Using ISO Metadata and OGC Catalog Services to find data on Protected Sites

OER4SDI

Open Educational Resources for Spatial Data Infrastructures

This tutorial will help you understand how metadata is provided and used in SDIs. In particular, we will look at metadata based on ISO standards, as well as interfaces for querying and accessing metadata according to the OGC specification of Catalog Services for the Web (OGC CSW).

Tobias Krumrein, Albert Remke
Institute for Geoinformatics, University of Münster

September 2024

ein Kooperationsprojekt, empfohlen durch:

DIGITALE HOCHSCHULE NRW

gefördert durch:

Ministry of Culture and Science of the State of North Rhine-Westphalia

1. Overview

This tutorial will help you understand how metadata is provided and used in SDIs. In particular, we will look at metadata based on ISO standards, as well as interfaces for querying and accessing metadata according to the OGC specification of Catalog Services for the Web (OGC CSW).

After completing this tutorial, you will be able to query and use OGC Catalog services via Python and QGIS.

The tutorial is structured as follows:

The tutorial takes about 90 Minutes for reading and viewing the provided materials, downloading the software and for conducting the hands-on exercises and tasks.

This OER is primarily designed to be used by students in Geoinformatics, Geomatics and similar study programs. It is also useful for students of other study programs and for practitioners who want to enhance their understanding of SDI concepts and technologies. Some basic knowledge of web technologies such as HTTP and Web Services is required. However, you will be able to follow and find links to further resources if needed. Your computer should have 8 GB of usable RAM and 2 GB of usable disk space to download and use the software for this tutorial.

This Tutorial has been developed at the Institute for Geoinformatics, University of Münster. Authors are Tobias Krumrein and Albert Remke. The latest version of this tutorial is always available on GitHub. We hope you will use GitHub issues to provide feedback and suggest improvements.

The OER4SDI project has been recommended by the Digital University NRW and is funded by the Ministry of Culture and Science NRW.

2. Background

Before we can begin with the exercises, we need to explain some terms and technical concepts. In doing so, we will focus on the use of ISO metadata and classic OGC catalog services (CSW). In the last chapter, we will briefly discuss other approaches for providing and using metadata.

2.1 The role of Metadata in SDIs

Metadata is everywhere in the modern world, especially in the digital world.

Imagine you are in your gallery on your phone and look through your photos. So each picture is the main data itself. But if you click on the details there is much more information about the photo like the resolution or the time the photo was taken. And this information about the data is called the metadata. It helps you to organize, store, find and understand data like the photos in your gallery without it you couldn't sort them by the time it was taken. You can think of metadata as little tags or labels attached to the main data. But metadata isn't just for photos, you can attach metadata to everything even outside the digital world like the author, title or genre of a book. While you think about other areas where metadata is integrated, you can see how important and present metadata is.

Therefore it's not surprising that metadata is playing a crucial role in Spatial Information Infrastructures (SIIs) as well. SIIs are very large distributed systems designed to facilitate the discovery, access, and use of geospatial information, often on a regional, national, or international scale. A term used synonymously is Spatial Data Infrastructure (SDI), which puts more emphasis on ‘data’ as the base material of information.

Many SIIs are focused on public sector information. Their purpose is to make data that is collected for a specific purpose (e.g. management of nature reserves) available and usable for any other applications and for any other users such as researchers, companies, and private individuals. These infrastructures are always managed by one or more organizations on the basis of binding regulations and agreements between the respective stakeholders.

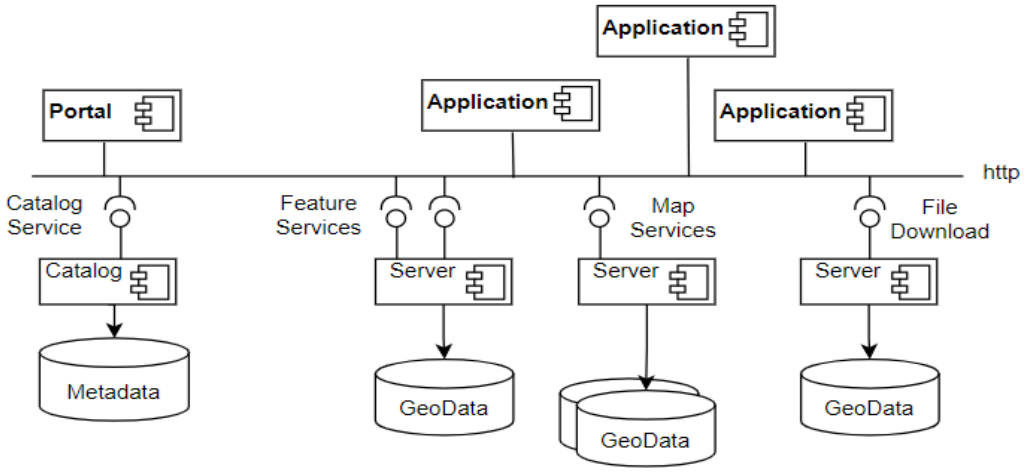

Fig.1: Technical components of Spatial Information Infrastructures

The picture (Fig. 1) shows a simplified overview on the technical components and interfaces of an SII. The core content is the geospatial data that resides somewhere in the infrastructure. A mapping and surveying organization may provide topographic maps and geodata on administrative units, an environmental agency may provide data on protected sites, a XYZ office may provide access to data on 'whatever', and so on. In the best case the data is contributed in well-known data models and formats via web services (e.g. data access, data processing, data visualization) that can be accessed through well-known protocols free of charge from anyone, anywhere, and at any time.

Metadata in SDIs provide essential information about spatial data resources, including their content, quality, and spatial extent, enabling users to find and evaluate data that meets their specific needs. It's like a detailed map that helps users navigate through the vast landscape of spatial data, ensuring they can locate and access the right information for their specific needs with confidence and ease. It forms the basis of an SDI architecture.

In this context, you may have heard of the publish-find-bind (PFB) paradigm, in which metadata plays a crucial role:

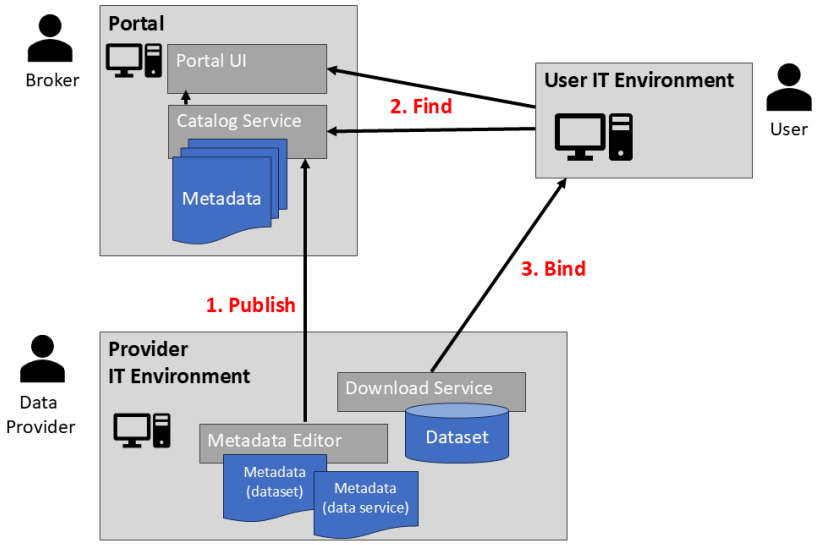

Fig. 2: Metadata supporting the Publish-Find-Bind pattern of open distributed systems

In an SII, all content providers publish their datasets and services by uploading metadata to one or more catalogs that are provided by information brokers (also called clearing houses). The metadata contains information such as title and abstract, spatial extent, quality, data formats, supported interfaces, and a point of contact.

The broker, for example the service center of the national spatial data infrastructure in Germany (GDI-DE), operates the catalog and a portal website with a user interface to the catalog. The catalogs may harvest metadata from other catalogs and thus form a system of federated catalogs with at least one central catalog and portal that provides access to all datasets and services that are available in that specific SII.

Users typically use the portal website to search for and find geospatial datasets and services. The metadata contains all information that is needed for online-access to these resources (binding). They could use the catalog service directly, but without a convenient user interface this is not really advantageous. The catalog interface is made for machine-to-machine interactions, e.g. for synchronizing catalogs or connecting to editing and metadata management systems.

As you can see, metadata and metadata catalogs are essential elements of an SII architecture that enable the efficient provision and use of (geo)information resources across countries and organizational boundaries.

