**spaceapplications** *SERVICES*

**OBEOS – Linked Data Prototype Software User Manual**

# OBEOS

# Linked Data Prototype Software User Manual

**Title** : OBEOS
Linked Data Prototype Software User Manual

**Abstract** : This document describes the usage of the Linked Data Prototype by an end-user for discovering EO dataset series and EO datasets registered in remote catalogues and datastores and navigating in the available metadata.

# Table of Contents

# List of Figures

# List of Tables

**OBEOS – Linked Data Prototype Software User Manual**

# 1    Introduction

## 1.1    Purpose of the Document

The document describes the usage of the Web-based user interface (WebUI) of the Linked Data Prototype software, initially developed in ESA project OBEOS.

## 1.2    References

**Table 1 – References**

| Id | Reference |
|---|---|
| [EO-LD] | **EO Metadata Discovery using Linked Data**<br>OBEOS-SPB-D3100.1<br>OGC Discussion Paper referenced OGC 16-074, Version 0.2.0, 25-Mar-2016 |
| [LDP-SCF] | **OBEOS Linked Data Prototype Software Configuration File**<br>OBEOS-SA-D5220.4, Issue 2.0.0, 8-Apr-2016 |
| [LDP-SRelD] | **OBEOS Linked Data Prototype Software Release Document**<br>OBEOS-SA-D5220.7, Issue 2.0.0, 8-Apr-2016 |
| [CKAN] | **CKAN – The Open Source Data Portal Software**<br>http://ckan.org/ |
| [ISO-19115] | **Geographic information — Metadata — Part 1: Fundamentals**<br>Reference: ISO 19115-1:2014(en)<br>https://www.iso.org/obp/ui/#iso:std:53798:en |
| [INSPIRE] | **INSPIRE Glossary**<br>http://inspire.ec.europa.eu/glossary |
| [W3C-3Store] | **Triple Store**<br>https://www.w3.org/2001/sw/Europe/events/20031113-storage/positions/rusher.html |
| [GEOSS] | **Global Earth Observation System of Systems**<br>http://www.earthobservations.org/geoss.php<br>http://www.geoportal.org/web/guest/geo_home_stp |
| [GEMET] | **GEneral Multilingual Environmental Thesaurus**<br>http://www.eionet.europa.eu/gemet |
| [GCMD] | **Global Change Master Directory**<br>http://gcmd.nasa.gov/ |
| [CSCDA] | **Copernicus Space Component Data Access**<br>https://spacedata.copernicus.eu/ |

**OBEOS – Linked Data Prototype Software User Manual**

### 1.3 Glossary of Terms

**Table 2 – Glossary of Terms**

| | |
|---|---|
| Dataset | Identifiable collection of data [ISO-19115]<br><br>In the EO Community, a dataset is typically called "product".<br><br>A CKAN Dataset is a collection of data resources (such as files), together with a description and other information, at a fixed URL. Datasets are what users see when searching for data. [CKAN] |
| Dataset Series | Collection of datasets sharing the same product specification [ISO-19115]<br><br>In the EO Community, a dataset series is also called "collection" or "dataset" (in CSCDA). |
| Information | Information is "meaningful data". While it's not entirely clear what the word "meaningful" is supposed to mean in this context, dictionaries tend to say that something is meaningful if it is significant, relevant, material, valid, or important.<br>(http://www.praxiom.com/iso-definition.htm) |
| Interpretation | The assignment of meaning to abstract symbols in a logical system. |
| Knowledge | Knowledge is a collection of information and a justified belief that this information is true with a high level of certainty.<br>(http://www.praxiom.com/iso-definition.htm) |
| Metadata | Information describing spatial data sets and spatial data services and making it possible to discover, inventory and use them [INSPIRE Directive]<br><br>NOTE A more general definition provided by [ISO-19115] is "data about data". |
| Object | An object is any entity that is either conceivable or perceivable.<br><br>Objects can be real or imaginary and could be material or immaterial. Examples include products, services, systems, organizations, people, practices, procedures, processes, plans, ideas, documents, records, methods, tools, machines, technologies, techniques, and resources.<br>(http://www.praxiom.com/iso-definition.htm) |
| Ontology | Representation of a set of concepts within a domain and the relationships between those concepts [Wikipedia], [INSPIRE] |
| RDF Database | A storage tool for the persistent storage, indexing, and query access to RDF data. It is the W3C's definition of *Triple Store* applied to RDF. |
| Triple Store | A triple store is designed to store and retrieve identities that are constructed from triplex collections of strings (sequences of letters). These triplex collections represent a subject-predicate-object relationship that more or less corresponds to the definition put forth by the RDF standard. [W3C-3Store] |

The terminology used to refer to EO products and collections of EO products may be confusing as in some cases a particular EO product is considered as a dataset (in the EO community) and in some cases, a dataset identifies a collection of EO products.

The table, below, indicates the terminology in use depending on the context:

**OBEOS – Linked Data Prototype Software User Manual**

**Table 3 – Datasets and Dataset Services Terminology Mapping**

| EO Community | Dataset Series Collection | Dataset Product |
|---|---|---|
| **CKAN** | Dataset | Resource |
| **DCAT** | `dcat:Dataset` | - |
| **VoID** | `void:Dataset` | - |
| **SPARQL-SD** | `sd:Dataset` | |

## 1.4    Acronyms and Abbreviations

The following acronyms are used in this manual. It is however not necessary to be familiar with them to make use of the OBEOS Linked Data Prototype software.

| | |
|---|---|
| CKAN | Comprehensive Knowledge Archive Network |
| CSCDA | Copernicus Space Component Data Access |
| EO | Earth Observation |
| ESA | European Space Agency |
| GCMD | Global Change Master Directory |
| GEMET | GEneral Multilingual Environmental Thesaurus |
| GEOSS | Global Earth Observation System of Systems |
| HTML | HyperText Markup Language |
| HTTP(S) | HyperText Transfer Protocol (Secure) |
| ISO | International Standards Organisation |
| LDP | Linked Data Prototype |
| OBEOS | Ontology-Based Earth Observation Search |
| OGC | Open Geospatial Consortium |
| RDF | Resource Description Framework |
| URI | Uniform Resource Identifier |
| URL | Uniform Resource Locator |

# 2 Software User Manual

## 2.1 Requirements

The OBEOS Linked Data Prototype User Interface is a Web-based software application accessible on-line using a classical Internet browser. It is advised to use a recent browser to access the application. The Web browser must have cookies and JavaScript support enabled.

