogo, motore di ricerca, interfacce macchina e per utenti;
- progettazione di dettaglio e realizzazione del software per la produzione di mappature tra schemi/profili di MetaDati;
- progettazione di dettaglio e realizzazione del componente di ingestione e persistenza dei dati e metadati;
- progettazione di dettaglio e realizzazione delle funzioni di ricerca avanzata ottimizzate;
- progettazione di dettaglio e realizzazione dei servizi web per l'accesso e il controllo delle funzionalità di tutti i moduli implementati;
- progettazione di dettaglio e realizzazione dell'Harvester;
- progettazione di dettaglio e realizzazione dell'Exposer.

Alla fine della fase saranno conseguiti i seguenti risultati:

- documento di progetto di ogni modulo software;
- sistema "BE-LINDA", composto da una infrastruttura software e dal complesso dei servizi previsti dalla fornitura;
- sistema "BE-HUB", composto da una infrastruttura software e dal complesso dei servizi previsti dalla fornitura;
- ambiente di sviluppo: le Ditte aggiudicatrici dovranno rilasciare l'ambiente di sviluppo configurato e tutti i codici sorgente prodotti, corredati da adeguata documentazione.

### 3.1.3 – Fase di test

L'obiettivo della fase consiste nella esecuzione di un piano di test della piattaforma fornita. È auspicabile l'adozione di test incrementali parallelamente agli sviluppi da svolgere congiuntamente con il team della stazione appaltante.

In questa fase saranno condotte le seguenti attività:

- individuazione del campione di utenza pilota di accesso ai servizi;
- formazione e addestramento dei soggetti coinvolti nella sperimentazione circa l'uso dei servizi e la gestione e manutenzione dell'ambiente prodotto;
- erogazione del servizio BE-LINDA;
- erogazione del servizio BE-HUB;
- valutazione dei risultati.

Alla fine saranno conseguiti i seguenti risultati:

- piano di erogazione/gestione del servizio;
- rapporto di validazione e valutazione dei risultati: in particolare le Ditte aggiudicatrici dovranno rilasciare alla stazione appaltante l'ambiente di testing, tutti i test di unità prodotti, tutto lo "scaffolding" di test di sistema.

### 3.1.4 – Rilascio

L'obiettivo della fase consiste nell'erogazione del sistema BE-LINDA e BE-HUB che, pur essendo lotti separati, devono essere tra loro integrati per cui le Ditte aggiudicatrici devono impegnarsi all'integrazione dei sistemi in tutte le fasi del progetto dalla progettazione alla fase del loro rilascio. In questa fase saranno condotte le seguenti attività:

- installazione e configurazione dei sistemi presso il Data Centre messo a disposizione dall'Università del Salento (anche su server di replica – test di ripristino);
- rilascio del software e di manuali operativi utili alle operazioni di installazione e configurazione, ripristino e gestione dei differenti moduli e servizi;
- collaudo del software: tale attività dovrà essere accompagnata da un piano di collaudo strutturato in una sezione generale, una di tutto il materiale utilizzato per l'esecuzione del servizio, della proprietà intellettuale del software prodotto e di tutti i diritti che ne derivano.

L'aggiudicatario di ciascun lotto assume, inoltre, la responsabilità piena ed esclusiva nei confronti dei terzi che dovessero rivendicare diritti di autore su parti e/o elementi del materiale utilizzato, del software e delle piattaforme utilizzate.

### 3.1.5 – Schedulazione temporale delle fasi Lotto 1 BE-LINDA

#	Task	M1	M2	M3	M4	M5	M6	M7	M8
1.1	Pianificazione del progetto	X	X						
1.2	Milestone 1: rilascio prima versione del documento di progettazione che include la lista di moduli da rilasciare ai mesi M3-M4 e ai mesi M6-M7	X							
1.3	Sviluppo servizi di catalogo e interfacce		X	X	X	X	X		
1.4	Sviluppo software mappature		X	X	X	X	X		
1.5	Ingestione e persistenza metadati		X	X	X	X	X		
1.6	(Meta) Data curation		X	X	X	X	X		
1.7	Integrazione GSD e GNA		X	X	X	X	X		
1.8	Miglioramento funzioni di accessibilità		X	X	X	X	X		
1.9	Interfaccia BE-HUB		X	X	X	X	X		
1.10	Primo rilascio			X	X				
1.11	Harvesting dei record provenienti dal popolamento iniziale di BE-HUB				X	X	X	X	
1.12	Fase di test			X	X	X	X	X	
1.13	Secondo rilascio						X	X	
1.14	Formazione							X	X

### 3.1.6 – Schedulazione temporale delle fasi Lotto 2 BE-HUB



#	Task	M1	M2	M3	M4	M5	M6	M7	M8
2.1	Pianificazione del progetto	X	X						
2.2	Milestone 2: rilascio prima versione del documento di progettazione che include la lista di moduli da rilasciare ai mesi M3-M4 e ai mesi M6-M7	X							
2.3	Sviluppo servizi Data Portal e interfacce		X	X	X	X	X	X	
2.4	Ingestione e persistenza dati e metadati		X	X	X	X	X	X	
2.5	Sviluppo Harvester		X	X	X	X	X	X	
2.6	Sviluppo Exposer		X	X	X	X	X	X	
2.7	Interfaccia BE-LINDA		X	X	X	X	X	X	
2.8	Primo rilascio			X	X				
2.9	Popolamento Data Portal				X	X	X	X	
2.10	Fase di test			X	X	X	X	X	
2.11	Secondo rilascio						X	X	
2.12	Formazione							X	X

### 3.1.7 - Termini e luogo di consegna ed installazione

I termini di consegna ed installazione dei beni e servizi di cui alla sezione 2, sono da intendersi in giorni naturali e consecutivi decorrenti dal giorno successivo alla sottoscrizione del contratto. Il rilascio dei servizi con la messa in produzione e la formazione dovrà avvenire seguendo i tempi indicati nella sezione 3.1.5.

La consegna e l'installazione dei beni e servizi della fornitura dovrà essere effettuata presso gli indirizzi indicati in tabella, in accordo con il Responsabile Unico del Procedimento:

Prodotto	Luogo di consegna e installazione
<b>LOTTO 1: BE-LINDA</b>	Università del Salento – CAMPUS ECOTEKNE via per Monteroni, sn, 73100 - LECCE
<b>LOTTO 2: BE-HUB</b>	Università del Salento – CAMPUS ECOTEKNE via per Monteroni, sn, 73100 - LECCE

## 4 – Obblighi delle Ditte aggiudicatrici

- Per consentire il corretto svolgimento del progetto entro i termini indicati nel presente capitolato, le Ditte aggiudicatrici dovranno avere il gruppo operativo presso la sede del DiSTeBA, costituito almeno dal 50% delle unità di personale previste per la componente ICT dello staff impegnato nel progetto.
- Tutte le persone coinvolte nello svolgimento delle attività dovranno, quindi, operare in stretto coordinamento con lo staff preposto allo svolgimento delle attività del progetto LifeWatchPlus, con attività in sede DiSTeBA o con analogo orario di lavoro dedicato al progetto da svolgersi nella sede della Ditta aggiudicatrice.



UNIONE EUROPEA  
Fondo Sociale Europeo  
Fondo Europeo di Sviluppo Regionale



UNIVERSITÀ  
DEL SALENTO

- Per tutta la durata delle attività saranno necessari aggiornamenti sul progresso delle attività mediante riunioni tra il personale delle Ditte aggiudicatrici e quello di progetto. La cadenza delle riunioni sarà mensile. Sarà consentito l'utilizzo di strumentazione digitale per effettuare le riunioni (Skype, Webex, ecc.), tuttavia sono previsti dei face-to-face meeting, con cadenza trimestrale, per le quali le Ditte aggiudicatrici dovranno garantire la presenza fisica di almeno due persone coinvolte nel progetto. Il linguaggio utilizzato durante le riunioni sarà l'italiano. Al termine della riunione dovrà essere stilata apposita minuta sempre in italiano.
- Il tempo rimanente dovrà essere utilizzato in affiancamento al personale di progetto, per la messa a punto di tutte le procedure, mediante debugging del software e delle funzionalità necessarie alla corretta integrazione delle piattaforme con le restanti componenti software previste nell'ambito di LifeWatchPlus.
- Tutte le attività di consulenza, inclusa l'installazione e la personalizzazione del software necessario, saranno condotte sui sistemi hardware di cui il progetto LifeWatchPlus dispone presso la sede dell'Università del Salento.
- Le Ditte aggiudicatrici dovranno fornire adeguata descrizione tecnica con le soluzioni progettuali ed implementative dettagliate utilizzando gli strumenti comuni dell'Ingegneria del Software quali UML, etc. o di Basi di Dati quali Modello Entità/Relazioni, etc.
- Le Ditte aggiudicatrici dovranno riportare inoltre la sequenza temporale di tutte le attività. Tutto il lavoro svolto dovrà essere documentato mediante deliverables di progetto come da capitolo "3 - Piano di Progetto" da consegnare al responsabile del progetto, con cadenza bimestrale, a partire dalla data di aggiudicazione.
- Le Ditte aggiudicatrici dovranno utilizzare strumenti per la gestione dei progetti (es. Pivotal Tracker) che garantisca la collaborazione in tempo reale su un backlog condiviso con l'ente appaltante e che identifichi e stabilisca le varie priorità.
- I moduli software dovranno essere tutti corredati da manuale di installazione su macchine Linux/Unix e manuale utente.

