Ecological Informatics 51 (2019) 15–24

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

DEIMS-SDR – A web portal to document research sites and their associated data

Christoph Wohner[a,⁎], Johannes Peterseil[a], Dimitris Poursanidis[b,⁎], Tomáš Kliment[c], Mike Wilson[d], Michael Mirtl[a], Nektarios Chrysoulakis[b]

[a] Environment Agency Austria, Spittelauer Lände 5, 1090 Wien, Austria
[b] Foundation for Research and Technology–Hellas (FORTH), Institute of Applied and Computational Mathematics, N. Plastira 100, Vassilika Vouton, Iraklion GR70013, Greece
[c] KLIMETO, Južná 698/4, Rožňava, Slovakia
[d] Centre for Ecology and Hydrology, Library Avenue, Lancaster Environment Centre, Lancaster, Bailrigg LA1 4AP000, United Kingdom of Great Britain and Northern Ireland

ARTICLE INFO

Keywords:
DEIMS-SDR
Metadata
Protected areas
Research sites
Long term observations
Ecosystem status
Persistent identification

ABSTRACT

Climate change and other drivers are affecting ecosystems around the globe. In order to enable a better understanding of ecosystem functioning and to develop mitigation and adaptation strategies in response to environmental change, a broad range of information, including in-situ observations of both biotic and abiotic parameters, needs to be considered. Access to sufficient and well documented in-situ data from long term observations is therefore one of the key requirements for modelling and assessing the effects of global change on ecosystems. Usually, such data is generated by multiple providers; often not openly available and with improper documentation. In this regard, metadata plays an important role in aiding the findability, accessibility and reusability of data as well as enabling reproducibility of the results leading to management decisions. This metadata needs to include information on the observation location and the research context. For this purpose we developed the Dynamic Ecological Information Management System – Site and Dataset Registry (DEIMS-SDR), a research and monitoring site registry (https://www.deims.org/) that not only makes it possible to describe in-situ observation or experimentation sites, generating persistent, unique and resolvable identifiers for each site, but also to document associated data linked to each site. This article describes the system architecture and illustrates the linkage of contextual information to observational data. The aim of DEIMS-SDR is to be a globally comprehensive site catalogue describing a wide range of sites, providing a wealth of information, including each site's location, ecosystems, facilities, measured parameters and research themes and enabling that standardised information to be openly available.

1. Introduction

Global environmental change has a variety of consequences for ecosystems and the services they provide for human societies (Carpenter et al., 2006; Isbell et al., 2011). Prominent examples are the reduction in the terrestrial carbon sink due to climate change (Ciais et al., 2005; Pan et al., 2011; Seidl et al., 2014), or the dramatic loss of species, particularly caused by conversion and degradation of their habitats (Butchart et al., 2010; Pereira et al., 2010; Ceballos et al., 2015; Newbold et al., 2015). In order to address and assess the causes, the magnitude and the effects of these changes, detailed data is required (Verheyen et al., 2016). This encompasses environmental observation and monitoring data, either remotely-sensed or through in-situ measurements as well as environmental/ecological modelling, relating to biodiversity management. In order to ensure targeted management of locations which receive protection because of their recognised natural and/or ecological values (called 'protected areas'), especially with regard to a changing environment and the need to meet environmental protection goals, the integration of in-situ data, remote sensing products, models and model outputs is needed (Nagendra et al., 2013).

To facilitate the findability and reuse of in-situ data, sufficient and standardised documentation is required (Michener et al., 1997). In this respect, science and evidence-based decision-making need access to data that is: (1) well documented, (2) quality controlled and (3) provides clear rules on usage. Applying the FAIR principles is one of the major goals in order to fulfil these needs. The FAIR Data Principles are a

⁎ Corresponding authors.
E-mail addresses: christoph.wohner@umweltbundesamt.at (C. Wohner), dpoursanidis@iacm.forth.gr (D. Poursanidis).

https://doi.org/10.1016/j.ecoinf.2019.01.005
Received 3 October 2018; Received in revised form 2 January 2019; Accepted 6 January 2019
Available online 21 February 2019

set of guiding principles to make data Findable, Accessible, Interoperable and Reusable (Wilkinson et al., 2016). These principles provide guidance for scientific data management and stewardship and are relevant to all stakeholders in the current digital landscape. They directly address data producers and data publishers to promote the maximum use and reuse of research data.

This not only needs to include the description of data objects but also the context of each observation, e.g. the research sites or research infrastructures (RIs) where the measurements have been taken. For any place-based observation, information on the observation facilities (e.g. research sites) is an intrinsic and important part of the description, discovery and reuse of data and expertise (Mirtl et al., 2018).

RIs, networks or research projects funded by the European Union (EU) are often required to produce standardised metadata records in order to meet project requirements. For networks that lack their own system, a light-weight solution is needed to fulfil metadata requirements. DEIMS-SDR (https://www.deims.org/) is designed to meet those needs. The aim of DEIMS-SDR is to be a catalogue of in-situ observation or experimentation facilities covering all (terrestrial) biomes, allowing data associated with each site to be documented and accessible to the public and particularly to scientists.

DEIMS-SDR is implemented as a web-based information portal for integrated ecological information that comprises detailed descriptions of sites where research is carried out, including the technical infrastructure, ecosystem properties and research activities and foci. Additionally, the generated datasets and information about the personnel involved can be included in the documentation. The available information is exposed through various metadata formats and standardised services. DEIMS-SDR also provides a common and openly available catalogue for the distinct identification of observation facilities (e.g. sites, stations, sensors), that makes it possible to reference those facilities in data objects deposited in any data repository. Within the International Long Term Ecological Research Network (ILTER) as well as its regional group LTER-Europe, DEIMS-SDR is used as the central site catalogue. This not only includes the general documentation of the research site and network but also workflows for the accreditation of research sites within the network. ILTER is a global network aiming to support long term ecosystem research by providing research sites and data. It is organised as a globally distributed network and infrastructure of long-term research sites organised in member networks (Mirtl et al., 2018).

DEIMS-SDR is an extended fork of the DEIMS system developed by the United States LTER network (US LTER) (Gries et al., 2010). DEIMS (in this paper commonly referred to as 'DEIMS Core') is an installation profile for the Content Management System (CMS) Drupal 7. It features both contributed modules and custom modules built by US LTER that are tailored to the description of ecological data. DEIMS Core is deployed as a GUI-based web application.

After initially using DEIMS Core, the requirements of both ILTER and LTER-Europe outgrew the capabilities of the original system and thus additional components were developed and deployed, allowing for more comprehensive documentation of relevant research objects. Subsequently, additional functionality for the documentation of research sites, networks and sensors was developed, and new software was added in order to provide more standardised interfaces to increase interoperability. Furthermore, the GUI (Fig. 1) has been enhanced based on the needs of the projects the system was used in and the subsequent user requirements, thus forming the platform called DEIMS-SDR.

Apart from acting as the central site catalogue for both LTER-Europe and ILTER since 2012, DEIMS-SDR is also a key component of the data tools used in the 'eLTER' project, an EU Horizon 2020 (H2020) project aimed to advance the European network of Long-Term Ecosystem Research sites and socio-ecological research platforms to provide high quality services for multiple use of a distributed research infrastructure (eLTER, 2018). It is used as both a site registry and metadata editor for measurements within EcoPotential, a large EU H2020 project that focuses its activities on a targeted set of internationally recognised protected areas, blending Earth Observations from remote sensing and field measurements, data analysis and modelling of current and future ecosystem conditions and services (EcoPotential, 2018). For this project an additional metadata model was developed and implemented to include campaign data, referred to as 'Data Products' (further described in 2.1.2 Data Product metadata model).

