



Artificial Intelligence for Earth Observation

Manolis Koubarakis
National and Kapodistrian
University of Athens



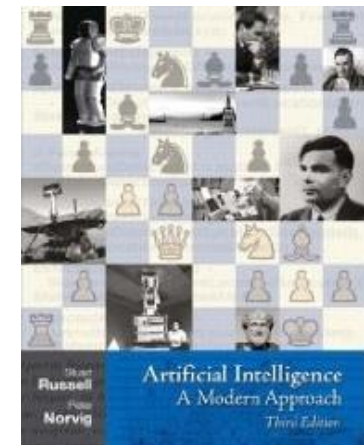
“Copernicus and Artificial Intelligence” workshop
28 January 2020
Brussels, Belgium

Outline

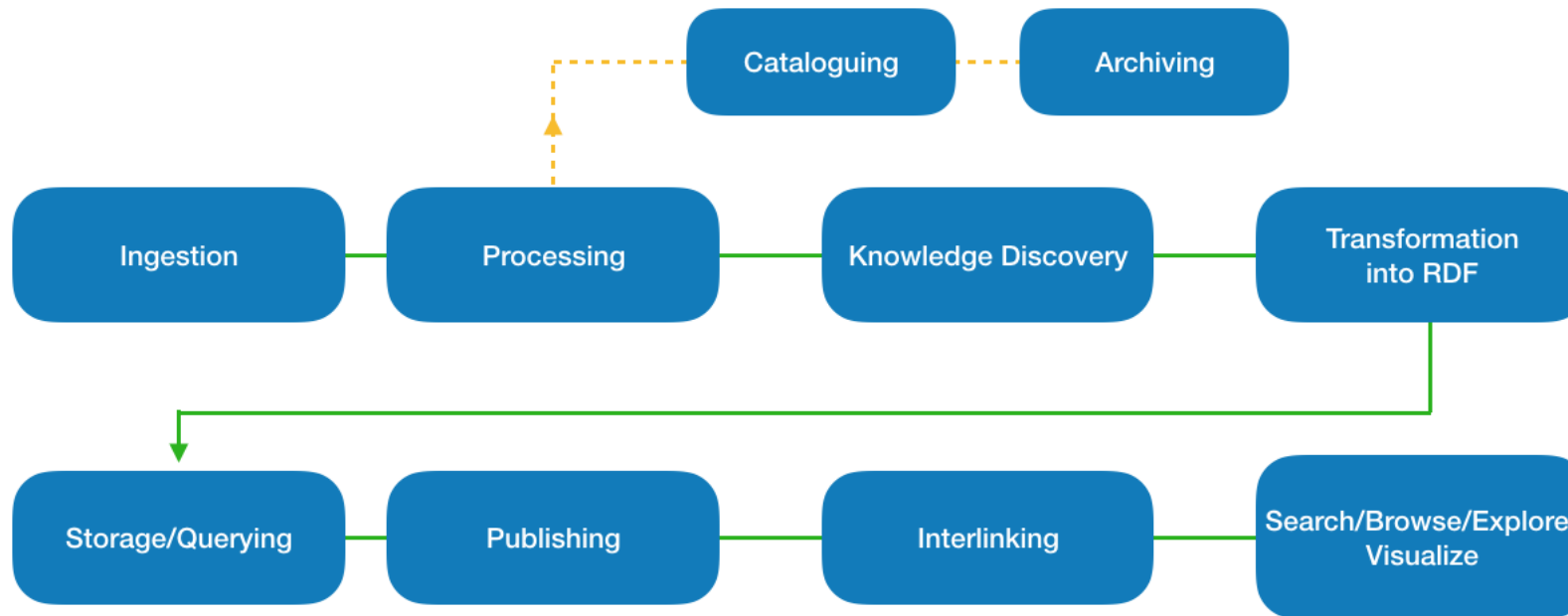
- A textbook view of AI technologies
- A data science pipeline for big linked Earth observation data
- AI techniques in the Earth observation domain
- The ExtremeEarth and AI4EU projects
- Conclusion

Artificial Intelligence: a Textbook View

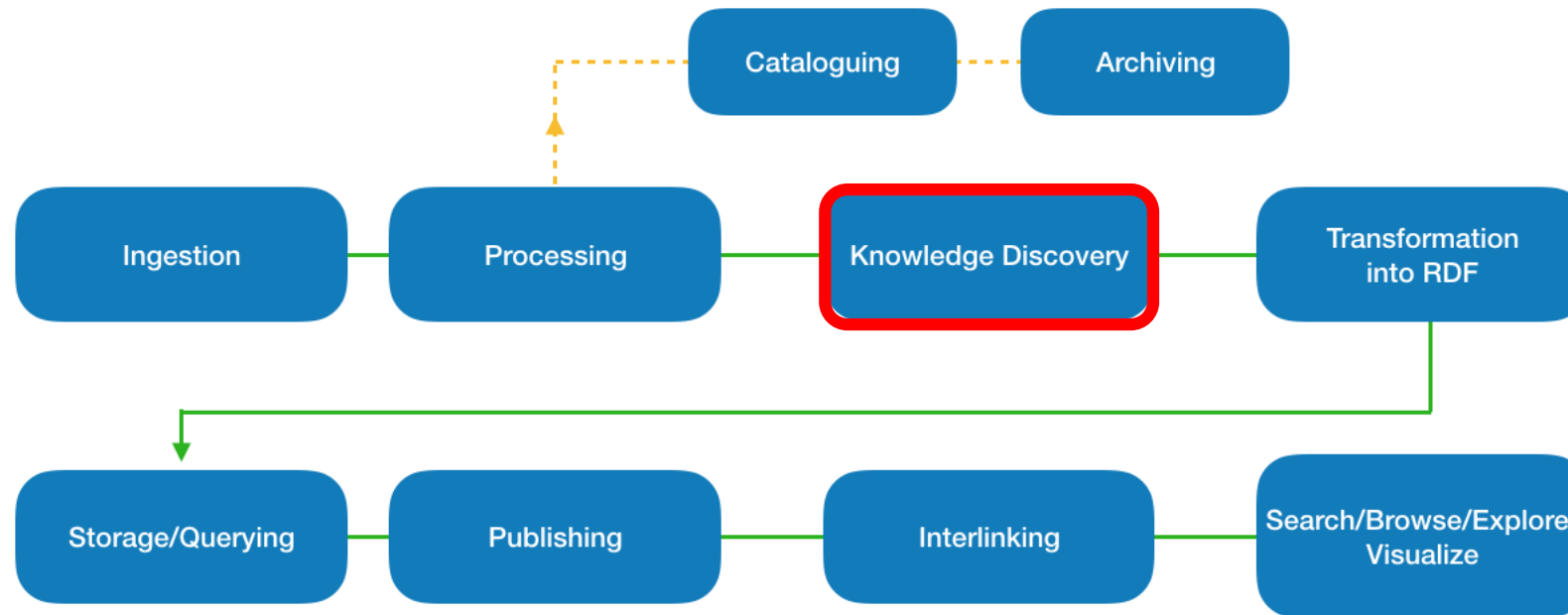
- Problem solving (search, heuristics)
- Constraint satisfaction
- **Knowledge representation and reasoning (esp. ontologies, knowledge graphs and linked data)**
- Planning
- Uncertainty (probabilistic reasoning, decision theory)
- **Machine learning (esp. deep learning)**
- **Natural language processing (esp. question answering)**
- **Perception (esp. computer vision)**
- Robotics