## 5 - Indicazione dei soggetti coinvolti

Per l'Università del Salento:

- un responsabile tecnico per la valutazione delle offerte e per il collaudo della fornitura erogata;
- Prof. Alberto Basset, in qualità di responsabile scientifico dell'obiettivo realizzativo "Rafforzamento del Service Centre e del Data Portal di LifeWatch-ITA" (O.R. 3) del progetto LifeWatchPlus;



UNIONE EUROPEA  
Fondo Sociale Europeo  
Fondo Europeo di Sviluppo Regionale



UNIVERSITÀ  
DEL SALENTO

- Prof. Alessandro Chiarucci, in qualità di responsabile scientifico dell'obiettivo realizzativo "Realizzazione di un Hub nazionale per le risorse di dati su biodiversità ed ecosistemi" (O.R. 4) del progetto LifeWatchPlus;
- un responsabile della comunicazione per tutti gli aspetti relativi alla presentazione grafica e all'organizzazione dei contenuti;
- un responsabile dell'infrastruttura LifeWatch per il supporto tecnico nelle varie attività del progetto;
- altro personale dell'Università del Salento in qualità di responsabile amministrativo del procedimento e secondo necessità.

Per le Ditte aggiudicatrici:

Per il LOTTO 1:

- un responsabile di progetto, con almeno 15 anni di esperienza nella conduzione di commesse di fornitura di servizi di consulenza ICT nell'ambito di progetti ICT complessi e dottorato di ricerca o documentata esperienza di ricerca;
- un responsabile tecnico, con qualifica di informatico o ingegnere informatico e con almeno 8 anni di documentata esperienza nella progettazione e realizzazione di sistemi e servizi informatici e nella conduzione tecnica di progetti ICT complessi;
- un system integrator con almeno 8 anni di esperienza documentata di sistemi di data warehouse e data mining;
- due sviluppatori Front-end con almeno 3 anni di esperienza documentata nell'implementazione di layout ed elementi grafici in siti internet per desktop e per mobile, conoscenza documentata di HTML, JavaScript (JS) e fogli di stile CSS;
- sei sviluppatori Back-end con almeno 3 anni di esperienza documentata su architetture e linguaggi server side: Java, Php, Ruby-on-rails, ecc.

Per il LOTTO 2:

- un responsabile di progetto con curriculum di system integrator con almeno 8 anni di esperienza documentata di sistemi di data warehouse e data mining;
- uno sviluppatore Front-end con almeno 3 anni di esperienza documentata nell'implementazione di layout ed elementi grafici in siti internet per desktop e per mobile, conoscenza documentata di HTML, JavaScript (JS) e fogli di stile CSS;
- uno sviluppatore Back-end con almeno 3 anni di esperienza documentata su architetture e linguaggi server side: Java, Php, Ruby-on-rails, ecc.;
- un data scientist, con almeno 5 anni di esperienza documentata nel settore dell'ecologia, ed un gruppo di non meno di 4 data scientist con almeno 2 anni di esperienza per la raccolta ed organizzazione dei dati messi a disposizione dai diversi gruppi della JRU.

Per entrambi i lotti e per tutto il personale proposto il concorrente dovrà produrre i curricula. Per i profili della componente ICT, i curricula dovranno, inoltre, attestare il possesso dei requisiti precedentemente specificati. Sarà ritenuta premiante la dimensione del gruppo di lavoro e la sua disponibilità già in fase di presentazione dell'offerta.

## 6 –Requisiti di carattere generale

La soluzione individuata per l'attuazione dell'infrastruttura telematica alla base dei progetti BE-LINDA e BE-HUB, oltre a soddisfare le esigenze applicative indicate già in fase progettuale, dovrà essere impostata nel rispetto dei principi di modularità, estendibilità e scalabilità:

- Modularità: la modularità della soluzione è data da un'architettura aperta in cui le responsabilità e le interfacce di ciascun componente sono chiaramente identificate, e dove, nel rispetto di tali responsabilità ed interfacce, i componenti possono essere sostituiti singolarmente con soluzioni equivalenti, garantendo così la necessaria flessibilità al cliente.

- Estendibilità: intesa sia dal punto di vista delle funzionalità da offrire agli utenti e sia dal punto di vista degli strumenti di gestione. Nuovi servizi e nuove entità potranno essere aggiunti in modo da integrarsi senza sforzo con l'architettura esistente.
- Scalabilità: il sistema realizzato sarà in grado di scalare all'aumentare del traffico in termini di numero di utenti che visiteranno il Portale.

Per entrambi i lotti si dovrà inoltre rispettare la grafica LifeWatch e il requisito del “responsive design”.

## 6.1 Il ciclo di vita del dato e l'architettura di sistema

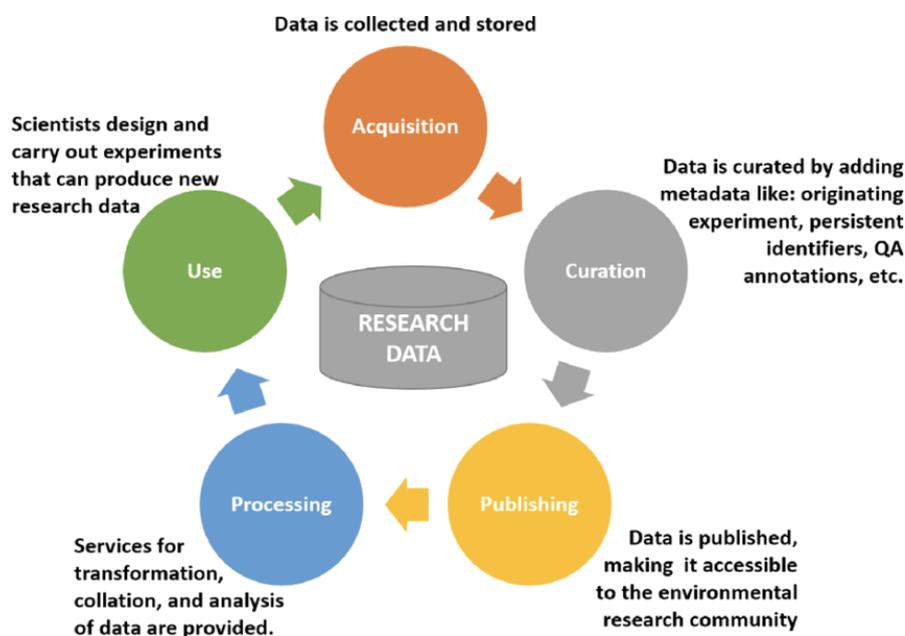


Figura 1. Ciclo di vita del dato

I prodotti BE-LINDA e BE-HUB sono stati ideati e pensati a supporto del ciclo di vita del dato così come descritto dal Reference Model del progetto europeo ENVRI [<http://www.envri.eu>]. In figura 1 è riportato il modello di un processo che copre la durata dei prodotti dei dati della ricerca scientifica, dalla progettazione alla raccolta, cura, processamento/analisi, pubblicazione e riuso. Tale ciclo vale sia per il dato che per il metadato ed illustra come questo è gestito dalle Infrastrutture di ricerca. In particolare il software BE-HUB dovrà supportare le prime tre fasi del ciclo di vita: l'acquisizione, la cura del dato e la sua pubblicazione. Il software BE-LINDA si concentrerà sulla pubblicazione del metadato e sulla sua discoverability.

Per quanto concerne la parte di data curation fondamentali saranno la re-ingegnerizzazione della Global Names Architecture (GNA) e le funzionalità per l'individuazione e correzione di errori nomenclaturali/tassonomici, sintattici o semantici all'interno dei dataset trattati dal software BE-HUB.

Premesso che l'architettura definitiva dei prodotti BE-LINDA e BE-HUB, dei loro servizi e funzionalità verrà definita a seguito dell'analisi di dettaglio e sarà il risultato delle attività di analisi e disegno

previste dal piano di lavoro, i capitoli successivi descriveranno i principali moduli che BE-LINDA e BE-HUB dovranno offrire agli stakeholder. In particolare, per ciascun modulo si descriveranno le principali caratteristiche e macrofunzionalità.

Il sistema, articolato e complesso nell'insieme, dovrà essere costituito da una serie di componenti infrastrutturali (framework) sulle quali saranno realizzate le funzionalità di base e i servizi applicativi offerti agli utenti del sistema, dalle interfacce che consentono l'interoperabilità verso sistemi esterni (applicativi dei vari nodi Europei di ricerca, come ad esempio i portali dei nodi nazionali), dagli stessi utenti che con ruoli e scopi differenti interagiscono con la piattaforma.

La motivazione principale nella scelta di questo approccio è rappresentata dalla possibilità di separare attraverso interfacce ben delineate i diversi strati elaborativi, così da poter adottare le soluzioni appropriate nell'implementazione delle diverse funzionalità, conservando un elevato livello di efficacia in termini di integrazione e garantendo un maggior grado di manutenibilità e scalabilità del sistema complessivo alla variazione dei parametri dimensionali.

