

# 1. Module and functional design

The main modules of scientific data resources network open service platform include: data resources display, data resources search, public support service tools, vocabulary specification, application demonstration.

Table 1.1 Module Introduction

module name	module covered
Data Resource Display	Display dataset, resource entity list and details
Data Resource Search	Search data sets, resource entities
utility support service tool	Show the introduction of common support service tools and jump links.
glossary specification	Display the common vocabulary specification introduction and jump links.
application demonstration	Display application demonstration node information, application demonstration data set and entity information.

## 2. prototyping

### 2.1 Data Resource Display

Data resources display includes home page and details page.

According to the layout from top to bottom, the home page displays the following contents: title, navigation bar, search box, platform statistical information, knowledge atlas of disciplinary datasets and dataset list. See below.

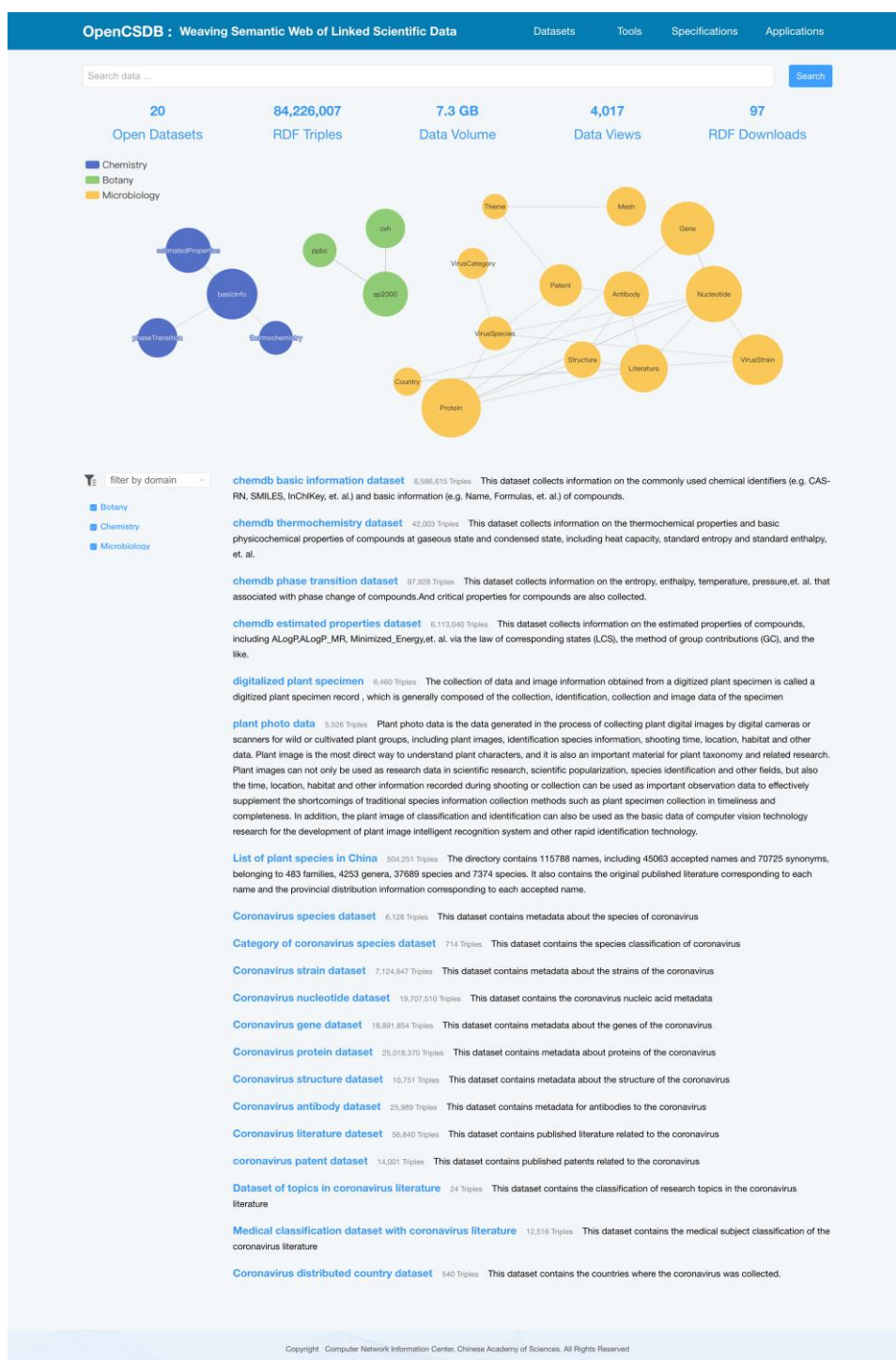


Figure 1 Home Page

Platform statistics are statistics for Open Datasets, RDF Triples, Data Volume, Data Views, RDF Downloads.

The data set list displays the title, RDF data volume and basic introduction. The subject area or publishing organization can be filtered and displayed through the check box on the left.

The knowledge graph of disciplinary dataset shows the association relationship of all datasets in the site. The node size represents the volume of the dataset, and the distance between nodes represents the close association between the two datasets.