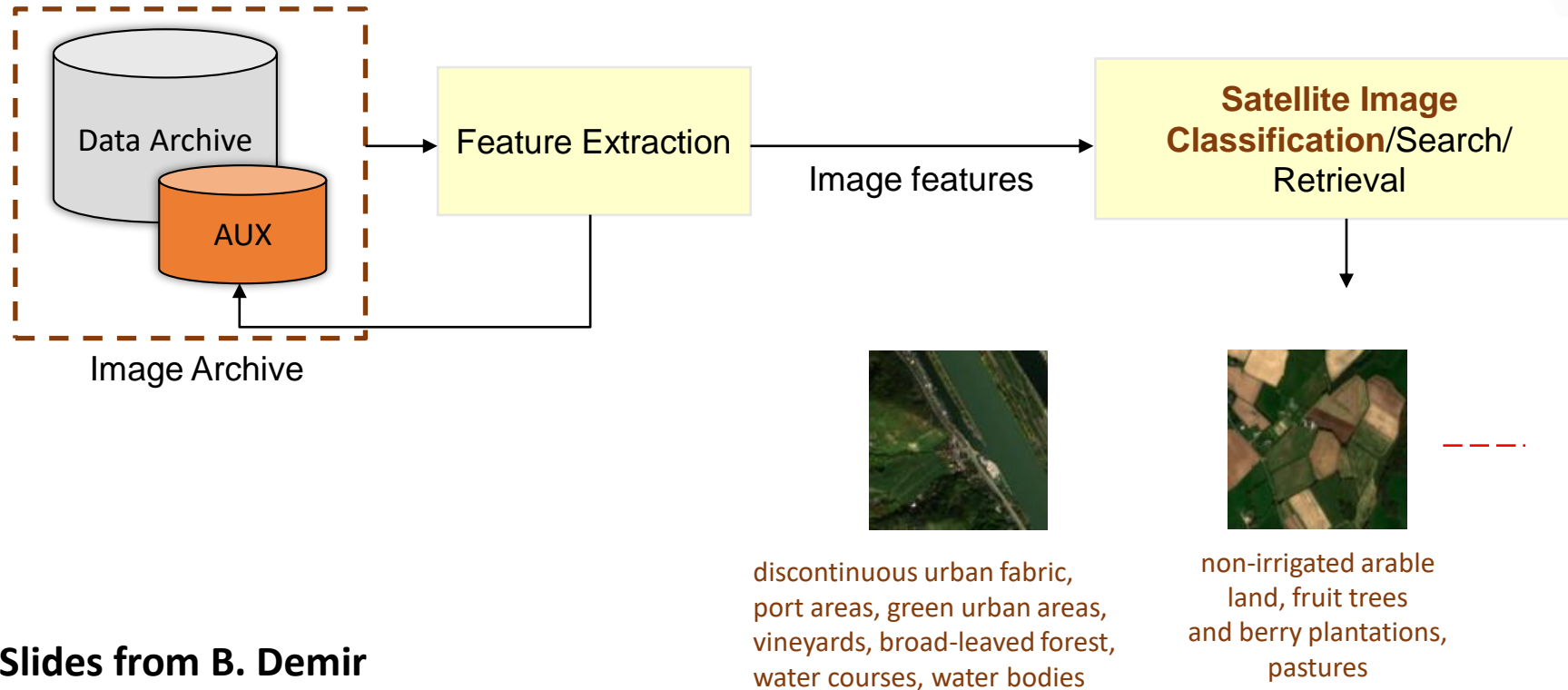
A Data Science Pipeline for Big Linked EO Data (IEEE GRSM 2016)



A Data Science Pipeline for Big Linked EO Data (cont'd)



Knowledge Discovery from Satellite Images



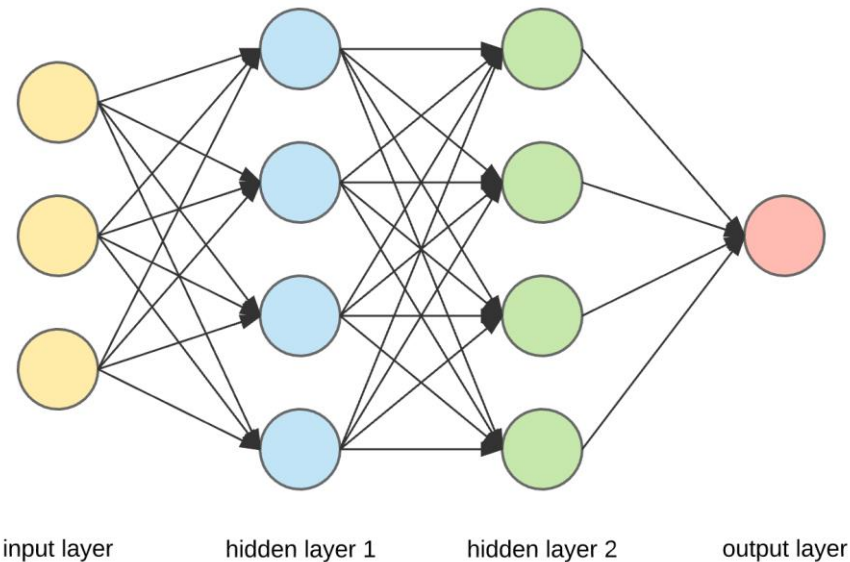
Slides from B. Demir

Satellite Image Classification Techniques

- Maximum likelihood classifiers
- K-nearest neighbor classifiers
- Decision trees, genetic algorithms based classifiers
- Artificial neural networks (multilayer perceptrons, radial basis function neural networks, deep neural networks)
- Kernel-based methods (i.e., regularized radial basis function neural networks, support vector machines)
- Graph matching algorithms

