



DICE

Data Infrastructure Capacity for EOSC

B2FIND

flexible and user-centred Discovery Portal
that acts on Community's needs

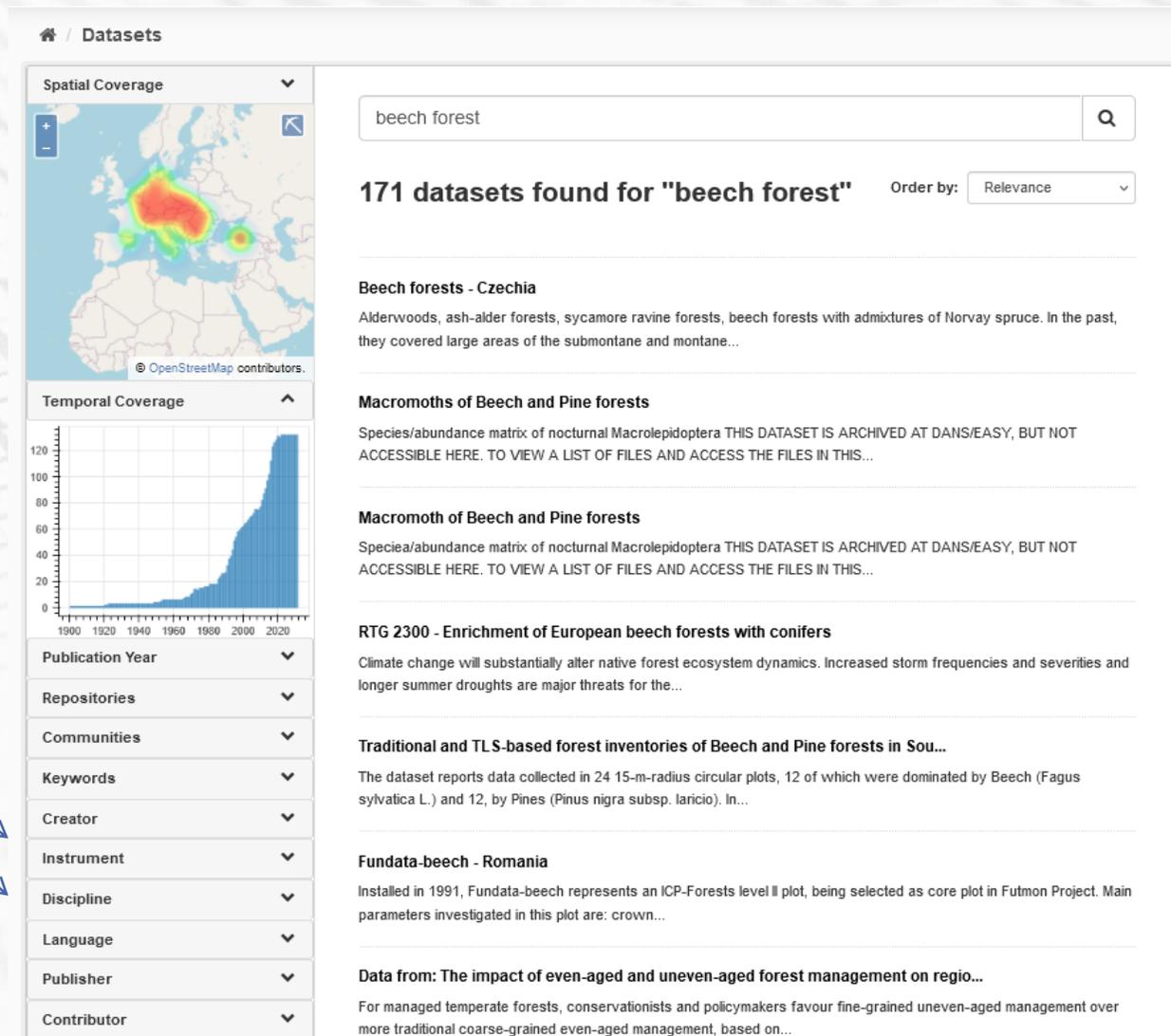
Claudia Martens / DKRZ (EUDAT)



The DICE project has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 under Grant Agreement no. 101017207