2.2 ISO metadata standards and how are they used in the context of INSPIRE

To ensure the effective exchange and use of geodata, it is important to standardize data models for metadata. The International Organization for Standardization (ISO) has developed a series of metadata standards, including ISO Standard 19115, which defines a conceptual data model for the description of geospatial data and associated metadata. This model provides a standardized framework for describing the content, quality and spatial extent of geodata as well as information on ownership, access and usage restrictions (see Fig.3).

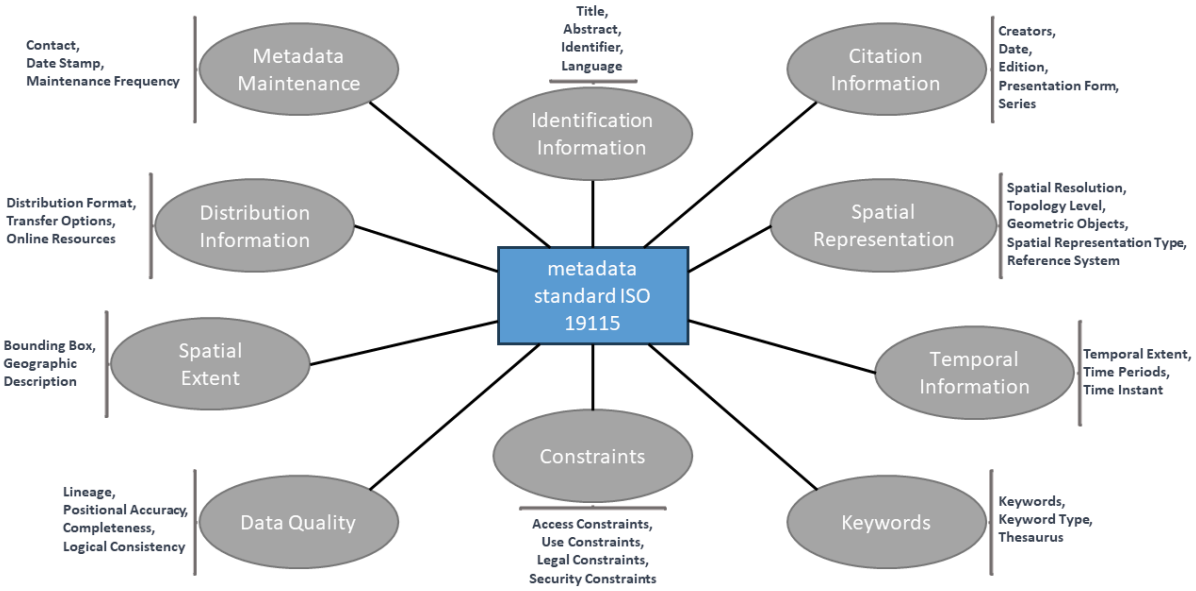

Fig. 3: ISO 19115 Metadata Elements, Overview

To facilitate the implementation of ISO metadata standards, ISO has also developed encoding standards, such as ISO 19139, which provides a standardized syntax and structure for encoding metadata as XML.

These standards are often quite general and comprehensive and need to be adapted to the requirements of a specific SDI. This is done by deriving specific versions of the metadata standards that comply with all the mandatory rules of the standard and are otherwise tailored to the respective requirements through restrictions and additions (a version of a standard derived in this way is called a profile of the standard).

In our tutorial, we will focus on metadata whose data model is derived from the ISO 19115 standard. Ultimately, there are many similarities between the existing standards.

ISO standards play a special role in Europe, as the European Commission has decided to use ISO metadata standards for the Infrastructure for Spatial Information in Europe (INSPIRE) and has developed its own INSPIRE profile for this purpose. INSPIRE requires Member States to create

national metadata catalogs that comply with ISO metadata standards to enable interoperable access to spatial data resources. By standardizing metadata, INSPIRE aims to promote the sharing and reuse of spatial data across Europe, facilitating the interoperability of data and metadata across borders and sectors. The INSPIRE metadata catalog brings together metadata from the national spatial data infrastructures of the member states.

2.3 OGC catalog services (CSW)

When it comes to accessing metadata catalogs using web technologies, the standards of the Open Geospatial Consortium (OGC)[1] are important. The OGC develops de facto standards that are intended to make geoinformation FAIR - FAIR stands for findable, accessible, interoperable and reusable.[2]

The term "Catalog Services for the Web" (CSW) refers to an OGC interface standard for accessing metadata catalogs in which information about geodata and geodata services is maintained.[3] The CSW standard defines a set of operations that can be executed by metadata catalogs. These operations include querying metadata as well as retrieving and manipulating metadata records.

Catalogs can also interact and form a federated system of catalogs in which search queries are forwarded or metadata is replicated. For example, the INSPIRE catalog service interacts in this way with the catalogs of the SDIs of the member states. In Fig. 4 you can see the structure of the network of SDI catalogs in Germany. Each catalog you see there belongs to a specific geoinformation infrastructure. The smooth cooperation of these catalogs would not be possible without the agreement on common standards.

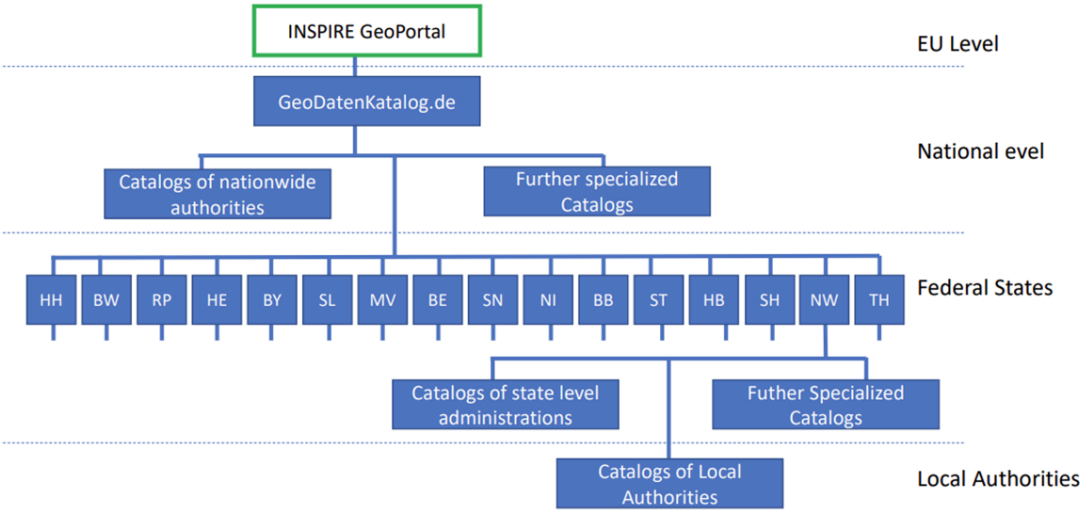

Fig. 4: Network of SDI catalogs in Germany

The ability to share metadata is a key capability of catalog services.

[1] Open Geospatial Consortium, OGC Standards, WebSite: https://www.ogc.org/standards/
[2] Go Fair Initiative, FAIR Principles, WebSite: https://www.go-fair.org/fair-principles/
[3] Open Geospatial Consortium, OGC Catalog Services, WebSite: https://www.ogc.org/standard/cat/

The CSW standard enables searching for information resources using metadata such as keywords, geographic coverage or other search criteria, as well as retrieving metadata records in different encodings, such as XML or JSON.

Compared to the major search engines such as Google, Bing or Yahoo, there are clear differences. These search engines are designed to automatically crawl and index countless websites with unstructured data so that users can search for information across the entire World Wide Web. Metadata catalogs are usually classic database applications for managing and providing structured metadata about data and data services. While the major search engines still struggle to efficiently support the search for data sets and data services, metadata catalogs are clearly lagging behind in terms of automated data preparation and intelligent search mechanisms.

CSW catalog services typically support most of the following operations:

- **GetCapabilities**: this operation allows querying the available functions and properties of the CSW, including the supported operations and the data schema.
- **DescribeRecord**: This operation allows you to retrieve information about the available metadata records, including the metadata schema and the supported metadata elements.
- **GetRecords**: This operation enables the retrieval of metadata records based on specific criteria such as keywords, spatial extent or time intervals.
- **getRecordById**: This operation enables the retrieval of metadata records based on their unique identification number.
- **transaction**: This operation supports the addition, update or deletion of metadata records in the CSW.
- **harvest**: This operation is used to collect metadata from other CSWs in order to import them into another catalog.