Current Emphasis on Deep Neural Networks

- Convolutional neural networks
- Recurrent neural networks
- Generative adversarial networks



The Importance of Big Training Datasets

- Deep learning architectures need a **very large amount of training data** for being effective.
- Computer vision image archives like ImageNet **cannot be reused** in Remote Sensing due to:
 - Differences in spatial resolution between computer vision images and remote sensing images
 - A much higher number of spectral bands in the case of remote sensing images
 - Different semantic content

The BigEarthNet Archive



- The BigEarthNet archive contains **590,326 Sentinel-2 image patches** with multiple land cover annotations (19 classes from CORINE land cover 2018 – slightly modified).
- See <http://bigearth.eu/>.



urban fabric,
marine waters,
industrial or commercial
units



urban fabric,
arable land,
mixed forest



coniferous forest,
mixed forest,
transitional
woodland/shrub,
inland waters



urban fabric,
arable land,
pastures,
marine waters



land principally
occupied by agriculture
with significant areas of
natural vegetation,
mixed forest,
transitional
woodland/shrub,
inland waters

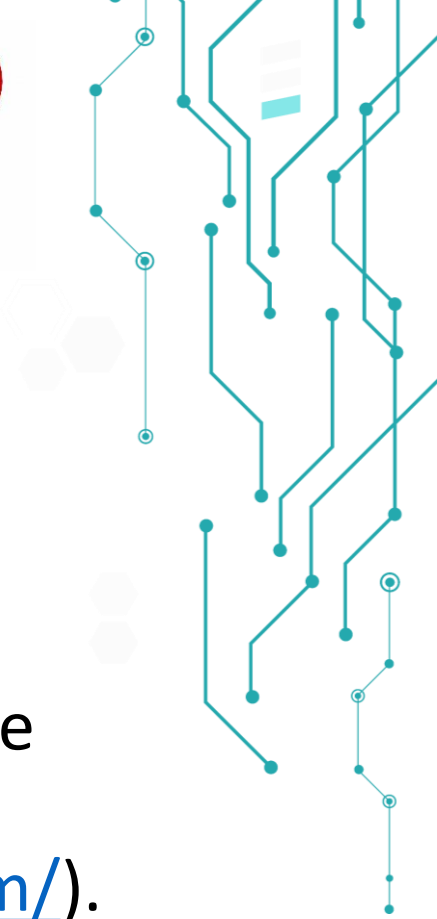


arable land,
land principally
occupied by agriculture
with significant areas of
natural vegetation,
mixed forest

Training CNNs on BigEarthNet

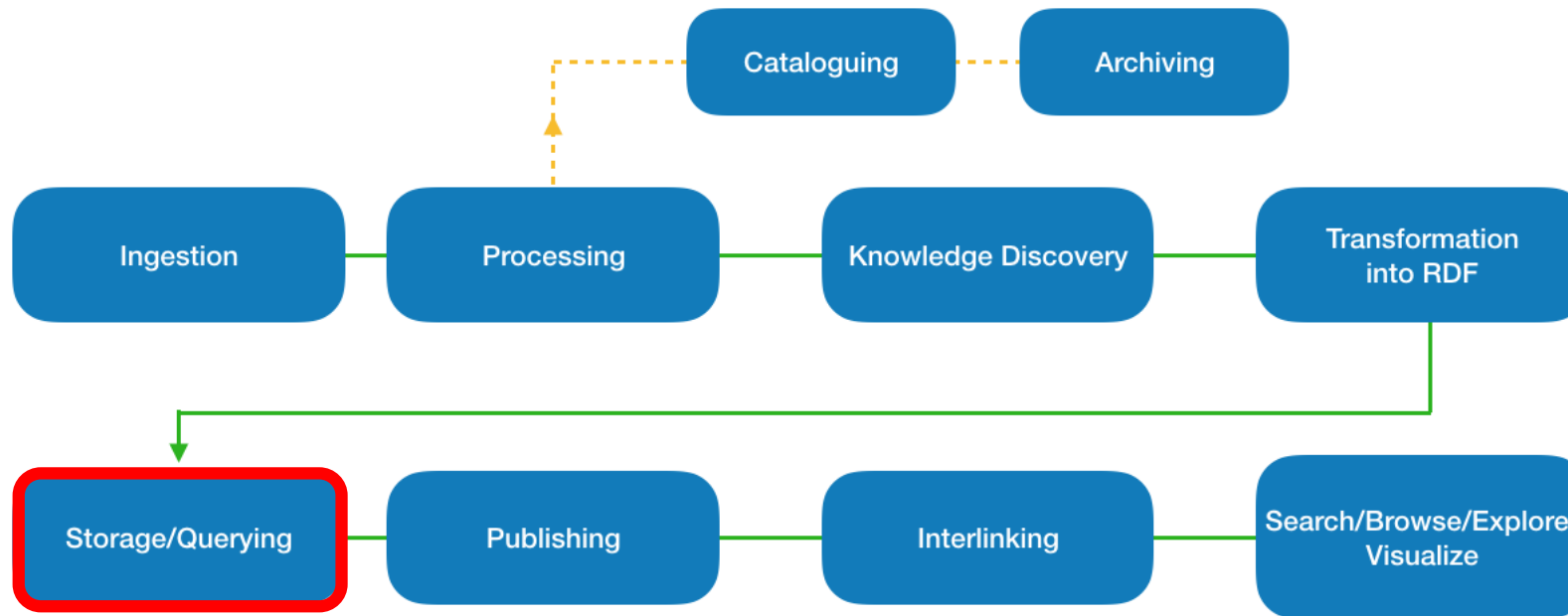
- Sumbul et al. show that two CNN architectures (ResNet50 and ResNet152) trained from scratch on BigEarthNet **significantly outperform** the same architectures when these are pre-trained on ImageNet and then used on BigEarthNet images.
- See <https://arxiv.org/pdf/2001.06372.pdf> .