Click the four different tabs in the navigation bar, and the responses are as follows: Datasets navigate to the dataset list page, Tools, Specifications, Applications enter the tool set page, specification page, application demonstration page respectively.

Enter the search term and click "Search" to enter the data resource search page, displaying the data search list. See for details.

Click any title on the dataset list page at the bottom of the home page to enter the dataset details page. If you click on the first data set, you will go to the following page.

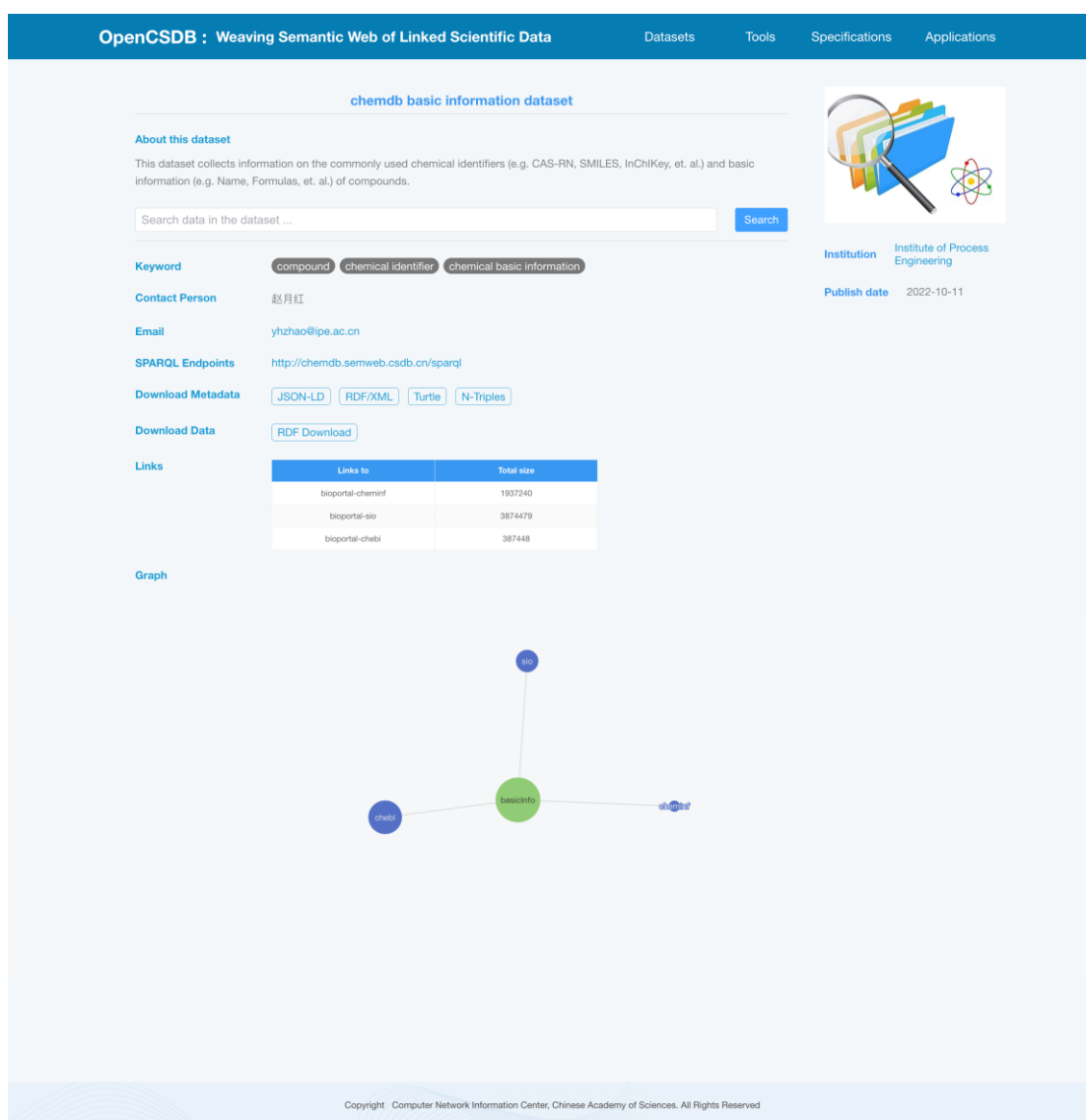


Figure 2 Data Set Detail Page

The dataset details page displays the following information from top to bottom: dataset title, dataset introduction, keywords, contacts and contact information, SPARQL Endpoints, metadata download link, data download link, etc. Among them, metadata download link provides the following four ways: JSON-LD, RDF/XML, Turtle, N-Triples; data download link provides RDF download way.

In the details page of the dataset, enter the search term

and click "Search" to enter the data resource search page and display the data search list. See 2.2 for details.

The data details page provides the number of associations between this dataset and other datasets. For example, this dataset has 1937240, 3874479 and 387448 associations with international open data sources bioportal-cheminf, bioportal-sio and bioportal-chebi in LOD, respectively.

At the bottom of the details page, the association map of this dataset is displayed. By clicking, you can enter the details page of this dataset separately.