The CSW standard has been implemented in a number of tools, such as QGIS MetaSearch or ESRI Geoportal Server. These tools provide software solutions for managing and accessing metadata catalogs. We will take a closer look at QGIS in Chapter 4.

Most of the classic spatial data infrastructures such as INSPIRE and the national spatial data infrastructures in Europe currently use the OGC CSW standard. However, with “OGC API Records” a promising RESTful interface standard has already emerged in the new generation of OGC APIs, which is very likely to succeed the CSW in the coming years.

3. Practical examples for using Metadata, Catalog Services and Portals

After all the reading we can now put our new knowledge into action.

First we’ll investigate how Geoportals can be used to search for and find geospatial datasets and services and how they use catalog services to accomplish this. Then we’ll have a closer look at QGIS

which is capable of interacting with catalog services in a plug-and-play mode, trying to understand what happens “under the hood”. Last not least we’ll use Jupyter notebooks and Python code to check how we can interact with catalog services on the level of programming languages.

We’ll use the following software configuration for our exercises:

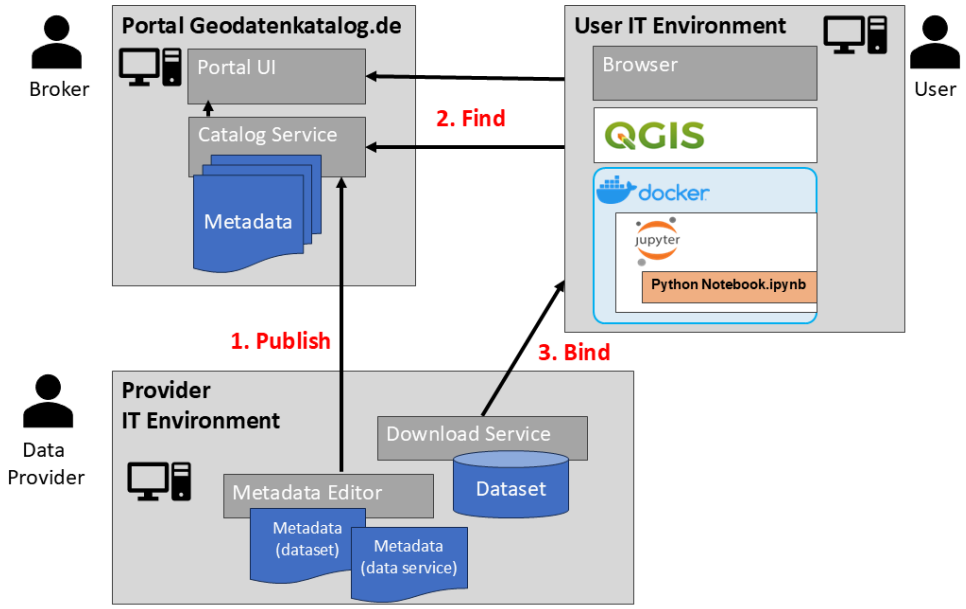

Fig.5: IT components used for the exercises

The portal that we will use is the central geoportal of the German national Spatial Data Infrastructure: https://geoportal.de/. The metadata catalog (called ”Geodatenkatalog”) is available at: https://gdk.gdi-de.org/gdi-de/srv/ger/catalog.search.

The provider’s IT environment for providing access to datasets and services is unknown to us. The metadata of those resources will contain all we need to interact with the provider’s infrastructure.

In your own IT environment you will use your browser, QGIS and a python notebook which will be executed by a Jupyter server that runs in a docker container on your computer. No worries, we’ll guide you through all steps that are necessary to get the software up and running.

3.1 Wind Farm Planning as a use case

Using the example of wind farm planning, we will see why metadata is needed and how it is provided and used in practice.

Imagine you work in a consulting company that supports clients in the planning, construction and operation of wind farms. Your company has been contracted by a group of investors to propose solutions for the repowering and expansion of an existing wind farm in Lorup, a small community in the Emsland district, Lower Saxony.

In fact, many areas are not suitable as sites for wind farms, since, for example, nature conservation concerns, the well-being of the local population, or competing usage interests have to be taken into account. Before a wind farm planning is approved by the responsible authorities, it must therefore be checked whether it complies with legal regulations and whether competing claims for use have been sufficiently taken into account. Citizens and institutions affected by the impact of the planning have the right to appeal against it and to demand a court decision.

For example, the following exclusion criteria, among others, must be met for wind farm sites:

- Nature reserves, bird sanctuaries, FFH areas are to be excluded.
- Water protection areas (category I + II), smaller than 2000 ha are to be excluded
- Wetlands (More, swamps) 10 ha or larger are to be excluded
- A distance of 800 m from closed residential areas is to be kept
- Campsite, facility for sports, leisure and recreation + 400 m buffer is to be excluded
- Air traffic control facilities (radar and ground navigation facilities) + 3,000 m buffer
- Roads and roadside areas (freeways +40m, otherwise +20m) are to be excluded
- …

So it is clear, you need high quality up-to-date geospatial data to be able to find the areas that are suitable for wind farm sites. In this tutorial we want to focus on data on Nature reserves.

But how and where to find them?
So, let’s start searching for datasets on Nature reserves (Naturschutzgebiete).

3.2 Accessing metadata via geoportals

In our first exercise, we want to take a closer look at how geoportals can be used to search for data and data services and what role catalog services play in this.

We are looking for data on protected areas (such as nature reserves) in the region of Lorup, Lower Saxony, which we need for our virtual wind farm planning project. For this purpose we use the "Geodatenkatalog" of the National Spatial Data Infrastructure Germany, GDI-DE. The web address of the portal can be found here: https://gdk.gdi-de.org/gdi-de/srv/ger/catalog.search#/home.

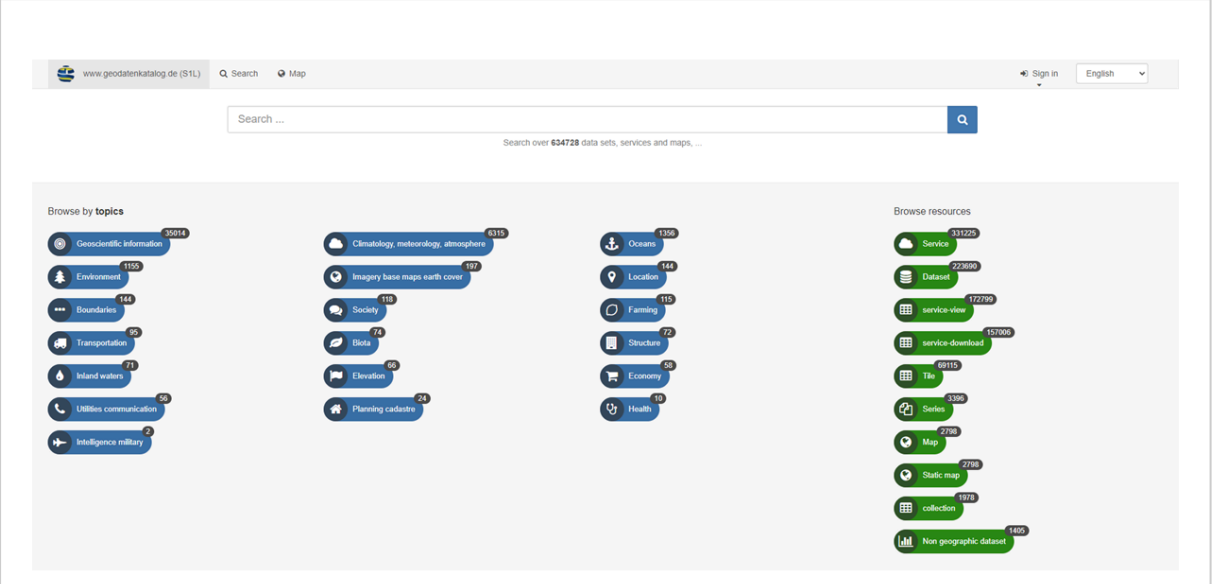

Fig.6: Geodatenkatalog GDI-DE homepage, topic overview

This is the homepage of the Geodatenkatalog. You can change the language in the top right corner. However, not all content is translated. If needed, use your browser's built-in translation features to translate the content into your own language. Here you'll find metadata of more than 600.000 available datasets, services and maps, too many to go through one by one. Luckily, we can filter by topic categories. In our wind farm planning context the topic category "Environment" might be the right one to go for. When selecting this category, the portal's interface will send a getRecords request to the GDI-DE catalog service which runs in the background, asking for metadata records with the topicCategory "Environment".