Open Problems



- **Even larger training sets** (e.g., include more countries, utilize existing geospatial data sources like OpenStreetMap).
- **New deep neural network architectures.**
- **Distributed deep learning** on infrastructures offering very large storage, virtual machines and GPUs. See the **Hops platform** of European startup LogicalClocks (<https://www.logicalclocks.com/>).
- **AutoML** (neural nets designing neural nets).

A Data Science Pipeline for Big Linked EO Data (cont'd)



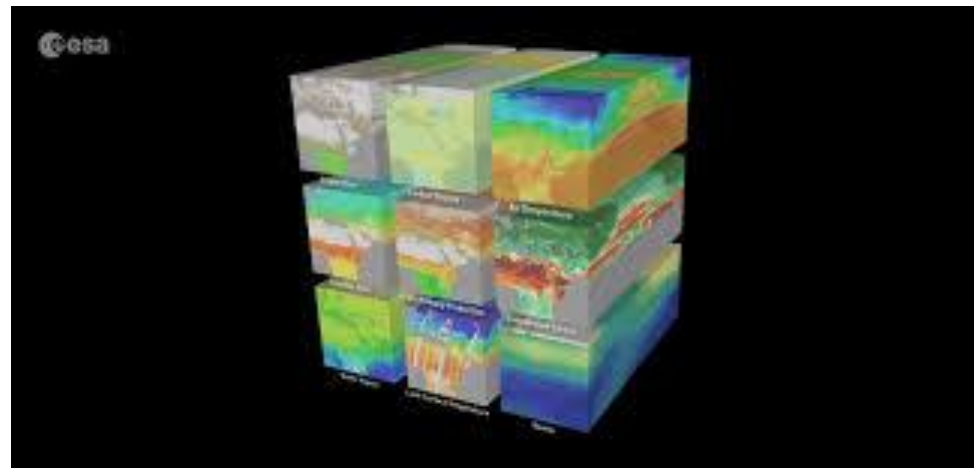
Array DBMS

- **Rasdaman** (Peter Baumann)
- **MonetDB/SciQL** (Martin Kersten and Stefan Manegold)
- SciDB (Mike Stonebraker)
- Paradigm4 (startup of Mike Stonebraker)



Data Cubes

- Open Data Cube (Australia)
- Swiss Data Cube
- Earth System Data Lab Data Cube



Spatiotemporal RDF Stores and Ontology-Based Data Access Systems

- Parliament
- **Strabon**
- uSeekM
- **RDF4J**
- **Virtuoso**
- AllegroGraph
- Stardog
- Oracle Spatial and Graph
- **GraphDB**
- **OntopSpatial**



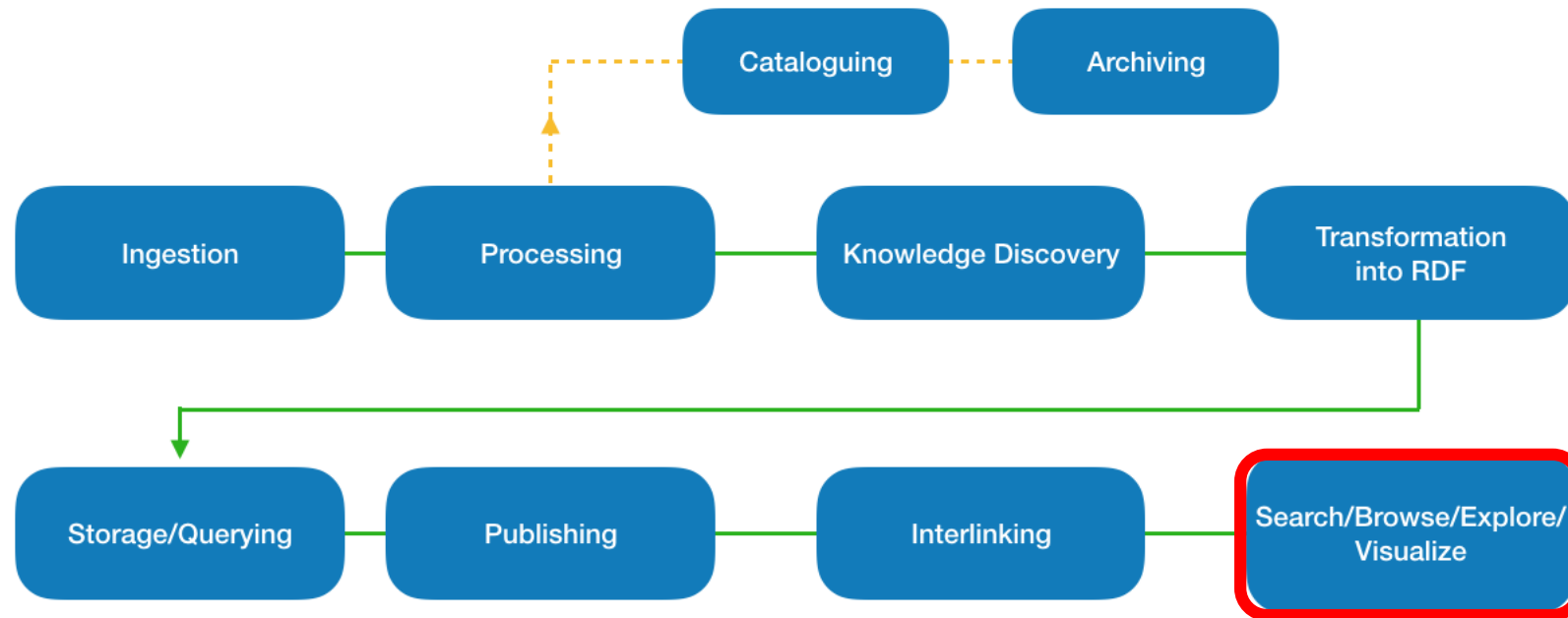
ontop
Spatial

Open Problems

- The best of the previous RDF stores and OBDA systems can scale only up to **hundreds of GBs** of geospatial data.
- **How can we scale to the PBs** of geospatial data in a satellite archive or a national cartographic agency?
 - There is work on relational geospatial big data systems (based on Hadoop, Spark and key-value stores) that can be exploited.



