
Strumenti e servizi per la gestione dei dati prodotti dalla Rete

Autori

Alessandro Oggioni¹, Caterina Bergami², Martina Zilioli¹, Paolo Tagliolato¹

Affiliazione

¹ Consiglio Nazionale delle Ricerche, Istituto per il Rilevamento Elettromagnetico dell'Ambiente (CNR-IREA), Via Bassini 15, 20133 Milano, Italia.

² Consiglio Nazionale delle Ricerche, Istituto di Scienze Marine (CNR-ISMAR), Via Gobetti 101, 40129 Bologna, Italia.

Nel volume prodotto in occasione del quinquennale della Rete LTER-Italia del 2011 è stato dedicato un intero capitolo alla descrizione della “Infrastruttura Spaziale di Dati nei siti LTER” (Oggioni *et al.* 2011). Sebbene essa fosse solo agli albori, era già chiaro che le Reti, Nazionale ed Europea, lavorassero congiuntamente per proporre modalità di gestione e curatela del dato in linea con le legislazioni e le pratiche internazionali. Da allora, le attività intraprese sono state numerose e, soprattutto nell'ambito dei due progetti europei Life+ EnvEurope ed eLTER H2020, è stato possibile implementare strumenti per l'archiviazione, la ricerca e la distribuzione di dati, e relativi metadati, specificamente pensati per la ricerca ecologica di lungo termine.

Qui di seguito verranno descritti gli strumenti realizzati, gli obiettivi, le modalità e le ragioni per cui un dato, una volta raccolto, debba essere condiviso.

Se da un lato la cura dei dati e la stesura di metadati adeguatamente strutturati sono processi chiave, dall'altro la disponibilità nella rete di entrambi permette un miglior utilizzo, un reale confronto e un'efficace analisi anche quando essi sono raccolti in ambiti progettuali diversi.

Il lavoro intrapreso nella comunità LTER è destinato a fornire dati controllati e affidabili a supporto della ricerca, per l'elaborazione di politiche ambientali efficaci e per la valutazione del loro impatto. Al fine di promuovere lo scambio e la condivisione di informazioni, i dati devono essere descritti utilizzando schemi di metadato specifici e vocabolari controllati. Ciò richiede un'adeguata descrizione dei dataset acquisiti dai siti LTER, dei servizi forniti e dei sistemi di osservazione, nonché lo sviluppo di infrastrutture informatiche che consentano la ricerca e l'accesso ai dati attraverso la rete (*web-based*).

La gestione dei dati per la maggior parte dei siti LTER è responsabilità di chi gestisce il sito stesso (Mirtl *et al.* 2018). Ciò crea un'elevata eterogeneità di formati e anche di strumenti di gestione.

Per tentare di superare questa eterogeneità, la Rete LTER europea fornisce, ad uso di tutti i network che ne fanno parte, una serie di strumenti e metodi (Zilioli *et al.* 2021) a sostegno della comunità LTER per la gestione dei metadati e dei dati di diverse entità della rete di ricerca (siti, network, dataset, e persone) (Fig. 1). Questo consente di ottenere uniformità nelle pratiche di gestione di metadati e dati a livello dei nodi nazionali e regionali mediante norme e servizi comuni.

Di seguito vengono presentati e brevemente descritti alcuni degli strumenti a disposizione della comunità LTER e utilizzati anche dalla Rete LTER-Italia; ognuno di essi assicura diversi livelli di interoperabilità (sintattica, strutturale, semantica e legale).

Citare questo capitolo come segue: Oggioni A., Bergami C., Zilioli M., Tagliato P. (2021). Strumenti e servizi per la gestione dei dati prodotti dalla Rete, p. 33-39. DOI: 10.5281/zenodo.5572354. In: Capotondi L., Ravaioli M., Acosta A., Chiarini F., Lami A., Stanisci A., Tarozzi L., Mazzocchi M.G. (a cura di) (2021). La Rete Italiana per la Ricerca Ecologica di Lungo Termine. Lo studio della biodiversità e dei cambiamenti, pp. 806. DOI: 10.5281/zenodo.5570272.