~ 1.036 mio records
 37 Repositories
 29 Communities
 > 130 harvesting endpoints

Spatial/Temporal Coverage, →
 PublicationYear,
 Repositories, Communities,
 Keywords, Creator,
 Instrument, Discipline, →
 Language, Publisher,
 Contributor, ResourceType,
 OpenAccess



The screenshot shows the DICE Datasets search results for the query "beech forest". The interface includes a search bar with the query, a dropdown menu for "Spatial Coverage" showing a map of Europe with a heatmap, and a dropdown menu for "Temporal Coverage" showing a bar chart of publication counts over time. Below these are several filter menus for "Publication Year", "Repositories", "Communities", "Keywords", "Creator", "Instrument", "Discipline", "Language", "Publisher", and "Contributor". The search results list 171 datasets found for "beech forest", ordered by Relevance. The first few results are:

- Beech forests - Czechia**: Alderwoods, ash-alder forests, sycamore ravine forests, beech forests with admixtures of Norway spruce. In the past, they covered large areas of the submontane and montane...
- Macromoths of Beech and Pine forests**: Species/abundance matrix of nocturnal Macrolepidoptera THIS DATASET IS ARCHIVED AT DANS/EASY, BUT NOT ACCESSIBLE HERE. TO VIEW A LIST OF FILES AND ACCESS THE FILES IN THIS...
- Macromoth of Beech and Pine forests**: Speciea/abundance matrix of nocturnal Macrolepidoptera THIS DATASET IS ARCHIVED AT DANS/EASY, BUT NOT ACCESSIBLE HERE. TO VIEW A LIST OF FILES AND ACCESS THE FILES IN THIS...
- RTG 2300 - Enrichment of European beech forests with conifers**: Climate change will substantially alter native forest ecosystem dynamics. Increased storm frequencies and severities and longer summer droughts are major threats for the...
- Traditional and TLS-based forest inventories of Beech and Pine forests in Sou...**: The dataset reports data collected in 24 15-m-radius circular plots, 12 of which were dominated by Beech (*Fagus sylvatica* L.) and 12, by Pines (*Pinus nigra* subsp. *laricio*). In...
- Fundata-beech - Romania**: Installed in 1991, Fundata-beech represents an ICP-Forests level II plot, being selected as core plot in Futmon Project. Main parameters investigated in this plot are: crown...
- Data from: The impact of even-aged and uneven-aged forest management on regio...**: For managed temperate forests, conservationists and policymakers favour fine-grained uneven-aged management over more traditional coarse-grained even-aged management, based on...

- ▶ long way from the launch of B2FIND in 2014 until today
- ▶ several new versions of B2FIND ingestion code
- ▶ several new machines, including reingest of all metadata
- ▶ ever evolving system!