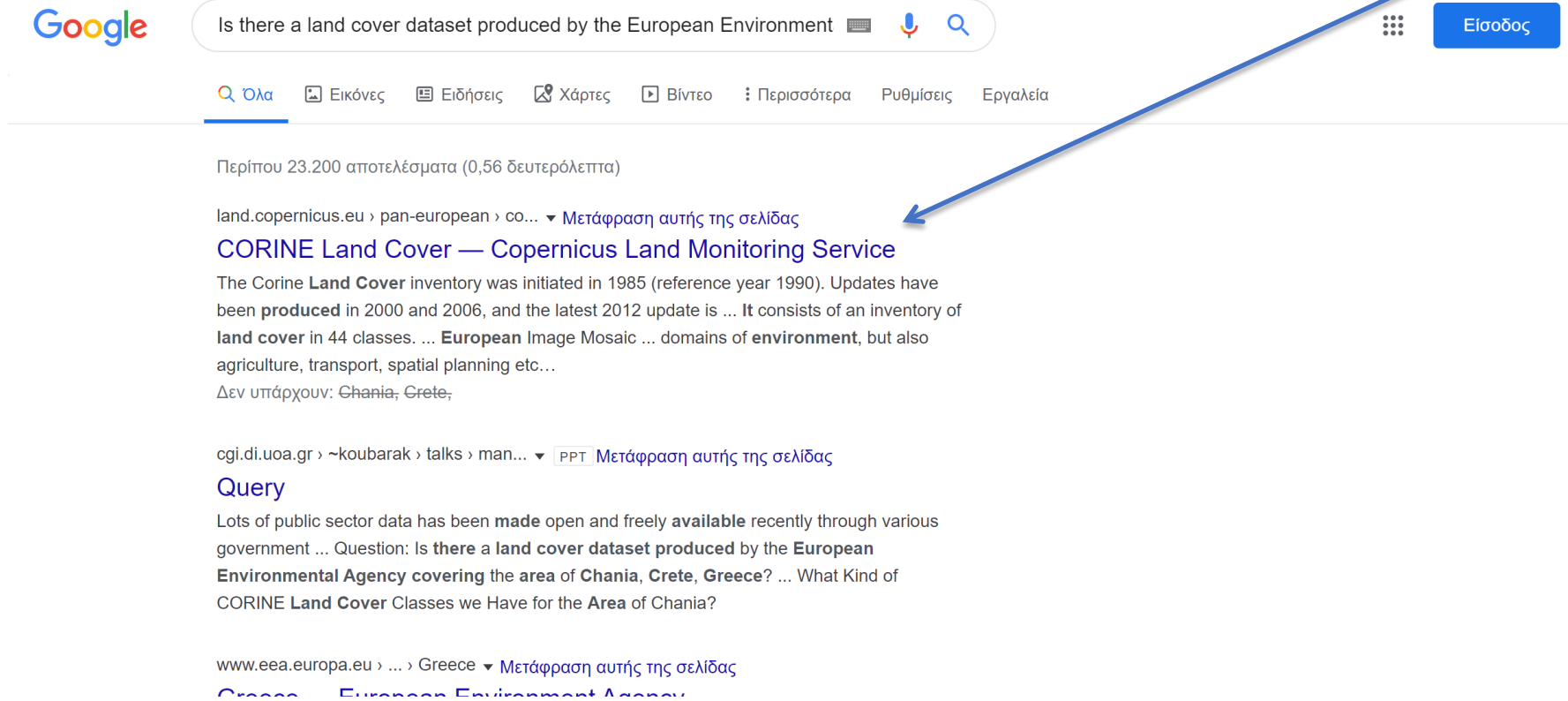
A Data Science Pipeline for Big Linked EO Data (cont'd)



Let us search for a land cover dataset

- **Question:** Is there a land cover dataset produced by the European Environmental Agency covering the area of Chania, Crete, Greece?
- Google it!

Answer



Google search results for the query: "Is there a land cover dataset produced by the European Environment". The top result is from land.copernicus.eu, titled "CORINE Land Cover — Copernicus Land Monitoring Service". A blue arrow points from the word "Perfect!" to this result. The snippet below the title reads: "The Corine Land Cover inventory was initiated in 1985 (reference year 1990). Updates have been produced in 2000 and 2006, and the latest 2012 update is ... It consists of an inventory of land cover in 44 classes. ... European Image Mosaic ... domains of environment, but also agriculture, transport, spatial planning etc... Δεν υπάρχουν: Χανιά, Crete,". The second result is from cgi.di.uoa.gr, titled "Query", with a snippet: "Lots of public sector data has been made open and freely available recently through various government ... Question: Is there a land cover dataset produced by the European Environmental Agency covering the area of Chania, Crete, Greece? ... What Kind of CORINE Land Cover Classes we Have for the Area of Chania?". The third result is from www.eea.europa.eu, titled "Greece — European Environment Agency".

Perfect!

Google Dataset Search

- Google has deployed a **Dataset Search** service and issued **guidelines for annotating public datasets** so they can more easily discovered by search engines.
- See <https://datasetsearch.research.google.com/> .
- **So if one follows their guidelines, then ...**

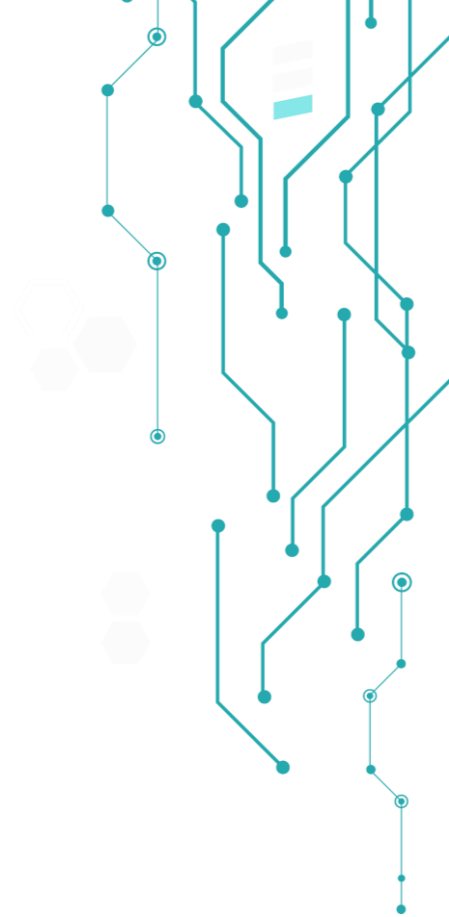
... One Can Use Google Dataset Search to Find the Dataset ...