Because the user interface runs entirely in an Internet browser, it can run on virtually any Operating System, even mobile ones.

## 2.2 Conventions

The following textual and graphical notations are used in the remaining of this document:

1. Page and graphical component names are written in bold face, at the minimum the first time they are mentioned. For example: the **Home Page**, the **Search** field.

2. Buttons and hyperlink labels are shown in white boxes. For example: | Datasets | link, | Insert ontology concept | button.

3. Example values typed in form fields are shown in grey boxes. For example, | `landsat nepal 2015` | could be searched by the user in CKAN.

Screenshots are used in this manual to illustrate the textual explanations. The screenshots usually show fragments of the pages in order to focus on the items of interest and to avoid the need of reducing their size. For the same reason, the components provided by the portal (including the header and the footer) will not be shown in each figure.

It must be noted also that this user manual describes the CKAN portal as it appears after a fresh installation of the CKAN software and the OBEOS LDP extension. Specific CKAN installations may however look differently if they have been customized to display different page snippets or to use alternate styles (colours, fonts, icons, etc.), for example.

## 2.3 Introduction to CKAN

This section provides a short overview of the CKAN software and more particularly its main concepts and its standard user interface. It is not the purpose of this user manual to provide a detailed description of the CKAN built-in features. To obtain this information we invite you to visit the official CKAN web site and more particularly the Feature Tour page at the following address: http://ckan.org/features/.

The CKAN community describes the CKAN software as follows[1]:

> *CKAN is the world's leading open-source data portal platform.*
>
> *CKAN is a complete out-of-the-box software solution that makes data accessible and usable – by providing tools to streamline publishing, sharing, finding and using data (including storage of data and provision of robust data APIs). CKAN is aimed at data publishers (national and regional governments, companies and organizations) wanting to make their data open and available.*

---

[1] Text fragment extracted from the "About" page of CKAN.

CKAN is an open data portal, which means its main objective is to allow end-users discover and possibly access the data that have been previously registered and uploaded by a sufficiently privileged user. Although the goal is to promote the publication of open data, the portal also permits to keep data private, that is, only accessible to specific users.

CKAN is a general purpose discovery portal backed by a registry that provides tag (keyword) based and textual indexing. There is no constraint on the nature of the items registered in CKAN. The portal only imposes the organization of these items on two levels: Datasets and Resources. These concepts are further explained in section 2.3.1, below.

CKAN has the ability to harvest the data published through other CKAN instances. When this feature is used, remote records are duplicated in the local repository. The CKAN portal is not a broker, meaning it is not foreseen to allow searching dynamically in remote catalogues.

Extensions can be installed separately to add various capabilities to CKAN. Example extended features are the support for geospatial data, the preview of different types of resources, the support for an OGC CSW compliant API, and the discovery of remote catalogues using DCAT documents.

### 2.3.1 CKAN Concepts

The primary searchable data type in CKAN represents **Datasets**. CKAN datasets may represent any kind of information. They are however mostly used to represent collections of data or documents. CKAN Datasets contain **Resources**. These resources typically represent a particular document or a data file. The datasets and resources information created using the Web interface is stored and indexed in a metadata registry.

Uploaded or harvested resource files are stored in a file store. These files are accessible to the users via the search result lists. CKAN integrates **Viewers** which allow previewing in the Web interface the content of resource files, provided they are encoded in a supported format.

CKAN Datasets may be given one or more keywords, or **Tags**. These tags can be freely typed-in by the person who configures the dataset in the portal.

CKAN also comes with a built-in **User** management module. An anonymous visitor may search and access publicly available datasets. A user must be registered and authenticated to create and manage datasets and resources, or to access private datasets. User authorizations to create, edit and publish depend on their given role.

**Organization**s may be defined to organize users and datasets. Datasets that are kept private are only visible to the signed in members of their owning publishing Organization.

**Group**s may be defined to create and manage collections of datasets. Groups may for example be used to organize the datasets per type, project, theme, region or time of interest, etc.

Organizations and Groups are searchable textually in their dedicated pages (not covered in this document).

The CKAN built-in search mechanism allows searching for datasets by text fragment or tag. In addition, search result lists may be filtered by organization, group, data format, and applicable license. These filters are also referred to as "**Facets**" in the CKAN literature.

### 2.3.2 CKAN Pages Layout

All the pages of the CKAN portal are using the same layout. In particular, the following areas are displayed, from top to bottom:

- The top-most bar shows a $\boxed{\text{Log in}}$ and a $\boxed{\text{Register}}$ link that allows users to authenticate and register into the system, respectively (see Figure 1). Note: Although these features are available by default in a freshly installed CKAN instance, they may also be disabled

**OBEOS – Linked Data Prototype Software User Manual**

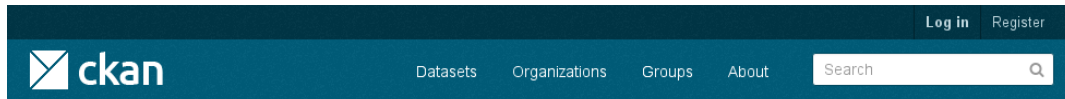independently by an administrator. In that case, one or both of these links are not displayed.



**Figure 1 – CKAN Page Header with Log in Link**

Depending on their given role, authenticated users may access the administration, creation, edition, and publication pages of CKAN. Authentication is not required for discovering the datasets and the resources, and for accessing the data hosted in CKAN.