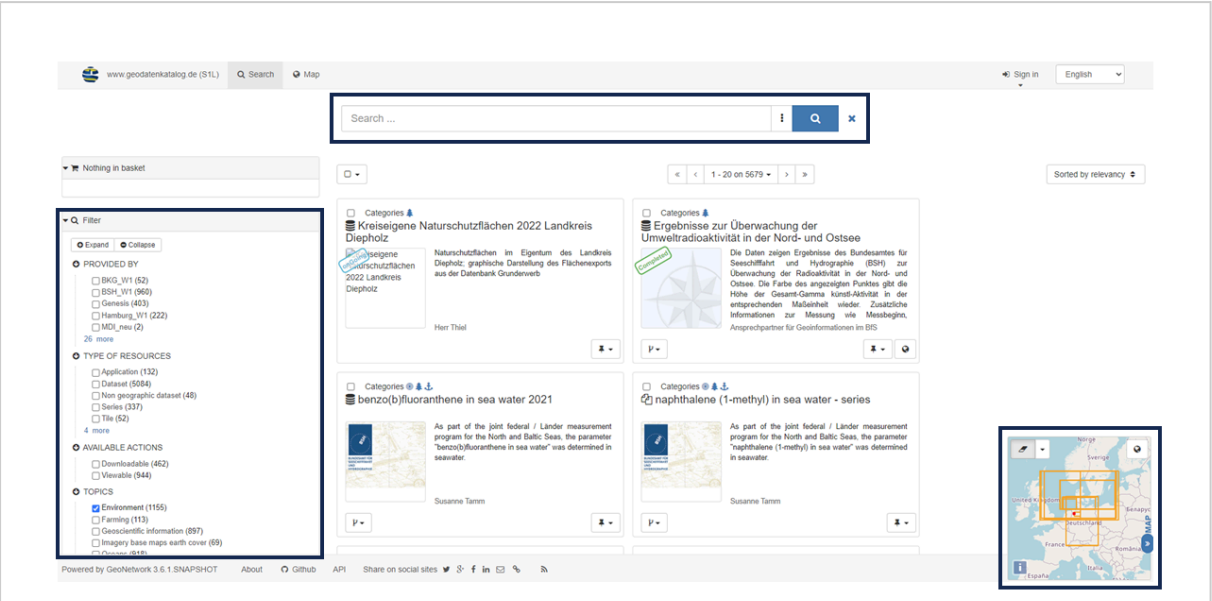

Fig. 7: Geodatenkatalog GDI-DE, Metadatasets in the topic Category "Environment"

Nice! We have filtered out many datasets and services. But we need to further narrow down the result set before we can look into the very details of the metadata. We have several options that we can use: the search bar at the top, a filter section where we can add more filters and a map tool on the bottom right corner. Let's try out the map tool first.

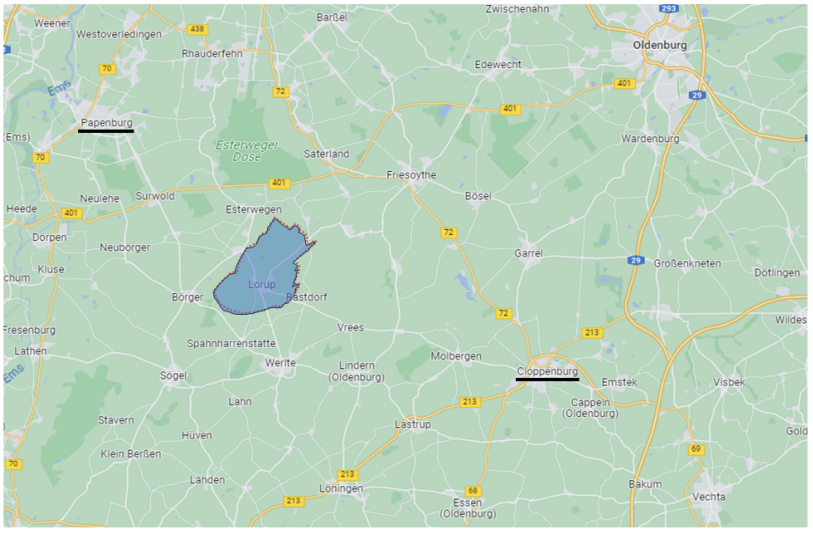

Fig. 8: Geodatenkatalog GDI-DE, searching by defining a bounding box

The region of the municipality "Lorup" that we are interested in is located between Papenburg and Cloppenburg in the district of Emsland. Try to zoom to this region and draw a rectangle that represents our area of interest. Use the criterion "intersects" to avoid that datasets are excluded that are relevant but cover a larger area. In the background the portal will send a getRecords request with a bounding box parameter as an additional filter argument.

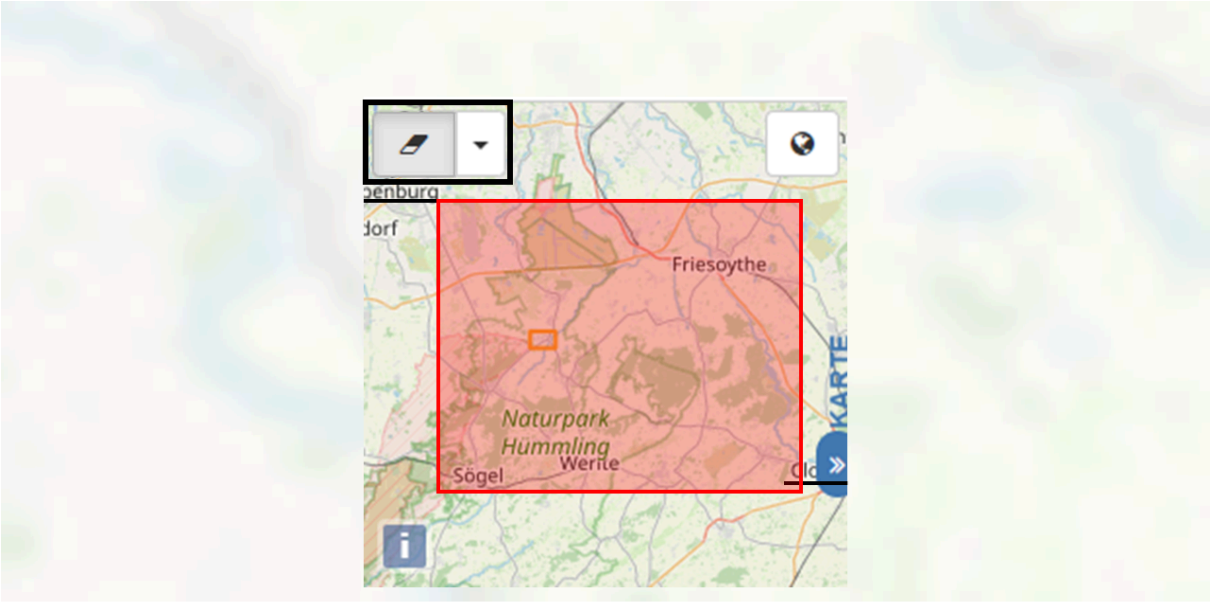

Fig. 9: Geodatenkatalog GDI-DE, draw rectangle for specifying the bounding box

There are still too many datasets and services in our result set.