Furthermore, DEIMS-SDR is recommended as a service to generate site identifiers in the technical specifications document for the National Emission Ceilings (NEC) Directive, Article 10 (4a) 'data requirement on the location of the monitoring sites and the associated indicators' (NEC Directive Technical Specifications, 2018). The NEC directive is an EU directive setting national emission reduction commitments for Member States and the EU for a number of air pollutants.

Metadata published by DEIMS-SDR is provided by CC-BY-NC 4.0 International license to enable a free use of the information. The data has also already been used for the analysis of the (socio-)ecological representativeness of LTER sites in Europe (Mollenhauer et al., 2018).

In this paper, however, we focus on three specific objectives: 1) to develop a method for efficient generation of standardised information for sites and their associated data using DEIMS-SDR, (2) making the generated data available in a number of formats through different services and therefore ensuring interoperability (3) providing easy access to the data and persistently identifying it. Thus, DEIMS-SDR not only serves to increase the visibility of sites but also to allow comparative research on research sites, their research foci and their representativeness.

2. Methods

2.1. Data models and data architecture

DEIMS-SDR allows information about sites, observation locations, sensors, datasets, data products, persons and networks to be stored. Each information entity type has a respective data model, detailing all component fields, their definitions, dependencies, list of values and number of selectable values, which are detailed in the following sections. This also includes relations between the different information types, allowing them to be linked and subsequently allowing the generation of different metadata formats based on these relations and the derived compound information (Fig. 2).

Output in comparatively simple metadata formats, like SensorML, can be generated by using the available sensor information. By using the relations between the DEIMS-SDR information entities and thus combining the available information, outputs in more complex metadata formats can be generated dynamically, utilising international standards for geospatial metadata, e.g. ISO 19115-1:2014 Geographic information – Metadata – Part 1: Fundamentals, previously also known as ISO 19115:2003 Geographic information – Metadata (ISO 19115, 2014); for conceptual definition and ISO/TS 19139:2007 Geographic information – Metadata – XML (ISO 19139, 2007) schema implementation for XML schema implementation of ISO 19115. These standards provide information models and encoding rules covering data about the identification, extent, quality, spatial and temporal aspects, content, spatial reference, portrayal, distribution, and other properties of digital geographic data and services. They define the minimum set of metadata required to serve most applications (discovery, fitness for use, access, transfer and use), as well as a broad set of optional elements to allow more extensive and hierarchically interlinked descriptions of resources.

Another supported format is the INSPIRE application schema for Environmental Monitoring Facilities, in this paper commonly referred to as 'INSPIRE EF' (INSPIRE EF, 2018). INSPIRE ('Infrastructure for Spatial Information in the European Community') is an EU directive that sets a general framework for spatial data infrastructures in order to

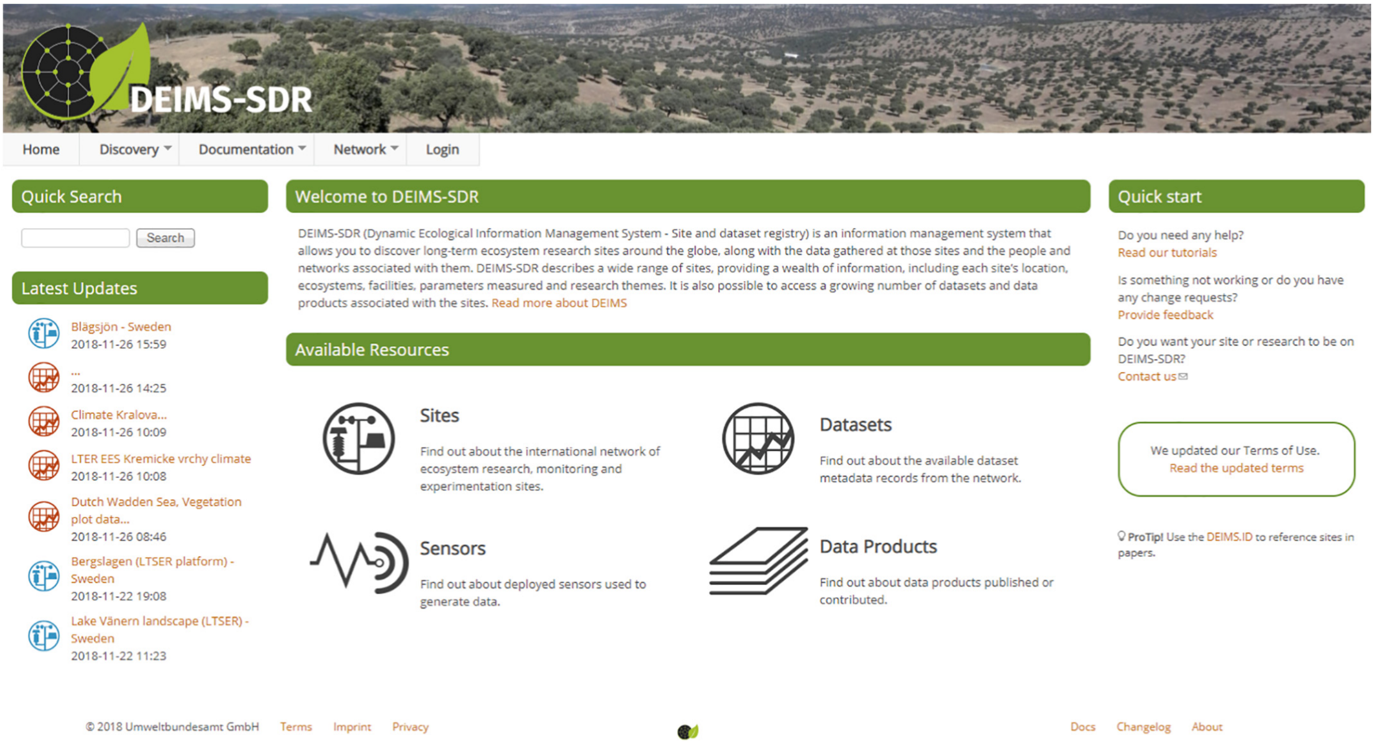

Fig. 1. The web portal and GUI of DEIMS-SDR.

support the European Community environmental policies and activities which may affect the environment. The INSPIRE framework includes dedicated data formats, such as INSPIRE EF, which have to be implemented by the EU member states until October 2020 (INSPIRE Directive, 2009).