harvesting

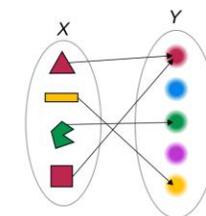
OAI-PMH,
CSW,
RestAPI

search



mapping

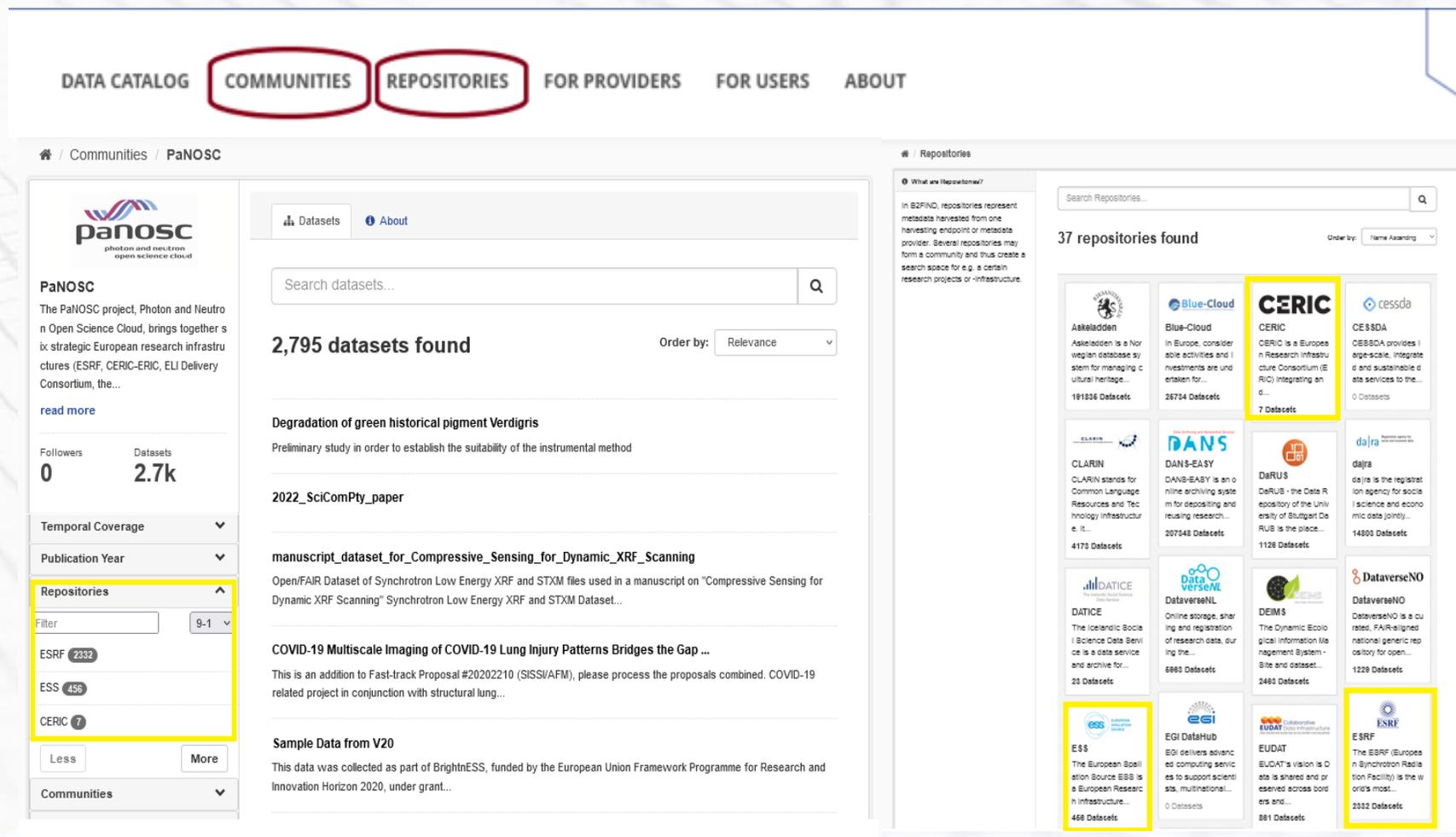
- DublinCore,
Datacite,
OpenAire,
EUDAT Core
- ISO 19115/19139
[INSPIRE],
DDI 2.5
- community
specific



Showcase: PaNOSC

➤ enhancing visibility of research infrastructures, projects and facilities in Photon and Neutron research

➤ enabling a common 'search space' – new structure of web GUI



The screenshot displays the PaNOSC web interface. At the top, navigation links include 'DATA CATALOG', 'COMMUNITIES', 'REPOSITORIES', 'FOR PROVIDERS', 'FOR USERS', and 'ABOUT'. The 'COMMUNITIES' and 'REPOSITORIES' links are highlighted with red circles.

The main content area is divided into two sections: 'Communities / PaNOSC' and 'Repositories'.

Communities / PaNOSC:

- PaNOSC:** The PaNOSC project, Photon and Neutron Open Science Cloud, brings together six strategic European research infrastructures (ESRF, CERIC-ERIC, ELI Delivery Consortium, the...)
- Followers:** 0
- Datasets:** 2.7k
- Temporal Coverage:** Filterable
- Publication Year:** Filterable
- Repositories:** Filterable (9-1)
- ESRF:** 2332
- ESS:** 456
- CERIC:** 7
- Less / More:** Buttons to toggle repository visibility.