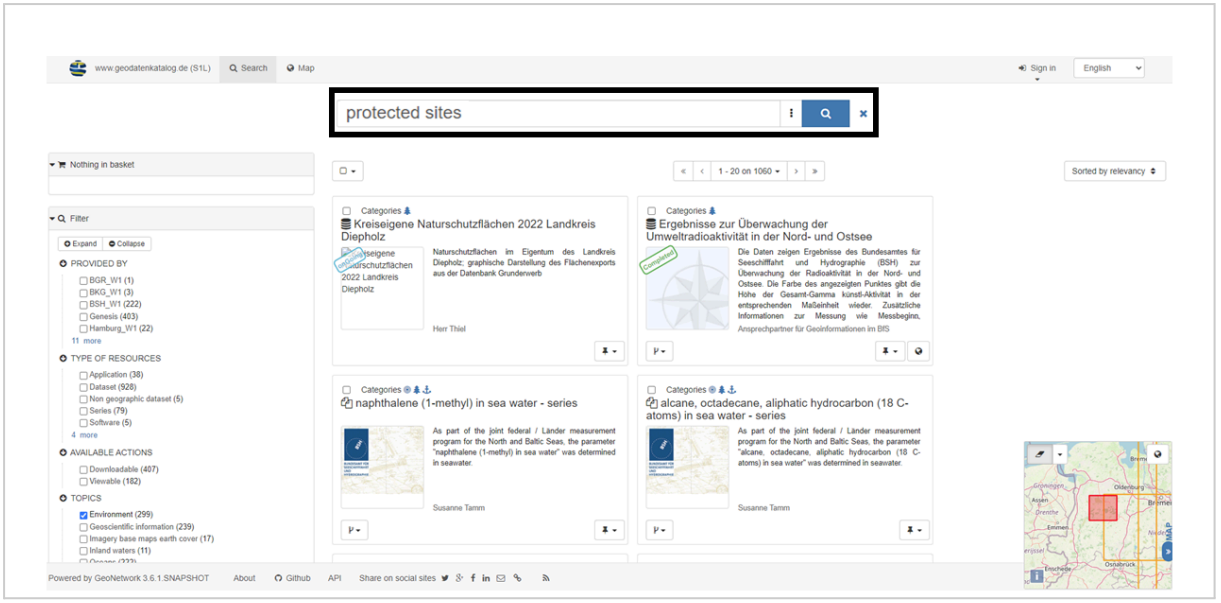

Fig. 10: Geodatenkatalog GDI-DE, searching by defining a bounding box

Searching for specific words such as „Protected Sites“, “INSPIRE”, “Niedersachsen” will create a more specific selection that better fits your interest. Let’s try filtering more until we have about 20-30 results remaining before investigating the metadata in more detail.

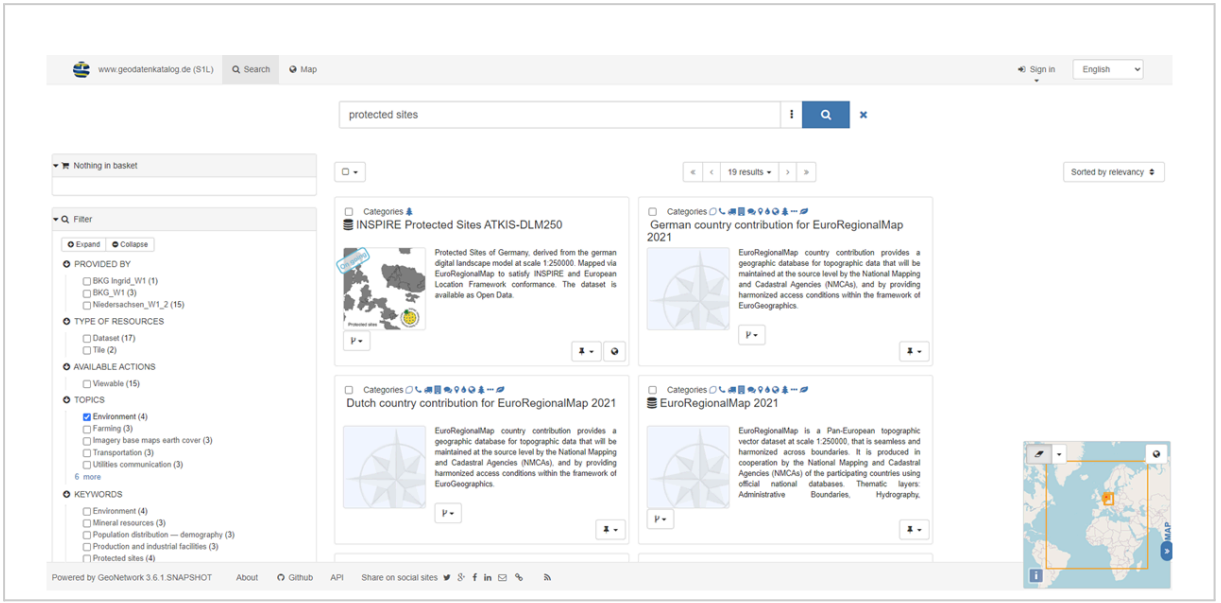

Fig. 11: Geodatenkatalog GDI-DE,Results after filtering with searching bar

That worked out quite well; we only have around 20 datasets left. The dataset „Naturschutzgebiete (NSG)“, seems to fit quite well for our use case since it is authoritative data reporting the extent and designation of protected Sites in Lower Saxony. Let’s select it to analyze it in more detail.

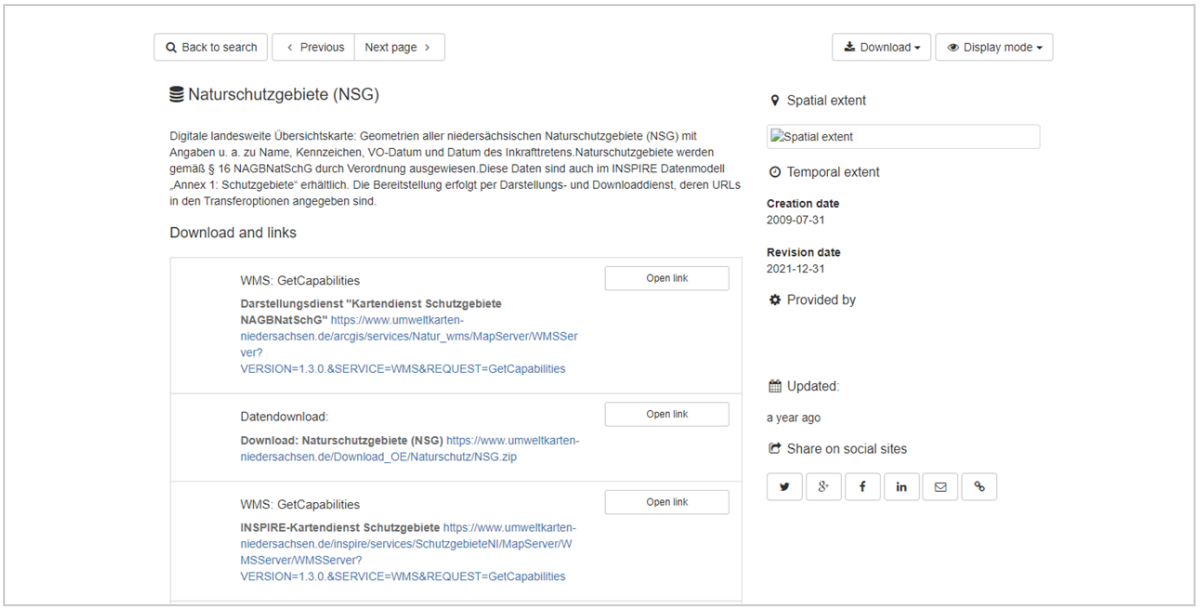

Fig. 12: Inspecting the metadata on the dataset “Naturschutzgebiete (NSG)” 1/4

The geoportal uses a request getRecordByID to retrieve the full metadata set from the catalog and provides a user-friendly HTML view that allows us to easily understand the dataset's attributes, such as its name, date of creation, and format. The abstract gives us a good overview on the content of the resource. Directly below you’ll find a list of links to access points for view and download services..

One of the most important attributes of information sources is the spatial and temporal coverage as well as the up-to-dateness of the available data. A dataset can fit your topic perfectly, but if it is not up to date - especially when it comes to protected areas - it is completely useless. So always pay attention to the date of the last modification.