By pulling together the specific information entities (sensor, person,

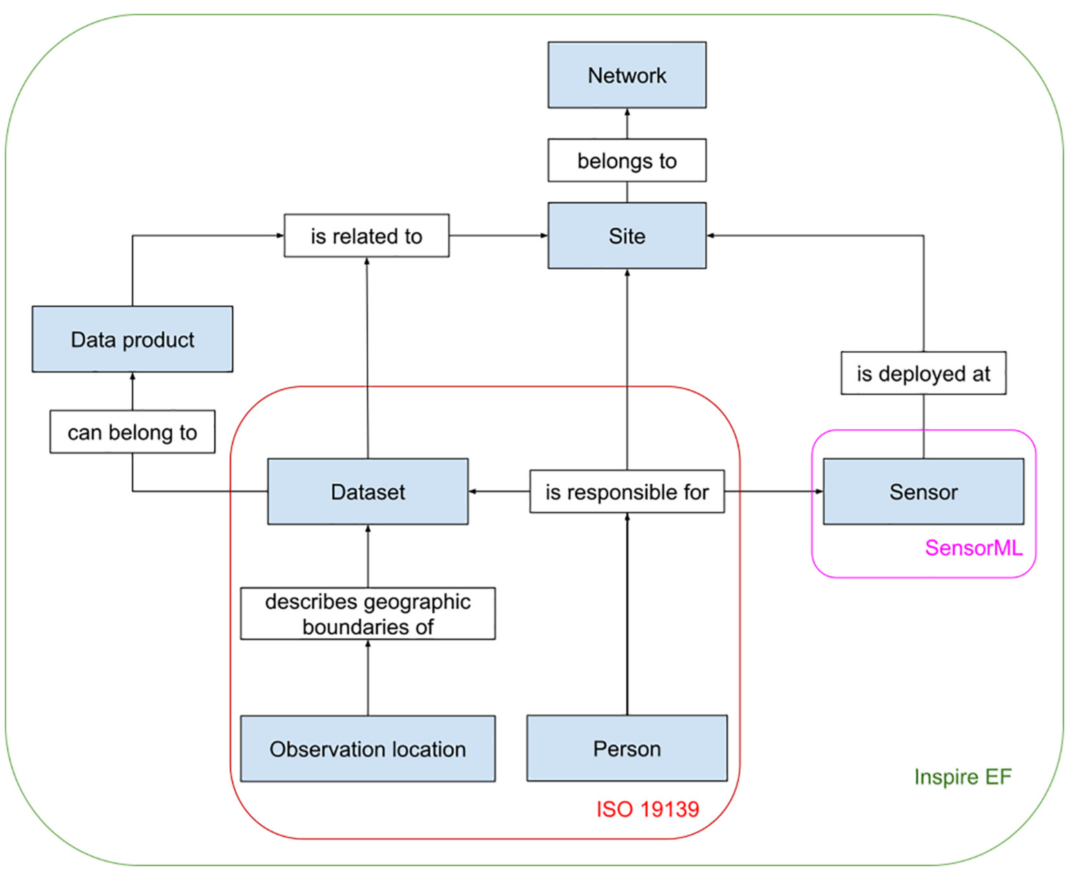

Fig. 2. Information entity relationships and metadata representation.

geographic feature, data product and site) INSPIRE EF records can be generated. INSPIRE EF data records contain information about research and/or monitoring facilities as well as the actual observations. In the case of DEIMS-SDR the inclusion of observation data is realised by linking to ISO 19139 records of available datasets. It is, therefore, possible to programmatically fetch all available information about datasets and subsequently all information connected to the sites by recursively parsing the records and their references.

In the following sections the metadata models of DEIMS-SDR are described in greater detail. All metadata models are stored as JavaScript Object Notation (JSON) files to also ensure machine-readability and maintainability and are implemented in a relational database management system (further described in 3.1. System architecture).

2.1.1. Site metadata model

The Site metadata model describes the metadata elements of a site. A 'Site' is defined as an in-situ observation or experimentation facility, delimited in space, but varying in size and complexity of the internal organisational and observational design, for the collection of data covering e.g. biogeophysical, biotic or socio-ecological characteristics. The metadata model is based on the requirements defined by target stakeholder groups (e.g. ILTER/LTER-Europe) and research projects (e.g. EnvEurope (EnvEurope, 2018), ExpeER (ExpeER, 2018), eLTER and EcoPotential) with the basic metadata elements being in line with the data specification of INSPIRE EF. It defines metadata elements about the organisation (e.g. contact, information and networks), the location, the observation characteristics (e.g. climate, habitats) or available equipment. Additionally, there are fields about the focus and design of a site, network affiliation and information about data policies and data management. The Site metadata model is the most comprehensive and complex of DEIMS-SDR data models. This extent allows the capture of a wide range of relevant information about a site that exceeds the scope of the ILTER community or the EcoPotential project, thus making it possible to incorporate sites outside of those communities and projects. Fig. 3 presents a simplified version of the Site metadata model, listing all of its content fields and indicating whether or not they are mandatory or recommended for a site record.

Apart from geographic and environmental characteristics, the model supports indicating which networks and projects a site belongs to, contact points for a site (e.g. institution or a person) as well as site hierarchies, e.g. umbrella sites consisting of subsites that may in turn consist of plots. It is also possible to reference external data collections and data portals of sites.

In addition to the mandatory fields (site name, site manager, country and network/project affiliation), a set of recommended fields based on the requirements of the ILTER and LTER-Europe communities was defined that forms the basis for calculating the completeness of a site record as a percentage, based on the amount of information provided for the recommended fields. That set of fields can be applied to sites outside of ILTER and LTER-Europe as well, which allows easier assessment of the quality of any given site record. Currently the following types of information are recommended to be provided:

- General characteristics ('General Site Description', 'Coordinates', 'Site Type', and 'Size')
- Climatic characteristics ('Mean Annual Air Temperature' and 'Sum Annual Precipitation')
- Topographic characteristics ('Elevation Range (minimum – maximum)')
- Ecosystem characteristics ('Biome' and 'Ecosystem and Land Use')
- Scientific characteristics ('Purpose of Site', 'Research Topics', 'Design of Observation', 'Scale of Observation', 'Design of Experiments', 'Scale of Experiments' and 'Observed parameters')
- Operation characteristics ('Year Site was established', 'Site Status (active, inactive, closed)', 'Permanent Operation', 'Accessible All Year', and 'Permanent Power Supply')
- Data management ('Data Request Format' and 'Data Storage Location')
- Metadata information ('Metadata provider')

Whenever feasible, existing vocabularies were used for the provision of lists of values for each field, implementing semantically meaningful terms identified by HTTP URIs. This was the case for parameters and research topics where the Environmental Thesaurus was used (EnvThes, 2018), which is managed by LTER-Europe. However, controlled vocabularies haven't been implemented for the field definitions themselves. The provided list of values for the field 'Ecosystem and land use' is an assignment of the main habitat type based on a list of main habitat types (or ecoregions) according to Olson et al. (2001). Other lists of values for fields, e.g. like the instrumentation and infrastructure components, have been defined based on the needs of the ILTER and LTER-Europe communities and currently lack the means to resolve to a controlled vocabulary, which will be addressed in future releases.

2.1.2. Data product metadata model

Within the EcoPotential project, a metadata specification for data products has been developed and implemented allowing for easy and lightweight documentation of the in-situ data collection campaigns provided by observation facilities and protected areas (Poursanidis et al., 2017). The Data Product metadata model describes the metadata elements of a 'Data Product', which is defined as the summary of all datasets collected and compiled during and after a measuring or observation campaign. The concept of a 'Data Product' was introduced in order to allow for a summarised description of a series of datasets. Using this, a brief overview of available data provided by a protected or research site can be created without a full description of every single dataset.

This model is a simplified version of the dataset model with additional fields that facilitate indicating the (digital and legal) availability of the generated data. Data Product information is exported as ISO 19139 and is also part of INSPIRE EF records.

2.1.3. Dataset metadata model

The core element for scientific workflows is the dataset or data stream. A 'Dataset' is defined as a collection of data values provided as a single data stream, file or service and documented by accompanying metadata. The Dataset metadata model is based on requirements defined by target stakeholder groups identifying the minimum set of metadata elements required to publish a valid metadata document under the INSPIRE directive. This is summarised in the LTER community profile (Poursanidis et al., 2017) building the basis for the technical implementation. The metadata elements were selected based on the Ecological Metadata Language (EML) (Michener et al., 1997) version 2.1.1 but also include elements required to ensure the discovery of ISO 19115 compliant metadata (Kliment and Oggioni, 2011) following the requirements of the INSPIRE metadata regulation. The aforementioned 'Observation location' is a part of the dataset description. It reflects the description of a location where research is conducted and could also be referred to as a 'station' or 'installation'. It includes a name, a textual description, and coordinates or the boundaries. The purpose of the observation location is to reduce the effort required to create multiple datasets that feature the same spatial extent.