Repositories:

- Search:** Search datasets... (2,795 datasets found)
- Order by:** Relevance
- Results:**
 - Degradation of green historical pigment Verdigris:** Preliminary study in order to establish the suitability of the instrumental method.
 - 2022_SciComPty_paper:** manuscript_dataset_for_Compressive_Sensing_for_Dynamic_XRF_Scanning. Open/FAIR Dataset of Synchrotron Low Energy XRF and STXM files used in a manuscript on "Compressive Sensing for Dynamic XRF Scanning" Synchrotron Low Energy XRF and STXM Dataset...
 - COVID-19 Multiscale Imaging of COVID-19 Lung Injury Patterns Bridges the Gap ...:** This is an addition to Fast-track Proposal #20202210 (SISIS/AFM), please process the proposals combined. COVID-19 related project in conjunction with structural lung...
 - Sample Data from V20:** This data was collected as part of BrightnESS, funded by the European Union Framework Programme for Research and Innovation Horizon 2020, under grant...

37 repositories found:

- Askeledden:** 191836 Datasets
- Blue-Cloud:** 26734 Datasets
- CERIC:** 7 Datasets
- CESSDA:** 0 Datasets
- CLARIN:** 4173 Datasets
- DANS:** 207348 Datasets
- DaRUS:** 1128 Datasets
- dajra:** 14800 Datasets
- DATICE:** 23 Datasets
- DataVerseNL:** 6983 Datasets
- DEIMS:** 2483 Datasets
- DataverseNO:** 1229 Datasets
- ESS:** 488 Datasets
- EGI DataHub:** 0 Datasets
- EUDAT:** 861 Datasets
- ESRF:** 2032 Datasets