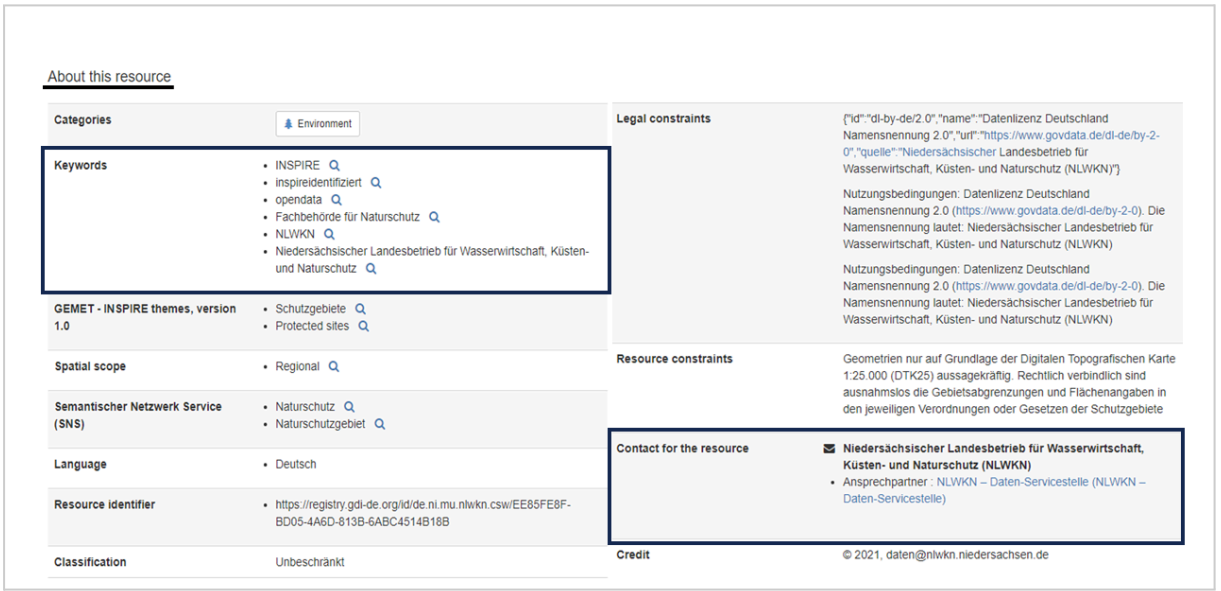

Fig. 13: Inspecting the metadata on the dataset “Naturschutzgebiete (NSG)” 2/4

Fig. 14: Inspecting the metadata on the dataset “Naturschutzgebiete (NSG)” ¾

Fig. 15: Inspecting the metadata on the dataset “Naturschutzgebiete (NSG)” 4/4

If you scroll further down, you will see more detailed information on the dataset, e.g. on the source, the coordinate reference system, points of contact and the unique ID[4] of the metadata set, which had been used by the portal to retrieve the metadata record from the catalog using the CSW operation getRecordByID.

In addition to the HTML interface you have the possibility to download the plain .xml file of the metadata. Use “download metadata” to download the XML file and open it in our preferred browser[5].

[4] ID of the metadata record on Naturschutzgebiete (NSG): EE85FE8F-BD05-4A6D-813B-6ABC4514B18B
[5] we’ve done so for you so that you can find the dataset as well at our GitHub repository: here

Try to find some of the information that was also displayed in the HTML representation. Keep in mind that the Geoportal UI gets its information from the .xml file, and not all metadata elements contained therein are necessarily displayed.

Below you will find an excerpt from the .xml file in case you have problems opening it:

```
<gmd:MD_Metadata xmlns:gmd="http://www.isotc211.org/2005/gmd"
xmlns:gco="http://www.isotc211.org/2005/gco"
xmlns:srv="http://www.isotc211.org/2005/srv"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:gfm="http://www.isotc211.org/2005/gfm"
xmlns:base="http://www.disy.de/preludio2/base"
xsi:schemaLocation="http://www.isotc211.org/2005/gmd
http://schemas.opengis.net/iso/19139/20060504/gmd/gmd.xsd"
uuid="EE85FE8F-BD05-4A6D-813B-6ABC4514B18B">
 <link type="text/css" rel="stylesheet" id="dark-mode-custom-link"/>
 <link type="text/css" rel="stylesheet" id="dark-mode-general-link"/>
 <style lang="en" type="text/css" id="dark-mode-custom-style"/>
 <style lang="en" type="text/css" id="dark-mode-native-style"/>
 <style lang="en" type="text/css" id="dark-mode-native-sheet"/>
 ...
 <gmd:identificationInfo>
  <gmd:MD_DataIdentification uuid="093121e3-f872-4aa2-b9db-40ee82cc1a18">
   <gmd:citation>
    <gmd:CI_Citation uuid="60e4561f-7aea-49f3-93cb-1c66ba52f406">
     <gmd:title>
      <gco:CharacterString>Naturschutzgebiete (NSG)</gco:CharacterString>
     </gmd:title>
     <gmd:alternateTitle>
      <gco:CharacterString>Naturschutzgebiete_DTK25</gco:CharacterString>
     </gmd:alternateTitle>
     <gmd:date>
      ...
     </gmd:date>
     <gmd:date>
      ...
     </gmd:date>
     <gmd:identifier>
      <gmd:MD_Identifier uuid="cb1b4ff1-bf68-4e5a-aa84-73043f65501f">
       <gmd:code>
<gco:CharacterString>https://registry.gdi-de.org/id/de.ni.mu.nlwkn.csw/EE85FE8F-BD05-4A6D-813B-6ABC4514B18B</gco:CharacterString>
```

```
          </gmd:code>
         </gmd:MD_Identifier>
        </gmd:identifier>
        <gmd:citedResponsibleParty>
         …
        </gmd:citedResponsibleParty>
        <gmd:presentationForm>
         …
        </gmd:presentationForm>
       </gmd:CI_Citation>
      </gmd:citation>
      <gmd:abstract>
       <gco:CharacterString>Digitale landesweite Übersichtskarte: Geometrien aller
niedersächsischen Naturschutzgebiete (NSG) mit Angaben u. a. zu Name, Kennzeichen, VO-Datum
und Datum des Inkrafttretens.Naturschutzgebiete werden gemäß § 16 NAGBNatSchG durch
Verordnung ausgewiesen.Diese Daten sind auch im INSPIRE Datenmodell „Annex 1: Schutzgebiete“
erhältlich. Die Bereitstellung erfolgt per Darstellungs- und Downloaddienst, deren URLs in den
Transferoptionen angegeben sind.</gco:CharacterString>
      </gmd:abstract>
```

Did you identify some elements of the dataset description from the portal in this .xml metadata file?

We should keep in mind that XML is mainly intended for machine-to-machine communication, e.g. with QGIS, Python or other catalogs. But sometimes it is important to read the textual content of XML files and search for specific information. The first part of the .xml file describes the structure of the .xml sheet. The metadata information about the record you are looking for is then described in <gmd:identificationInfo>. Here you will find information about the title, the summary or the identifier of the data set. The suffix </gmd:> stands for the namespace "geospatial metadata".

We have now seen how the portal functions and the underlying metadata catalog can be used to find out about the information resources available in an SDI.

In the next step, we will take a closer look at how GIS systems such as QGIS support access to CSW metadata catalogs

3.3 Accessing metadata via QGIS MetaSearch

3.3.1 QGIS Installation

If you already have QGIS 2.0 or higher installed you can skip this part and go directly to 3.3.2, because the plugin MetaSearch, which you're using later, is included by default.

For downloading QGIS please follow the next steps:

a) Go to the QGIS Website https://www.qgis.org/en/site/forusers/download.html

b) Choose your operating system from the options that are listed and download QGIS

c) After downloading run the installer and follow the prompts to install QGIS on your computer

d) Activate the metaSearch Plugin via >plugins >manage and install plugins

3.3.2 QGIS MetaSearch

This exercise is about exploring how QGIS can be used to connect to existing SDI catalogs to search for resources and integrate them into the GIS environment. For this purpose we will use the QGIS plugin “MetaSearch”.

Please open QGIS and start a new project.

Use the menu bar to navigate to the option >Web >MetaSearch.

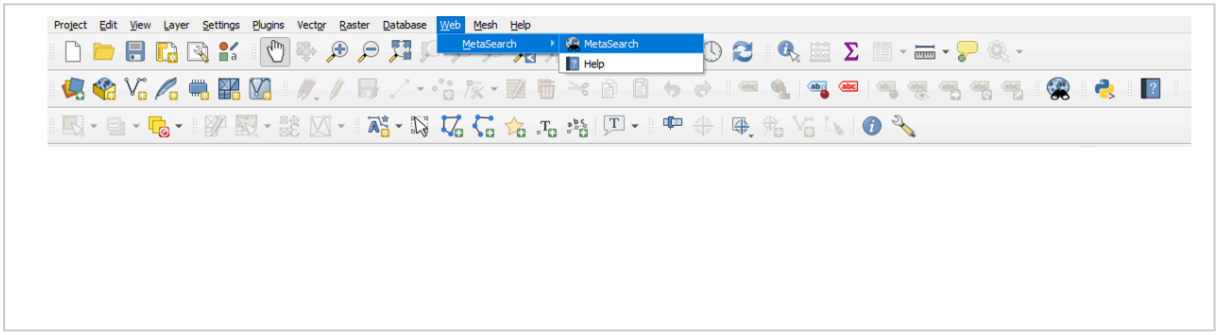

Fig. 16: Navigating to the plugin MetaSearch via the QGIS Menubar

It should open a new window with the MetaSearch tool. Navigate to “Services” (Dienste) and create a new connection with “New”.