Authenticated users see their name in this bar, as shown on Figure 2. The user name links to the user profile page. The icons displayed next to the name give access to, from left to right, the user dashboard, the user settings page, and the log-out function.
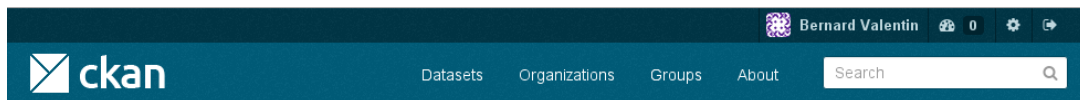


**Figure 2 – CKAN Page Header with User Name**

- The menu bar is displayed right below the login bar. This shows, from left to right:
  - The CKAN logo which, when pressed, allows navigating back to the home page.
  - A link to the Datasets discovery page.
  - A link to the Organizations discovery page.
  - A link to the Groups discovery page.
  - A link to the About page, which gives a short introduction to CKAN and provides links to the official web site.
  - A 1-field Dataset search form.
- The area located below the menu bar is the page body. Its content depends on the visited page. For example, in the Home page, the main area includes a dataset search form and highlights the most popular groups, organizations, and datasets.
- The page footer, shown in Figure 3, provides internal and external links as well as a language selector.
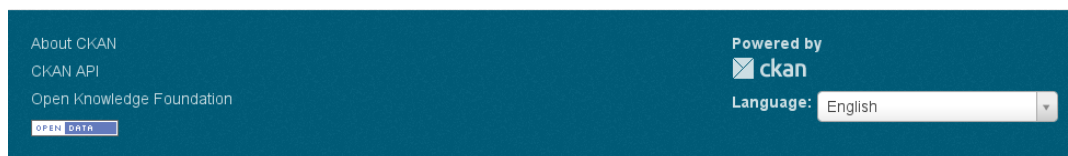


**Figure 3 – CKAN Page Footer**

The following section gives an overview of the CKAN typical search and access page flow.

**OBEOS – Linked Data Prototype Software User Manual**

### 2.3.3    CKAN Discovery and Access Pages

Figure 4, below, depicts the different pages included in CKAN to support the search and access of datasets and resources. The directed arrows show the main navigation links that are available between the pages. The green/blue boxes represent the pages and components pre-existing in the CKAN software. The orange/yellow boxes represent customizations and additions brought by the OBEOS Linked Data Prototype extension.
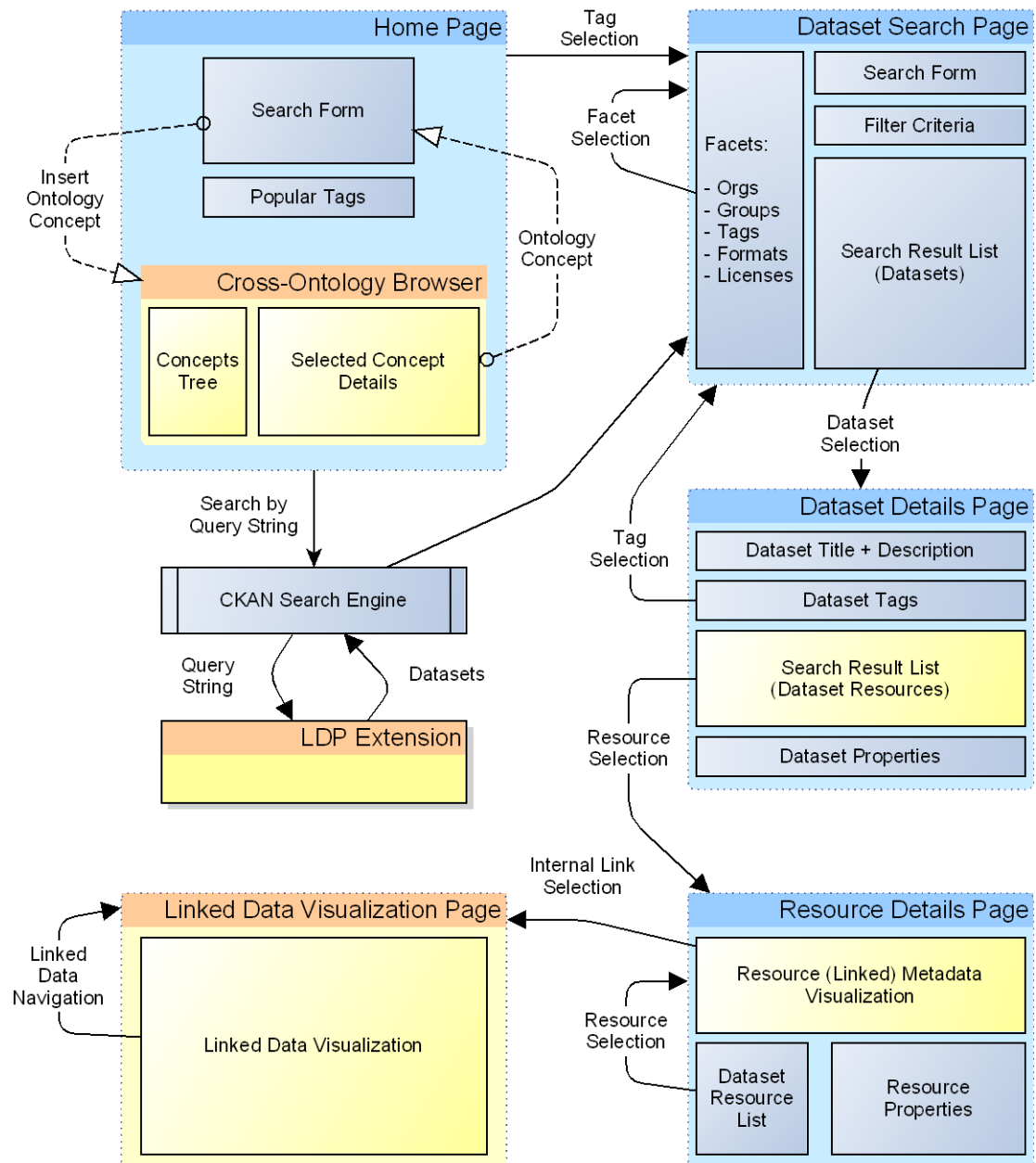


**Figure 4 – CKAN Search and Access Page Flow**

**OBEOS – Linked Data Prototype Software User Manual**

The constituents of each page and the main navigational flow are as follows:

1.  The **Home Page** is the first page displayed when navigating to CKAN. Its main constituents are a dataset search form and the popular groups, organizations and datasets.