Pulling together the dataset and observation location information allows for the generation of dataset metadata records in EML, Biological Data Profile (BDP), a profile of the FGDC/CSDGM metadata standard, intended to support the collection and processing of biological data, as well as ISO 19115 and ISO 19139. The minimum information necessary for a dataset is based on the minimum requirements for ISO 19115, 19139 and EML. A custom Drupal module was developed to dynamically generate XML encoded metadata documents from the DEIMS-SDR Dataset entity type following the EML 2.1.1 and ISO19139 standards.

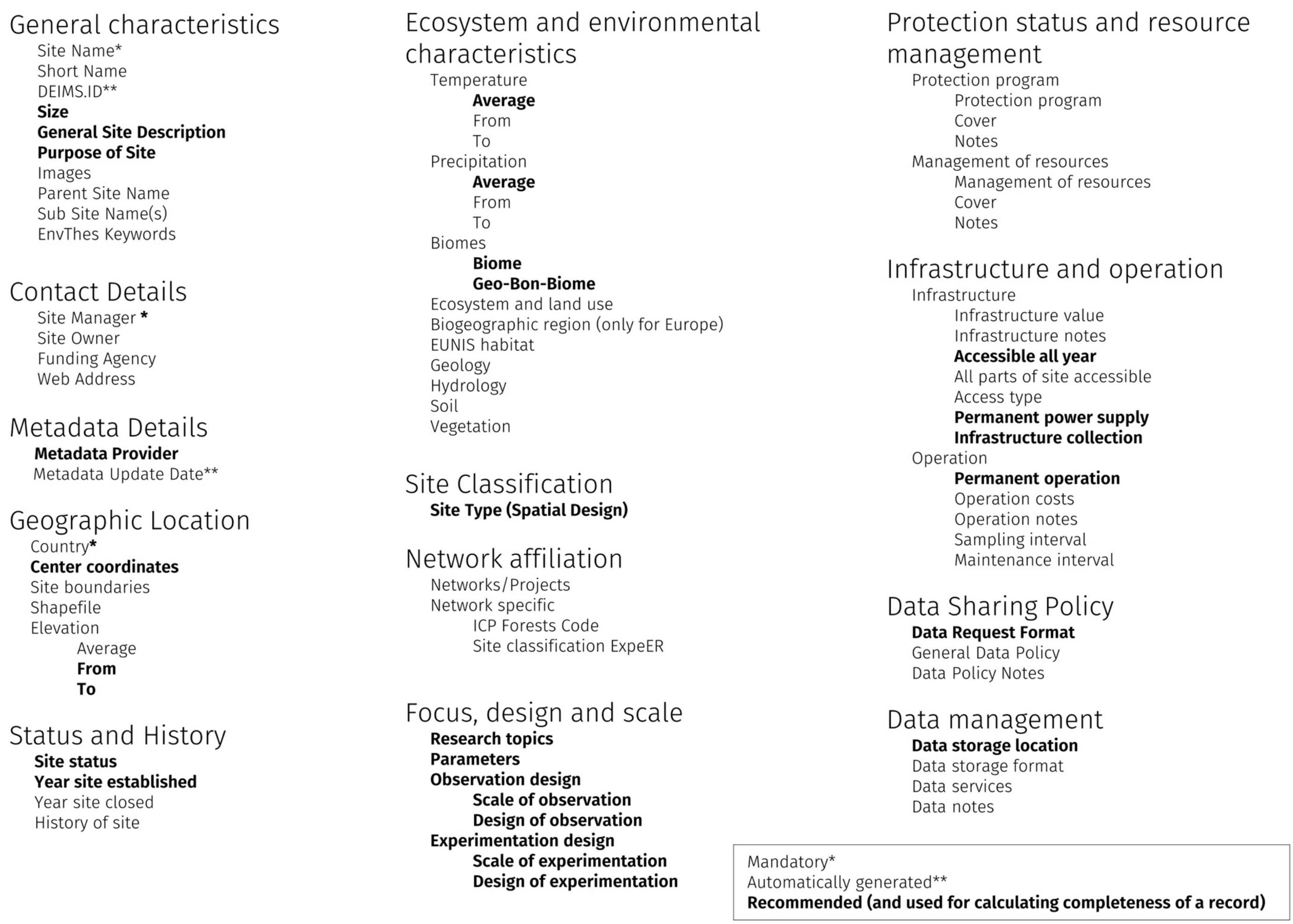

Fig. 3. Simplified Site metadata model content.

2.1.4. *Other metadata models*

The Sensor metadata model describes the metadata elements of a device, module, or subsystem whose purpose is to detect events or changes in its environment or its observation time series. The model enables the generation of both SensorML 2.0 and subset information for INSPIRE EF. SensorML is an Open Geospatial Consortium (OGC, 2018) standard providing standard models and an XML encoding for describing sensors and measurement processes (SensorML, 2018).

The Network data model is designed to collect basic information on a network itself as well as the associated information provided within DEIMS-SDR and is based on the requirements defined by the ILTER and LTER-Europe networks. This includes the links to associated people, sites and datasets, making it possible to calculate metrics for each network (e.g. size and research foci of a site network). For networks managed within DEIMS-SDR, the affiliation of research sites can be managed and the status of affiliation can be changed. The Network metadata model therefore especially applies to ILTER/LTER member networks with the only mandatory information being the network name. There is no dedicated metadata export for network information, however, if linked to a site, parts of it are used to generate INSPIRE EF records.

The Person metadata model describes the metadata elements of any individual (e.g. researcher or administrative person) related to a digital research object (e.g. site, station, sensor, or dataset) described in DEIMS-SDR. It provides a source for metadata provider, contact points and site managers (basic contact information). The only required information is the full name; all other information is optional. Analogous to network information there is no dedicated metadata export for person data. It is, however, used for every metadata export whenever a person is referenced. In addition to these models, it is also possible to add projects and organisations consisting only of a name and a URL.

2.2. *Persistent identification of sites*

A common requirement for the registration of in-situ observation or experimentation sites is to create persistent identifiers for each site, while also providing information about the setup of observation devices and general technical capabilities. Many of the research sites are used by different networks or research infrastructures since basic infrastructure components (e.g. permanent energy supply or measurement towers) are needed in order to implement observation campaigns. Colocation takes place e.g. at the ‘Hyytiälä SMEAR II LTER’ station in central Finland (DEIMS-SDR Hyytiälä, 2018), which is listed as both an eLTER RI component and an Integrated Carbon Observation System (ICOS) station (‘SMEAR II-ICOS Hyytiälä’) (ICOS Site Catalogue, 2018). A common registry and persistent identifiers are needed in order to unambiguously identify such observation facilities and allow for network specific documentation of the relevant components.

In earlier versions of DEIMS-SDR, unique identification was achieved through the LTER site codes, an alphanumeric code consisting of the acronym of the national or member network and an integer number, issued by the respective LTER member network, and the DEIMS-SDR base URL. However, after opening DEIMS-SDR to sites and networks outside of the LTER community this approach was no longer viable, therefore a new system independent from network affiliation was implemented.