L'architettura logica del sistema è rappresentata in Figura 2.

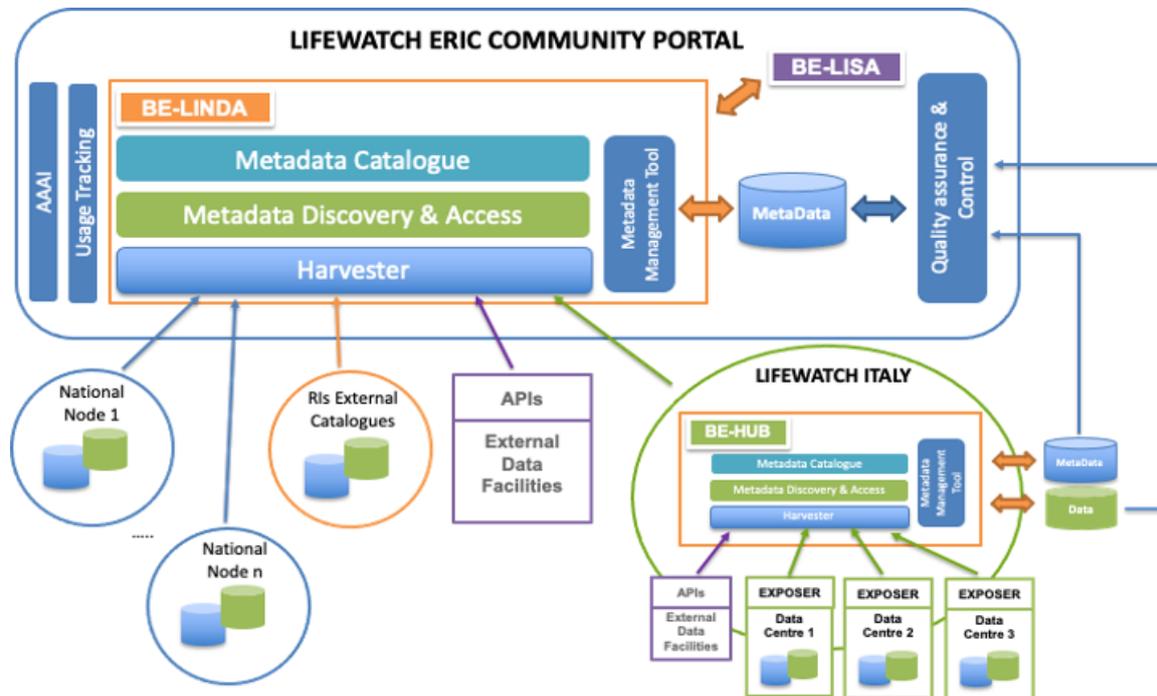


Figura 2. Architettura logica del sistema

## 7 – Requisiti Tecnici Minimi LOTTO1: BE-LINDA

L'infrastruttura di ricerca LifeWatch ERIC ha come obiettivo principale di divenire il punto di accesso principale alle basi di conoscenza del dominio della biodiversità e degli ecosistemi. A tal fine LifeWatch ERIC offrirà strumenti per la discovery delle risorse (dati, servizi di analisi e calcolo, Virtual Research Environment, ecc.) e strumenti per la pubblicazione e condivisione di queste.

Allo stato attuale negli ultimi anni si è lavorato alla messa a punto di una prima versione del catalogo dei metadati delle varie risorse gestite da LifeWatch ERIC, versione che è online all'indirizzo <https://metadatalogue.lifewatch.eu> (Figura 3).

L'attuale catalogo dei metadati è basato sulla versione di GeoNetwork 3.10.4, e gestisce le risorse:

- Virtual Research Environments, con un profilo di metadati basati sullo standard ISO 19139
- Services, con un profilo di metadati basati sullo standard ISO 19139
- Workflows, con un profilo di metadati basati sullo standard ISO 19139
- Research Sites, con un profilo di metadati basati sullo standard ISO 19139
- Dataset, con un profilo di metadati basato sullo standard EML 2.2.0

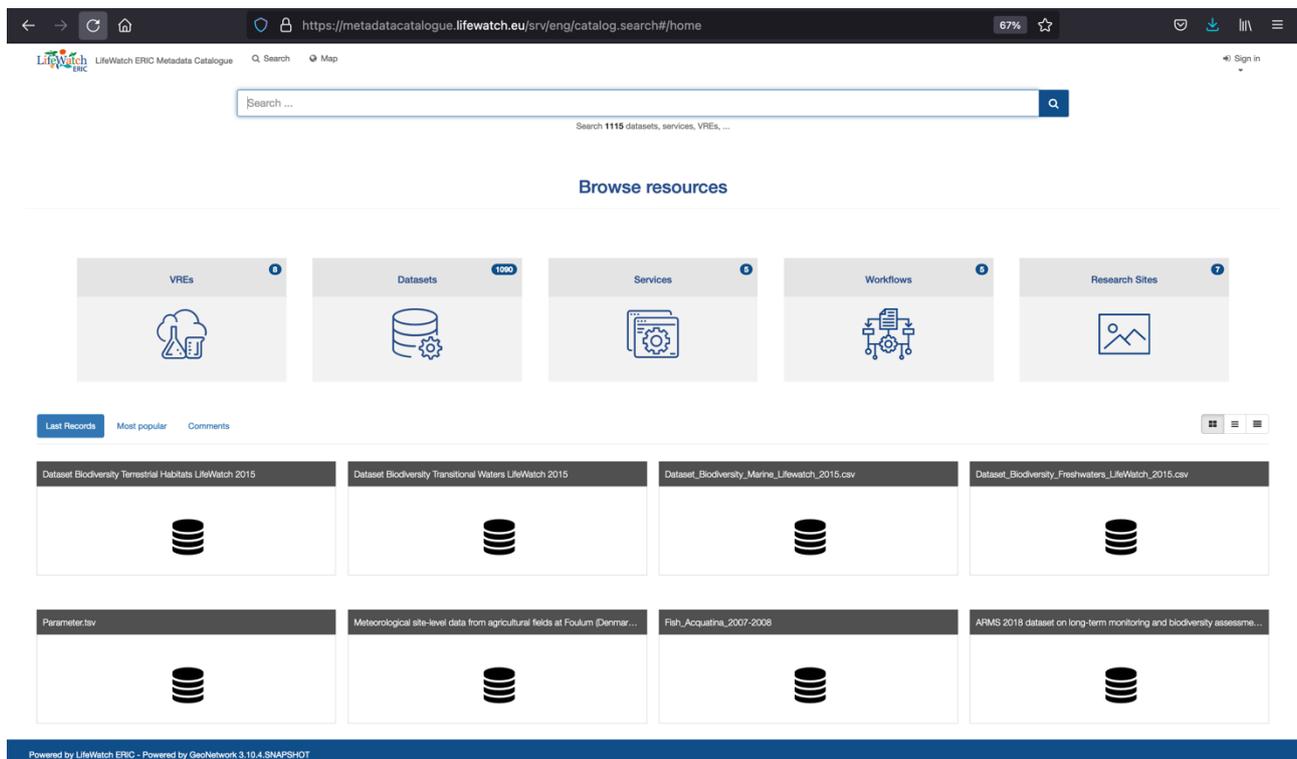


Figura 3. LifeWatch ERIC MetaData Catalogue

La richiesta prevista da capitolato per la piattaforma BE-LINDA prevede quindi un'attività di re-ingegnerizzazione, potenziamento ed implementazione del catalogo dei metadati fornendo i componenti e le funzionalità descritte nei paragrafi successivi. L'utilizzo di soluzioni e tecnologie innovative (come, ad esempio, l'automazione nell'uso di risorse semantiche per la compilazione dei metadati, l'uso di strumenti visuali per la creazione di query semantiche sui metadati, l'uso dell'Artificial Intelligence per ottimizzare la ricerca dei dati, ecc.) costituirà punteggio premiante ai fini della valutazione dell'offerta tecnica.

Sarà valutato positivamente anche il grado di FAIRness delle soluzioni proposte per soddisfare le funzionalità richieste.