    The Linked Data Prototype (LDP) extension adds an ⸢Insert ontology concept⸣ button that allows revealing the **Ontology Browser**.

    The **Home Page** is further described in section 2.5.2 and the **Ontology Browser** is described in section 2.5.3.

2.  The **Dataset Search Page** displays a list of datasets. It is in particular accessible from the **Home Page**, when a textual or tag-based search is performed, or when ⸢Datasets⸣ is selected in the menu bar. In the first case, the list contains the datasets matching the search query. In the latter case, the list shows the most relevant (popular) datasets.

    Controls are displayed in the left margin which allows adding and removing organizations, groups, tags, data format and licenses terms in the search criteria.

    Click on one of the listed datasets to navigate to its **Dataset Details Page**.

    The **Dataset Search Page** is further described in section 2.5.4.

3.  The **Dataset Details Page** shows, among other information, the properties of the selected dataset, the groups it belongs to, the history of the dataset record, and the list of its resources that are matching the original query.

    Clicking on a resource entry allows accessing its **Resource Details Page**.

    The **Dataset Details Page** is described in section 2.5.5.

4.  The **Resource Details Page** shows the properties of the selected resource. If the resource type is supported by CKAN (or one of its extensions), a Web-based visualization of the resource data may be revealed inside the page. For example CSV files may be visualized in tables or graphs, and geospatial features (e.g. GeoJSON documents) may be visualized on interactive maps.

    The OBEOS LDP extension not only adds the possibility to visualize the resources data encoded in RDF (Linked Data) but also allows navigating to related nodes by following the RDF links.

    The **Resource Details Page** is described in section 2.5.6.

### 2.4 Introduction to the OBEOS Linked Data Prototype

The OBEOS Linked Data Prototype extension for CKAN extends the capabilities of the portal as follows:

**Configuration and discovery of remote EO catalogues**

The Linked Data Prototype brings the possibility to register remote catalogues and datastores whose capabilities are then discovered automatically. This mechanism requires the remote services to publish their capabilities and content in a specific manner, namely using the DCAT, OpenSearch and SPARQL protocols.

In the context of OBEOS, these capabilities are applied to Strabon, an RDF datastore that exposes a SPARQL interface, and the FedEO Gateway that exposes an OpenSearch endpoint.

**Selection of concepts in ontologies**

By default, CKAN searches for datasets using the terms entered by the users. The matching datasets are thus the ones that contain one or more of the search terms in their textual properties, such as the title and the description text.

The Linked Data Prototype allows the users to specify terms that are automatically resolved into low-level search criteria. These search criteria are then used to augment the search query.

The terms that have a special meaning are provided as concepts in ontologies. These concepts may be selected using an embedded Ontology Navigation tool.

**Advanced user search query interpretation**

In additional to the ontology concepts, the query resolution mechanism mentioned above is also capable of recognizing toponyms and temporal expressions expressed in natural language.

The result of a user query analysis may thus contain spatial, temporal, and other data type specific constraints. It is the combination of all these constraints that are used by the Linked Data Prototype extension to search for matching datasets and resources.

**Discovery of EO dataset series (as CKAN datasets) and EO datasets (as CKAN resources)**

The result of the user query analysis is used to search for EO dataset series and EO datasets registered in the remote catalogues and datastores mentioned above.

Both SPARQL and OpenSearch endpoints are supported. The remote services and the data they contain are however required to be compliant with the OGC Discussion Paper "EO Metadata Discovery using Linked Data" (see [EO-LD]).

Retrieved EO dataset series are presented as CKAN datasets in the user interface, and EO datasets are presented as CKAN resources.

**Visualization and navigation in Linked Data documents**

The EO dataset series and EO dataset metadata obtained from the remote services are encoded as Linked Data.

The Linked Data Prototype allows visualizing this data in a Web-based user-readable manner. Moreover, the URIs of the linked entities are displayed as links, allowing to navigate to these objects (internal navigation). This allows, for example, navigating from an EO dataset to its parent dataset series, or to its related platform or instrument entry.

When a subject URI is also a valid URL (i.e. it indicates an Internet protocol such as "`http`", "`https`", or "`ftp`"), this URL is also made available for navigating to the new target (external navigation).

### 2.5 End-to-End Discovery Scenario

#### 2.5.1 Introduction

This section describes in details the different pages and components that are involved by a typical search in the OBEOS Linked Data Prototype application. It is decomposed into the following steps:

1. The user accesses the **Home Page** and inspects the dataset search form. This includes a new Insert ontology concept button.

**OBEOS – Linked Data Prototype Software User Manual**

2. The user has clicked on this button, which reveals the **Cross-Ontology Browser**, embedded in the home page. A concept is selected and inserted in the search form and a search is initiated.

3. The list of matching CKAN datasets is displayed on the **Dataset Search Page**.

4. The user selects one of these datasets, which reveals the **Dataset Details Page**. This page provides the dataset properties as well as the list of its resources.

5. The user selects one of these resources, which reveals the **Resources Details Page**. This page provides the resource details, the visualization of the metadata encoded as Linked Data, and links for navigating in the available data.

6. The user has clicked on one of the internal links available in the Linked Data, which reveals the **Linked Data Visualization and Navigation Page**. In this page, the user may continue his exploration of the Linked Data cloud.

### 2.5.2    Step 1: Home Page and Search Form

The CKAN **Home Page** appears when you navigate to the OBEOS Linked Data Prototype Web site.