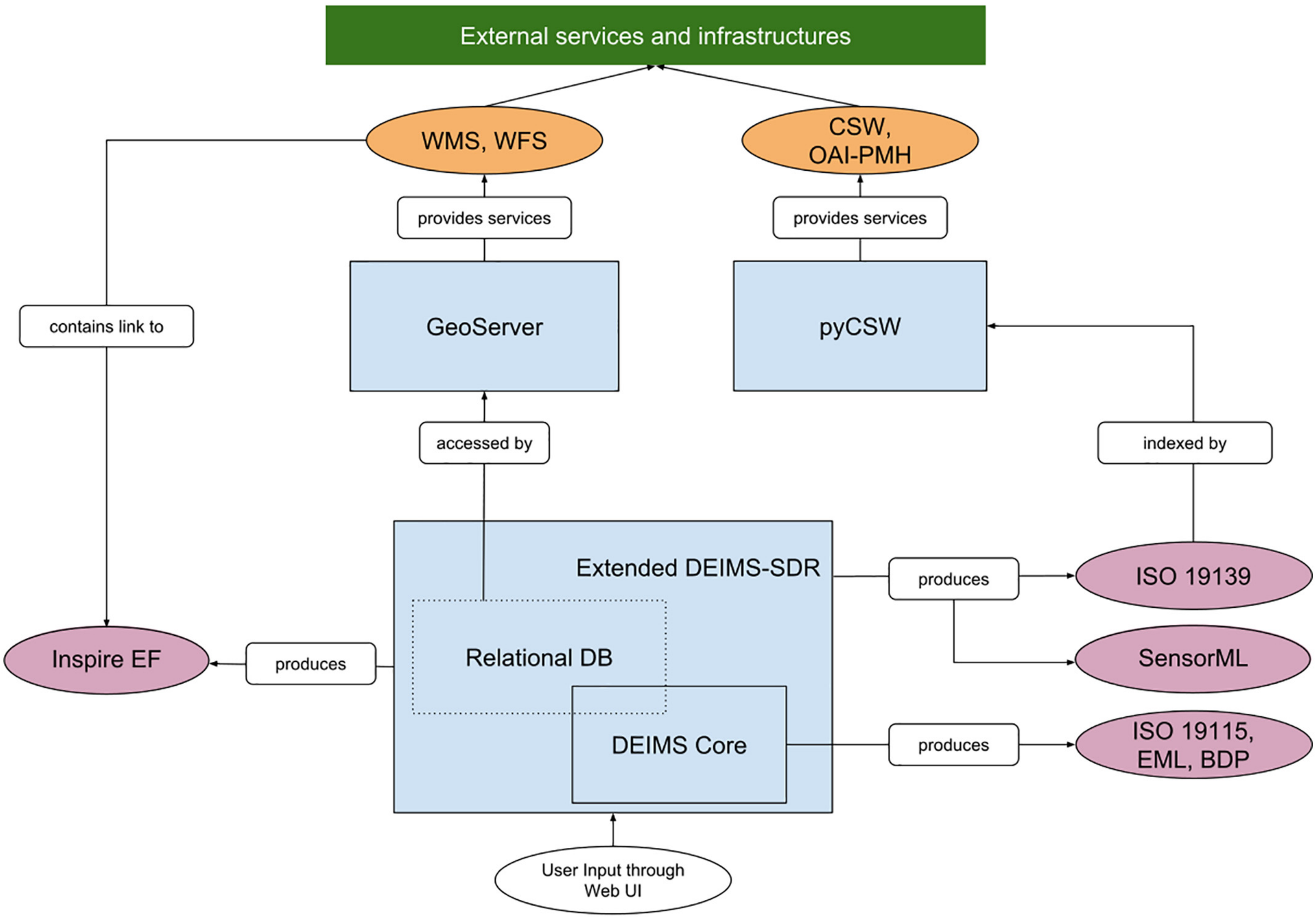

Fig. 4. System components and dependencies of DEIMS-SDR web platform.

For this purpose, DEIMS-SDR now generates a universally unique identifier (UUID) v4 (RFC, 2005) using PHP code. For each site, a UUID is generated only once upon the creation of the record. The generated UUIDs are appended to the base URL ('deims.org') and thus form a unique and resolvable identifier, e.g. https://deims.org/663dac80-211d-4c19-a356-04ee0da0f0eb for the 'Hyytiälä SMEAR II LTER' research site in Finland. The neutral nature of this identifier, called 'DEIMS.ID', allows cross-RI identification of research sites, thus creating more sustainable and usable site metadata records. Due to the longevity of DEIMS-SDR, guaranteed due to its role in the LTER community, the persistence of the issued identifiers is also ensured. The DEIMS.ID, being persistent, unique and resolvable, therefore fulfils the 4th principle of the Joint Declaration Of Data Citation Principles and can be used for data citation (Data Citation Synthesis Group, 2014).

3. Implementation

3.1. System architecture

DEIMS-SDR currently consists of four main components fulfilling different needs and offering linked services (Fig. 4), which are described in greater detail in 3.2.1. Services and applications.

1. DEIMS Core: Provides a basic user interface for storing information about datasets and people and for generating respective metadata records in EML, BDP and ISO 19115. It is based on an installation profile for the Open-Source-CMS Drupal 7 (Gries et al., 2010) implementing a relational database management system embedded in the Drupal PHP framework itself.
2. DEIMS-SDR: An extension of DEIMS Core featuring additional custom-built modules that allow the storing of information about research sites, data products and sensors and the generation of ISO19139, SensorML and INSPIRE EF metadata records. Additional search and display functionalities and an interface to directly upload datasets to the open eScience data sharing platform B2SHARE (Ardestani et al., 2015) have been developed for Drupal 7. DEIMS-SDR uses the same relational database management system as DEIMS Core, currently a MySQL database. However, it is likely that the database management system will be replaced with a different system, such as PostgreSQL, during the next major update. However, at the time of writing no decision has been made.
3. WebGIS Service: An instance of GeoServer, an OGC compliant implementation of a number of open standards such as Web Feature Service (WFS), Web Map Service (WMS), and Web Coverage Service (WCS), is connected to the aforementioned relational database of DEIMS-SDR and exposes geographic information of sites as view and download services with additional rudimentary information about research sites, e.g. name, DEIMS.ID and the URL to the INSPIRE EF record.
4. Catalogue Service: DEIMS-SDR periodically loads all available metadata records into a pyCSW instance. pyCSW is an OGC Catalogue Service for the Web (CSW) server implementation written in Python that exposes records through standardised services, making them harvestable by other systems.

Additionally, there are a number of Python and shell scripts that ease the process of moving information from one subsystem to another.

3.1.1. Services and applications

Due to the usage of pyCSW and GeoServer, a variety of OGC services are supported that allow data from DEIMS-SDR to be queried. Currently, the following services are supported:

- Web Feature Service (WFS) – an interface allowing requests for geographical features across the web using platform-independent calls
- Web Map Service (WMS) – a standard protocol developed by the Open Geospatial Consortium for serving georeferenced map images over the Internet
- Catalogue Service for the Web (CSW) – a standard for exposing a catalogue of geospatial records in XML on the Internet
- Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) – a protocol developed for harvesting (or collecting) metadata descriptions of records in an archive.

The relation and interfaces of all software components are illustrated in Fig. 4. The support of WMS and WFS allows for on the fly usage of site information in desktop GIS applications, such as QGIS, or online applications using JavaScript libraries, like OpenLayers or Leaflet. CSW and OAI-PMH ensure interoperability with other metadata catalogues, like other instances of pyCSW or GeoNetwork. This allows data on DEIMS-SDR to be harvested by external systems, e.g. the GEOSS Portal (re3data.org: GEOSS Portal; editing status, 2018) or DataONE (re3data.org: DataONE, 2018). Additionally, simple CSV exports can be generated covering basic information about research sites and datasets.

4. Results

A summary of all available records on DEIMS-SDR as at August 2018 (Table 1) shows that the largest proportion of records is site records, closely followed by person and dataset records, which are usually linked to site records as well. The remaining record types, data products and sensors represent only a small fraction of all available records. This means that on average there is less than one dataset and person associated per site. It can therefore be derived that the main usage of DEIMS-SDR revolves around the site registry.

By using the WFS capabilities of DEIMS-SDR, the locations of all site records that provide point coordinates were compiled and used for further analysis. The resulting map of site locations is presented in Fig. 5.