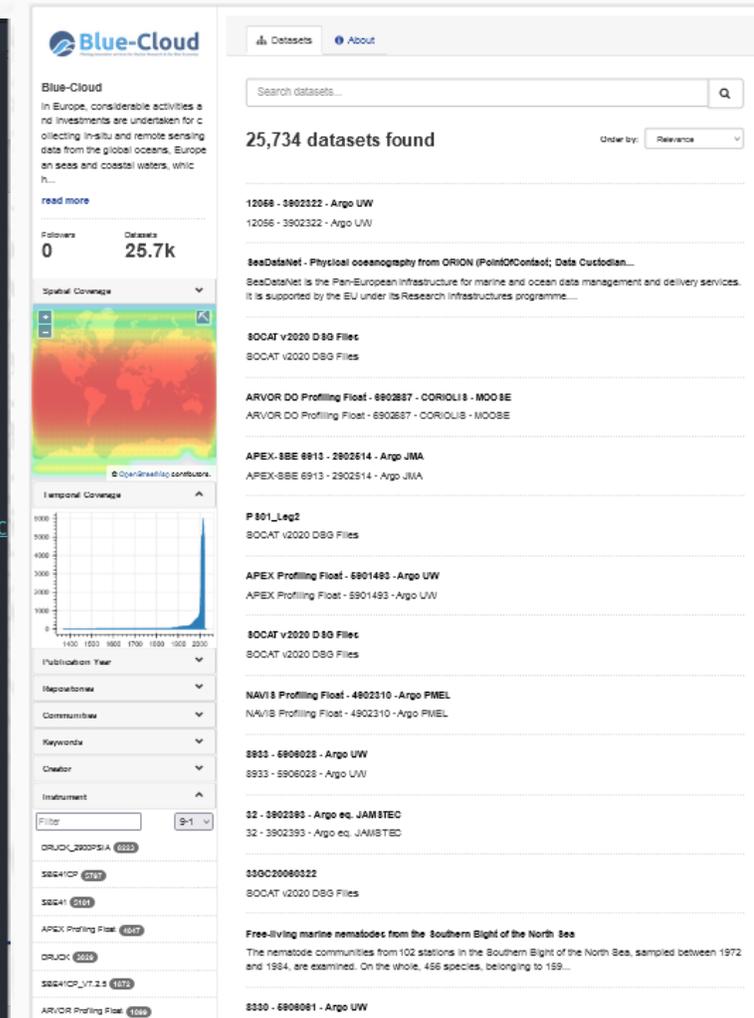
Showcase: Blue-Cloud

- including harvesting via RestAPI iterating through 'collections' and 'datasets'
- community-specific mapping including <instrument> and <temporalCoverage>
- but <discipline> fixcoded to 'Marine Science'

```

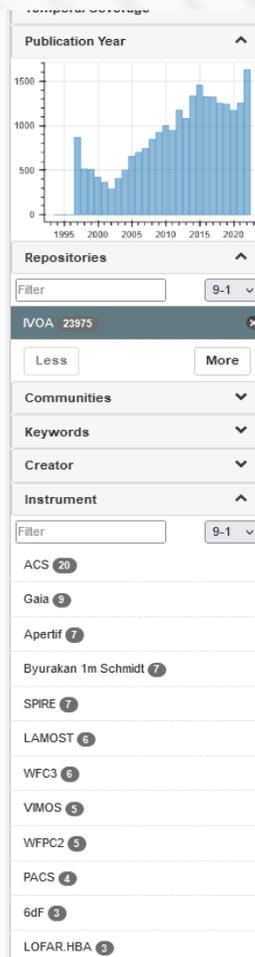
{
  "Abstract": "8547 - 2903365 - Argo eq. JMA",
  "Bounding_Box_EastLongitude": 137.08099,
  "Bounding_Box_NorthLatitude": 25.1,
  "Bounding_Box_SouthLatitude": 22.948,
  "Bounding_Box_WestLongitude": 127.817,
  "Identifier": "0019473E11FA3840EECB57432B909C86C8C0B36D",
  "Instruments": [
    "DRUCK",
    "SBE41"
  ],
  "Keywords": [
    "49",
    "TWR",
    "ARGO"
  ],
  "Last_Update": "2022-09-12T00:09:48",
  "OnlineResourceUrl":
  "https://data.blue-cloud.org/search-details?step=-0120019473E11FA3840EECB57432B909C86C8C0B36D",
  "Organisations": [
    "JMA"
  ],
  "Parameters": [
    "SUBSURFACE PRESSURE",
    "SUBSURFACE SALINITY",
    "SUBSURFACE TEMPERATURE"
  ],
  "Platforms": [
    "8547"
  ],
  "Source": "EuroArgo - Argo",
  "Temporal_Extent_Begin": "2019-05-07T07:50:03",
  "Temporal_Extent_End": "2020-07-31T10:07:43",
  "Title": "8547 - 2903365 - Argo eq. JMA"
}

```



The screenshot shows the Blue-Cloud search interface. On the left, there is a search bar and a 'Search datasets...' input field. Below the search bar, it displays '25,734 datasets found' and 'Order by: Relevance'. The main content area shows a list of search results, including details for various datasets such as '12068 - 3902322 - Argo UW', 'SeaDataNet - Physical oceanography from ORION (PointOfContact; Data Custodian...', 'SOCAT v2020 D8G Files', 'ARVOR DO Profiling Float - 6902857 - CORIOLIS - MOOSE', 'APEX-SBE 6913 - 2902514 - Argo JMA', 'P801_Leg2', 'APEX Profiling Float - 6901493 - Argo UW', 'SOCAT v2020 D8G Files', 'NAVIS Profiling Float - 4902310 - Argo PMEL', '8933 - 6906028 - Argo UW', '32 - 3902390 - Argo eq. JAMSTEC', '330C20080322', and 'Free-living marine nematodes from the Southern Bight of the North Sea'. The interface also includes a 'Spatial Coverage' map and a 'Temporal Coverage' line graph.

- ▶ displaying more information using EUDAT Core Metadata Schema
- ▶ including a metadata element `<instrument>` = facet in B2IFND
- ▶ including UAT in `<disciplines>` = assign records to the 'right' discipline



Publication Year

Repositories

Filter 9-1

IVOA 23975

Less More

Communities

Keywords

Creator

Instrument

Filter 9-1

- ACS 20
- Gaia 9
- Aperitif 7
- Byurakan 1m Schmidt 7
- SPiRE 7
- LAMOST 6
- WFC3 6
- VIMOS 5
- WFPC2 5
- PACS 4
- 6dF 3
- LOFAR.HBA 3

Call spectroscopy of SMC red giants. IV.
This paper represents a major step forward in the systematic and homogeneous study of Small Magellanic Cloud (SMC) star clusters and field stars carried out by applying the...

Ly-Alpha Forest spectra simulation analysis. I.
We present moderate-resolution (~50km/s FWHM) spectra of the Ly-alpha forest for seven quasars with redshifts ranging from 2.53 to 3.13, obtained with the Blue Spectrograph and...

Main galaxy groups from SDSS-DR5
We investigate the correlations between galaxy properties and environment. From the main galaxy sample of the SDSS Data Release 5, we have extracted 11 163 groups at the linking...