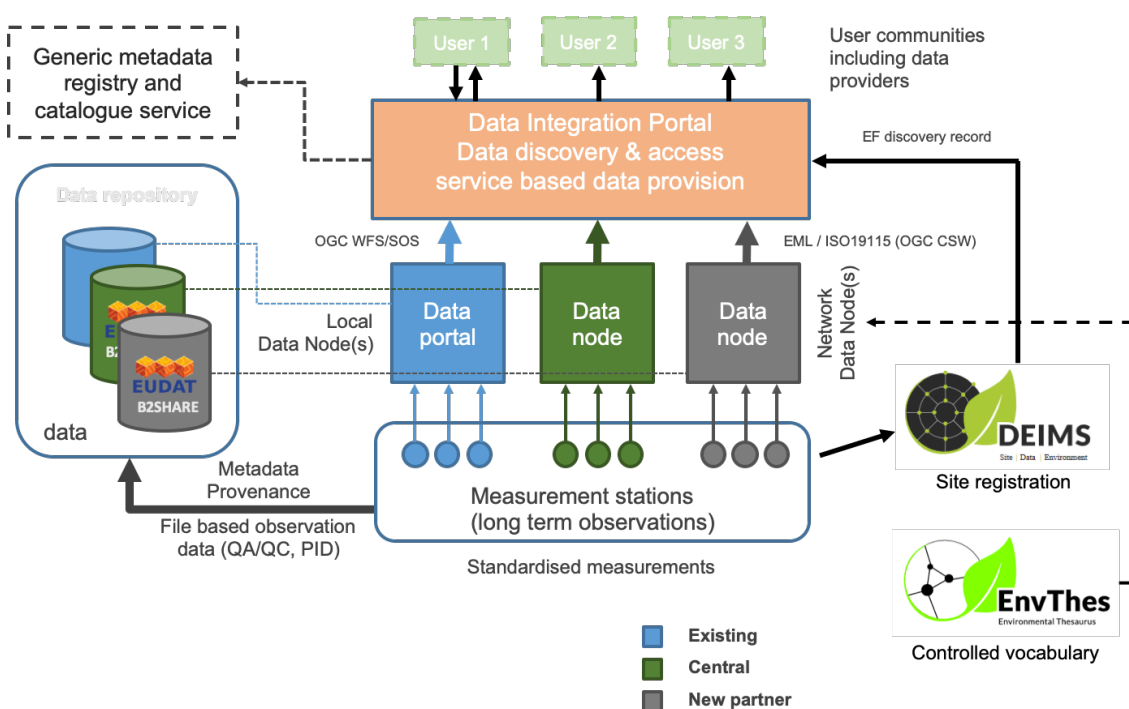


Fig. 1 - eLTER Information System (da Peterseil et al. 2018)

Dynamic Ecological Information Management System – Site and Dataset Registry (DEIMS-SDR)

DEIMS-SDR² è il registro ufficiale dei siti e dei dataset di tutta la Rete LTER Internazionale (ILTER) (Wohner *et al.* 2019). È un sistema sviluppato dall'Università del Nuovo Messico, l'Università di Puerto Rico, l'Università del Wisconsin e da Palantir.net ed esteso poi da LTER Europa al fine di descrivere al meglio i dati ecologici e le strutture di monitoraggio ambientale. Nella attuale versione DEIMS-SDR consente di descrivere, i siti di ricerca LTER e le entità ad essi collegati (Fig. 2): i network a cui i siti afferiscono, i dataset raccolti nei siti e le persone che lavorano nei siti.

DEIMS-SDR consente una descrizione dettagliata dei singoli siti di ricerca con informazioni sulle attività di ricerca, sulle infrastrutture tecniche (e.g. dispositivi sensoristici, strumentazioni per il campionamento) presenti e sui parametri ambientali misurati.