Figure 5, below, shows the Home Page as it appears after a fresh installation of CKAN and the OBEOS extension. CKAN instances may however be customized to display different panels or to have a different style.
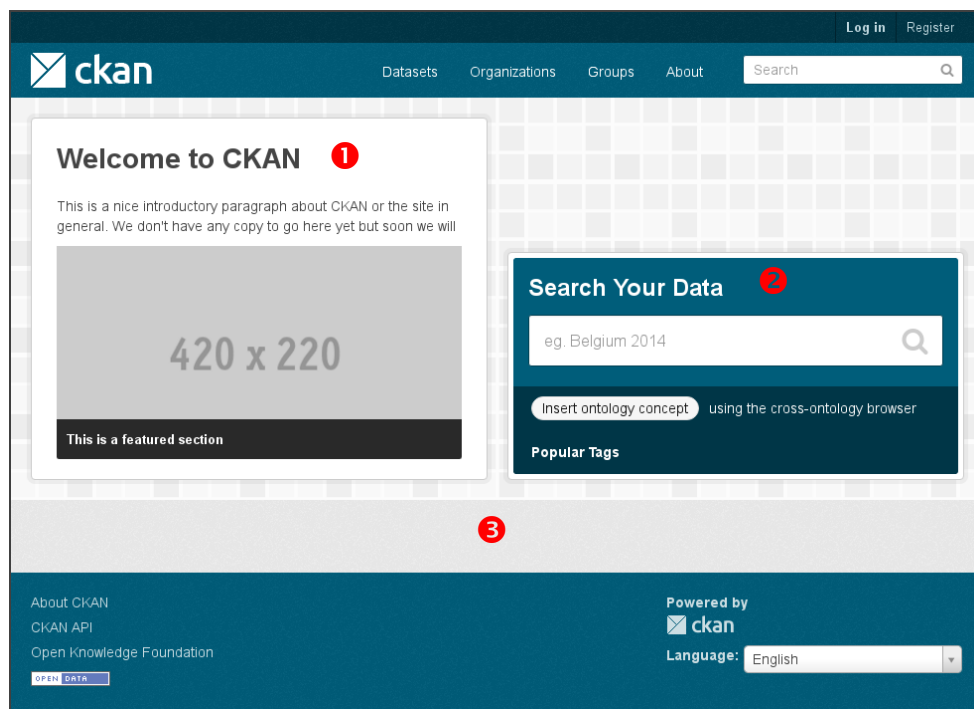


**Figure 5 – CKAN Home Page**

**OBEOS – Linked Data Prototype Software User Manual**

Besides the common panels already described in section 2.3.2, and situated at the top and the bottom of each page, the CKAN Home Page contains the following sections:

❶ A *Welcome* notification and a featured selection (empty until some content is added in CKAN) are displayed on the left side.

❷ The CKAN Dataset search form is located on the right side. This search form allows you to enter a query string expressing search criteria.

By default, the CKAN search form only allows entering a query string manually. The Ontology Browser extension adds an Insert ontology concept button to the form, as shown on Figure 6.
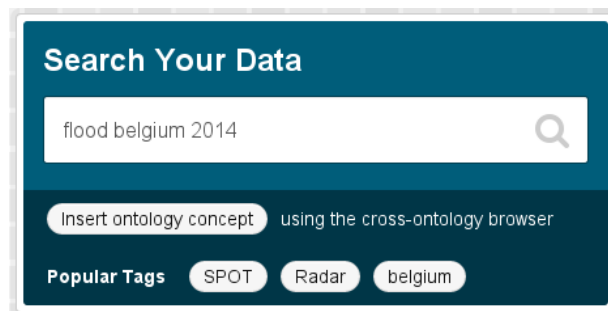


**Figure 6 – CKAN Dataset Search Form**

The "Popular Tags" control lists shows the tags that are the most popular among all the tags used by the publishers to characterize their datasets.

❸ The page wide area situated below the search form is used to show the most popular datasets per organization and group. It is also used by the Linked Data Prototype to embed the Cross-Ontology Browser revealed when the Insert ontology concept button is pressed.

### 2.5.3 Step 2: Ontology Navigation and Concept Selection

The Ontology Browser tool, represented in Figure 7 gives you the possibility to navigate in the supported ontologies and to insert selected semantic terms in the search form. It allows exploiting the knowledge contained in the ontologies by providing relevant information for each concept and highlighting the connections between different (but related) concepts belonging to the same or other ontologies.

The Browser supports the following four ontologies:

• ESA's CSCDA, the Copernicus Space Component Data Access ontology.

• GEOSS, the Global Earth Observation System of Systems ontology.

• GEMET, the GEneral Multilingual Environmental Thesaurus.

• NASA's GCMD, the Global Change Master Directory.

The Cross-Ontology Browser is an embedded Web-based component. It is hidden by default and revealed by selecting the Insert ontology concept button located in the search form on the Home Page (see Figure 6).

The following sections describe the features of the Cross-Ontology Browser.

**OBEOS – Linked Data Prototype Software User Manual**

### 2.5.3.1 Ontology Concepts Discovery

Figure 7, represents the Cross-Ontology Browser tool. It shows the details of the "Hydrological Disaster" concept selected in the CSCDA (ESA's Multi-Domain Thesaurus) ontology.
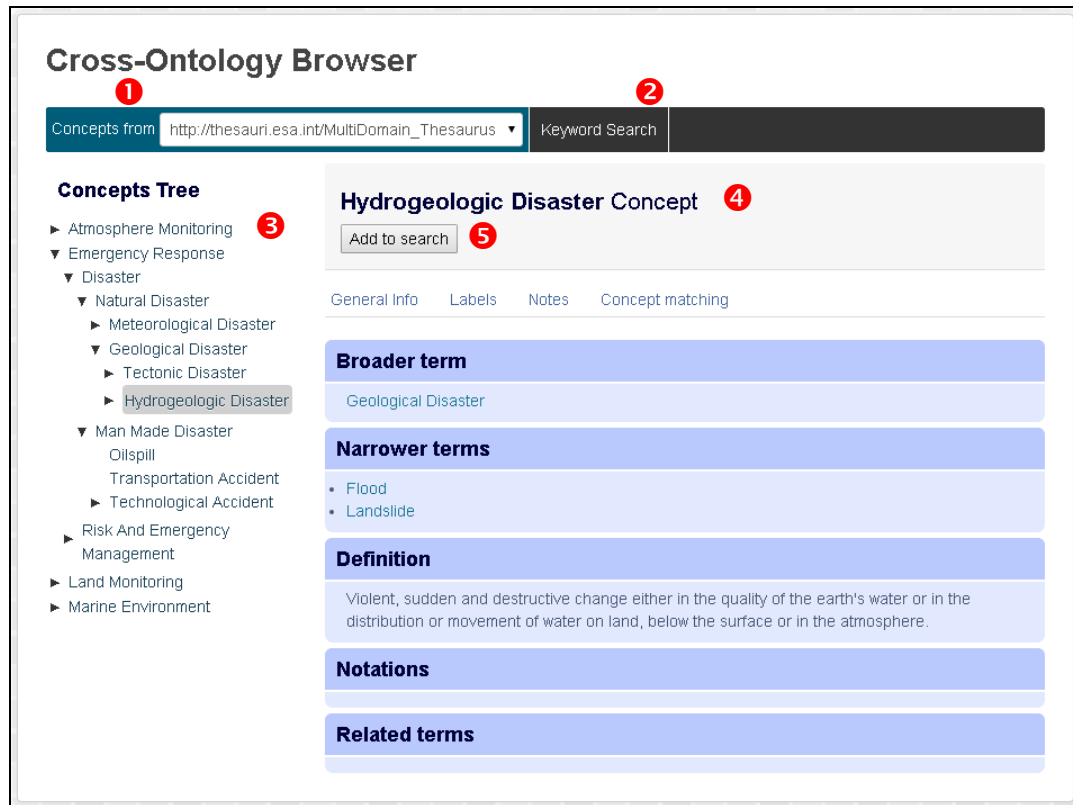