### 7.1 Servizio di Catalogo, motore di ricerca, interfacce macchina e per utenti

Il potenziamento riguarderà le seguenti funzionalità:



UNIONE EUROPEA  
Fondo Sociale Europeo  
Fondo Europeo di Sviluppo Regionale



UNIVERSITÀ  
DEL SALENTO

- Ricerca delle risorse del catalogo di Metadati prevedendo diverse tipologie di interfaccia di ricerca (semantica e non). Anche l'informazione dello schema del dato dovrà essere indicizzata e ricercabile tramite filtri specifici.
- Creazione di strumenti di personalizzazione mediante interfaccia web (lato amministratore) della ricerca semplice e avanzata (sia con search box che con facets). A titolo di esempio, dovrà essere almeno possibile per l'amministratore personalizzare lo strumento di ricerca semplice e avanzata direttamente da interfaccia, per aggiungere, modificare, perfezionare e/o eliminare i criteri di ricerca impostati senza intervenire sul codice. Inoltre, dovrà essere possibile effettuare la ricerca anche mediante mappa: selezionando direttamente l'area geografica disegnando il quadrato di ricerca e/o inserendo i valori specifici delle coordinate.
- Esposizione dei MetaDati secondo diversi possibili schemi e formati (es. LW JSON, LW XML, ecc.), compresi quelli di scambio di dati geospaziali e sensoristici (es. GML, O&M, geoTIFF, SHP, ecc.) sia mediante interfaccia web (GUI) che mediante l'uso di API di tipo REST. È fondamentale in questo caso gestire anche il versioning e la sicurezza delle API che consentiranno di interrogare in modo programmatico il catalogo.
- Realizzazione di servizi di dati Geospaziali e sensoristici e visualizzazione di Mappe, secondo interfacce standard di servizio di catalogo, basandosi su un Server G.I.S. dedicato. Dovrà essere possibile visualizzare spazialmente i record di MetaDati presenti sul catalogo basandosi sulle informazioni di copertura geografica presenti nei MetaDati. La visualizzazione spaziale dovrà essere interattiva: la mappa dovrà essere navigabile (spostamento, zoom-in, zoom-out, ecc.) e il click su una determinata risorsa dovrà offrire una visione in anteprima dei dettagli del record selezionato con la possibilità di cliccare su un link per la visualizzazione dei dettagli (la visualizzazione della landing page della risorsa). La visualizzazione su mappa dovrà essere disponibile in home page (per la ricerca spaziale) e in ogni landing page dei vari record.
- Produzione e visualizzazione di statistiche descrittive attraverso interfacce user-friendly, quindi anche in forma grafica, di dati geospaziali provenienti da qualunque altra componente dell'infrastruttura o da servizi esterni (ad es. Google analytics). A titolo di esempio, dovrà essere possibile creare e gestire le statistiche del catalogo che consentano di analizzare le caratteristiche principali (es. counter dei vari tipi di record, numero di visite, numero di download, record più popolari, keyword più cercate, ricerche più frequenti, ecc.) scegliendo il tipo di grafico da utilizzare, le etichette degli assi, i colori e le altre caratteristiche grafiche e formali, ecc.
- Creazione e caricamento di nuovi schemi/profilo di MetaDati con diverse modalità. Dovrà essere possibile creare un nuovo profilo di MetaDati da zero (specificando mediante interfaccia grafica i dettagli dei vari campi costituenti il profilo) o partendo da un profilo esistente e personalizzarlo in base alle necessità (importando o copiando nell'editor il template in XML o altro formato). Per ogni singolo attributo dovrà essere possibile specificarne il titolo, una breve descrizione (che diventerà il tooltip nell'editor), se è multivalore, se è obbligatorio oppure opzionale, se è composto, se è associato a una codelist, il tipo di dato richiesto in fase di compilazione (numero, stringa, link, ecc.), le eventuali regole di validazione associate, ecc. Tale funzionalità dovrà essere utilizzabile mediante interfaccia web, senza dover intervenire a livello di codice.
- Creazione dell'interfaccia di comunicazione con l'Harvester del bene BE-HUB per ospitare i MetaDati da esso forniti. In particolare sarà cura della ditta aggiudicatrice garantire la corretta

interconnessione e integrazione con il bene BE-HUB assicurando in fase di deploy il corretto harvesting delle risorse di metadati.

- Creazione dell'interfaccia di comunicazione con l'Help Desk di LifeWatch ERIC che dovrà consentire all'utente di aprire nuovi ticket associati al catalogo. Le richieste così generate saranno così annotate per consentire l'assegnazione automatica all'agente o al gruppo di agenti opportuno.
- Gestione dell'autenticazione e dell'autorizzazione degli utenti. Dovrà essere possibile per l'amministratore gestire le fasi di registrazione e login degli utenti specificando il protocollo da utilizzare (es. SAML v2) per garantire l'integrazione con il Single Sign On (SSO) di LifeWatch ERIC, e i campi da richiedere per la registrazione di nuovi utenti. Dovrà inoltre essere possibile creare gruppi (a cui associare loghi e altri dettagli) con appropriati diritti di accesso (lettura e/o scrittura) su specifici record o su gruppi di record. Si dovrà partire dai ruoli di default previsti da GeoNetwork e garantire la possibilità di poterli personalizzare in base alle necessità.

### 7.2 Integrazione di un modulo SW per la produzione di mappature tra schemi/profilo di MetaDati nei moduli Harvester e Servizio di Catalogo

Attraverso l'implementazione di un'interfaccia grafica si dovrà consentire all'amministratore di effettuare il mapping da uno schema/profilo di MetaDato all'altro senza la necessità di dover intervenire a livello di codice ma mediante collegamento tra gli elementi dello schema di partenza e quelli dello schema di arrivo. Dovrà essere possibile creare una trasformazione da zero (mediante collegamento tra elementi dello schema) o partire da una trasformazione esistente (importando il file o copiandolo nell'editor) e personalizzarla in base alle necessità. Il modulo software dovrà garantire inoltre la generazione automatica (con possibile download) del file di trasformazione ottenuto per effettuare il mapping da uno schema/profilo di MetaDato all'altro (es. da ISO 19139 a EML 2.2.0, da EML 2.1 a EML 2.2.0, ecc.). È fondamentale in questo caso gestire anche il versioning dei vari profili di MetaDati.

### 7.3 Ingestione e persistenza metadati

Tale modulo prevede tutte le attività volte:

- All'ampliamento delle tipologie e delle categorie di metadati ospitabili dal catalogo (es. di natura biomolecolare, geospaziale, sensoristica, ecc.). Si dovrà garantire l'ampliamento delle tipologie di risorse ospitabili dal catalogo con conseguente aggiornamento automatico di tutti i componenti di backend (creazione e caricamento del template di riferimento per la nuova risorsa, gestione dell'editor per la metadattazione, ecc.) e di frontend (landing page del record di metadati associato alla nuova risorsa, aggiornamento dell'home page e dei vari menu, scelta dell'icona rappresentante la nuova risorsa, ecc.).
- Alla gestione dell'assegnazione di identificatori persistenti alle risorse (a tal fine LifeWatch ERIC ha stretto un accordo con DATACITE per la generazione di DOI). Si dovrà garantire il supporto alla gestione dell'assegnazione del DOI mediante un opportuno workflow redazionale che preveda tutte le fasi necessarie per la valutazione e la validazione del record specifico da pubblicare. Una volta ottenuto il DOI, dovrà essere automaticamente integrato tra i metadati della risorsa specifica a prescindere dal tipo di metadata standard considerato.
- Al versioning dei metadati durante tutto il loro ciclo di vita.

- All'ottimizzazione del flusso di ingestione dei MetaDati che faccia anche uso di funzioni semantiche per il data cleaning in fase di importazione dei metadati (vedi sezione 7.5).
- Alla gestione dei log per monitorare e risalire a chi ha creato/modificato/eliminato cosa e alla gestione della concorrenza per evitare l'aggiornamento simultaneo di uno stesso record (dovrà essere possibile ad esempio notificare quando un record è in corso di aggiornamento da parte di un altro utente).
- All'ottimizzazione dell'editing tool per la metadattazione attraverso l'introduzione di utilities (es. auto-completamento dei campi; creazione facilitata dei metadati attingendo dal dataset caricato - es. estensione temporale, tassonomica, spaziale, o i semplici attributi -; annotazione con risorse semantiche; inserimento di riferimenti a pubblicazioni o persone - es. ORCID; possibilità di duplicare sezioni dell'editor - es. replicare i dettagli di una persona in più sezioni del profilo dei metadati -; ecc.).
- All'integrazione con la piattaforma semantica (BE-LISA) della funzionalità di annotazione degli schemi dei metadati e dei loro contenuti mediante l'inserimento di riferimenti a risorse semantiche (URI di risorse RDF servite con tecnologie linked data/sparql) a fronte delle relative etichette testuali presentate nell'interfaccia all'utente.
- **Alla validazione dei singoli record di metadati. La validazione dovrà riguardare lo schema/profilo specifico, le raccomandazioni associate allo standard utilizzato e le regole associate allo standard utilizzato. Dovrà essere gestita in base a diversi livelli di priorità utilizzando ad esempio colori distinti per meglio identificare gli errori che non consentono la pubblicazione del record, gli errori associabili a warning che non pregiudicano la pubblicazione del record ma che forniscono precise indicazioni, suggerimenti e raccomandazioni per la pubblicazione di un record di metadati di qualità che rispetti i principi FAIR, ecc. La validazione automatica dovrà poter essere supportata da quella manuale effettuata dagli utenti con ruolo di reviewer seguendo un opportuno workflow redazionale per l'approvazione o il rifiuto del record di metadati specifico. Il workflow dovrà essere corredato inoltre da un sistema di notification che invierà automaticamente le email necessarie per la gestione dell'intera fase di validazione.**
- **All'integrazione con i servizi di meta (data) curation (vedi sezione 7.5).**
- All'integrazione con l'hub nazionale dei dati di biodiversità ed ecosistemi (BE-HUB).

7.4 Attivazione di servizi web per l'accesso e il controllo delle funzionalità dei moduli di cui sopra.

Lo sviluppo di tali servizi web dovrà essere articolato secondo i protocolli e gli standard internazionali di riferimento. La documentazione dei servizi dovrà essere esposta su web in forma sia human- che machine-readable secondo gli standard e le prassi internazionali.

