

Fare sistema: il valore dei dati acqua-salute

La silenziosa rivoluzione dell'acqua
I dati, gli strumenti e il dialogo per una governance sostenibile

Carmen Ciciriello
WHOW project coordinator
Celeris

23.03.2023

1.

Water Health Open knowWledge (WHOW) Project



ARIA



Celeris



CNR-ISTC



ISPRA



Obiettivi



- 1.** Accrescere la comprensione delle dinamiche relative alle acque e all'impatto sulla salute, attraverso lo sviluppo di un **grafo della conoscenza** europeo che colleghi **dati armonizzati** secondo modelli condivisi.
- 2.** Sviluppare **servizi innovativi** con il riutilizzo dei dati disponibili come Linked Open Data, attraverso il **WHOW toolkit**.
- 3.** Supportare l'implementazione di **normative e policy**.

Concetti principali



Linked Open Data

Linked Data (LD) rappresenta un paradigma di accesso ai dati che ha lo scopo di facilitare l'accesso all'informazione da parte di agenti automatizzati (software). Il termine Linked Open Data (LOD) aggiunge ai LD la connotazione di dati aperti, liberamente accessibili.

RDF

Al fine di rendere il dato non solo accessibile, ma anche comprensibile ad un agente automatizzato (associare, cioè, una "**semantica**" al dato) è tuttavia necessario disporre di un modello generale per la rappresentazione dei dati, costituito dal Resource Description Framework (RDF). Ulteriori linguaggi, sempre espressi secondo il modello definito da RDF, consentono poi di definire schemi di dati con elevata espressività: RDF Schema (RDFS) e Web Ontology Language (OWL).

Ontologia

Un'ontologia rappresenta un framework che consente di modellare strutture dati con un livello di dettaglio più preciso. Inoltre, le ontologie consentono di superare l'interpretazione "a mondo chiuso" dei modelli precedenti e di realizzare strutture dati che possano essere interpretati secondo la logica "a mondo aperto" che contraddistingue il Web Semantico. **Linked Data e ontologie rappresentano gli strumenti per accedere e strutturare, rispettivamente, dati connessi tra loro e caratterizzati da una semantica specifica.**

Knowledge Graph (KG)

Un KG consente di collegare le descrizioni di entità (oggetti reali, eventi, situazioni o concetti astratti) alle relazioni che intercorrono tra gli stessi, dove le descrizioni hanno una struttura formale che permette di processarle in maniera efficiente e non ambigua. Un KG semantico è tipicamente rappresentato in RDF.

Principi FAIR

I principi guida FAIR sono stati concepiti per garantire che tutte le risorse digitali possano essere *reperibili, accessibili, interoperabili e riutilizzabili* dalle macchine e dagli esseri umani. Questi principi descrivono le caratteristiche, gli attributi e i comportamenti che consentono una migliore gestione dei dati.

Milestones



**Programma di
Co-creazione**

Maggio 2021

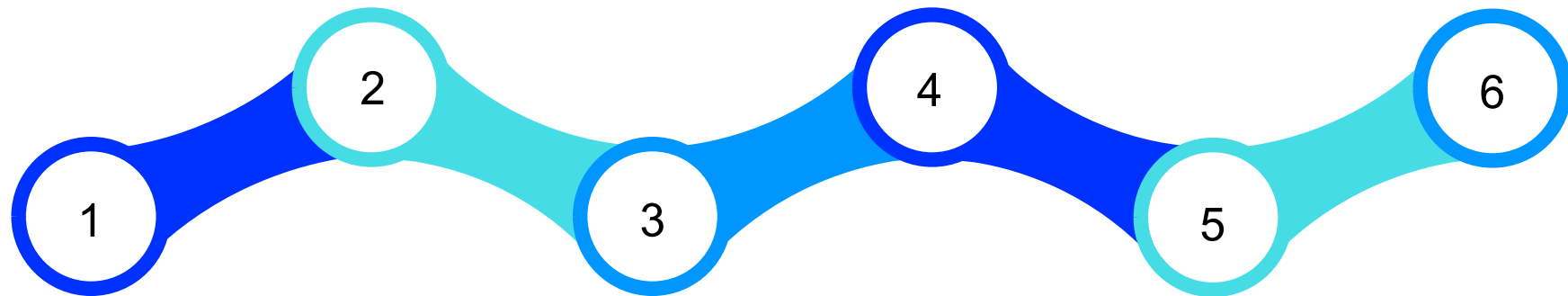
Data pre-processing

**Data pipeline
SDGs e KPIs**

Maggio 2022

**Knowledge Graph
Sostenibilità**

Agosto 2023



Inizio progetto

Settembre 2020

**Casi d'uso
Architettura**

Dicembre 2021

Validazione (MQA)