## **2.2 Data Resource Search**

On the homepage of the portal, enter the search term in the search box, click "Search" to enter the data resource search page and display the data search list.

If the search term is "water", the following data entity list can be retrieved, and all the entity contents related to "water" such as compounds, microorganisms and plants can be obtained across three different subject fields. as shown.

The screenshot displays the OpenCSDB interface with the search term 'water'. The left sidebar shows domain filters: Botany, Chemistry, and Microbiology. The main content area lists search results with the following attributes: label name, RDF:type, associated resource, organization domain, and data set.

Label Name	RDF:type	Associated Resource	Organization Domain	Data Set
water	compound:7732-18-5	CHEBI:24431	Institute of Process Engineering	chemdb basic information dataset
Water	Mesh:D014867	ncov:Mesh	Institute of Microbiology	Medical classification dataset with coronavirus literature
Waste Water	Mesh:D062065	ncov:Mesh	Institute of Microbiology	Medical classification dataset with coronavirus literature
Fresh Water	Mesh:D005618	ncov:Mesh	Institute of Microbiology	Medical classification dataset with coronavirus literature
Water Supply	Mesh:D014881	ncov:Mesh	Institute of Microbiology	Medical classification dataset with coronavirus literature
descriptor:water solubility	CHEMINF:000258		Institute of Process Engineering	chemdb estimated properties dataset
descriptor:water InChIKey	CHEMINF:000399		Institute of Process Engineering	chemdb basic information dataset
descriptor:water IUPAC Name	CHEMINF:000382		Institute of Process Engineering	chemdb basic information dataset

Figure 3 Resource Search List Page

The data list displays the following attributes of the retrieved data resource entity: label name, RDF:type, associated resource, organization domain, and data set.

Click URI to jump to URI resolution display page of external data source, such as CHEMBI: 24431, etc.

Click institution to enter the specific analysis place of institution vocabulary, such as [www.example.com](http://semweb.casdc.cn/resource/casorgs#IM).

Click on the dataset to enter the dataset details page.

Click the title of any search result to enter the entity resource details page. For example, click on the first search

result to get the following page, as shown in.

OpenCSDB : Weaving Semantic Web of Linked Scientific Data

DatasetsToolsSpecificationsApplications

water compound:7732-18-5

chemdb basic information dataset

SIO-000008

descriptor:water\_CAS\_registry\_number

rdf:type (type) CHEMINF:000446 (CAS registry number)

SIO-000300 (has value) 7732-18-5

SIO-000008

descriptor:water\_IUPAC\_Name

rdf:type (type) CHEMINF:000382 (IUPAC Name)

SIO-000300 (has value) water

SIO-000008

descriptor:water\_InChIKey

rdf:type (type) CHEMINF:000389 (InChIKey)

SIO-000300 (has value) XLYOFNQVPUJNP-UHFFFAOYSA-N

SIO-000008

descriptor:water\_molecular\_formula

rdf:type (type) CHEMINF:000395 (molecular formula)

SIO-000300 (has value) H2O

SIO-000008

descriptor:water\_molecular\_weight

rdf:type (type) CHEMINF:000334 (molecular weight)

SIO-000300 (has value) 17.9994

rdf:type (type) CHEBI:24431 (chemical entity)

rdfs:label (label) water

chemdb basic information dataset



Institution

Institute of Process Engineering

Publish date

2022-10-11

chemdb estimated properties dataset

SIO-000008

descriptor:water\_formal\_charge

rdf:type (type) CHEMINF:000118 (formal charge descriptor)

SIO-000300 (has value) 0

SIO-000008

descriptor:water\_hydrogen\_bond\_acceptor\_count

rdf:type (type) CHEMINF:000245 (hydrogen bond acceptor count)

SIO-000300 (has value) 1

SIO-000008

descriptor:water\_hydrogen\_bond\_donor\_count

rdf:type (type) CHEMINF:000244 (hydrogen bond donor count)

SIO-000300 (has value) 1

SIO-000008

descriptor:water\_solubility

rdf:type (type) CHEMINF:000258 (solubility)

SIO-000300 (has value) 1.472

SIO-000008

descriptor:water\_surface\_area

rdf:type (type) CHEMINF:000229 (surface area descriptor)

SIO-000300 (has value) 11.3

chemdb thermochemistry dataset

SIO-000008

descriptor:7732-18-5\_formation\_entropy\_8720

rdf:type (type) CHEMINF:000060 (formation entropy descriptor)

SIO-000300 (has value) 69.95

SIO-000221 (has unit) J/(mol\*K)

SIO-000008

descriptor:water\_formation\_entropy\_19713

rdf:type (type) CHEMINF:000241 (formation enthalpy descriptor)

SIO-000300 (has value) -241.83

SIO-000221 (has unit) kJ/mol

SIO-000008

descriptor:water\_formation\_entropy\_19714

rdf:type (type) CHEMINF:000241 (formation enthalpy descriptor)

SIO-000300 (has value) -241.826

SIO-000221 (has unit) kJ/mol

SIO-000008

descriptor:water\_formation\_entropy\_19715

rdf:type (type) CHEMINF:000241 (formation enthalpy descriptor)