A ogni sito viene attribuito un codice DEIMS.ID³, che permette di identificarlo e citarlo in maniera univoca e persistente, esattamente come si farebbe con il DOI di una pubblicazione.

Attraverso DEIMS-SDR è inoltre possibile accedere direttamente ai metadati relativi ai dataset o ad altri prodotti della ricerca che fanno riferimento a un determinato sito. I singoli siti possono utilizzare questo strumento per descrivere, attraverso uno schema di metadato comune, i dataset prodotti e condivisi attraverso repository generalisti o più specifici (es. Zenodo, Pangae, etc...). La Rete LTER-Italia suggerisce di utilizzare Zenodo, in quanto è un archivio per i tutti i dati riguardanti le scienze della vita, le scienze naturali e le scienze ingegneristiche, è ospitato dal CERN e finanziato dall'Unione Europea. Zenodo ha una serie di funzionalità utili a supporto della scienza aperta e i suoi contenuti sono indicizzati su "Thomson Reuters Data Citation Index".

DEIMS-SDR, grazie alle sue specifiche, assicura i livelli di interoperabilità sintattica, strutturale e legale dei metadati.

² <https://deims.org>

³ <https://deims.org/docs/deimsid.html>

Search ...

Latest Updates

-  Mar Piccolo of Taranto phyto**biomass from 2011 to 2019**
2020-02-03 16:16
-  Simon Stevin Research Vessel - Belgium
2020-01-31 13:36
-  VLIZ Thornton Buoy - Belgium
2020-01-31 12:15
-  Wytham: solar radiation - 1991-2015 (UK Environmental Change Network (ECN))
2020-01-31 09:46
-  Wytham: precipitation - 1991-2015 (UK Environmental Change Network (ECN))
2020-01-31 09:45
-  Wytham: air temperature - 1991-2015 (UK Environmental Change Network (ECN))
2020-01-31 09:44
-  Wytham: UK Environmental Change Network (ECN) Stream Water Chemistry data: 1992-2015
2020-01-31 09:43

Welcome to DEIMS-SDR

DEIMS-SDR (Dynamic Ecological Information Management System - Site and dataset registry) is an information management system that allows you to discover long-term ecosystem research sites around the globe, along with the data gathered at those sites and the people and networks associated with them. DEIMS-SDR describes a wide range of sites, providing a wealth of information, including each site's location, ecosystems, facilities, parameters measured and research themes. It is also possible to access a growing number of datasets and data products associated with the sites. [Read more about DEIMS](#)

RESOURCES

-  Sites
-  Datasets
-  Sensors
-  Activities

Fig. 2 - Pagina introduttiva di DEIMS-SDR

Data Integration Portal (DIP)

DIP⁴ è il catalogo principale della comunità LTER attraverso cui è possibile ricercare, accedere e visualizzare le fonti di dati forniti da tutti i nodi della Rete (Fig. 3).

I dati forniti come servizio OGC SOS possono essere dinamicamente visualizzati e scaricati. DIP assicura l'interoperabilità sintattica e strutturale di dati e metadati.