Giugno 2023

Co-creazione



Oltre 100 stakeholder coinvolti

Meeting tematici per definire i casi d'uso e individuazione delle criticità dei dati

Hackathon organizzati regolarmente per definire l'architettura, le ontologie e i vocabolari

Incontri con Regioni interessate a riutilizzare output di progetto

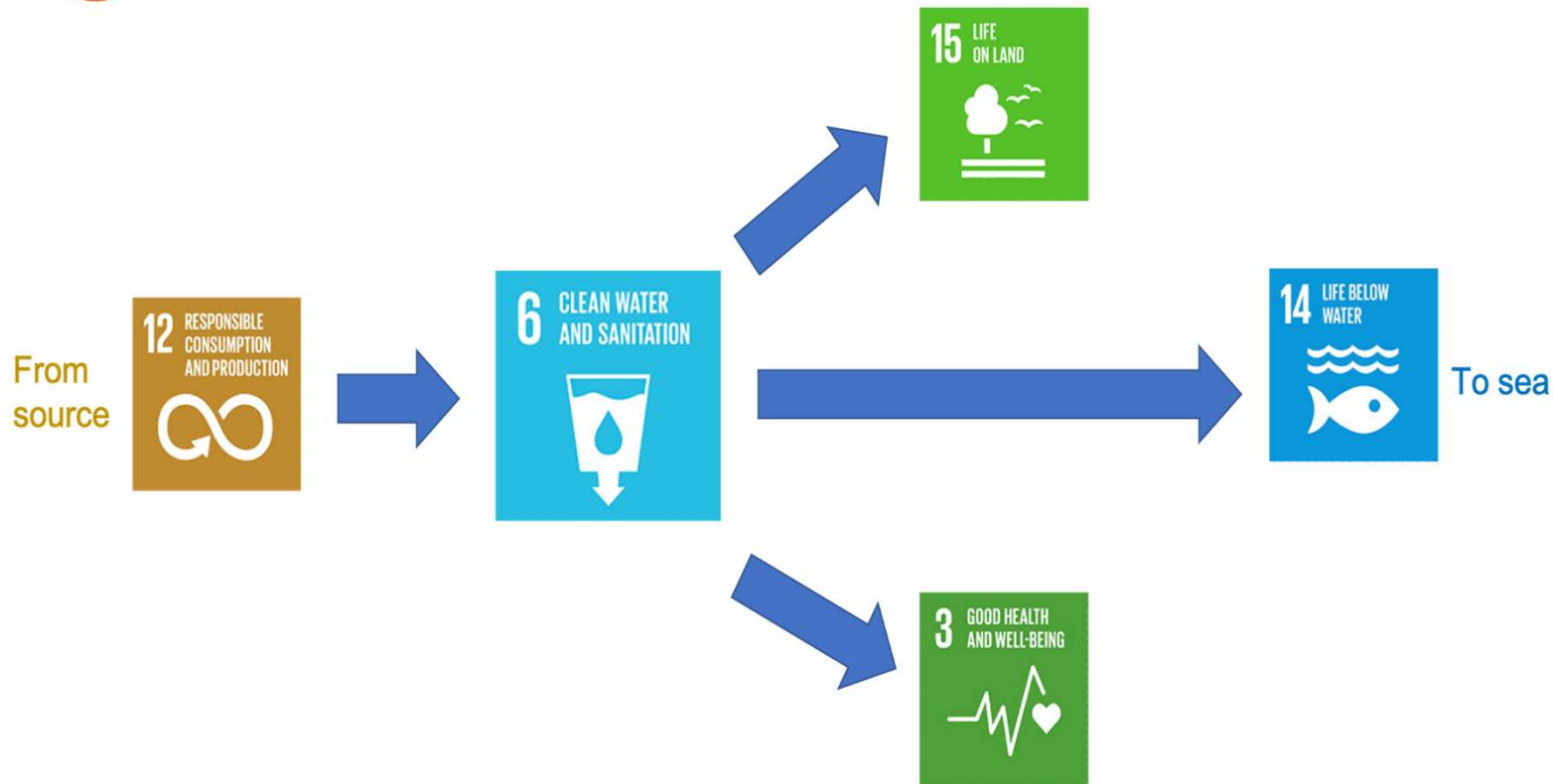


SDGs – Source to Sea



UN 2030 Agenda and Sustainable Development Goals

Reducing emerging pollutants to improve water quality



Source: Stockholm International Water Institute (SIWI)

Casi d'uso



Esposizione umana a inquinanti chimici e biologici nelle acque marine

Collegando dati sul bio-accumulo e esposizione umana a inquinanti chimici e biologici (ad es. *Ostreopsis Ovata*).