**Figure 7 – Cross-Ontology Browser: Hierarchical Concept Browsing**

The tool contains the following parts:

Two tabs are displayed right below the title bar:

❶ • The first tab allows navigating in the ontologies and inspecting the properties of the selected concept. This tab includes a drop-down list that allows jumping to one of the supported ontologies. This list is not clickable if the tab is not selected. Selecting an entry in the list automatically updates the hierarchical representation, showing the top-most concepts of the selected ontology.

❷ • The second tab allows searching for concepts by name (label) in any of the supported ontologies (as explained in section 2.5.3.2, below).

❸ The hierarchical representation of the concepts belonging to the selected ontology is displayed on the left side. The top-level concepts give an overview of the different thematic branches of an ontology. The concepts which have child (narrower) concepts are shown with a small triangle. Click on this triangle to expand or collapse the sub-trees.

**OBEOS – Linked Data Prototype Software User Manual**

❹ The right side is occupied by the concept viewer, which shows all the available information about the selected concept, including the concept labels and definition, broader, narrower and related concepts, and so on.

From the selected concept, the viewer allows navigating to other related concepts, either from the same or from different ontologies, in particular using the | Concept matching | panel.

❺ The | Add to search | button provided below the context label copies the selected concept into the search query. Concepts already present in the search string are not inserted again.

### 2.5.3.2 Ontology Concepts Textual Search

In case the user is unable to find a suitable concept by navigating the ontologies, the Cross-Ontology Browser provides also a keyword search mechanism. The keyword search enables the users to find concepts in a quicker way, since apart from searching a match between a keyword and concept labels, it searches also other types of annotations, such as definitions. Pagination is used to group the results in pages of ten results, so it is easier for users to navigate within the search results.

To access the concept search form, click on the | Keyword Search | tab at the top.

The search form contains a single input field and a | Search | button. Results of the search are displayed below the form. Ten concepts are shown at a time. Navigate to the previous or next page using the | « | and | » | buttons.

Figure 8, below, shows the result of using the concept search feature of the tool. In this example, the keyword "flood" has been searched for and the concept "Flood Alert" has been selected.

**OBEOS – Linked Data Prototype Software User Manual**



**Figure 8 – Cross-Ontology Browser: Searching for Concepts by Label**

### 2.5.4 Step 3: Dataset Search

The actual search in CKAN is triggered by hitting the "Enter" key in the search input field, or by clicking on the magnifier icon located in the same field.

The default behaviour of CKAN is to perform a textual search of the words in its internal catalogue. This behaviour remains as the OBEOS extension performs its external search without preventing CKAN to perform its internal search. Search results are merged before being rendered on the screen. The search result list thus shows both internal and external entries in a seamless manner.

Figure 9 shows the body of the **CKAN Dataset Search Page**. This includes the following components:

❶ The breadcrumb string shown at the top indicates that the current page is meant to display a collection of CKAN datasets, possibly matching filtering criteria: 🏠 **/ Datasets**

❷ The left margin shows the property types that may be used to filter the datasets. These are, by default: Organizations, Groups, Tags, Formats and Licenses. The lists of filtering values are created dynamically from the entries already present in the CKAN repository. Only the values applicable to the filtered list are displayed. The page is refreshed each time the user selects or deselects a value in the filter.

After a fresh install, the repository is empty and thus CKAN has no values to propose to filter the datasets.

The figure depicts a situation that should not happen in a stock version of CKAN: no filtering criteria are proposed despite the datasets list is not empty. The reason here is the CKAN catalogue is empty and the listed datasets have all been discovered in remote catalogues.

**OBEOS – Linked Data Prototype Software User Manual**

The main area shows, from top to bottom:

① A search form shows the query string that has been applied to search for datasets. The only input field allows issuing a refined or a new query without moving back to the Home page.

It must be noted that when this form is used to initiate a new search, CKAN takes only the new string into account and forgets about the filter criteria selected in the left margin.

② The amount of matches and the search string, if any, are provided below the search field.

At the right side, a drop-down menu allows choosing the sorting criteria. Relevance (default), Name Ascending , Name Descending , and Last Modified are proposed. The selection of a new sort criterion automatically refreshes the page.

③ The following selected information is displayed for each matching CKAN dataset:

- the title, displayed in bold face,
- the description, truncated to about two lines if too long,
- the format of its resources (consolidated list),
- the associated tags, if any.

④ If more than 20 datasets (default value) match the filter criteria, the list is cut into pages of 20 entries. Navigation controls provided at the bottom of the page allow navigating to the previous and next pages, or directly accessing a particular page.