Tali servizi saranno necessari anche per il riuso e l'integrazione delle funzionalità del Catalogo in altri strumenti dell'infrastruttura di LifeWatch come ad esempio i Virtual Research Environment.

### 7.5 (Meta) Data Curation

Si richiede la progettazione e lo sviluppo di opportuni servizi per verificare la consistenza, congruenza e correttezza semantica, sintattica e tassonomica nell'intero ciclo di vita del metadato. Il set minimo dei servizi di data curation che dovranno essere sviluppati è il seguente:



- **servizi di check tassonomico come ad esempio l'utilizzo della Global Names Architecture che interroga i principali servizi come WoRMS, Catalogue of Life, ecc. (vedi sezione 7.7);**
- **servizi di check sintattico come ad esempio l'utilizzo di correttori ortografici; di algoritmi per il controllo numerico e di formato; di algoritmi per il check degli URL; ecc.**
- **servizi di check semantico come ad esempio l'utilizzo di algoritmi di controllo nomenclaturale/tassonomico, l'interrogazione di ontologie, tesauri, vocabolari controllati**

Tali servizi dovranno essere configurabili e utilizzabili mediante interfaccia web, senza dover intervenire a livello di codice.

Sono altresì richiesti opportuni strumenti ed interfacce utente/macchina per la valutazione della qualità del (meta)dato basate su indicatori appropriati ed efficaci. La Ditta aggiudicatrice in sede di offerta dovrà proporre quali indicatori sviluppare e con quali strumenti gestirli.

#### 7.6. Global Species Databases (GSDs)

Per Global Species Databases (GSDs) si intendono banche dati tassonomiche la cui scala di copertura territoriale è estesa a tutto il mondo o macroregionale. Tali banche dati sono gestite da gruppi di ricerca che lavorano in determinati settori. Ad esempio, si possono ricordare: WoRMS - <http://www.marinespecies.org>, che si occupa della biodiversità marina, Fish Base - <http://www.fishbase.org>, sui pesci; Mammals species of the World - <https://www.departments.bucknell.edu/biology/resources/msw3/>, per i mammiferi, ecc.).

Per l'infrastruttura LifeWatch è importante poter prendere in considerazione queste banche dati sia per incrementare la completezza dei propri dati, sia per validarne la qualità, sia infine per offrire un protocollo di scambio e/o condivisione di dati con esse. Ad esempio, la possibilità di ricercare una specie, ottenendone l'inquadramento tassonomico più aggiornato, le eventuali sinonimie e la presenza nelle GSDs attualmente esistenti, insieme alla possibilità di visualizzare le varie informazioni in griglie comparative da cui sia possibile estrarre quelle di maggior interesse, può far comprendere come LifeWatch possa supportare il mondo della ricerca.

I macro-blocchi funzionali di cui il sistema di integrazione con le GSDs dovrà essere dotato, così da garantire l'interoperabilità con la piattaforma di ricerca europea LifeWatch, sono:

- un sistema di connessione alle GSDs. Tale macro-blocco deve contenere astrazioni come "apertura connessione", "chiusura connessione", "invio dati a connessione", "ricezione dati da connessione";
- un sistema di servizi di ricerca;
- un sistema di cruscotti di comparazione;
- un sistema di servizi di export di dati.

Le componenti sopra elencate saranno impiegate per la realizzazione di una piattaforma che soddisfi i seguenti requisiti minimi:

- Connessione a sorgenti GSDs esistenti. Dovrà essere sviluppata un'interfaccia di connessione uniforme a livello logico. Le sorgenti GSDs concrete dovranno poter essere estese sia in termini di formato, sia in termini di sistema di gestione dei dati originari;
- Ricerca dei dati nelle sorgenti GSDs. Dovrà essere possibile ricercare informazioni nelle GSDs con opportuni motori di ricerca (supporto almeno della full boolean search). Le ricerche devono poter essere fatte sia sui dati, sia sui metadati. Dovrà essere possibile accedere ai relativi metadati a partire da dati ritrovati e, viceversa, accedere ai dati a partire dallo stack di metadati.



UNIONE EUROPEA  
Fondo Sociale Europeo  
Fondo Europeo di Sviluppo Regionale



UNIVERSITÀ  
DEL SALENTO

Dovrà essere possibile memorizzare risultati di ricerche (per raffinamenti successive). Dovrà essere possibile eseguire operazioni di unione, intersezione, differenza di risultati di ricerche.

- Esportazione dei dati dalle sorgenti GSDs nella LifeWatch Knowledge Base. Dovrà essere possibile estrarre dati dalle GSDs ed importarli all'interno di una LifeWatch Knowledge Base.
- Comparazione e correzione. I dati provenienti dalle GSDs dovranno essere utilizzati per procedure di data cleaning per mantenere la coerenza e la qualità della LifeWatch Knowledge Base. Dovranno essere previste procedure automatiche e semi-automatiche di comparazione e correzione dei dati.

### 7.7. Global Names Architecture (GNA)

Attraverso la propria piattaforma l'infrastruttura LifeWatch, rende disponibili dataset comprendenti misure riferite a organismi, popolazioni e comunità aventi differente origine. La diversità delle fonti pone a questa struttura, come a qualsiasi aggregatore di dati, importanti problemi di coerenza nomenclaturale. Tali problemi, se non risolti, possono riflettersi negativamente sulla effettiva possibilità di utilizzo dei dati. La nomenclatura scientifica, utilizzata pressoché universalmente in Biologia sistematica, costituisce una formidabile risorsa, un'etichetta attraverso la quale indicizzare ogni tipo di informazione relativa agli organismi. Il corretto uso della nomenclatura scientifica è normato da appositi codici internazionali: es. ICN (McNeill et al., 2012), ICZN (International Commission on Zoological Nomenclature, 1999), ICNB (Lapage et al., 1992); PhyloCode (Cantino & de Queiroz, 2010).

Purtroppo, l'uso della nomenclatura scientifica all'interno di sistemi informativi non è agevole (cfr. ad es. Vanden Berghe et al., 2015) a causa di due principali ordini di motivi (Patterson et al., 2010):

- **Errori di scrittura e trascrizione. Tali errori, eventualmente presenti in origine tendono ad essere moltiplicati nel processo di digitazione.**
- **Errori dovuti all'uso di nuove combinazioni di nomi. L'attribuzione di un organismo a un determinato taxon è il risultato di un processo al termine del quale all'organismo stesso viene inquadrato all'interno di un sistema classificatorio e gli viene assegnato un binomio scientifico esistente o nuovo. Poiché l'uso di nomi scientifici esistenti può essere ridefinito da chiunque e nuovi taxa vengono continuamente descritti e denominati, la nomenclatura scientifica risulta molto dinamica nel tempo ed è tutt'altro che infrequente l'insorgenza di casi di denominazioni diverse per uno stesso taxon (sinonimie) o l'uso di nomi scientifici identici per taxon differenti (omonimie).**

A fronte delle prospettive di mancanza di uniformità nei dataset ospitati sulla propria piattaforma, la soluzione proposta da LifeWatch-ITA è quella di costruire un proprio sistema di riferimento nomenclaturale interno chiamato "Global Name Architecture" (GNA). La GNA non deve essere costruita ex-novo, ma utilizzerà i Taxonomic Authority File (TAF) forniti da nomenclatori quali, ad esempio, WoRMS ([www.marinespecies.org](http://www.marinespecies.org)), Catalogue of Life ([www.catalogueoflife.org](http://www.catalogueoflife.org)), ecc. La GNA rappresenterà quindi un proxy parziale di due o più Taxonomic Authority File. La GNA dovrà:

- **contenere esclusivamente i nomi (e altre informazioni come l'ID, l'inquadramento tassonomico, ecc.) relativi ai taxa presenti nei dataset ospitati nel sistema di LifeWatch ITA;**



UNIONE EUROPEA  
Fondo Sociale Europeo  
Fondo Europeo di Sviluppo Regionale



UNIVERSITÀ  
DEL SALENTO

- **contenere esclusivamente i nomi scientifici presenti nella GNA saranno esclusivamente i nomi validi (accepted names) considerati tali nei TAF di riferimento al momento della consultazione o, in alternativa, da un board di tassonomi della LifeWatch community.**

La GNA di LifeWatch ITA sarà costruita aggiungendo nuovi nomi via via che nuovi dataset con nuovi taxa vengono aggiunti all'HUB. Poiché Worms, Catalogue of Life e in generale i nomenclatori sono sistemi dinamici in continuo aggiornamento, la GNA deve essere a sua volta sincronizzata attraverso "massive checks" che permettano l'aggiornamento dei nomi della GNA con quelli dei TAF. Questo permetterà di rispondere alla primaria necessità di mantenere aggiornato il thesaurus tassonomico in maniera armonica col contesto internazionale.