Perfect!

The screenshot shows a Google Dataset Search interface. The search bar contains 'CORINE Land Cover 2018'. Below the search bar are filters for 'Updated Date', 'Download Format', 'Usage Rights', and 'Free'. The results section shows '100+ datasets found'. The top result is highlighted with a blue arrow pointing to it from the word 'Perfect!' above. This result is '[LATEST VERSION] Copernicus Land Monitoring Service - Corine Land Cover' from 'data.europa.eu', updated on Jan 13, 2019. A blue button labeled 'Explore at data.europa.eu' is visible. Below the main result are two other entries: 'Corine Land Cover 2018' from 'hub.arcgis.com' (updated Dec 19, 2018) and 'Corine Land Cover 2018 (vector) - version 20, Jun. 2019' from 'sdi.eea.europa.eu'. The detailed view for the top result shows it was updated on Jan 13, 2019, provided by the Publications Office of the European Union, and includes a description of the CORINE Land Cover (CLC) inventory.

... on the EU Open Data Portal

The screenshot shows a web browser window with the URL `data.europa.eu/euodp/data/dataset/data_copernicus-land-monitoring-service-corine`. The page header includes the EU Open Data Portal logo and navigation links for Sitemap, Legal notice, Contact, and a language dropdown set to English (en). The breadcrumb trail reads: EUROPA > Open Data Portal > Data > Publisher > European Environment Agency > [LATEST VERSION] Copernicus Land... A navigation bar contains links for Home, Data, Applications, Linked Data, Visualisations - beta, Developers' corner, and About. The main content area features a search bar with the text "Search datasets..." and a magnifying glass icon. Below the search bar, there are radio buttons for "Show results with:" options: "all of these words", "any of these words", and "the exact phrase". To the right, a "Suggest a dataset" section asks "Is there data you would like to find on the portal?" and includes a "Make a suggestion" button. The dataset title is "[LATEST VERSION] Copernicus Land Monitoring Service - Corine Land Cover" with social media icons. The publisher is listed as "European Environment Agency". On the right side, there are sections for "Keywords" (land cover, land use) and "EuroVoc concepts".



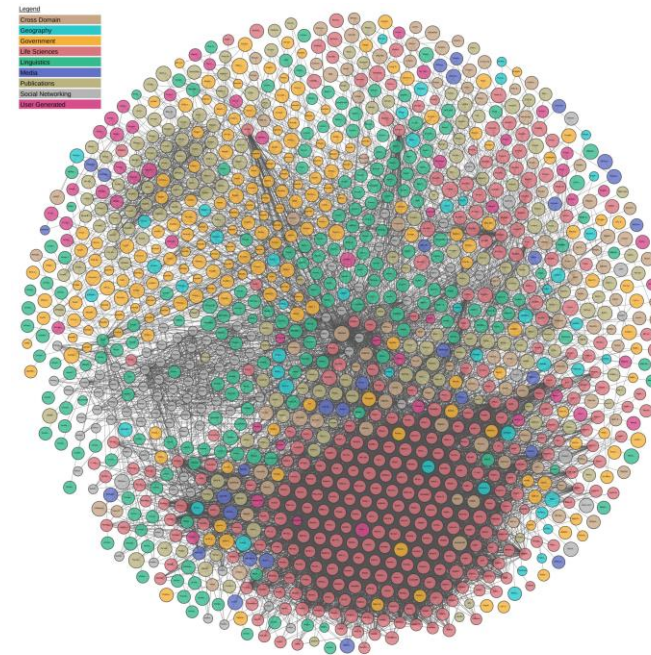
But This is Still Rather an Exception ...

- All Copernicus data and products still exists in different **data silos** (e.g., different EO archives or portals).
- Current search engines **do not index this data** in an effective way.



Main Objective of our Work in TELEIOS, LEO, Melodies and Copernicus App Lab

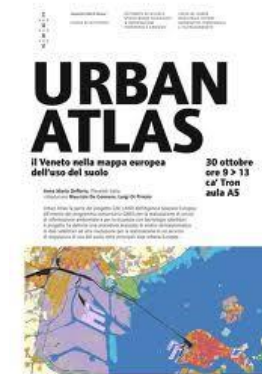
- Open up EO data silos by moving their data and/or metadata over to the **linked data paradigm**.



Examples of Linked Open EO Data

- CORINE land cover of the year 2012
- Urban Atlas of the year 2012

<https://ai.di.uoa.gr/#datasets>

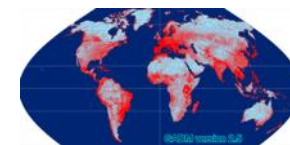


Examples of Interesting Linkages

- The CORINE land cover dataset can be usefully linked with the following datasets:
 - GeoNames
 - Global Administrative Areas
 - DBpedia
 - OpenStreetMap



GeoNames



What Kind of CORINE Land Cover Classes we Have for the Area of Chania?

PREFIX corine: <http://geo.linkedopendata.gr/corine/ontology#>

PREFIX strdf: <http://strdf.di.uoa.gr/ontology#>

SELECT ?lu (COUNT(?lu) AS ?instances)