SIO-000300 (has value) -265.83

SIO-000221 (has unit) kJ/mol

SIO-000008

descriptor:water\_formation\_entropy\_19716

rdf:type (type) CHEMINF:000241 (formation enthalpy descriptor)

SIO-000300 (has value) -265.83

SIO-000221 (has unit) kJ/mol

chemdb phase transition dataset

SIO-000008

descriptor:water\_boiling\_point\_15039

rdf:type (type) CHEMINF:000257 (boiling point descriptor)

SIO-000300 (has value) 373.17

SIO-000221 (has unit) K

SIO-000008

descriptor:water\_boiling\_point\_5106

rdf:type (type) CHEMINF:000257 (boiling point descriptor)

SIO-000300 (has value) 373.17

SIO-000221 (has unit) K

SIO-000008

descriptor:water\_freezing\_point\_5673

rdf:type (type) CHEMINF:000432 (freezing point descriptor)

SIO-000300 (has value) 273.2

SIO-000221 (has unit) K

SIO-000008

descriptor:water\_melting\_point\_5673

rdf:type (type) CHEMINF:000256 (melting point descriptor)

SIO-000300 (has value) 273.15

SIO-000221 (has unit) K

Graph

chemdb phase transition dataset

chemdb estimated properties dataset

chemdb basic information dataset

chemdb thermochemistry dataset



Copyright Computer Network Information Center, Chinese Academy of Sciences. All Rights Reserved

Figure 4 Entity List Page

The resource details page displays the RDF detailed analysis of data resources and the association relationship graph respectively.

As shown in the figure above, RDF triples are displayed for the content details of resources in the three datasets, and secondary parsing is performed for some attributes.

On the right is information about the organization to which the data entity belongs.

## **2.3Utility Support Service Tools**

Click "Tools" on the navigation bar of the platform to enter the toolsets page of the Scientific Data Association Web Portal.

This page shows two common support service tools of scientific data association network system, namely scientific data ontology publishing tool and scientific data thesaurus publishing tool.



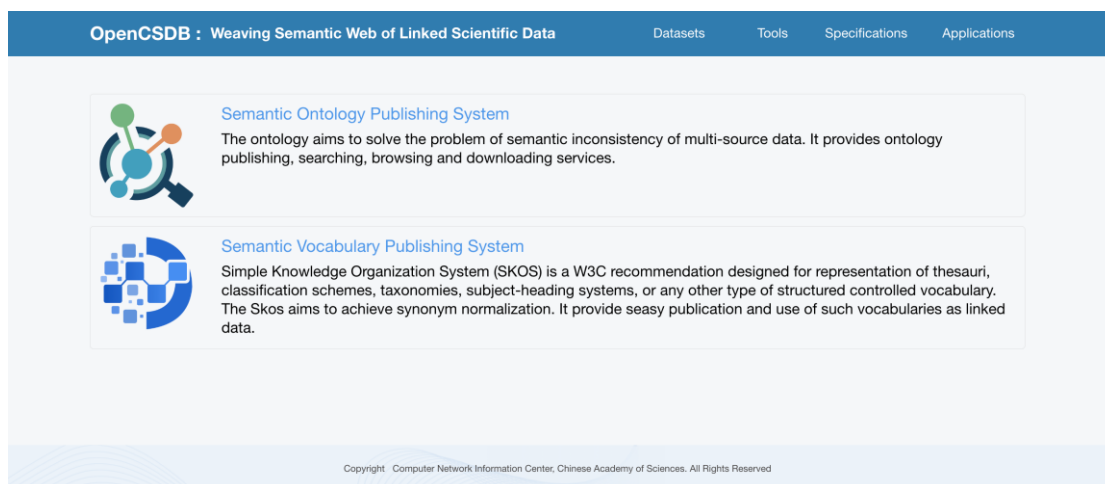


Figure 5 Utility Support Services Tool List Page

This page displays the toolset logo, title, and profile information, respectively. Click the tool title on the tool set list page, jump to the corresponding link and enter the system home page of 2 tools respectively. The jump links are <http://ols.semweb.csdb.cn/> and [www.example.com](http://www.example.com). <http://skos.semweb.csdb.cn/>

## 2.4 glossary specification

Click "Specifications" on the navigation bar of the platform to enter the vocabulary specification page of the scientific data association network portal platform. This page presents four common supporting vocabulary specifications for scientific data association networks.