La Global Name Architecture rappresenta un componente fondamentale dell'architettura BE-LINDA/BE-HUB. I macro-blocchi funzionali di cui il sistema dovrà essere dotato sono:

- Sviluppo di un Global Resources Discovery System (GBRDS) per avere un servizio di registrazione comune delle sorgenti dati coinvolte, in maniera tale da condividere nomi comuni sulle risorse dati con altre iniziative sulla biodiversità o tematiche ad essa correlata;
- Definizione di un set di standard per lo scambio e la pubblicazione dei dati tassonomici e di nomenclatura lavorando a stretto contatto con altri gruppi quali quello di Biodiversity Information Standards (TDWG);
- Utilizzo di un Integrated Publishing Toolkit come un sistema per pubblicare dati tassonomici e di nomenclatura in maniera tale che in collaborazione con il GBRDS, si fornisca un framework per lo scambio e la pubblicazione di dati per il GNA;
- Sviluppo di servizi di indicizzazione tassonomica basati sul GNA.

I componenti sopra elencati saranno impiegati per la realizzazione di una piattaforma che soddisfi i seguenti requisiti minimi:

- Popolamento iniziale di contenuti. Tutti i dati dovranno essere disponibili in inglese, italiano e, per quanto attiene alla nomenclatura scientifica dei taxa, in latino. Altre lingue comunitarie devono poter essere aggiunte in seguito senza richiedere l'intervento di tecnici programmatori.
- Capacità di inserimento e pubblicazione dati. La piattaforma deve consentire l'inserimento dei termini, dei nomi vernacolari e di una breve descrizione del termine, in modalità multilingua. I dati inseriti devono essere pubblicati nel dizionario. Deve essere consentita la possibilità di modificare e/o eliminare i termini con le relative descrizioni.
- API per software tool. Tutte le funzionalità del servizio devono essere disponibili in opportune librerie, per essere integrate in altri software. Le API devono essere fornite con interfaccia Web Services, quali SOAP o REST.

Modularità ed installazioni multiple. La piattaforma dovrà essere dotata di moduli funzionali attivabili secondo le specifiche necessità. Ogni modulo dovrà essere comprensivo di manuale di installazione, per consentire di installare e configurare il servizio più volte senza costi aggiuntivi, e manuale utente.

## 8 – Requisiti Tecnici Minimi LOTTO 2: BE-HUB



Il catalogo dei MetaDati di LifeWatch ERIC descritto nel bene BE-LINDA, dovrà esser alimentato in maniera automatica da Cataloghi di dati e metadati, implementati e gestiti a livello nazionale dai vari nodi del network LifeWatch. Il fine del prodotto BE-HUB è quindi quello di implementare un HUB nazionale di LifeWatch Italia (LW ITA) per le risorse di dati su biodiversità ed ecosistemi di diversa risoluzione, tipologia e scala spaziale e temporale.

Anche in questo caso LifeWatch Italia ha già avviato una sperimentazione con il tool open source MetaCat (<https://github.com/NCEAS/metacat>). L'attuale installazione del DataPortal di LifeWatch Italia è raggiungibile all'indirizzo <https://dataportal.lifewatchitaly.eu>. (Figura 4).

La sperimentazione ha messo in luce differenti limiti dello strumento ed in particolare:

- la mancanza di profilazione degli accessi, dando la possibilità solo ai proprietari dei dataset di modificare i relativi MetaDati;
- la mancanza di un'interfaccia di amministrazione per la gestione dei dataset ed in particolare la loro eliminazione;
- la mancanza di gestione dei permessi e regole di visibilità sui metadati e sui dataset (privato, sola lettura, ecc.);
- la mancanza di strumenti utili per la citazione (es., esporta citazione);
- la totale assenza di processi di (meta) data curation per garantire la qualità del dato (a tal fine si richiede che la soluzione ideata utilizzi i servizi indicati e sviluppati nel lotto BE-LINDA – vedi sezione 7.5)
- la gestione grezza dei gruppi;
- **la dipendenza dalla rete di DATA-ONE, con cui si condividono i dati, per la gestione e l'utilizzo delle API e per la creazione di sub-portali per i progetti con i relativi dataset associati;**
- l'utilizzo esclusivo di Google Maps con tutte le relative restrizioni e costi in caso di alto volume di dati;
- la mancanza di creazione e gestione di statistiche descrittive attraverso interfacce user-friendly (es., numero di dataset pubblicati nell'ultimo anno, numero di visite, numero di download, dataset più popolari, keyword più cercate, ricerche più frequenti, ecc.);
- la mancanza di indicatori di qualità del (meta)dato;
- la non completa gestione del metadata schema EML 2.2.0 (inflexibilità e modifiche possibili solo con interventi diretti sul codice);
- la mancanza di un sistema di gestione delle notifiche all'amministratore alla creazione di un nuovo dataset;
- la mancanza di un sistema di autenticazione proprietario (è attualmente gestito da ORCID);
- la mancanza di un "workflow redazionale" che supporti tutte le fasi del ciclo di vita di un record: creazione, validazione, pubblicazione, modifica, eliminazione e archiviazione.

Si richiede quindi che la Ditta aggiudicatrice, partendo dai risultati dell'attuale sperimentazione, proponga una soluzione che superi i limiti sopra esposti e soddisfi i requisiti di seguito elencati procedendo alla personalizzazione del software METACAT o proponendo alternativamente un software sviluppato ad hoc che soddisfi le esigenze dell'ente appaltante. Saranno ritenute premianti soluzioni innovative basate su strumenti e risorse semantiche che supportino l'utente nella compilazione (autocompilazione) dei metadati, nell'individuazione e correzione di errori (ed in tutti i processi di data curation in generale), nel facilitare la ricerca dell'utente, ecc.

Sarà valutato positivamente anche il grado di FAIRness delle soluzioni proposte per soddisfare le funzionalità richieste.

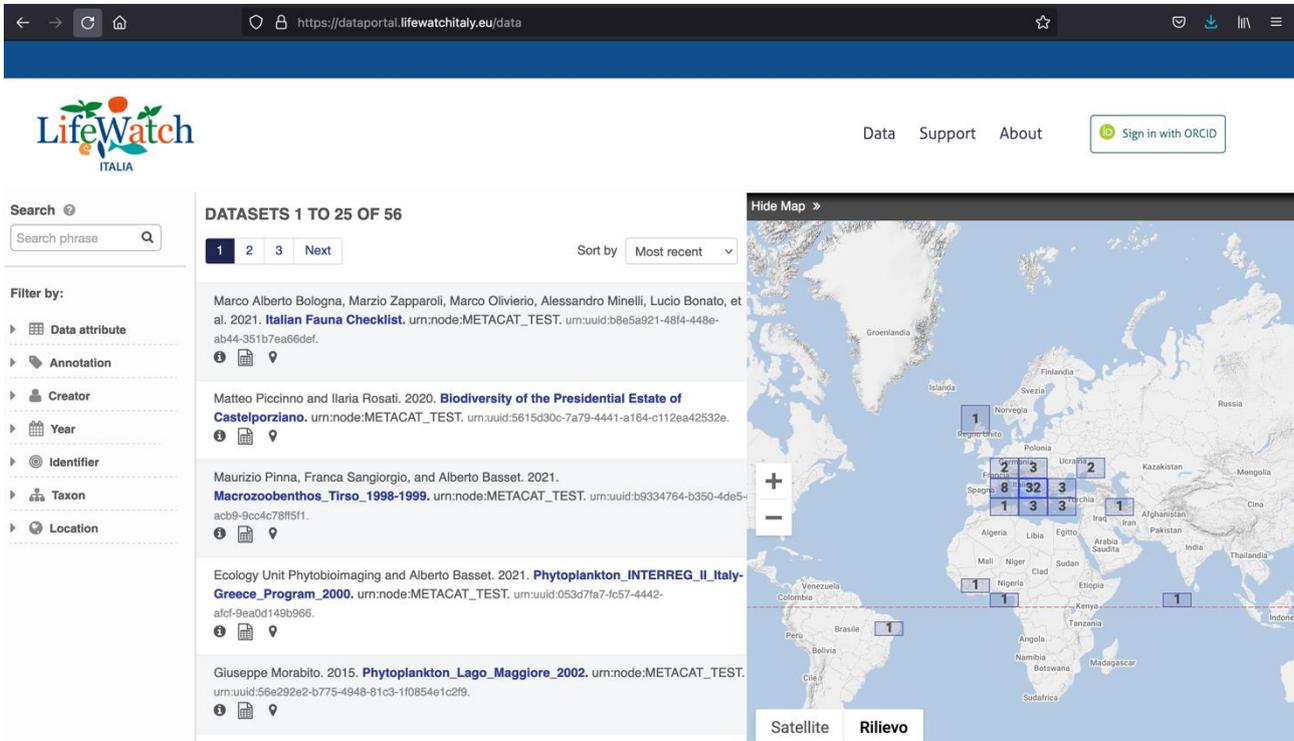


Figura 4. LifeWatch Italy Data Portal

Il software BE-HUB dovrà includere i seguenti moduli:

### 8.1 Interfacce di ricerca

Tale modulo riguarderà la creazione di strumenti di personalizzazione mediante interfaccia web (lato amministratore) della ricerca semplice e avanzata (sia con search box che con facets) consentendo l'interrogazione dell'HUB anche attraverso opportune API.