Figure 9 shows the second page of a search result list. Because the result list contains 25 datasets, the second page contains le last 5 entries.
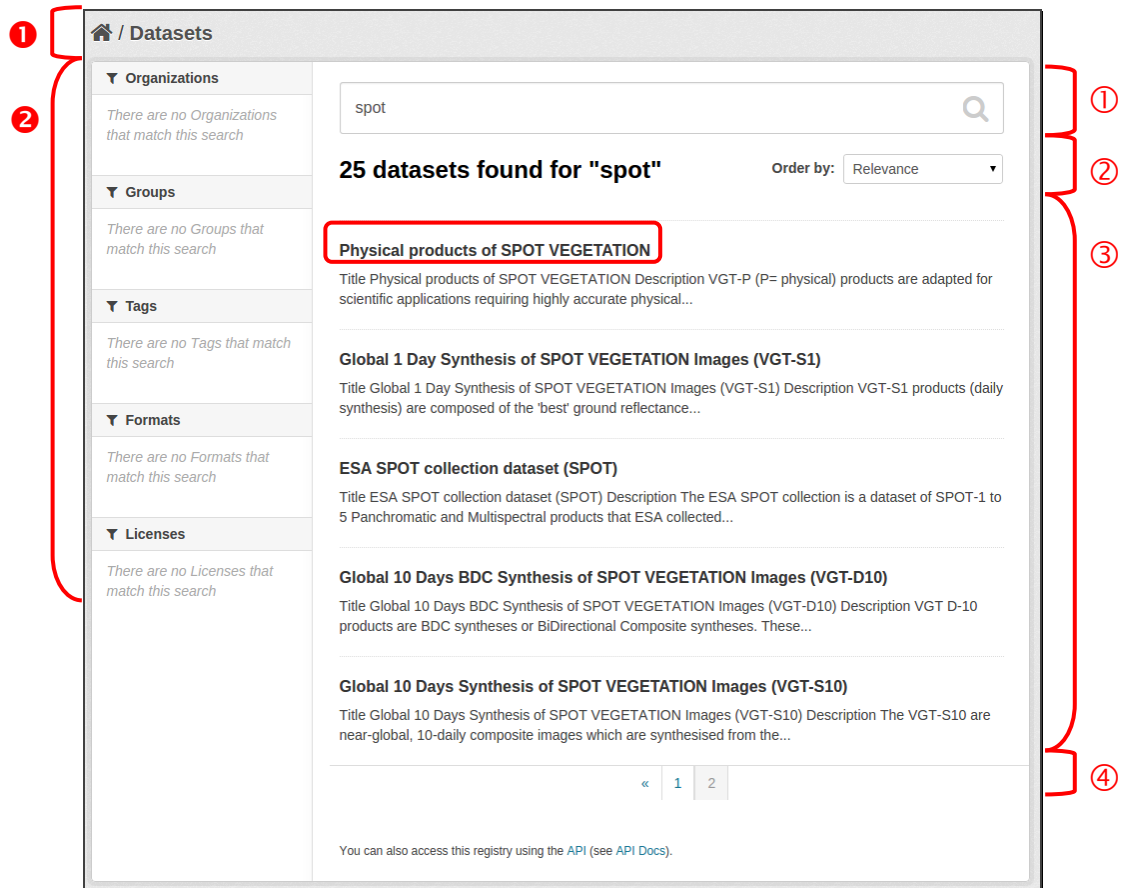
**Figure 9 – CKAN Dataset Search Page**

To obtain the properties of a given dataset and identify the resources that belong to it, click on the dataset title (e.g. the red rectangle in Figure 9). The dataset properties and resources are then displayed in the **Dataset Details Page**, described in the following section.

**OBEOS – Linked Data Prototype Software User Manual**

### 2.5.5  Step 4: Dataset Details

The **Dataset Details Page** provides detailed information about a given CKAN Dataset. In the OBEOS Linked Data Prototype software, a CKAN Dataset represent an EO Dataset Series whose metadata have been obtained from a remote catalogue.

The page shows the dataset properties and the list of its resources (see Figure 10, below). Its constituents are as follows:

❶  At the top, the breadcrumb string is extended with the title of the selected dataset, truncated to about 30 characters:  🏠 / Datasets / **LANDSAT-8 Level 1T (precision …**

The left margin shows, from top to bottom:

❷  The full title of the current dataset.

❸  Links to social web sites that facilitate the dissemination of the dataset.

❹  The license(s) applicable to the content of the dataset resources.

The main area shows, from top to bottom:

①  A row of tabs that allow accessing different types of information, namely:

  - Dataset corresponds to the content shown in Figure 10,
  - Groups lists the groups the dataset belongs to, and
  - Activity Stream shows the dataset edition history.

②  The full title and description text of the dataset.

③  The resources belonging to the current dataset and <u>matching the original user query</u>. Figure 10 shows that three resources comply with search query **landsat nepal** . A search for **landsat china** , for example, would have resulted in a list of 25 resources displayed in the exact same page.

The resources table is paginated, showing pages of 10 entries by default. Controls allow changing the page size (the Show list proposes 10, 25, 50, or 100 entries), navigating through the pages, and filtering the resources (using the Search field).

Each dataset resource represents a product of the current dataset. The table includes the following information:
  - The **resource** identifier,
  - The sensing **start** and **end** time,
  - A **thumbnail** image, if available.

The identifier, start time and end time columns may be sorted.

Clicking on a resource identifier (see for example the red rectangle in Figure 10) opens the dedicated **Resource Details Page** (see section 2.5.6, below).

④  Additional information about the dataset organized in a two-column Field / Value table.

**OBEOS – Linked Data Prototype Software User Manual**



**Figure 10 – CKAN Dataset Details Page**

**OBEOS – Linked Data Prototype Software User Manual**

To obtain the properties of a given dataset resource, click on the resource identifier (e.g. the red rectangle in Figure 10). The resource properties are then displayed in the **Resource Details Page**, described in following section.

### 2.5.6 Step 5: Resource Details

The **Resource Details Page** provides detailed information about a given CKAN Dataset Resource. In the OBEOS Linked Data Prototype software, a CKAN Dataset Resource represents an EO Dataset (or Product) whose metadata have been obtained from a remote catalogue.