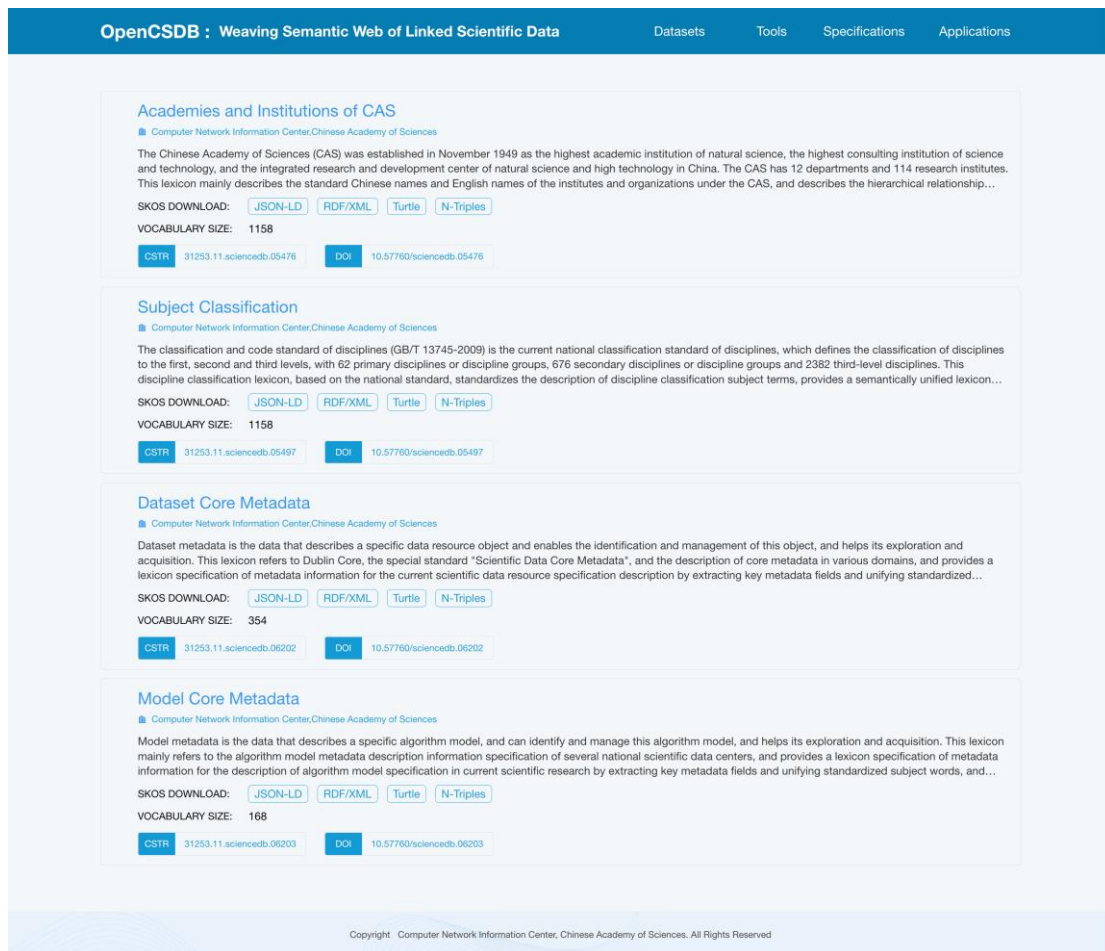


Fig. 6 List page of thesaurus

The thesaurus specification page displays the title, publisher, glossary profile, SKOS Download form, glossary size, glossary CSTR, and DOI identification number of each glossary. Among them, SKOS Download provides the following four ways: JSON-LD, RDF/XML, Turtle, N-Triples.

## 2.5 application demonstration

Click "Applications" on the navigation bar of the platform to enter the application demonstration page of the scientific data association network portal platform. This page shows three

current domain application demonstrations of the Scientific Data Link Web Portal.

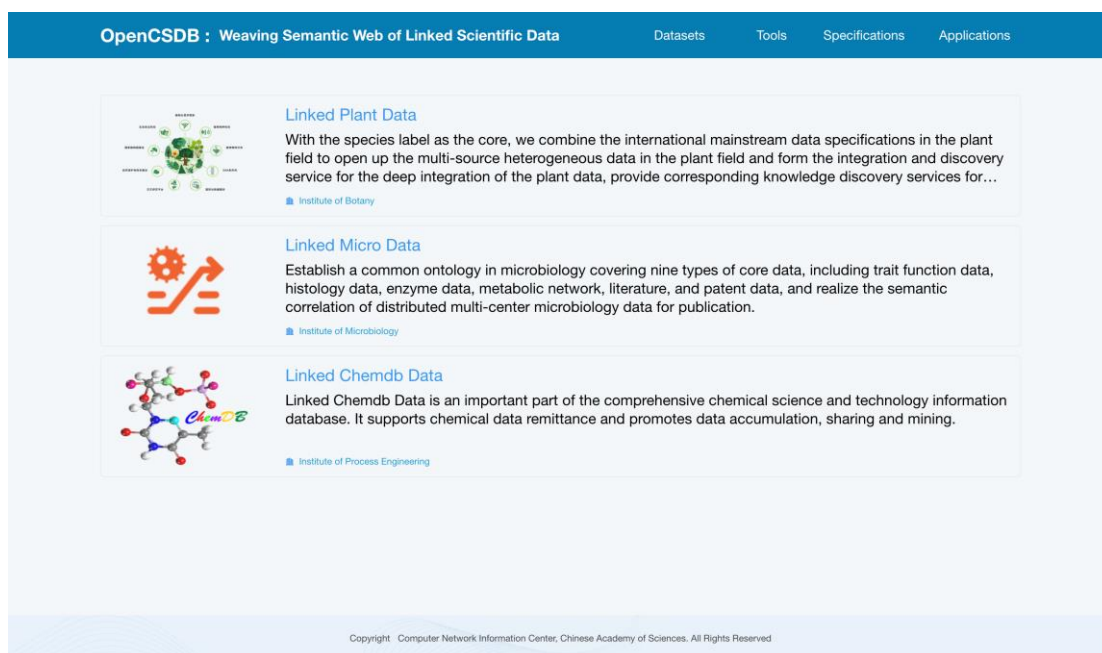


Figure Application Demonstration Page

The application demonstration page displays the title, profile, Logo, profile, and demonstration organization of each application demonstration.