Esempi di datasets: CDR-Eionet, Annuario dei dati ambientali (ISPRA)

Acqua per consumo umano

Collegando dati sull'uso e qualità dell'acqua per il consumo umano, misurati secondo i parametri UE microbiologici, chimici e fisici, con dati sulle malattie legate all'acqua e patogeni

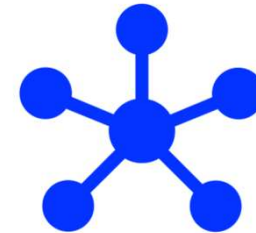
Esempi di datasets: campionamenti per acque potabili, dati analitici di corpi idrici (laghi/fiumi); health conditions by age

Eventi estremi

Collegando dati meteorologici, osservazioni marine e siccità a dati sulla salute umana, alterazione del ciclo idrologico, elementi a rischio (beni culturali)

Esempi of datasets: Weather sensor data, Weekly inflows/outflows per basin

KG – problemi e soluzioni



Dati - problemi

- Mancanza di dati
- Dati chiusi
- Eterogeneità
- Applicazione principi FAIR

Soluzioni

- Programma di co-creazione
- Fare leva su comunità open data
- Definizione di semantica comune

Architettura – problemi

- Nessuna componente centralizzata
- Dati esterni (e.g., Copernicus)

Soluzioni

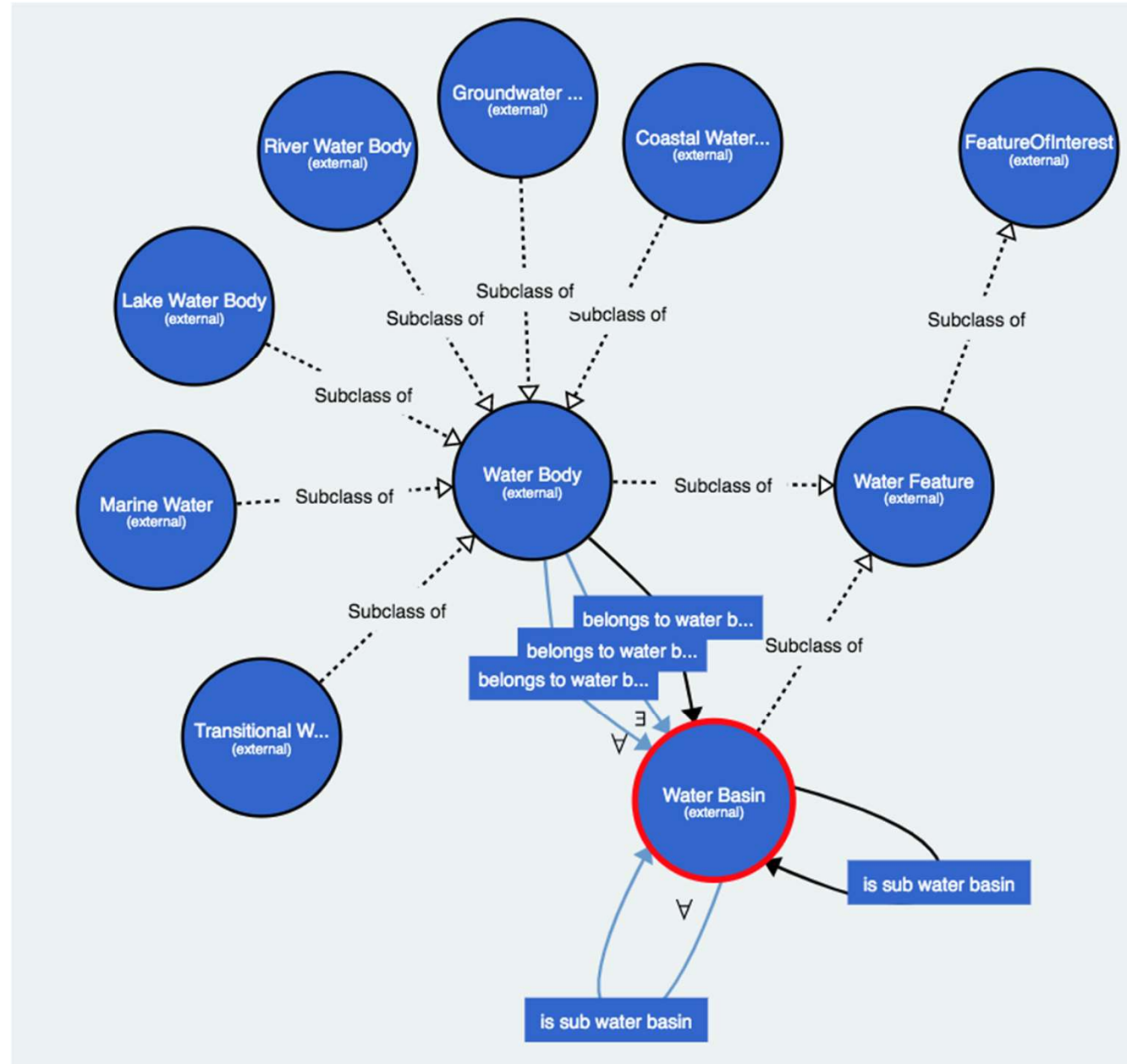
- Definizione di un'architettura di riferimento implementata da ogni data provider
- Utilizzo di knowledge graph virtuali