These site locations were then intersected with the ESRI World Regions dataset (ESRI, 2018) in order to determine the overall coverage of world regions. Out of the 1019 records with proper point coordinates, only 919 could be successfully intersected with the world regions dataset due to the fact that some are located on islands not covered by the dataset or are marine and/or coastal sites, which are also often not covered by the ESRI World Regions dataset. The resulting figure (Fig. 6) illustrates the coverage of each world region.

At the low end are the scarcely represented regions such as Asiatic and European Russia, Middle, Northern and Western Africa (with only one site each) or regions not covered at all (Central Asia, Micronesia and Polynesia). The highest geographic coverage of regions is in Western Europe (with 210 sites) followed by Northern Europe (157), Southern Europe (151), Eastern Europe (116) and Eastern Asia (104). This illustrates the strong use of DEIMS-SDR by European networks and research infrastructures. Based on all published sites (as at August 2018), the mean completeness value (described in 2.1.1 Site Metadata Model) for a site record is 89%. For sites located in Europe the mean completeness value is 93%. Where the network affiliation was provided in site records, it was also possible to calculate mean completeness values for networks. For accredited LTER-Europe sites, the mean completeness value is 98% and for some national LTER networks, such as LTER Switzerland, LTER Greece, LTER Slovenia and LTER Slovakia, all the network's sites are 100% complete. In general networks and sites located outside of Europe are predominantly below average in terms of site record completeness, further illustrating the focus of DEIMS-SDR on European sites, networks, research infrastructures and projects.

Table 1
Published DEIMS-SDR records as at August 2018 (n = 2967).

Record type	Number of published records (relative share)
Site	1042 (35.12%)
Person	907 (30.56%)
Dataset	893 (30.10%)
Data Product	104 (3.51%)
Sensor	21 (0.71%)

5. Discussion

The current system architecture of DEIMS-SDR has been developed within a number of European scale projects and networks and aims to meet major current stakeholder requirements. It is currently an ideal compromise between feature richness and maintainability with the available resources. Using an Open-Source CMS has proven useful insofar as the maintenance requirements of parts of the system, especially the user interface, are minimal. Security issues, which are often a reason for concern for Open-Source CMS, have not been an issue so far due to the semi-open user account generation and regular update cycles for both Drupal and the server itself, e.g. the operating system and installed packages. Security penetration scans that check for known vulnerabilities, and a regular backup schedule, reduce potential data loss to approximately one day, even in the case of a security breach. This further decreases potential security issues of an Open-Source CMS.

The modular setup of DEIMS-SDR allows for easier replacement of single components and maintainability, which is likely to be pursued and expanded further in the future. Downsides of the current setup include challenging data migration to new major releases of Drupal and the overall complexity of the entire software stack, that currently makes it impractical to publish the entire codebase on a dedicated code platform. However, certain modules and scripts are stored in dedicated, openly available code repositories. Due to the dispersed code, substantial community input and developer participation is not to be expected, except for the vocabularies, such as EnvThes, that DEIMS-SDR uses. The possibility of replacing Drupal with a different CMS or a custom-built solution is currently being debated for the next major release. The goal would be to allow easier data migration and handling, and overall simplification of the system and architecture, thus allowing for more centralised code management in a dedicated repository.

Other tasks for future releases include the evaluation and revision of the metadata models. This is especially true for the site metadata model in order to better capture information about the accessibility of datasets and to facilitate linking to existing repositories or catalogues. The increased usage of controlled vocabularies, to follow World Wide Web Consortium (W3C) semantic interchange recommendations and thus ensure better interoperability, is also planned. Other future work might involve exposing schema.org compliant metadata as JSON-LD for better harvesting by major search engines.

While DEIMS-SDR also allows registering of datasets, it is not its core functionality, Thus, there are no plans for it to become a European or global registry of datasets; other platforms are better suited for this purpose. As a consequence no major changes or updates for the dataset functionality of DEIMS-SDR are currently planned.

DEIMS-SDR is positioned as central site registry for any long term observation and experimentation facilities. Nevertheless, currently several other site catalogues implemented by the different research infrastructures exist, e.g. ICOS Site Catalogue (ICOS Site Catalogue, 2018), TERENO Site Catalogue (TERENO, 2018), INTERACT Site

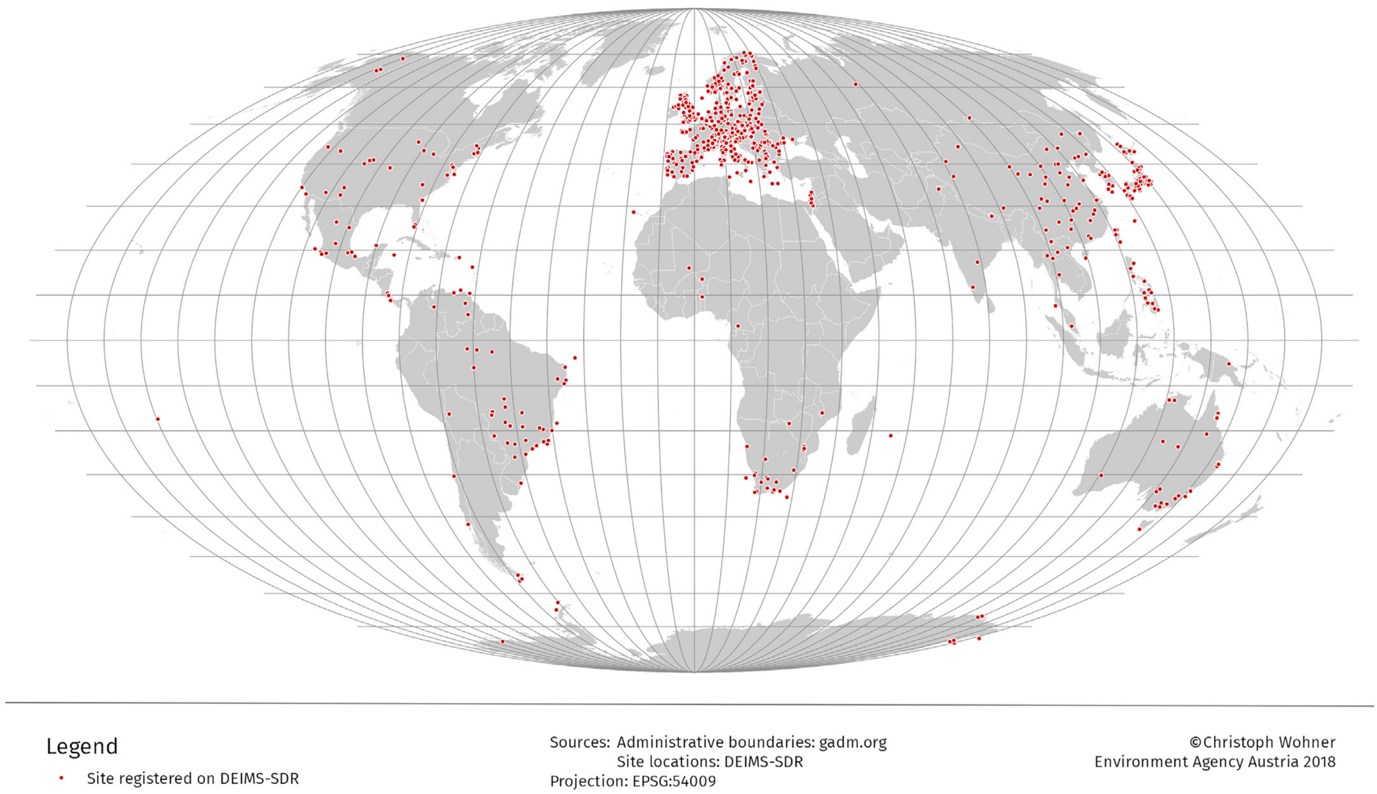

Fig. 5. Global distribution of sites registered on DEIMS-SDR as at August 2018 (n = 1019).