Click on the title of each demonstration to go to the specific portal page for that demonstration.

For example, click "Linked Plant Data" to enter the details page.

In the application demonstration list page, click "Linked Plant Data" to enter the plant association data portal page.

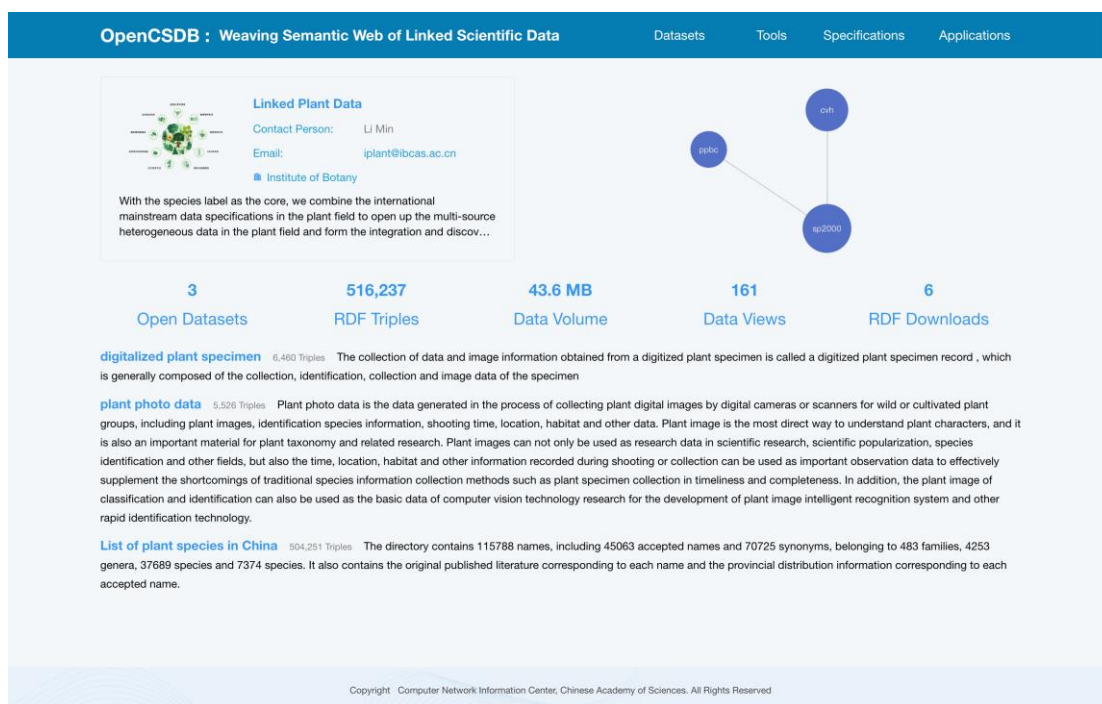


Figure Plant Application Demonstration Page

On the plant application demonstration page, the following information is displayed from top to bottom: demonstration name, contact person and Email, organization, introduction, demonstration data set association map, demonstration data statistics, and specific data set list included in the demonstration.

The application demonstration portal platform statistics are Open Datasets, RDF Triples, Data Volume, Data Views, and RDF Downloads statistics.

The data set list displays the title, RDF data volume and basic introduction. The subject area or publishing organization can be filtered and displayed through the check box on the left.

The knowledge graph of disciplinary dataset shows the association relationship of all datasets in the site. The node size represents the volume of the dataset, and the distance between nodes represents the close association between the two datasets. For example, the size and distance between three data sets published in the field of plant science represent the amount of data and the number of associations.

Click any title on the dataset list page at the bottom of the home page to enter the dataset details page. If you click on a data set, you will enter the page as shown in Figure 2.

### 3. database design

Using MongoDB database, contains tables:

Name	Code
Application Demonstration Portal Table	applications
Data Download Record Sheet	data_download
Data View Record Sheet	data_view
Data Set Information Table	dataset
entity class table	entityClass
entity attribute table	entityProperty
Data Set ES Search Log Table	es_log
thesaurus entity class list	specifications

### 3.1 Organization chart (applications)

field name	Chinese name	data type	description
_id	Primary key id	ObjectId	meaningless unique identifier
description	Data Center Description	String	
image_url	Image url address	String	Data centerlogo image url
institution	institution name	String	
institution_url	Organization URL Address	String	
title	portal name	String	
title_url	portal url address	String	
datacenter_id	Data Center Identification	String	
records	Number of resource entities in the field	Int32	