Ontologie

1. Monitoraggio delle acque
2. Condizioni meteo
3. Idrografia
4. Salute

La **rete ontologica** è organizzata in moduli ontologici che permettono di rappresentare:

Il monitoraggio delle acque, sia marine che interne, sotterranee e destinate al consumo umano; il monitoraggio delle condizioni meteorologiche; e il monitoraggio di specifici indicatori di salute.



Architettura – principi per il design



1. **Completamente distribuito e modulare:**

- Nessun componente centralizzato, tranne semantica e protocolli condivisi.
- La piena proprietà e il controllo dei dati sono preservati (attraverso i grafi virtuali).
- I fornitori di dati (ad es., agenzie ambientali e sanitarie) diventano nodi di WHOW adottando l'architettura di WHOW e distribuendo tutti o alcuni dei suoi componenti.
- La creazione collaborativa di grafi di conoscenza è abilitata dal toolkit WHOW.
- Applicazione principi FAIR.

2. **Multi dominio:** Architettura disegnata indipendente dal dominio

3. **Data quality:** Componenti architetturelari disegnate per abilitare la validazione dei dati

4. **Coinvolgimento della community:** Programma di Co-creazione e coinvolgimento interno degli enti coinvolti

Risultati



Semantic layer

- **Set di risorse semantiche che include le ontologie e linked open data per fornire un layer semantico comune**
- **Standard per rappresentare dati eterogenei provenienti da diverse fonti e domini (acqua e salute)**

Software components

- **Data consumers: strumenti per l'utilizzo dei dati e relativi data models**
- **Data providers: architettura tecnica che offre servizi software necessari per una gestione sostenibile dei processi**

2. Sostenibilità

Riutilizzo dei risultati



WHOW Data Providers



ISPRA

- Ontologie sviluppate
- Vocabolari controllati
- Dataset trasformati in Linked Open Data (es. ostreopsis, pesticidi)
- Dataset ripubblicati con ontologie consolidate (RON, RMN, Rendis)

ARIA

- Apertura di dataset (es. acque consumo umano)
- Contaminazione altri data provider (es. gestori idrici)
- Definizione standard
- Pulizia dei dati
- Implementazione architettura toolkit WHOW

AnTeA



I risultati del progetto WHOW
rappresentano possibili soluzioni
per:

1. **Interoperabilità** del sistema
AnTeA con il SINTAI e altri sistemi
1. Modello di **collaborazione** per
condividere competenze per il
successo di simili iniziative





Carmen Ciciriello
Director
CELERIS

Email: carmen.ciciriello@celeris.group.eu

LinkedIn: <https://www.linkedin.com/in/carmen-ciciriello-b89965/>

Grazie



 whowproject.eu
 info@whowproject.eu



Creative Commons Attribution 4.0
International (CC BY 4.0)



Co-financed by the Connecting Europe
Facility of the European Union