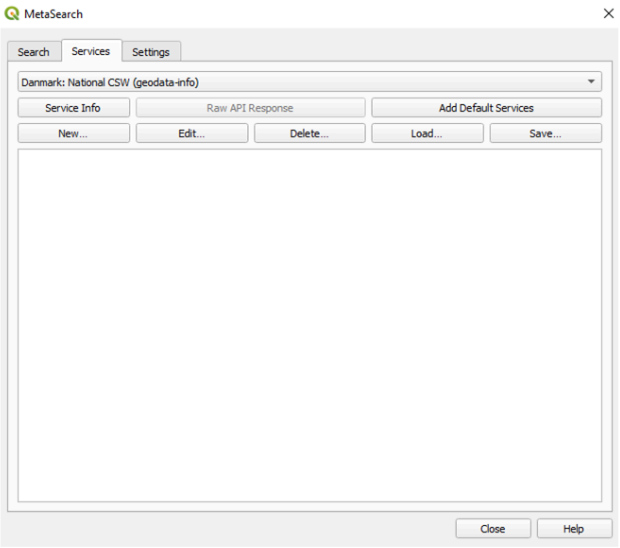

Fig. 17: QGIS MetaSearch New Service

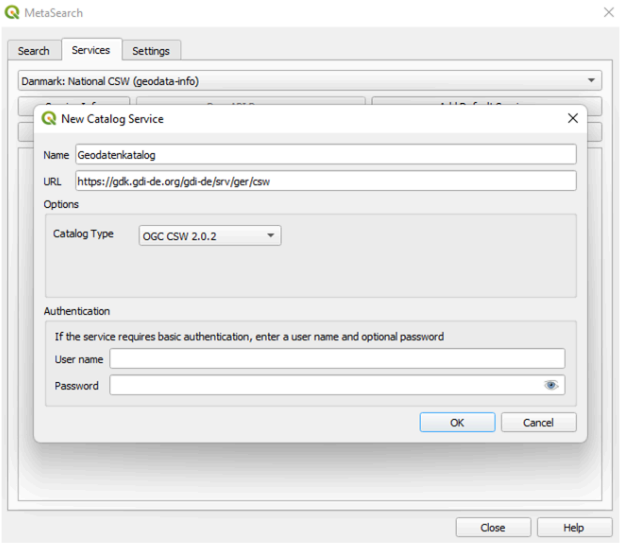

Fig.18: New MetaSearch connection

Now you can define a name for your new connection, type in the URL of the catalog service and confirm. We want to use the "Geodatenkatalog" of GDI-DE again, which has the url: https://gdk.gdi-de.org/gdi-de/srv/ger/csw

You can find the url by searching for the term "Geodatenkatalog" in the Geodatenkatalog.de and examining the metadata as you have done in the previous section.

Now we can access the CSW by activating the tab "Service Info" (Dienstinfo). QGIS will send a GetCapabilites request to the GDI-DE catalog service and display the information from the response message in a readable form.

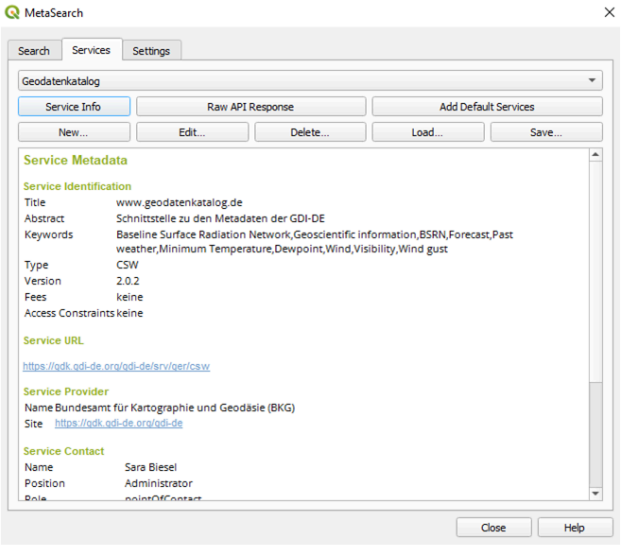

Fig. 19: Using the ServciceInfo option of MetaSearch

If this was successful, you're evidently connected to the Geodatenkatalog and ready to start with searching for datasets.

The tab “Search” provides you with a form that you can use to specify your search request. QGIS will use the content of the form to generate a getRecords request with all the query parameters and send it to the CSW catalog service. The results from the getRecords response of the catalog service will be shown in the results section.

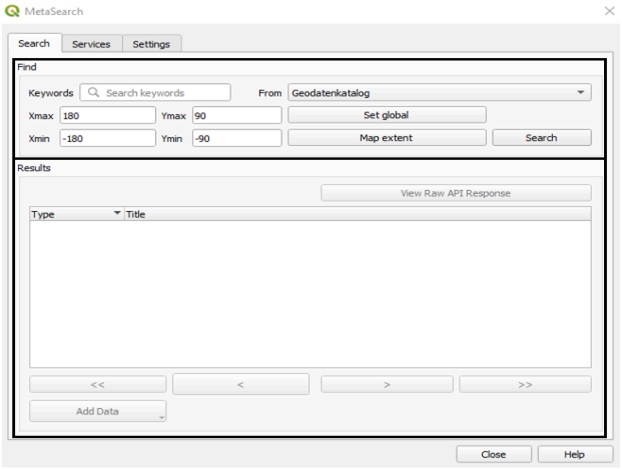

Fig. 20: Searching in MetaSearch

In the search section you have two options to filter in the datasets. You can search for keywords and use a BBox as a filter. Let's start filtering through keywords. You can also type in more than one keyword. Can you find the metadata record we found in the last exercise here as well?

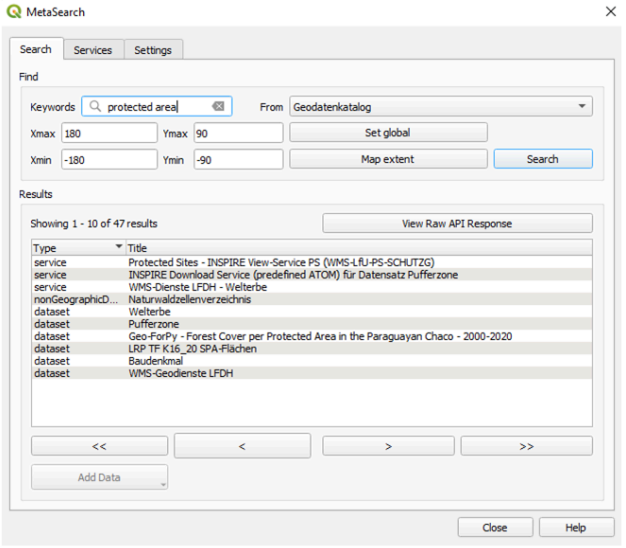

Fig. 21: MetaSearch searching for protected area

Double clicking an item opens a new window displaying details of the metadata record. In this case, QGIS has sent a getRecordByID message to the catalog which responded with the full XML encoded metadata record. Have you found the dataset that we used before? Just check the abstract or the Identifier (EE85FE8F-BD05-4A6D-813B-6ABC4514B18B).

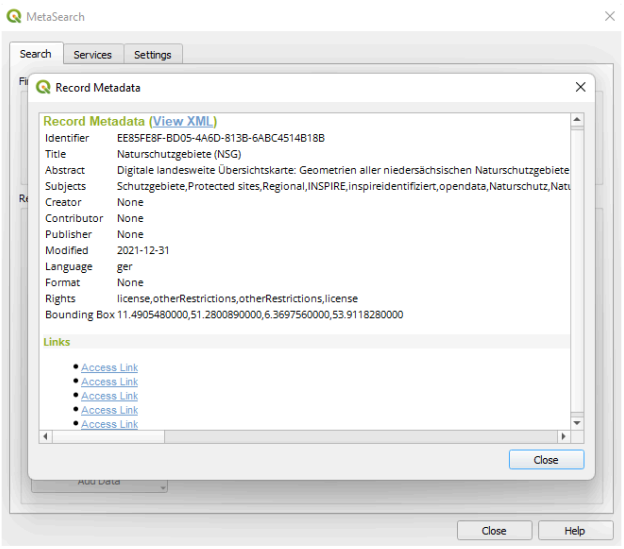

Fig. 22: MetaSearch Naturschutzgebiete (NSG)

You also have the option to view the XML encoded metadata with all its details that has been sent from the catalog service. As you can see, the XML file conveys exactly the same information on the NSG dataset but looks quite different due to the fact that the GetRecordsByID response has a specific encoding.