### 3.2 Data download record table (data\_download)

field name	Chinese name	data type	description
_id	Primary key id	ObjectId	meaningless unique identifier
file_type	Download File Type	String	
ip	Download clientip	String	
time	download time	Date	
url	Download file path	String	

### 3.3 Data view record table (data\_view)

field name	Chinese name	data type	description
_id	Primary key id	ObjectId	meaningless unique identifier
ip	BrowseIP	String	
time	browse time	Date	
url	Browse page address	String	

### 3.4 Data set information table (dataset)

field name	Chinese name	data type	description
_id	Primary key id	ObjectId	meaningless unique identifier
contact_email	contact email	String	
contact_name	contact name	String	
data_volume	Data Set Corresponding Entity File Mass	Int32	
description	data set describe	Object	
description.cn	Data Set DescriptionChinese	String	
description.en	Data Set DescriptionEnglish	String	
domain	the art	String	
identifier	identifier	String	
image	Data Set Thumbnail Path	String	
links	List of associated datasets	list	

links.target	Associated dataset name	String	
links.target_triples	Number of triples contained in the associated dataset	String	
links.value	the amount of data associated	String	
publish_time	Data Set Release Date	Date	
rdf_download_url	rdf entity file download address	String	
sparql	sparql endpoint link	String	
title	dataset name	String	
triples	the numb of triples contain in that dataset	String	
unit_name	Name of affiliated institution	String	
website	Affiliate links	String	
datacenter_id	Data Center Identification	String	
keywords	keywords	list	

### 3.5entityClasstable

field name	Chinese name	data type	description
_id	Primary key id	ObjectId	meaningless unique identifier
endPointsList	endpoint address list	Array	



label	entity name	String	
uri	URI address	String	

### 3.6entityProperty

field name	Chinese name	data type	description
_id	Primary key id	ObjectId	meaningless unique identifier
label	Chinese name	String	
label_en	English name	String	
org	affiliatedinstitutions	String	
uri	Completeuri	String	
uri_parse	Is uri resolvable?	Boolean	

### 3.7 Dataset ES Search Log Table (es\_log)

field name	Chinese name	data type	description
_id	Primary key id	ObjectId	meaningless unique identifier
create_time	retrieval time	Date	
dataset_identifier	data set identifier	String	
num_in_page	Number of entries per page	Int32	
page_num	page number	Int32	
resource_entity_uri	Uri	String	
title	title	String	

### 3.8 List of entity classes (specifications)

field name	Chinese name	data type	description
_id	Primary key id	ObjectId	meaningless unique

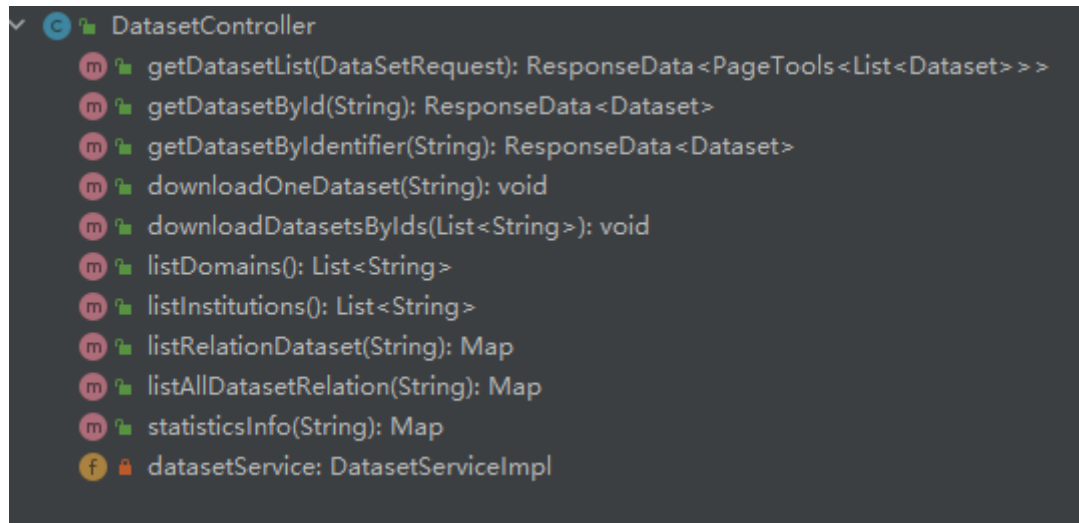
			identifier
cstr	cstr identifier	String	
description	vocabulary description	String	
doi	doi logo	String	
download_jsonld	SKOSDownload JSON-LD     Download Link	String	
institution	institution name	String	
institution_url	OrganizationURL Address	String	
title	thesaurus name	String	
title_url	thesaurusurl address	String	
vocabularyPageUrl	HTMLpage address	String	
download_nt	SKOS Download-N-Triples Download Link	String	
vocabulary_size	vocabulary	Int32	
download_rdf	SKOS Download-RDF/XML Download Link	String	
download_ttl	SKOS Download-Turtle Download Link	String	

## 4. criticalclass design

### 4.1 DatasetController.java

Class Function: Data Set Query Control Class

Class diagram:



Attribute Description:

attribute name	type	meaning
datasetService	DatasetServiceImpl	Data Set Query Related Service Layer

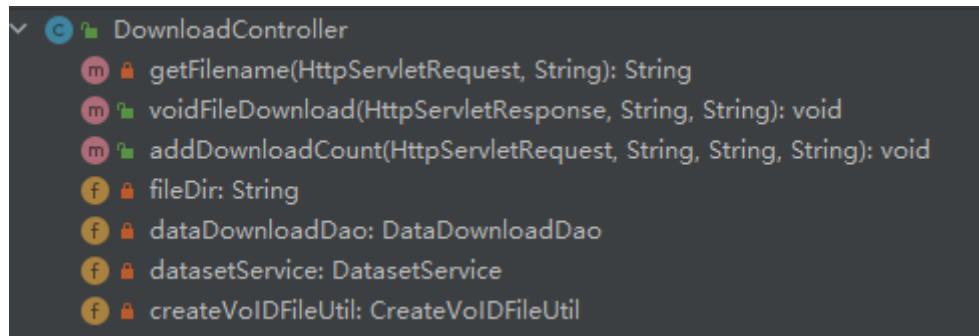
Method Description:

method name	method describes
getDatasetList	Get a list of datasets
getDatasetById	Query dataset detailsbased on datasetid
getDatasetByIdentifier	Query dataset detailsbased on datasetidentifier
downloadOneDataset	Download a datasetbased onid
downloadDatasetsByIds	Download multiple datasetsbased onid list
listDomains	Query all dataset domains
listInstitutions	Search all institutions
listRelationDataset	Lists the dataset information associated with this dataset
listAllDatasetRelation	List associations between all datasets
statisticsInfo	Home Statistics

## 4.2 DownloadController.java

Class Function: Data Set Metadata Download Control Class

Class diagram:



Attribute Description:

attribute name	type	meaning
dataDownloadDao	DataDownloadDao	Data Set Download Related Service Layer
createVoIDFileUtil	CreateVoIDFileUtil	Dataset Metadata Generation Tool Class

Method Description:

method name	method describes
getFilename	GetRDF file name
voidFileDownload	Download metadata description Void description file
addDownloadCount	Record downloads

## 4.3 ApplicationsController.java

Class Function: Data Center Portal Control Class

Class diagram:

```

ApplicationsController
  findApplications(ResourceListRequest): ResponseData<PageTools<List<Applications>>>
  getAppByDataCenterId(ResourceListRequest): ResponseData<ApplicationDetail>
  applicationsService: ApplicationsService

```

Attribute Description:

attribute name	type	meaning
applicationsService	ApplicationsService	Data Center Related Service Layer

Method Description:

method name	method describes
findApplications	Get a list of data center portals
getAppByDataCenterId	Get Data Center Portal Details

## 4.4 ResourceEntityController.java

Class Function: Resource Entity Control Class

Class diagram:

```

ResourceEntityController
  getResourceList(ResourceListRequest): ResponseData<PageTools<List<ResourceList>>>
  getResourceListByES(ResourceListRequest): ResponseData<PageTools<List<ResourceList>>>
  getResourceDetail(ResourceDetailRequest): ResponseData<ResourceDetail>
  listRelationResource(String): ResponseData<Map>
  exportToFileByGraph(ExportRequest): ResponseData<String>
  resourceEntityService: ResourceEntityService

```

Attribute Description:

attribute name	type	meaning
resourceEntityService	ResourceEntityService	Resource Entity Related Service Layer

Method Description:

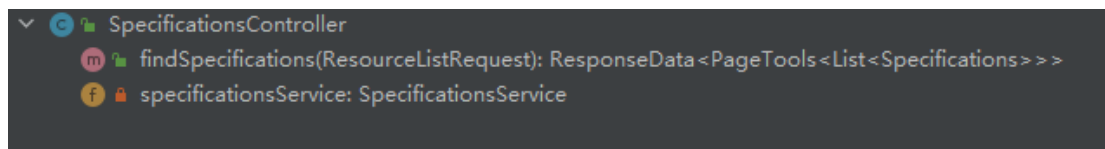
method name	method describes
-------------	------------------

getResourceList	Get a list of resources
getResourceListByES	Get resource list via ES
getResourceDetail	Get Resource Details
listRelationResource	Query the triples associated with the resource, including the sparql endpoint to which it belongs and the name of the dataset to which it belongs
exportToFileByGraph	Export ternary composition as file

## 4.5 SpecificationsController.java

Class Function: Vocabulary Processing Control Class

Class diagram:



Attribute Description:

attribute name	type	meaning
specificationsService	SpecificationsService	thesaurus-dependent Service layer

Method Description:

method name	method describes
findSpecifications	Get a list of thesauri

## 4.6 ViewController.java

Class function: pageview related control class

Class diagram:

```
▼ ViewController
  m addViewCount(HttpServletRequest, String, String): void
  f dataViewDao: DataViewDao
```

### Attribute Description:

attribute name	type	meaning
DataViewDao	dataViewDao	pageviews related Service layer

### Method Description:

method name	method describes
addViewCount	Record access requests