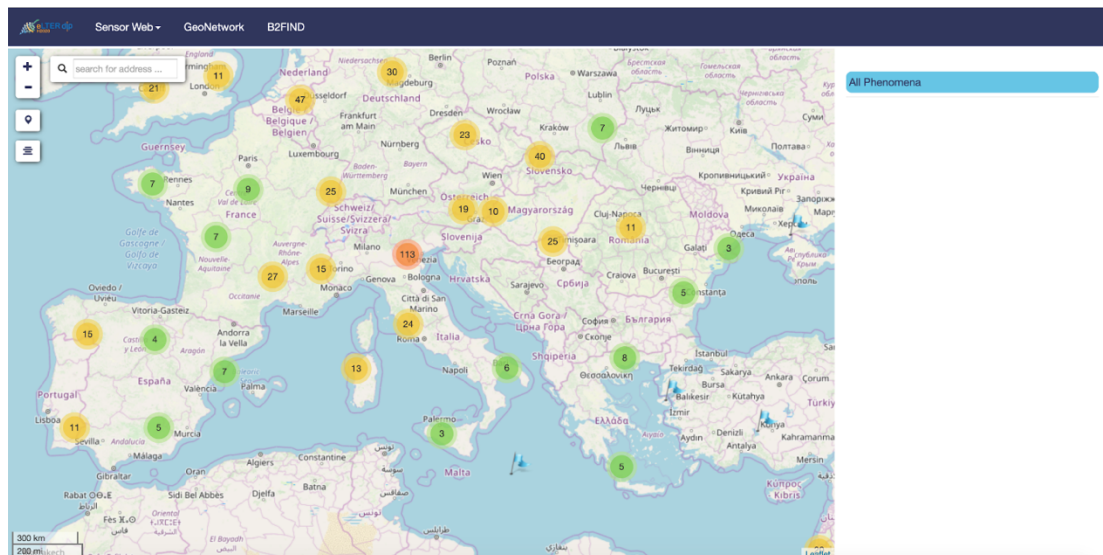


Fig. 3 - DIP Data Integration Platform

⁴ <http://dip.lter-europe.net>

Environmental Thesaurus (EnvThes)

EnvThes⁵ è il vocabolario controllato appositamente implementato per la comunità che si occupa di ricerca ecologica e costituisce il supporto semantico per la descrizione dei siti e dei dati raccolti (viene ad esempio utilizzato, in DEIMS-SDR, per l'autocompletamento delle parole chiave e dei parametri).

Esso è basato sul vocabolario controllato sviluppato dalle ReteILTER, è libero e aperto e rappresenta uno sforzo collettivo della comunità LTER, in quanto viene costantemente aggiornato ed è tradotto in molteplici lingue.

Geoinformation Enablement Toolkit (GET-IT)

GET-IT⁶ è una suite software che consente ai ricercatori esperti di dominio di abilitare un nodo per la distribuzione interoperabile di dati e metadati. Utilizzando norme pertinenti con le specifiche del web geospaziale (Open Geospatial Consortium – OGC⁷ e W3C⁸), esso consente la distribuzione, attraverso servizi web, dei dati sia con caratteristiche spaziali (e.g. immagini georiferite) sia osservazioni *in-situ* e la creazione di propri archivi di dati territoriali e agevola l'inserimento e la cura di dati e metadati (Menegon *et al.* 2018).

In particolare la nuova implementazione del software include:

- un client, denominato EDI, per la metadattazione, che consente la creazione e la convalida di metadati in conformità a diversi profili o modelli (Pavesi *et al.* 2016). EDI consente di collegare fonti di dati esterne che vengono rese disponibili come risorse semantiche;
- una pagina per la visualizzazione delle caratteristiche del sensore registrato;
- un gestore per le osservazioni raccolte da stazioni *in-situ*, che consente la registrazione di sensori utilizzando il profilo Sensor Metadata Language (SensorML)⁹;
- un'interfaccia di caricamento delle osservazioni raccolte *in-situ*, tramite semplice copia e incolla;
- un client di visualizzazione delle osservazioni caricate su mappa;
- un servizio per l'accesso tramite servizi web alle osservazioni *in-situ*.
- GET-IT assicura i livelli di interoperabilità sintattica, strutturale e semantica di dati e metadati.

Data Policy

Per Data Policy si intende l'insieme di principi più o meno vincolanti che regolano la gestione e la pubblicazione dei dati della ricerca all'interno di un'organizzazione o di una istituzione.