### 8.2 Ingestione e persistenza di dati e metadati

Tale modulo prevede tutte quelle attività volte:

- all'ampliamento delle tipologie di dati ospitabili dall'HUB (es. di natura biomolecolare, geospaziale, sensoristica, ecc.);
- alla gestione dell'assegnazione di identificatori persistenti alle risorse;
- al versioning di dati e metadati durante tutto il loro ciclo di vita;
- all'ottimizzazione dell'editing tool per la metadattazione attraverso l'introduzione di utilities (es. autocompletamento dei campi, creazione facilitata dei metadati attingendo dal dataset caricato - es. estensione temporale, tassonomica o spaziale-, annotazione con risorse semantiche, inserimento di riferimenti a pubblicazioni o persone, ecc.);
- alla creazione di servizi per il data cleaning tassonomico, sintattico e semantico:
  - **algoritmi per controllo sintattico, di formato e numerico;**

- re-ingegnerizzazione dell'algoritmo di controllo nomenclaturale/tassonomico;
  - re-ingegnerizzazione della Global Names Architecture (GNA).
- all'integrazione con la piattaforma semantica (BE-LISA) della funzionalità di annotazione degli schemi dei metadati e dei loro contenuti mediante l'inserimento di riferimenti a risorse semantiche (URI di risorse RDF servite con tecnologie linked data/sparql) a fronte delle relative etichette testuali presentate nell'interfaccia all'utente.

### 8.3 Harvesting di dataset remoti

I dataset utilizzati nell'ambito delle ricerche su biodiversità ed ecosistemi prodotti dai differenti gruppi di ricerca dovranno poter essere esposti tramite un opportuno tool (vedi sezione 8.4) e mediante interfaccia grafica di amministrazione resi disponibili all'interno del Data Portal dopo aver percorso le varie fasi del ciclo di vita soprattutto per i processi di Data Cleaning e Data Curation. Identificazione e autorizzazione dovranno essere gestiti con i meccanismi previsti da LW-ITA.

### 8.4 Implementazione di un Exposer di metadati e dati

In questo caso si fa riferimento ai dataset in possesso dei Data Provider. La componente software consentirà al nodo del Data Provider che già gestisce le sue risorse in maniera strutturata (ad esempio con un DBMS) di poterle esporre anche in un formato accessibile dall'Harvester descritto sopra. Esempio in tal senso è il tool IPT (Integrated Publishing Toolkit, <https://www.gbif.org/ipt>) utilizzato dal network GBIF per esporre i dati sul data portal.

### 8.5 Interfaccia BE-LINDA

La Ditta aggiudicatrice dovrà assicurare l'interconnessione/integrazione con il bene BE-LINDA, attraverso opportune interfacce in grado di esporre i metadati e i dati presenti in BE-HUB. Sarà cura della Ditta definire il formato di interscambio che si dovrà rifare a standard riconosciuti a livello internazionale, implementarlo e verificarne il corretto funzionamento in fase di rilascio con il bene BE-LINDA.

Tali servizi saranno necessari anche per il riuso e l'integrazione delle funzionalità del prodotto BE-HUB in altri strumenti dell'infrastruttura di LifeWatch come ad esempio i Virtual Research Environment.

### 8.6 Popolamento dei Cataloghi dei dati e dei Metadati

I cataloghi dei Dati e dei Metadati dovranno essere popolati con le risorse dati provenienti dalle istituzioni afferenti alla JRU di LifeWatch Italia e riportate nella seguente tabella.

Istituzione	Referente	Categoria Dati	Digitalizzati	Strutturati	~ record da gestire
UniSalento	<a href="mailto:franca.sangiorgio@unisalento.it">franca.sangiorgio@unisalento.it</a> <a href="mailto:alberto.basset@unisalento.it">alberto.basset@unisalento.it</a>	Ecosystem Parameters	Yes	No	5 Giga
UniFerrara	<a href="mailto:giuseppe.castaldelli@unife.it">giuseppe.castaldelli@unife.it</a> <a href="mailto:mattias.gaglio@unife.it">mattias.gaglio@unife.it</a>	Biological parameters	Yes	No	>10000 records



<b>UniParma</b>	<a href="mailto:giampaolo.rossetti@unipr.it">giampaolo.rossetti@unipr.it</a>	Biological parameters Water parameters	Yes/No	No	>1000 records
<b>CNR-IRSA</b>	<a href="mailto:angela.boggero@cnr.it">angela.boggero@cnr.it</a> <a href="mailto:marzia.ciampittiello@cnr.it">marzia.ciampittiello@cnr.it</a> <a href="mailto:gianluca.corno@cnr.it">gianluca.corno@cnr.it</a> <a href="mailto:ester.eckert@cnr.it">ester.eckert@cnr.it</a> <a href="mailto:diego.fontaneto@cnr.it">diego.fontaneto@cnr.it</a> <a href="mailto:alejandro.martinez@cnr.it">alejandro.martinez@cnr.it</a> <a href="mailto:andrea.lami@cnr.it">andrea.lami@cnr.it</a> <a href="mailto:stefano.mammola@cnr.it">stefano.mammola@cnr.it</a> <a href="mailto:aldo.marchetto@cnr.it">aldo.marchetto@cnr.it</a> <a href="mailto:nicoletta.riccardi@cnr.it">nicoletta.riccardi@cnr.it</a> <a href="mailto:michela.rogora@cnr.it">michela.rogora@cnr.it</a>	Biological parameters Genetic parameters Environmental parameters Ecosystem parameters	Yes/No	Yes/No	15 Giga
<b>UniTrieste</b>	<a href="mailto:martelst@units.it">martelst@units.it</a>	Biological parameters	Yes	Yes	>500000 records
<b>UniTorino</b>	<a href="mailto:sandro.bertolino@unito.it">sandro.bertolino@unito.it</a> <a href="mailto:antonio.rolando@unito.it">antonio.rolando@unito.it</a> <a href="mailto:danieledward.chamberlain@unito.it">danieledward.chamberlain@unito.it</a> <a href="mailto:enrico.caprio@unito.it">enrico.caprio@unito.it</a> <a href="mailto:crisrina.giacoma@unito.it">crisrina.giacoma@unito.it</a> <a href="mailto:claudia.palestrini@unito.it">claudia.palestrini@unito.it</a> <a href="mailto:enrico.barbero@unito.it">enrico.barbero@unito.it</a> <a href="mailto:angela.roggero@unito.it">angela.roggero@unito.it</a> <a href="mailto:matteo.negro@unito.it">matteo.negro@unito.it</a> <a href="mailto:marco.isaia@unito.it">marco.isaia@unito.it</a> <a href="mailto:francesca.barbero@unito.it">francesca.barbero@unito.it</a> <a href="mailto:simona.bonelli@unito.it">simona.bonelli@unito.it</a> <a href="mailto:emilio.balletto@unito.it">emilio.balletto@unito.it</a> <a href="mailto:francesca.bona@unito.it">francesca.bona@unito.it</a> <a href="mailto:livio.favaro@unito.it">livio.favaro@unito.it</a>	Ecosystem parameters	Yes	Yes/No	~ 200000 records
<b>UniFirenze</b>	<a href="mailto:annamaria.nistri@unifi.it">annamaria.nistri@unifi.it</a> <a href="mailto:l.cecchi@unifi.it">l.cecchi@unifi.it</a>	Biological parameters	Yes/No	Yes	100000 records
<b>UniSassari</b>	<a href="mailto:bmpadedda@uniss.it">bmpadedda@uniss.it</a> <a href="mailto:ctsatta@uniss.it">ctsatta@uniss.it</a> <a href="mailto:pulinasi@uniss.it">pulinasi@uniss.it</a>	Biological parameters Environmental parameters Water parameters	Yes/No	Yes	> 500000 records
<b>OGS</b>	<a href="mailto:bcataletto@inogs.it">bcataletto@inogs.it</a>	Atmospheric parameter Biological parameters Chemical parameters Ecosystem parameters Environmental parameters Physical parameters Water parameters	Yes	Yes	> 500000 records
<b>UniPadova</b>	<a href="mailto:carlotta.mazzoldi@unipd.it">carlotta.mazzoldi@unipd.it</a> <a href="mailto:francesca.cima@unipd.it">francesca.cima@unipd.it</a> <a href="mailto:lorenzo.zane@unipd.it">lorenzo.zane@unipd.it</a> <a href="mailto:leonardo.congiu@unipd.it">leonardo.congiu@unipd.it</a>	Biological parameters Genetic parameters	Yes	Yes	200000 records