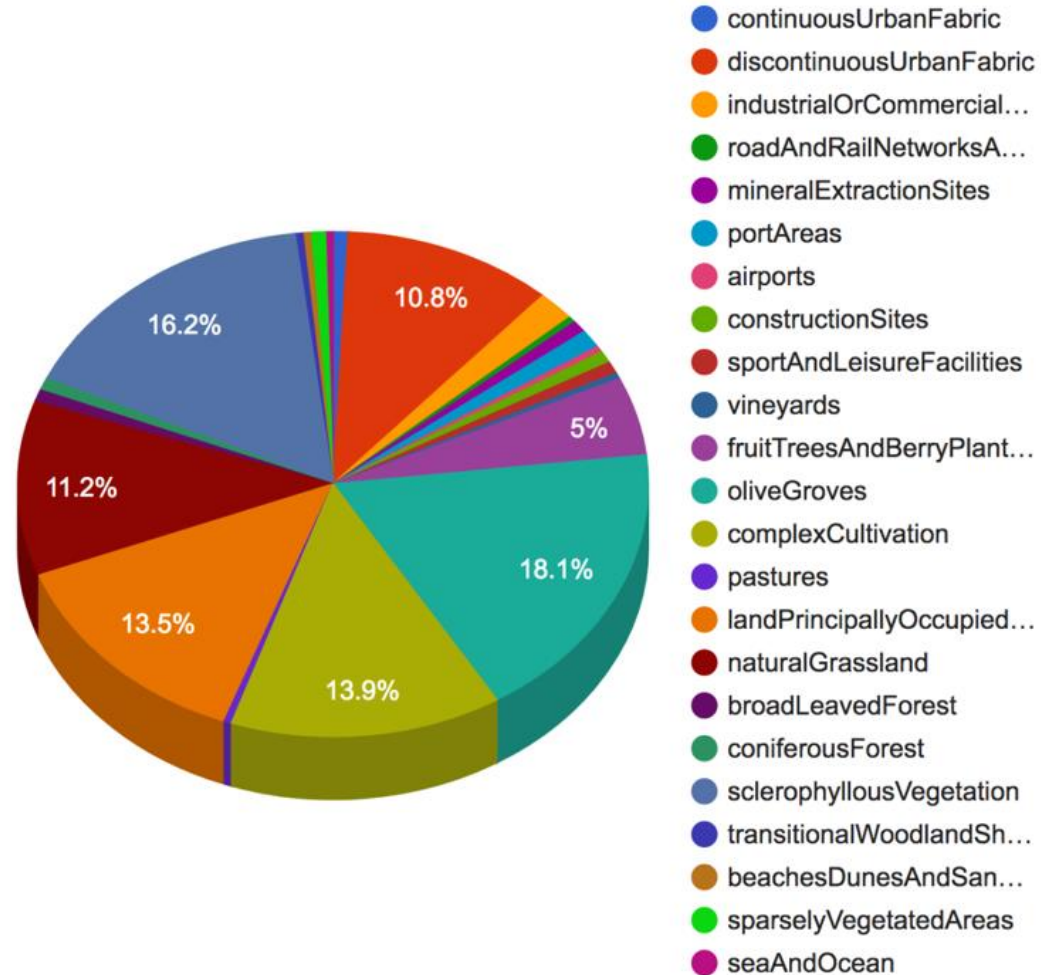
WHERE

```
{  
  ?area corine:hasLandUse ?lu .  
  ?area corine:hasGeometry ?geometry .  
  FILTER (strdf:intersects(?geometry, "POLYGON ((24.28 35.42, 23.89 35.42, 23.89 35.61,  
24.28 35.61, 24.28 35.42))"^^strdf:WKT)) .  
}
```