Radial velocities of 29 cataclysmic variables
We report follow-up spectroscopy of 29 cataclysmic variables from the Sloan Digital Sky Survey (SDSS), 22 of which were discovered by SDSS and seven of which are previously...

CONCH-SHELL catalog of nearby M dwarfs
We present an all-sky catalogue of 2970 nearby (d<~50pc), bright (J<9) M- or late K-type dwarf stars, 86% of which have been confirmed by spectroscopy. This catalogue will...

Variations in supernova yields
Theoretical nucleosynthetic yields from supernovae are sensitive to both the details of the progenitor star and the explosion calculation. We attempt to comprehensively identify...

Chandra observations of Monoceros R2
We report on the results of the Chandra ACIS-I observation on the central region of the Monoceros R2 cloud (Mon R2), a high-mass star-forming region (SFR) at a distance of...

72 WINGS nearby clusters luminosity functions
Using V band photometry of the WINGS survey, we derive galaxy luminosity functions (LF) in nearby clusters. This sample is complete down to M_V = -15.15, and it is homogeneous...

BR photometry in Coma, A2199 & AWM5 clusters
We present the results of a study on galaxy interactions, tides, and other processes that produce luminous fine-scale substructures in the galaxy clusters: Coma, Perseus, Abell...

ICRF3 radio-loud quasars with Gaia DR2 data
We investigate a sample of 3413 International Celestial Reference Frame (ICRF3) extragalactic radio-loud sources with accurate positions determined by very long baseline...