The page shows the dataset resource properties as well as a tabular representation of the record originally encoded in Linked Data (see Figure 11, below). The page contains the following elements, from top to bottom:

❶ At the top, the breadcrumb string includes the title of the selected dataset and the title of the selected resource. Titles are truncated if they are more than 30 characters long:

🏠 / Datasets / LANDSAT-8 Level 1T (precision … / **LC81430392015182LGN00**

❷ The title and description of the current dataset resource.

❸ The visualization of the resource properties and relations obtained as Linked Data (RDF) from the remote catalogue. The information is displayed by means of nested tables. The whole structure is encapsulated in a scrollable frame.

The example given in the figure shows the metadata of a Landsat-8 optical image intersecting Nepal boundaries. The metadata includes information about the time, the orbit, the processing (incl. processing centre, processor version, processing level), the cloud coverage, the illumination conditions, etc. Images (incl. icons and thumbnails) that can be identified in the properties are displayed in-line.

The available information varies depending on the mission and the product type.

The linked entities are displayed as clickable hyperlinks leading to the **Linked Data Navigation Page**. This page allows navigating in the Linked Data graph in a generic manner (see section 2.5.7, below).

The figure shows links to several data types: Observation, TemporalObject (phenomenon time), and Process (procedure). In the invisible section (reachable using the scrollbar), the document also provides links to, e.g., the platform (LANDSAT-8), the Instrument (OLI_TIRS) and the Parent Identifier (that is, the EO Dataset Series, or Collection: L1T).

The area located below the visualization panel is divided in two columns:

❹ On the left side, one can find the list of the dataset resources matching the original user query. This list provides a direct access to the description of the other resources removing the need to navigate back to the dataset page beforehand.

❺ Links to social web sites are provided below the resources list. These links facilitate the dissemination of the selected dataset resource.

❻ Additional information about the dataset resource is displayed on the right side by means of a two-column Field / Value table.

**OBEOS – Linked Data Prototype Software User Manual**



**Figure 11 – CKAN Resource Details Page**

**OBEOS – Linked Data Prototype Software User Manual**

### 2.5.7 Step 6: Navigation in Linked Data

The OBEOS Linked Data Prototype extends CKAN with a **Linked Data Visualization and Navigation Page** (RDF Viewer, in short). This displays information about an entity obtained as Linked Data (RDF graph) and allows navigating to the linked entities via hyperlinks.

Each entity is uniquely identified with a URI. This URI allows obtaining the entity information from the data source. It is common, and even recommended, to use valid and resolvable URLs as URIs. These URLs are then expected to refer to resources accessible on the Internet that represent the entity. These resources could be, for example, human readable Web pages, machine readable documents, or API calls (e.g. resources obtained through REST interfaces).

The entity properties are displayed in a two-column Property/Value table. When two or more values are linked through the same property type, these values are listed in a single cell.

Linked Data graphs may contain anonymous entities, referred to as "blank nodes". The standard does not allow navigating through blank nodes. For that reason, blank nodes are automatically detected and their own properties and links displayed by means of nested tables.

Entities identified on the page (e.g. by means of a title or a label) are displayed as traditional HTML hyperlinks. Clicking on one of these links allows navigating to the corresponding entity. If data about the target entity can be obtained as Linked Data, the page is refreshed and the data is visualized in a new table (possibly containing nested tables). Otherwise, the target Web page is displayed, replacing the Linked Data Visualization and Navigation page.

The Linked Data Prototype tries to obtain the target entity data in the configured external RDF databases, using the entity URI as selector. In addition, if the URI is a well-formatted URL, the LDP also retrieves the resource referred to by that URL. If the resource is Linked Data, the collected information is merged and presented as a single graph.

As a concrete example, let us consider a configured RDF Database that contains Landsat-8 products metadata. Each of these products is linked to an entity that represents the Landsat-8 platform. The URI of the platform entity is the URL of its related concept in the GCMD/KMS ontology[2].

On one hand, the platform entity exists in the RDF Database, is given properties (e.g. linked literal values), and is linked to other entities (including the Landsat-8 products).

On the other hand, the platform entity URI is a well-formed URL and the URL refers to an RDF graph served by GCMD. This is thus another means to obtain information about the Landsat-8 platform.

Should the ⊞Landsat-8⊞ link be clicked in the RDF Viewer, the LDP will:

- fetch Linked Data from the RDF Database (URI-based query),
- fetch Linked Data from GCMD (URL-based query),
- merge the two graphs, and
- render and display the consolidated graph by means of Property/Value tables.

The above mechanism also takes place if Linked Data may be obtained from one of the two sources only.

However, if no Linked Data can be obtained about the entity and the entity URI is a well-formed URL, the Web browser is re-directed to that URL. This allows navigating from the Linked Data Visualization and Navigation Page to, e.g., a human readable HTML page.

---

[2] http://gcmdservices.gsfc.nasa.gov/kms/concept/13e3a08a-0d28-4e3f-a306-a20d9fb4fff8

**OBEOS – Linked Data Prototype Software User Manual**

Figure 12, below, represents the Linked Data Visualization and Navigation page applied on the Landsat-8 L1GT Dataset Series. The page shows the dataset series properties and its relations with other entities, such as the Landsat-8 platform and the OLI_TIRS instrument.



**Figure 12 – RDF Viewer: Landsat L1Gt Dataset Series (Truncated)**

**OBEOS – Linked Data Prototype Software User Manual**

Figure 13 shows the properties of the Landsat-8 platform entry rendered by the RDF Viewer. The page integrates information originating from both the RDF Database configured in CKAN and the GCMD portal referred to by the platform identifier (i.e. the subject URI which is also a valid URL).

Information obtained from the GCMD portal includes the platform label (LANDSAT-8), the full-text description, the identification of its parent (broader) and its scheme (platforms).

Information obtained from the RDF Database is the Datasets and the Resources that are semantically linked to the Landsat-8 platform ("is platform of" predicate): L1GT, L1T, etc.



**Figure 13 – RDF Viewer: Landsat-8 Platform**