Catalogue (INTERACT, 2015), UKEOF Catalogue (UKEOF, 2018) and research projects trying to cover all requirements for their respective networks or research infrastructures. They often share the same research facilities (e.g. in the terrestrial domain) or research facilities are co-located within the same area. This can lead to the repeated or overlapping documentation of the same observation facility (e.g. RI site) in different site catalogues. Due to the more general nature of DEIMS-SDR, it features not only more sites than any of these catalogues, but also enables issuing of a persistent identifier, as well as indicating of project and network affiliation, and therefore contains information about overlaps and co-location.

However, even though DEIMS-SDR features a wider collection of

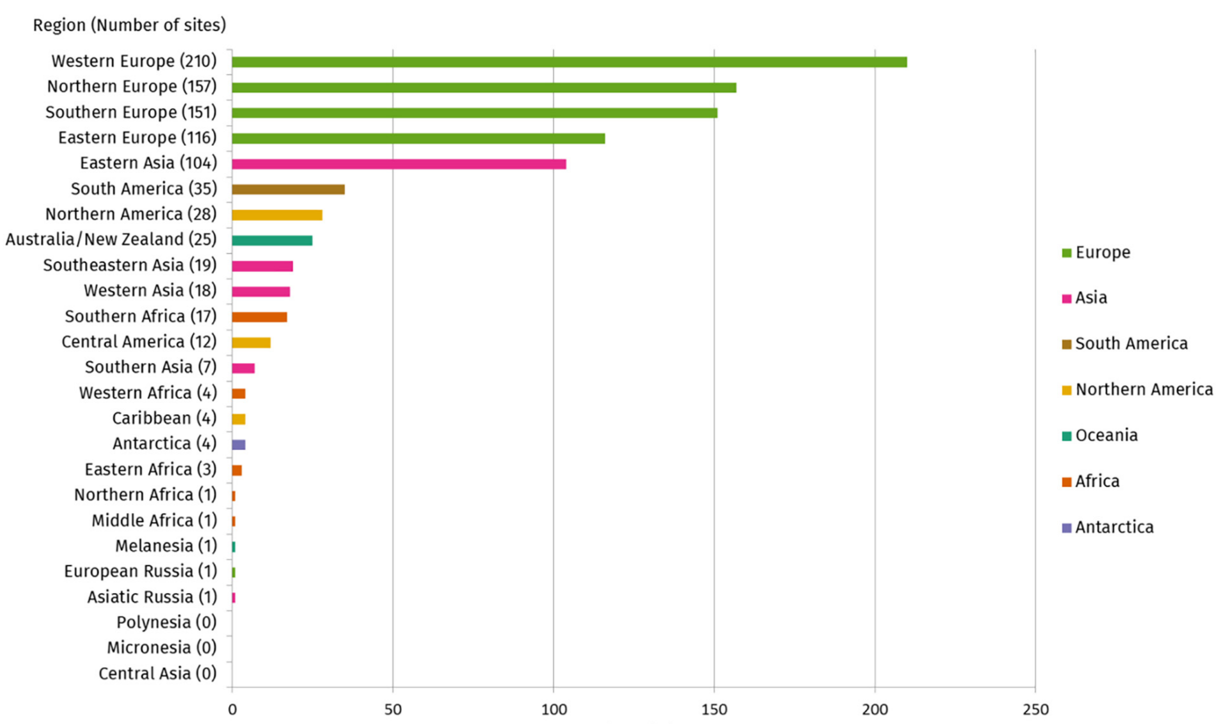

Fig. 6. Number of sites per region (n = 919).

site records than other catalogues, there are still regions only scarcely covered by records on DEIMS-SDR, such as European and Asiatic Russia, Northern, Western and Eastern Africa or regions with no coverage altogether, such as Micronesia, Polynesia and Central Asia (Figs. 5 and 6). In other regions such as the Americas, Eastern Asia, Australia and especially Europe, DEIMS-SDR has a satisfying geographic coverage. As a result of the usage of DEIMS-SDR in a number of EU projects, a high proportion of site records describe European sites. European representation is further boosted by the fact that DEIMS-SDR is recommended as a service to generate site identifiers for the National Emission Ceilings (NEC) directive. This increases the motivation of European users to keep their records up-to-date and explains the above average completeness of European site records. However, in order to achieve the aim of DEIMS-SDR to be a globally comprehensive site catalogue, more networks and projects, especially outside of Europe and the LTER networks, should be incorporated. In 2016, ILTER was invited to become Participating Organisation of the Group on Earth Observation (GEO). One of the main reasons was the existence of DEIMS-SDR and ILTER's experience in rolling out a global in-situ site registration process across 44 countries. ILTER's engagement in the GEO In-situ Foundational Task resulted in the adoption of DEIMS-SDR as a GEO pilot for a global registry of in-situ observation sites (2017–2019 GEO Work Programme, 2016). This substantially increases the potential of becoming a globally comprehensive site catalogue used across networks, projects and research infrastructures. In successfully building DEIMS-SDR and collecting the presented site information, we have achieved the set goals of collecting standardised information and making it available in a standardised manner. As a consequence, the existing information can be used for comparative analyses of registered sites and their networks, which is likely to be explored in follow-up papers.

Acknowledgements

DEIMS-SDR is a product of LTER-Europe and ILTER. This study was done in the context of the EcoPotential Project (grant agreement no 641762) and the eLTER project (grant agreement no 654359). Both projects have received funding from the European Union's Horizon 2020 research and innovation programme.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2019.01.005.