Discipline

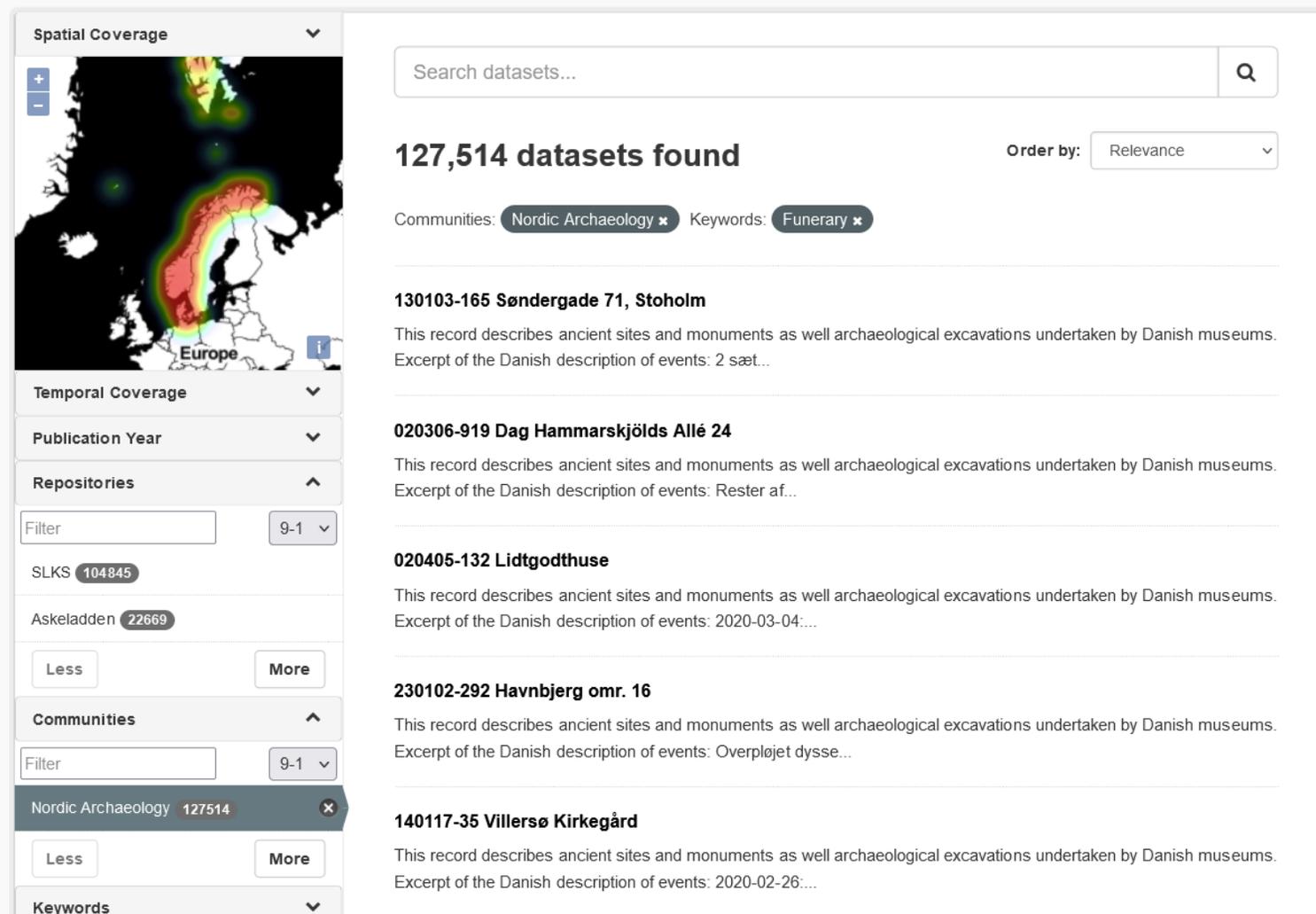
Filter 9-1

- Astrophysics and Astronomy 23970
- Natural Sciences 23051
- Physics 23049
- Observational Astronomy 19341
- Stellar Astronomy 11642
- Galactic and extragalactic Ast 8854
- Cosmology 4887
- Interstellar medium 3149
- Exoplanet Astronomy 2846
- Interdisciplinary Astronomy 2745

Less More

Showcase: Nordic Archaeology

- ▶ creating a “Nordic Archaeology” Community including two national data providers from Norway and Denmark
- ▶ using ArcGIS API for harvesting, community-specific mapping
- ▶ including a Jupyter Notebook for assigning multilingual terms (mapping of different concepts) within the ingestion process



Spatial Coverage

Search datasets...

127,514 datasets found Order by: Relevance

Communities: Nordic Archaeology Keywords: Funerary

130103-165 Söndergade 71, Stoholm
This record describes ancient sites and monuments as well archaeological excavations undertaken by Danish museums. Excerpt of the Danish description of events: 2 sæt...

020306-919 Dag Hammarskjölds Allé 24
This record describes ancient sites and monuments as well archaeological excavations undertaken by Danish museums. Excerpt of the Danish description of events: Rester af...

020405-132 Lidtgodthuse
This record describes ancient sites and monuments as well archaeological excavations undertaken by Danish museums. Excerpt of the Danish description of events: 2020-03-04:...

230102-292 Havnbjerg omr. 16
This record describes ancient sites and monuments as well archaeological excavations undertaken by Danish museums. Excerpt of the Danish description of events: Overpløjet dysse...

140117-35 Villersø Kirkegård
This record describes ancient sites and monuments as well archaeological excavations undertaken by Danish museums. Excerpt of the Danish description of events: 2020-02-26:...

Metadata Aggregator



- ▶ md schema based on EUDAT Core
(Datacite + temporalCoverage, Instrument, Discipline, Contact, several types of Identifier)
- ▶ md exposure via OAI-PMH
 - ▶ with md prefix euc (all we have)
 - ▶ with md prefix datacite (= OpenAire)

- ▶ md schema based on Datacite
(with multiple types of Identifier)
- ▶ harvesting only with md prefix datacite!

Sum up

- ▶ in theory metadata should be openly accessible, clearly defined and ‘interoperable’ (exposed and exchanged in standardised ways)
- ▶ in practice very often they are not!
- ▶ due to
 - ▶ diverging amount of resources («big» infrastructures vs. small research groups)
 - ▶ different views of what «the best way» for standardisation is
 - ▶ new technologies that need to be implemented
 - ▶ interesting concepts of «sustainability» (in particular for European Projects)
- making *all* (meta)data findable requires time and resources
- we may support data providers and make their (meta)data discoverable on their specific needs



Questions?

JOIN OUR COMMUNITY



DICEosc



/company/diceosc

Contact

DICE

<https://www.dice-eosc.eu/call-service-requests>

EOSC Portal

<https://marketplace.eosc-portal.eu/services/b2find>

EUDAT

<https://www.eudat.eu/contact-support-request>

email

b2find@dkrz.de