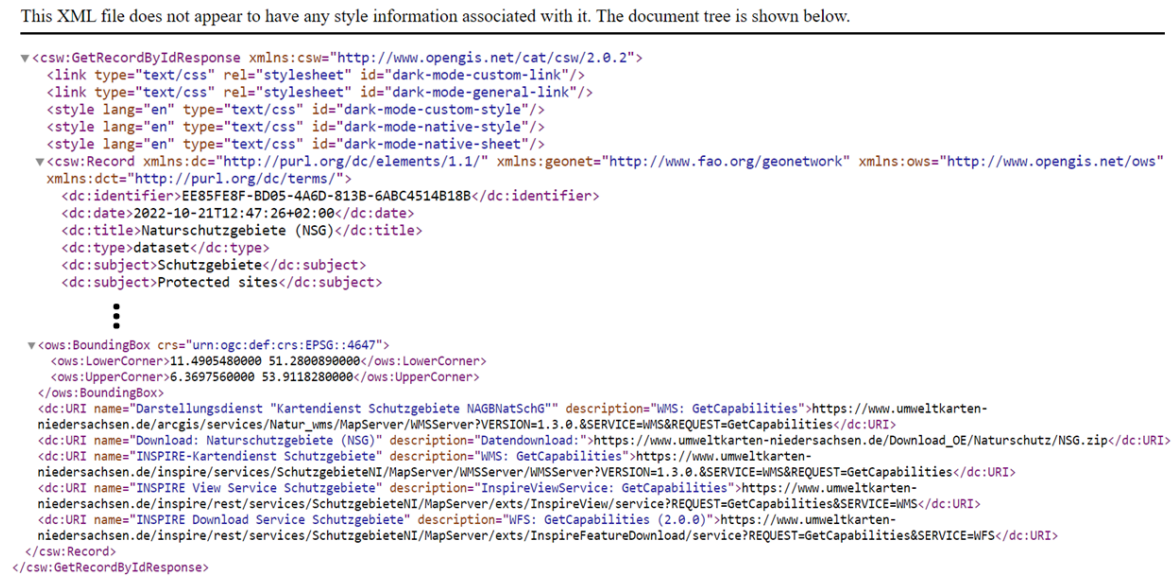

Fig. 23: MetaSearch XML Naturschutzgebiete (NSG)

As you can see, QGIS can interact with CSW catalog services in plug-and-play mode because it has implemented the OGC CSW interface standard and now only needs the web address of the service to establish the connection to the service. To do this, QGIS first uses the Get_Capabilities and DescribeRecords operations to retrieve all relevant details about the interface of the catalog instance. QGIS then uses your entries in the query form to search for metadata records that match your search criteria using the GetRecords() function.

So far, so good. Unfortunately, QGIS is not smart enough to understand the metadata and access the data sets and services directly via the described interfaces. Perhaps this will be possible in the future. Anyway, QGIS-Metasearch is a good example of how standard software can be enabled to interact with catalog services in plug-and-play mode by implementing the standardized interface specification.

3.4 Accessing metadata via Python from a Jupyter Notebook

In our last exercise, we will learn how to access standardized catalog services using a programming language (here: Python).

3.4.1 Installing Docker, Jupyter and a Python Notebook

The software environment that we need for this exercise is Docker:

a) Install Docker Desktop: Go to the official web site https://docs.docker.com/get-docker/ and follow the guidance which is provided there to install docker on your computer (8GB usable RAM recommended).
b) Download the GitHub Repository of this Tutorial (https://github.com/oer4sdi/OER-MetadataAccessVia-OGC-CSW) and unzip it in your wanted

working directory.
If you are working with GIT, you can also use git clone to download the repository.

c) Make sure Docker Desktop is up and running
d) Open a terminal and change to your working directory
e) Use "docker-compose up --build" to build and run the docker container

Once this is complete, your terminal output should look similar to this:

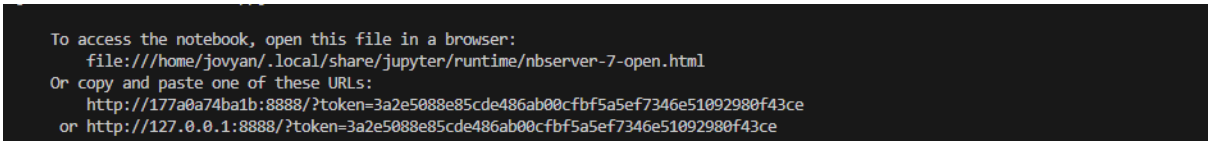

Fig. 24: Example of a correct response in the terminal after the docker container is running

To access the python notebook look out for the link http://127.0.0.1:8888/?token=... in your output (it should look like the one on the bottom in the picture above). Copy the link and the token (the latter will be created new every time you start the container). Paste the link to your preferred browser to access the Jupyter server which displays the notebooks that can be used.

3.4.2 Jupyter Notebook

After installation, please start the notebook "Accessing_Metadata_with-jupyter.ipynb" and work through it on your own. It will show you a different approach to accessing CSW metadata catalogs.

You will work with the Python package OWSLib, which supports interaction with OGC web services such as CSW, WMS or WFS. You will connect to the catalog using the same URL you used in QGIS. You will also use the same operations as in the previous exercises, with the difference that you will now have to configure and send these requests yourself and also receive and process the responses from the service by yourself.

(please work through the Notebook Accessing_Metadata_with-jupyter.ipynb)

Once you have completed this exercise, you can shut down and clean up your working environment.

To do this, you must enter "docker-compose down" in the terminal in which you created and started the Docker container. This will cause the Docker container to be shut down. If you also want to delete the docker image, enter "docker image rm oer_jupyter".

4. Summary and Notes on Related Topics

You have now learned how OGC catalog services support querying and downloading metadata from metadata catalogs. Portals provide you with a user interface that hides all the technical details of interacting with the catalog services that are actually used under the hood.

The key messages of this tutorial are:

- Metadata and metadata catalogs are key components of SDIs supporting the publish-find-bind paradigm of open distributed systems by making existing information resources discoverable and accessible.
- Metadata standards are essential means for (meta) data interoperability
- Standardized interfaces such as the OGC CSW allow software clients of any kind to connect to catalogs in a plug-and-play mode

But, the OGC CSW is not the only interface standard on stage..

Gaining Ground: OGC API Records

As for the CSW type of catalog services, be aware that it implements the classic style of OGC Web Services, heavily relying on sending remote procedure calls using XML encoded message formats. The new "OGC API Records" standard implements a more lightweight RESTful interface based on JSON encodings that is supposed to take over within the next few years. For more information we recommend visiting the OGC API workshop on this service type that is provided by the Open Geospatial Consortium: https://ogcapi-workshop.ogc.org/api-deep-dive/records/

Linked Data & DCAT

As well you should pay attention to DCAT - the data catalog vocabulary, which is based on the linked data principles. In its core, it consists of a set of well-defined terms that can be used for cataloging datasets and data services. It reuses to a maximum extent terms from already existing controlled vocabularies (such as the Dublin Core Vocabulary) and defines new ones only if evidently needed.
DCAT has its strengths in integrating heterogeneous metadata from various resources by describing the relation of (new) terms to existing ones which can be used to automate the integration of these metadata sources. The European Data portal https://data.europa.eu uses DCAT application profiles to harvest data from many metadata sources all over Europe. If you want to dive into this topic you might find this learning material helpful:https://github.com/oer4sdi/OER-DCAT.

And another one: STAC

The SpatioTemporal Asset Catalog (STAC) is an open standard for describing geospatial data in a JSON based machine-readable format. STAC is designed to handle both spatial and temporal aspects of geospatial assets (e.g. data, organized in files), making it particularly well-suited for Earth observation data, satellite imagery, and other time-series data. A really exciting new citizen in town. So if you

want to learn more about STAC we recommend visiting the tutorials on the official STAC web site at: https://stacspec.org/en/tutorials/intro-to-stac/

Hey, you made it! 😊

Now you should be able to answer the following questions:

- How can we use metadata for discovering datasets and services in an SDI?
- What is the role of the ISO and INSPIRE metadata standards?
- What is the role of OGC catalog services (CSW)?
- How can I access OGC catalog services with QGIS and/or Python Code?
- What other approaches are being used to support the discoverability of datasets and services?

We hope that you enjoyed reading and working with the tutorial and the practical examples and that you were able to deepen your understanding of SDIs.