In ambito LTER documenti riguardanti la politica sui dati sono stati prodotti e discussi per lungo tempo e, sebbene esistesse già un documento in forma di bozza a partire dal 2016, solo con la conclusione del progetto eLTER H2020 è stato prodotto un Deliverable interamente dedicato a “*Governance and data policy for sharing and publishing of data*” (Kunkel *et al.* 2019).

eLTER RI incoraggia la condivisione pubblica di dati e metadati, sottolineando la necessità che i dati raccolti a seguito della concessione di fondi pubblici, debbano essere resi disponibili online gratuitamente, con il minor numero possibile di restrizioni e su base non discriminatoria. Gli scienziati LTER dovrebbero fare ogni sforzo per rilasciare i dati in modo tempestivo, correlandoli di metadati accurati e completi per ridurre il numero di interazioni tra l'utente pubblico e il proprietario del dato.

La politica di condivisione dei dati in LTER mira a fornire orientamenti generali e raccomandazioni per lo sviluppo e l'attuazione di politiche locali a livello di rete o sito di ricerca e favorisce altresì l'inclusione della rete internazionale, delle reti nazionali o dei singoli nodi in framework e iniziative di reclutamento (*data mobilization*) e aggregazione di dati su scala globale come quella legata alle Essential

⁵ <http://vocabs.lter-europe.net/edg/tbl/EnvThes.editor>

⁶ <http://get-it.it>

⁷ <https://www.opengeospatial.org>

⁸ <https://www.w3.org>

⁹ <https://www.opengeospatial.org/standards/sensorml>

Biodiversity Variables (EBV) attivata da Group on Earth Observations Biodiversity Observation Network (GEO BON)¹⁰ (Zilioli *et al.* 2019). La Rete LTER-Italia ha recepito le indicazioni di eLTER e si è dotata di una data policy che è stata presentata per la prima volta durante l'Assemblea annuale della Rete, tenutasi a Bolzano a Maggio 2018. Nel documento vengono definiti i tipi di dati che la Rete produce e che può rendere accessibili e vengono proposte una serie di regole per la condivisione sia dei metadati sia dei dati, utilizzando gli strumenti informatici messi a disposizione della comunità LTER (vedi sopra). La data policy LTER-Italia è accessibile a questo link: <https://doi.org/10.5281/zenodo.3763378>.

Buone pratiche per la gestione dei dati in LTER-Italia

Facendo seguito a quanto indicato nelle Data Policy redatte da eLTER e LTER-Italy e a quanto descritto nei diversi deliverable del progetto eLTER H2020 reperibili on-line¹¹, qui di seguito vengono descritte alcune buone pratiche per la gestione dei dati raccolti in ambito LTER, seguendo il ciclo di vita dei dati (Fig. 4) dalla fase di curatela a quella di utilizzo e passando per la pubblicazione e l'analisi. Le pratiche di acquisizione sono specifiche per sito di ricerca e utilizzano metodi e protocolli definiti in base al tipo di ricerca svolta.

Seguire queste buone pratiche permette al ricercatore LTER di valorizzare i propri dati, rendendoli accessibili, citabili e riutilizzabili, dando ulteriore valore alla propria attività di ricerca e a quella di coloro che collaborano alle diverse fasi del ciclo di vita del dato.