	<a href="mailto:lpalmeri@unipd.it">lpalmeri@unipd.it</a> <a href="mailto:alessandro.vezzi@unipd.it">alessandro.vezzi@unipd.it</a>	Ecosystem parameters Water parameters			
<b>CNR-IBIOM DBBB-UNIBA</b>	<a href="mailto:g.pesole@ibiom.cnr.it">g.pesole@ibiom.cnr.it</a> <a href="mailto:m.santamaria@ibiom.cnr.it">m.santamaria@ibiom.cnr.it</a> <a href="mailto:b.fosso@ibiom.cnr.it">b.fosso@ibiom.cnr.it</a> <a href="mailto:balechbachir@gmail.com">balechbachir@gmail.com</a> <a href="mailto:carmela.gissi@uniba.it">carmela.gissi@uniba.it</a> <a href="mailto:f.deleo@ibiom.cnr.it">f.deleo@ibiom.cnr.it</a> <a href="mailto:ma.tangaro@ibiom.cnr.it">ma.tangaro@ibiom.cnr.it</a>	Genetic parameters	Yes	Yes	9 Tera
<b>SZN</b>	<a href="mailto:priscilla.licandro@szn.it">priscilla.licandro@szn.it</a>	Biological parameters Environmental parameters Genetic parameters Water parameters	Yes	Yes	10 Tera
<b>UniTorVergata</b>	<a href="mailto:valerio.sbordoni@uniroma2.it">valerio.sbordoni@uniroma2.it</a> <a href="mailto:gabriele.gentile@uniroma2.it">gabriele.gentile@uniroma2.it</a>	Biological parameters Genetic parameters Environmental parameters	Yes/No	Yes/No	50 Giga
<b>Arpa Umbria</b>	<a href="mailto:v.dellabella@arpa.umbria.it">v.dellabella@arpa.umbria.it</a>	Biological parameters Chemical parameters Water parameters	Yes	Yes	10 Giga
<b>UniTuscia</b>	<a href="mailto:cimmaruta@unitus.it">cimmaruta@unitus.it</a> <a href="mailto:darioangeletti@unitus.it">darioangeletti@unitus.it</a>	Biological parameters Genetic parameters Ecosystem parameters	Yes/No	Yes/No	10 Giga
<b>UniBari</b>	<a href="mailto:giuseppe.corriero@uniba.it">giuseppe.corriero@uniba.it</a>	Ecosystem parameters Water parameter	Yes	Yes	>200000 records
<b>IAMB</b>	<a href="mailto:ladisa@iamb.it">ladisa@iamb.it</a>	Biological parameters Ecosystem parameters Chemical parameters Environmental parameters	Yes/No	Yes/No	> 10 Giga
<b>UniBicocca</b>	<a href="mailto:paolo.galli@unimib.it">paolo.galli@unimib.it</a>	Ecosystem parameters	Yes	Yes	10000 records
<b>UniBo</b>	<a href="mailto:alessandro.chiarucci@unibo.it">alessandro.chiarucci@unibo.it</a>	Environmental parameters Genetic parameters Ecosystem parameters Water parameter Chemical parameters	Yes/No	Yes/No	>250000 records

<b>UniMi</b>	<a href="mailto:Raoul.Manenti@unimi.it">Raoul.Manenti@unimi.it</a>	Genetic parameters Ecosystem parameters Water parameter Environmental parameters	Yes/No	Yes/No	> 500000 records
<b>UniPa</b>	<a href="mailto:antonio.mazzola@unipa.it">antonio.mazzola@unipa.it</a>	Ecosystem parameters Water parameter Chemical parameters	Yes	Yes/No	>100000 records

Tale popolamento sarà utilizzato anche per validare l'architettura prodotta. Sarà cura della Ditta aggiudicatrice predisporre un piano di collaudo a tal fine che racchiuda almeno tre Data Providers fra quelli su elencati, l'interoperabilità con un Provider Internazionale come GBIF o WoRMS, la verifica delle differenti procedure di data curation e l'auto-compilazione dei relativi campi di metadati nel conseguente harvesting fatto dai cataloghi.

## 9 - Servizi di avviamento ed esercizio comuni ad entrambi i Lotti

### 9.1 - La manutenzione

L'avviamento del sistema informativo gestionale non è sufficiente a garantire il suo mantenimento e la sua produttività; per questo, l'attività di assistenza post-avviamento, che deve essere assicurata per un anno dopo il collaudo rappresenta il costante monitoraggio della funzionalità dello stesso. Tali competenze verranno trasferite dalle Ditte aggiudicatrici alle risorse interne della stazione appaltante, consentendo all'ente stesso di poter analizzare, governare e soprattutto di poter verificare i risultati finali in termini quantitativi e qualitativi. Il servizio di assistenza on-site dovrà garantire le prestazioni di seguito descritte:

- Assistenza telefonica per l'installazione di nuove release di aggiornamenti e correzioni rese disponibili dal Fornitore e dalle case produttrici coinvolte per i moduli software oggetto della presente fornitura.
- Assistenza telefonica e/o in collegamento remoto al Sistema per la soluzione di eventuali inconvenienti e difetti inerenti ai moduli software oggetto della presente fornitura.
- Risoluzione dei problemi "bloccanti" entro 12 ore decorrenti dal momento della segnalazione con servizio h24 7/7.
- Risoluzione dei problemi "severi" entro 48 ore lavorative decorrenti dal momento della segnalazione.
- Risoluzione dei problemi "minori" entro 72 ore lavorative decorrenti dal momento della segnalazione.
- Sviluppo di correzioni temporanee o soluzioni alternative.
- Forniture degli aggiornamenti e/o nuove release disponibili.



UNIONE EUROPEA  
Fondo Sociale Europeo  
Fondo Europeo di Sviluppo Regionale



UNIVERSITÀ  
DEL SALENTO

- Eliminazione di errori, anomalie e malfunzionamenti di qualunque tipo che dovessero evidenziarsi.
- Assistenza telefonica per problemi di utilizzo e installazione inerenti ai moduli software oggetto della presente fornitura.

## 9.2 - La formazione

Il processo di formazione che le Ditte aggiudicatrici dovranno assicurare al personale informatico dell'ente passerà attraverso diversi moduli, con percorsi differenti a seconda degli obiettivi stabiliti, per raggiungere l'operatività completa e diversificata dei tecnici dell'ente stesso.

L'alternanza di sessioni teoriche con quelle pratiche servirà alla verifica immediata di quanto appreso dalle singole risorse e dal servizio erogato. Le Ditte aggiudicatrici dovranno fornire tutto il materiale didattico (ivi inclusi i manuali utente e manuale tecnico delle piattaforme sviluppate) e strumentale occorrente per la perfetta riuscita dei corsi. Il personale docente dovrà essere di solida e documentata esperienza nell'insegnamento delle materie oggetto dei corsi stessi. A tal proposito le Ditte aggiudicatrici dovranno produrre idonea documentazione attestante l'esperienza del personale incaricato.

Il piano dei corsi deve essere così articolato:

- corso di formazione destinato al personale utente del sistema BE-LINDA per l'utilizzo dei servizi sviluppati. Il corso, della durata di almeno 40 ore da erogare in 5 giorni lavorativi consecutivi e rivolto a circa 20-30 persone con profilo di ricercatore, deve illustrare le funzionalità dei vari moduli, i dati richiesti e quelli forniti e deve mettere in grado l'operatore di poter svolgere autonomamente la propria attività.
- corso di formazione per l'amministrazione del sistema BE-LINDA che deve consentire al personale tecnico dell'ente di rendersi autonomo nella funzionalità di amministrazione della piattaforma; sarà rivolto a 5 tecnici informatici e avrà una durata di almeno 40 ore distribuite in 5 giorni lavorativi consecutivi.
- corso di formazione destinato al personale utente del sistema BE-HUB per l'utilizzo dei servizi sviluppati. Il corso, della durata di almeno 20 ore da erogare in 5 giorni lavorativi consecutivi e rivolto a circa 20-30 persone con profilo di ricercatore, deve illustrare le funzionalità dei vari moduli, i dati richiesti e quelli forniti e deve mettere in grado l'operatore di poter svolgere autonomamente la propria attività.
- corso di formazione per l'amministrazione del sistema BE-HUB che deve consentire al personale tecnico dell'ente di rendersi autonomo nella funzionalità di amministrazione della piattaforma; sarà rivolto a 5 tecnici informatici e avrà una durata di almeno 20 ore distribuite in 5 giorni lavorativi consecutivi.

Tutti i corsi di formazione devono essere tenuti presso i locali dell'Università del Salento.

## 9.3 - L'assistenza

Il servizio deve prevedere l'assistenza in tempo reale (dalle ore 9 alle ore 18, dal lunedì al venerdì, escluse le festività) sia telefonica che tramite e-mail per ogni tipo di problema riscontrato nell'uso del software. Il servizio di manutenzione dovrà comprendere anche l'aggiornamento della piattaforma software e le nuove release.

## Acronimi e definizioni

Acronimo	Definizione
ABCD	Access to Biological Collections Data
API	Application programming interface
EBI	European Bioinformatics Institute
EMBRC	European Marine Biological Resource Centre
EML	Ecological Metadata Language
ESGF	Earth System Grid Federation
ETL	Extract, Transform, Load
GBIF	Global Biodiversity Information Facility
GSD	Global Species Database
GNA	Global Names Architecture
HTC	High Troughput Computing: risolve problemi che possono essere suddivisi in molti task più piccoli che sono essenzialmente indipendenti (coarse-grained problems). Essi richiedono una grande quantità di risorse scarsamente accoppiate



HPC	High Performance Computing: risolve problemi che possono essere suddivisi in piccoli task che necessitano di scambiare informazioni e spesso richiedono sincronizzazione (fine-grained problems). Questi richiedono una grande quantità di risorse altamente accoppiate quali memoria, CPU, disco, ecc.
LSF	Load Sharing Facility
REST	REpresentational State Transfer
RFLP	Restriction Fragment Length Polymorphism
SLA	Service Level Agreement
SOA	Service Oriented Architecture
SOAP	Simple Object Access Protocol
WoRMS	World Register of Marine Species