References

2017–2019 GEO Work Programme, 2016. https://www.earthobservations.org/documents/geo_xiii/GEO-XIII-2-Inf-01(Rev1)_2016_Work_Programme_Progress-Report.pdf, Accessed date: 21 November 2018.
Ardestani, S.B., Hakansson, C.J., Laure, E., Livenson, I., Stranak, P., Dima, E., van de Sanden, M., 2015. B2SHARE: an open eScience data sharing platform. In: In 2015 IEEE 11th International Conference on e-Science. IEEE. https://doi.org/10.1109/escience.2015.44.
Butchart, S.H.M., Walpole, M., Collen, B., van Strien, A., Scharlemann, J.P.W., Almond, R.E.A., ... Watson, R., 2010. Global biodiversity: indicators of recent declines. Science 328 (5982), 1164–1168. https://doi.org/10.1126/science.1187512.
Carpenter, S.R., DeFries, R., Dietz, T., Mooney, H.A., Polasky, S., Reid, W.V., Scholes, R.J., 2006. ECOLOGY: enhanced: millennium ecosystem assessment: research needs. Science 314 (5797), 257–258. https://doi.org/10.1126/science.1131946.
Ceballos, G., Ehrlich, P.R., Barnosky, A.D., Garcia, A., Pringle, R.M., Palmer, T.M., 2015. Accelerated modern human-induced species losses: entering the sixth mass extinction. Sci. Adv. 1 (5), e1400253. https://doi.org/10.1126/sciadv.1400253.
Ciais, P., Reichstein, M., Viovy, N., Granier, A., Ogée, J., Allard, V., Valentini, R., 2005. Europe-wide reduction in primary productivity caused by the heat and drought in 2003. Nature 437 (7058), 529–533. https://doi.org/10.1038/nature03972.
Data Citation Synthesis Group, 2014. In: Martone, M. (Ed.), Joint Declaration of Data Citation Principles, https://doi.org/10.25490/a97f-egyk. Force11.
DEIMS-SDR Hyytiälä, 2018. https://deims.org/663dac80-211d-4c19-a356-04ee0da0f0eb accessed 21-November-2018.
EcoPotential, 2018. http://www.ecopotential-project.eu/, Accessed date: 21 November 2018.
eLTER, 2018. http://www.lter-europe.net/elter, Accessed date: 21 November 2018.
EnvEurope, 2018. http://www.enveurope.eu/, Accessed date: 21 November 2018.
EnvThes, 2018. http://vocabs.ceh.ac.uk/evn/tbl/envthes.evn, Accessed date: 21 November 2018.
ESRI, 2018. https://www.arcgis.com/home/item.html?id=84dbc97915244e35808e87a881133d09, Accessed date: 21 November 2018.
ExpeER, 2018. http://www.expeeronline.eu/, Accessed date: 21 November 2018.
Gries, C., San Gil, I., Vanderbilt, C., Garrit, H., 2010. Drupal Developments in the LTER Network. link. https://lternet.edu/wp-content/uploads/2017/12/2010-spring-lter-databits.pdf.
ICOS Site Catalogue, 2018. https://www.icos-cp.eu/node/83, Accessed date: 21 November 2018.
INSPIRE Directive, 2009. https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ%3AL%3A2009%3A148%3A0018%3A0026%3AEN%3APDF, Accessed date: 21 November 2018.
INSPIRE EF, 2018. Infrastructure for Spatial Information in the European Community – Environmental Monitoring Facilities. https://inspire.ec.europa.eu/theme/ef, Accessed date: 21 November 2018.
INTERACT, 2015. INTERACT Station Catalogue – 2015. In: Elger, K., Opel, T., Topp-Jorgensen, E., Hansen, J., Tairova, Z., Rasch, M. (Eds.), DCE Danish Centre for Environment and Energy. Aarhus University, Denmark (305 pp).
Isbell, F., Calcagno, V., Hector, A., Connolly, J., Harpole, W.S., Reich, P.B., Loreau, M., 2011. High plant diversity is needed to maintain ecosystem services. Nature 477 (7363), 199–202. https://doi.org/10.1038/nature10282.
ISO, 2007. ISO/TS 19139:2007 Geographic Information – Metadata – XML Schema Implementation. International Organization for Standardization, Geneva, Switzerland.
ISO, 2014. ISO 19115-1:2014 Geographic Information - Metadata – Part 1: Fundamentals. International Organization for Standardization, Geneva, Switzerland.
Kliment, T. & Oggioni, A. (2011). Metadatabase: EnvEurope Metadata specification for Dataset Level. EnvEurope (LIFE08 ENV/IT/000339) Project Report PD.A1.1.4 87pp. (link: http://www.enveurope.eu/misc/PD_1_1_4_Kliment_Metadatabase_201112_final_v1.0.pdf).
Michener, W.K., Brunt, J.W., Helly, J.J., Kirchner, T.B., Stafford, S.G., 1997. Nongeospatial metadata for the ecological sciences. Ecol. Appl. 7 (1), 330–342. https://doi.org/10.1890/1051-0761(1997)007[0330:nmftes]2.0.co;2.
Mirtl, M., Borer, E.T., Djukic, I., Forsius, M., Haubold, H., Hugo, W., ... 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, 1439–1462. https://doi.org/10.1016/j.scitotenv.2017.12.001.
Mollenhauer, H., Kasner, M., Haase, P., Peterseil, J., Wohner, C., Frenzel, M., Zacharias, S., 2018. Long-term environmental monitoring infrastructures in Europe: observations, measurements, scales, and socio-ecological representativeness. Sci. Total Environ. 624, 968–978.
Nagendra, H., Lucas, R., Honrado, J.P., Jongman, R.H.G., Tarantino, C., Adamo, M., Mairota, P., 2013. Remote sensing for conservation monitoring: assessing protected areas, habitat extent, habitat condition, species diversity, and threats. Ecol. Indic. 33, 45–59. https://doi.org/10.1016/j.ecolind.2012.09.014.
NEC Directive Technical Specifications, 2018. http://ec.europa.eu/environment/air/pdf/Technical%20Specifications%20NEC%20Article%209%20location%20and%20indicators%20final.docx, Accessed date: 21 November 2018.
Newbold, T., Hudson, L.N., Hill, S.L.L., Contu, S., Lysenko, I., Senior, R.A., Purvis, A., 2015. Global effects of land use on local terrestrial biodiversity. Nature 520 (7545), 45–50. https://doi.org/10.1038/nature14324.
OGC, 2018. Open Geospatial Consortium. http://www.opengeospatial.org/, Accessed date: 21 November 2018.
Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V., Underwood, E.C., ... Loucks, C.J., 2001. Terrestrial ecoregions of the world: a new map of life on earth a new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. BioScience 51 (11), 933–938. https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2.
Pan, Y., Birdsey, R.A., Fang, J., Houghton, R., Kauppi, P.E., Kurz, W.A., Hayes, D., 2011. A large and persistent carbon sink in the World's forests. Science 333 (6045), 988–993. https://doi.org/10.1126/science.1201609.
Pereira, H.M., Leadley, P.W., Proenca, V., Alkemade, R., Scharlemann, J.P.W., Fernandez-Manjarres, J.F., Walpole, M., 2010. Scenarios for global biodiversity in the 21st century. Science 330 (6010), 1496–1501. https://doi.org/10.1126/science.1196624.
Poursanidis, D., Peterseil, J., Wohner, C., Chrysoulakis, N., Wetzel, F., Alonso, J., Castro, P., Beierkuhnlein, C., Bernd, A., Zabala, A., Masó, J., Domingo, C., Vetaas, O., Bargmann, T., Bosch, S. (2017). D5.2 Metadata for Pre-existing Datasets. ECOPOTENTIAL Project (H2020 GANr. 641762) Deliverable. http://www.ecopotential-project.eu/images/ecopotential/documents/D5.2.pdf 116 pp. (accessed 21-November-2018).
re3data.org: DataONE; editing status (2018). 2017-11-13; re3data.org – Registry of Research Data Repositories. 10.17616/R3101G (accessed 21-November-2018).
re3data.org: GEOSS Portal; editing status (2018) 2017-07-24; re3data.org – Registry of Research Data Repositories. 10.17616/R34938 (accessed 21-November-2018).
RFC, 2005. Request for Comments 4122 section 4.4. http://www.rfc-editor.org/rfc/rfc4122.txt, Accessed date: 21 November 2018.
Seidl, R., Schelhaas, M.-J., Rammer, W., Verkerk, P.J., 2014. Erratum: Corrigendum: Increasing forest disturbances in Europe and their impact on carbon storage. Nature Climate Change 4 (10). https://doi.org/10.1038/nclimate2393. 930-930.
SensorML, 2018. https://www.opengeospatial.org/standards/sensorml, Accessed date: 21 November 2018.
TERENO, 2018. http://teodoor.icg.kfa-juelich.de/observatories/ELRV_Observatory?set_

language = en, Accessed date: 21 November 2018.
UKEOF, 2018. https://catalogue.ukeof.org.uk/, Accessed date: 21 November 2018.
Verheyen, K., De Frenne, P., Baeten, L., Waller, D.M., Hédl, R., Perring, M.P., Bernhardt-Römermann, M., 2016. Combining biodiversity resurveys across regions to advance global change research. BioScience 67 (1), 73–83. https://doi.org/10.1093/biosci/biw150.
Wilkinson, M.D., Dumontier, M., Aalbersberg, Ij.J., Appleton, G., Axton, M., Baak, A., Mons, B., 2016. The FAIR guiding principles for scientific data management and stewardship. Sci. Data 3, 160018. https://doi.org/10.1038/sdata.2016.18.