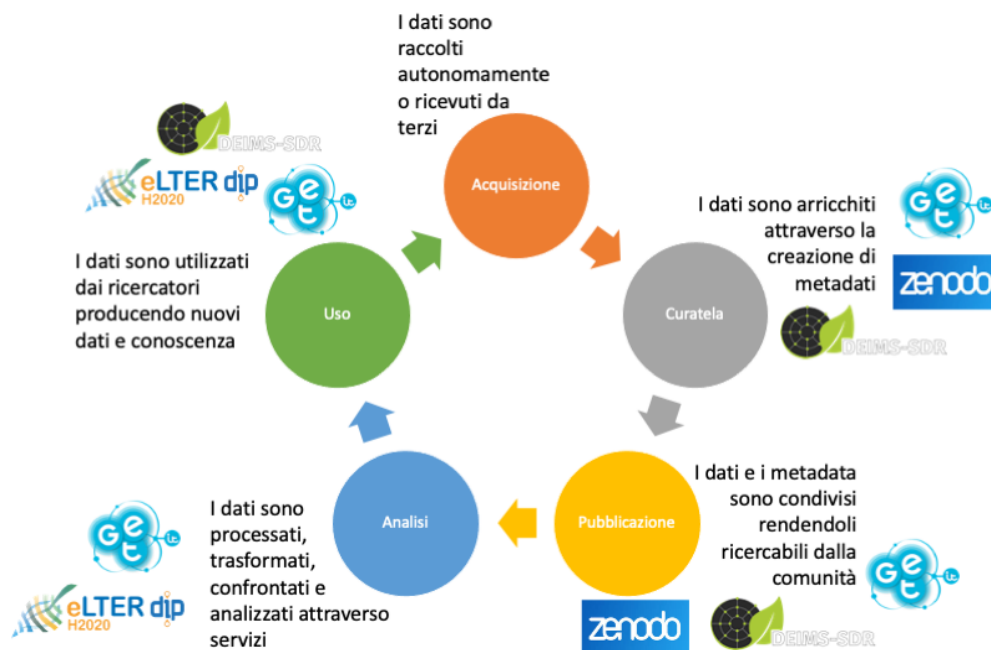


Fig. 4 - Ciclo di vita del dato dall'acquisizione all'uso e relativi strumenti per ogni pratica. Modificato da ENVRI Reference Model¹²

¹⁰ <https://geobon.org>

¹¹ <https://www.lter-europe.net/document-archive/elter-h2020-project-files>

¹² <https://confluence.egi.eu/display/EC/Model+Overview>

Nella tabella (Tab. 1) vengono elencate le fasi, dalla curatela all'uso del dato, descritte le operazioni possibili e indicati gli strumenti a disposizione.

Descrizione della pratica	Fasi del ciclo di vita (Fig. 4)	Strumenti messi a disposizione dall'infrastruttura eLTER
Metadattazione del dataset Metadattazione dei sensori	Curatela	DEIMS-SDR per la metadattazione del dato attraverso il dataset metadata model (DSMM, Kliment <i>et al.</i> 2012) ¹³ di LTER armonizzato con le INSPIRE metadata regulation ¹⁴ e realizzato usando Ecological Metadata Language (EML) ¹⁵ metadata specification; GET-IT (EDI) e DEIMS-SDR per la metadattazione dei dataset e dei sensori; EnvThes per arricchire semanticamente i metadati così che il loro significato sia armonizzato.
Pubblicazione del dataset Citabilità del dataset Pubblicazione del data paper ¹⁶	Pubblicazione	GET-IT e Zenodo per la pubblicazione del dataset così che sia ricercabile; <u>Zenodo per l'assegnazione di un DOI al dataset così che sia citabile.</u>
Analisi comparata di serie temporali	Analisi	<u>GET-IT</u> e <u>DIP</u> per la georeferenziazione del punto di acquisizione e la visualizzazione simultanea di più serie temporali.
Scaricamento e riuso del dataset	Uso	<u>GET-IT</u> , <u>DIP</u> e <u>DEIMS-SDR</u> per il download o l'accesso tramite servizi web al dataset e suo riutilizzo tenendo conto della licenza ad esso attribuita.

¹³ <https://deims.org/models>

¹⁴ <https://inspire.ec.europa.eu/documents/inspire-metadata-regulation>

¹⁵ <https://rdamsc.dcc.ac.uk/msc/m16>

¹⁶ Riviste per la pubblicazione dei data paper richiedono la pubblicazione del dataset su repository riconosciuti e la persistenza della risorsa dato attraverso l'uso di PID (es. DOI).

Bibliografia

- Kliment T. & Oggioni A. (2012). EnvEurope (LTER-Europe) Metadata Specification for Dataset Level.
- Kunkel R., Peterseil J., Oggioni A., Wohner Ch., Watkins J., Minić V., Sorg J. (2019). D3.2 Governance and data policy for sharing and publishing of data. <https://www.lter-europe.net/document-archive/elter-h2020-project-files/d3-2-governance-data-policy>.
- Menegon S., Oggioni A., Carrara P. (2018). GET-IT, Geoinformation Enabling ToolKIT starterkit®, Versione 1.0. Roma, CNR Pubblicazioni 2018. pp. 43, ISBN 978-88-8080-299-0. DOI: 10.5281/zenodo.3678070.
- Mirtl M., Borer E.T., Djukic I., Forsius M., Haubold H., Hugo W., Jourdan J., Lindenmayer D., McDowell W.H., Muraoka H., Orenstein D.E., Pauw J.C., Peterseil J., Shibata H., Wohner C., Yu X., Haase P. (2018). Genesis, goals and achievements of long-term ecological research at the global scale: a critical review of ILTER and future directions. *Sci. Total Environ.*, 626, p. 1439-1462, DOI: 10.1016/j.scitotenv.2017.12.001.
- Oggioni A., Manca D., Kliment T. & Bastianini M. (2012). Infrastruttura Spaziale di Dati nei siti LTER – Interoperabilità, archiviazione, condivisione, metadattazione. In R. Bertoni (Ed.), *La rete italiana per la ricerca ecologica a lungo termine (LTER-Italia)* (pp. 23-31). Roma: Aracne.
- Oggioni A., Wohner Ch., Watkins J., Ciar D., Schentz H., Lanucara S., Minić V., Škrbić S., Bodroški Z., Kunkel R., Sorg J., Kliment T., Sanchez F., Magagna B., Peterseil J. (2018). D3.1 eLTER State of the art and requirements. <https://www.lter-europe.net/document-archive/elter-h2020-project-files/d3-1-data-integration>.
- Pavesi F., Basoni A., Fugazza C., Menegon S., Oggioni A., Pepe M., Tagliolato P., Carrara P. (2016). EDI – A Template-Driven Metadata Editor for Research Data. *Journal of Open Research Software – JORS* 4. DOI: 10.5334/jors.106.
- Peterseil J., Watkins J., Minić V., Kunkel R., Oggioni A., Wohner Ch., Magagna B., Ciar D., Sorg J., Mirtl M., Crnojevic V. (2018). eLTER Information System-a European contribution to share scientific data from long term ecosystem research in Europe. <https://lternet.edu/wp-content/uploads/2018/03/2018DatabitsSpringIssue-web.pdf>.
- Tagliolato P., Fugazza C., Oggioni A., Carrara P. (2019). Semantic Profiles for Easing SensorML Description: Review and Proposal. *ISPRS Int. J. Geo-Inf.* 2019, 8(8), 340. DOI 10.3390/ijgi8080340.
- Wohner C., Peterseil J., Poursanidis D., Kliment T., Wilson M., Mirtl M., Chrysoulakis N. (2019). DEIMS-SDR – A web portal to document research sites and their associated data. *Ecological Informatics*, 51:15-24, ISSN: 1574-9541. DOI: 10.1016/j.ecoinf.2019.01.005.
- Zilioli M., Oggioni A., Tagliolato P., Pugnetti A., Carrara P. (2019). Feeding Essential Biodiversity Variables (EBVs): actual and potential contributions from LTER-Italy. *Nature Conservation* 34, 477-503. DOI: 10.3897/natureconservation.34.30735.
- Zilioli M., Bergami C., Carrara P., Fugazza C., Oggioni A., Pugnetti A. Tagliolato Acquaviva d’Aragona P. (2021). Enabling the Reuse of Long-Term Marine Biological Observations in Essential Variables Frameworks Through a Practical Approach. *Front. Mar. Sci.*, 8:645997. DOI: 10.3389/fmars.2021.645997.