GROUP BY ?lu

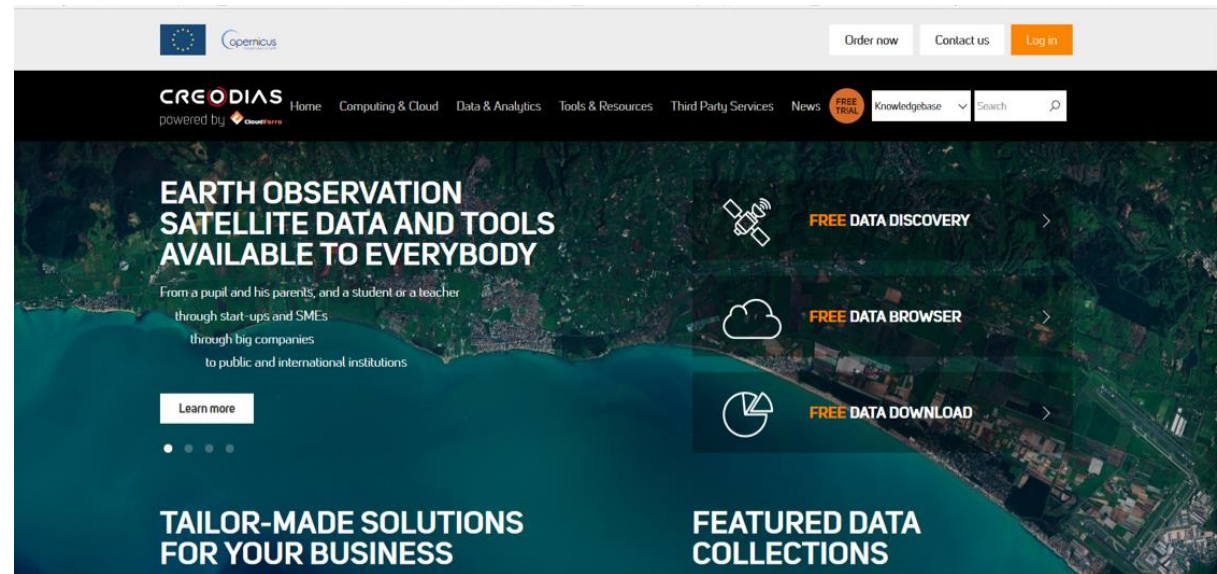


Answer

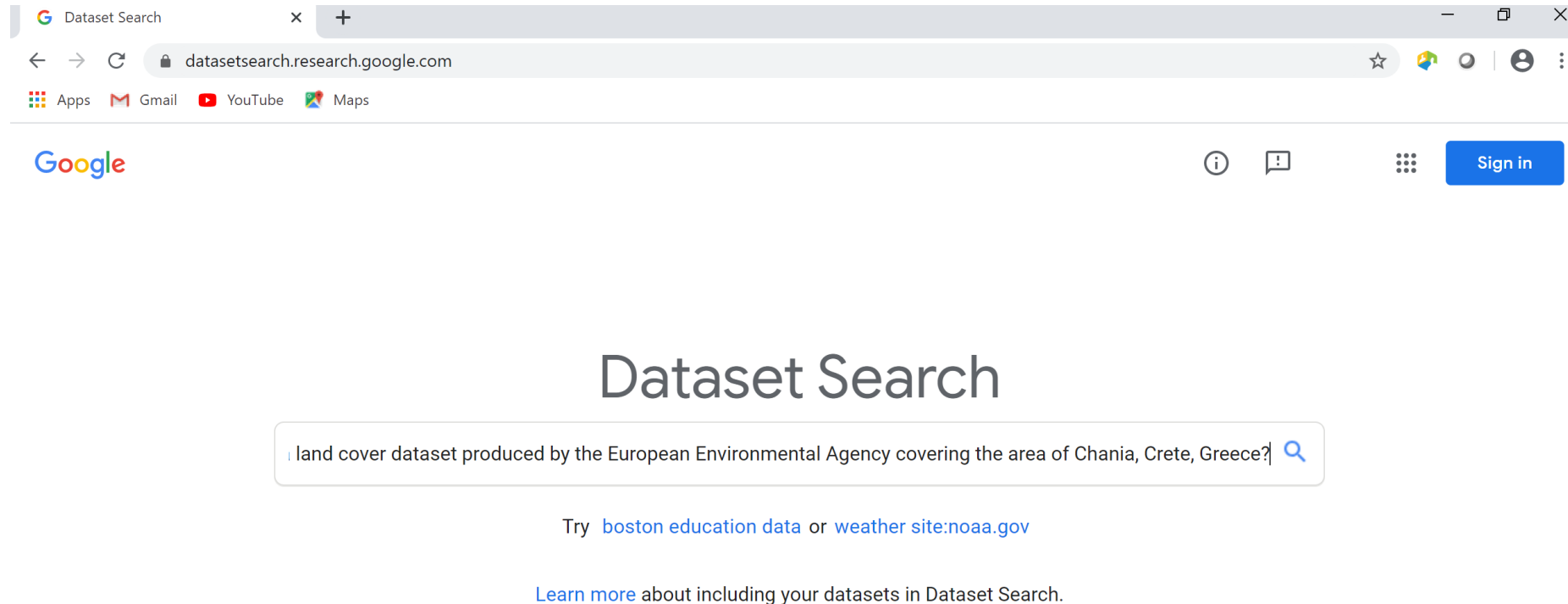


Copernicus Data and Information Access Services (DIAS)

- Five DIAS now in operation (and further development)
- **One of them uses linked data for its catalogue:**
<https://creodias.eu/>



But Notice: No Question Answering in Google Dataset Search (yet!)



Dataset Search

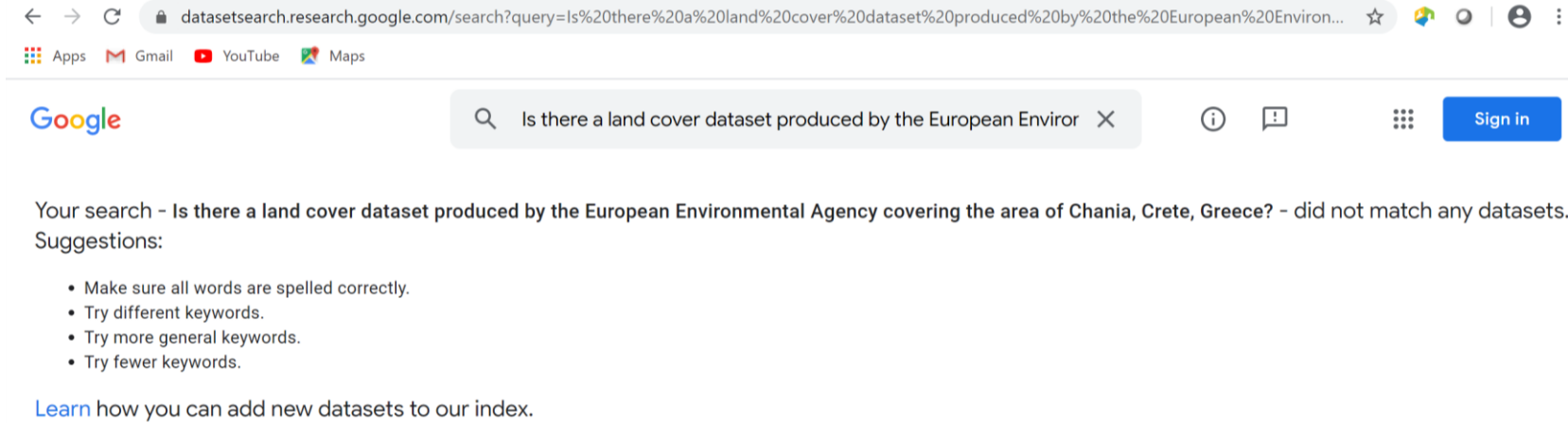
land cover dataset produced by the European Environmental Agency covering the area of Chania, Crete, Greece?

Try [boston education data](#) or [weather site:noaa.gov](#)

[Learn more](#) about including your datasets in Dataset Search.



Answer



datasetsearch.research.google.com/search?query=Is%20there%20a%20land%20cover%20dataset%20produced%20by%20the%20European%20Environ...

Apps Gmail YouTube Maps

Google

Is there a land cover dataset produced by the European Environor X

Sign in

Your search - Is there a land cover dataset produced by the European Environmental Agency covering the area of Chania, Crete, Greece? - did not match any datasets.

Suggestions:

- Make sure all words are spelled correctly.
- Try different keywords.
- Try more general keywords.
- Try fewer keywords.

[Learn](#) how you can add new datasets to our index.

Knowledge Graphs

- There is currently a lot of work in the development of **large knowledge graphs**:
 - Google Knowledge Graph
 - Bing Knowledge Graph
 - Wikidata
 - DBpedia
 - Yago
- There is also work on **scientific knowledge graphs**, some of them by publishers like Elsevier and Springer. See <https://projects.tib.eu/orkg/> for an ERC project in this area.

Earth Observation Knowledge Graphs

- There is currently **no work** on **Earth observation knowledge graphs** although ESA has funded a number of projects on ontologies for Earth observation (SMAAD, OTE, OTEG, RARE, ProdTrees and OBEOS).
- However, there is work on **environmental and EO ontologies**:
 - CSCDA
 - GEMET
 - NASA GCMD
 - GEOSS EO vocabulary
 - FedEO ontology
- There is also work on relevant **OGC standards**:
 - EO Metadata Profile of Observations and Measurements (2016)
 - EO Dataset Metadata Vocabulary (2018)
 - EO Product Metadata GeoJSON/JSON-LD Encoding (forthcoming)

Knowledge in the EO KGs

- **General knowledge** about satellite remote sensing and its applications:
 - Radar remote sensing is a kind of satellite remote sensing
 - Flood emergencies typically use radar images.
- **Knowledge about EO programs** like Copernicus and specific satellites like the Sentinels
 - Sentinel 1A is a radar satellite.
- **Knowledge about EO datasets**
 - CORINE land cover 2012 is a dataset produced using images from satellites ResourceSat-1 and RapidEye.
 - CORINE land cover 2012 covers the following countries: Albania, Austria, ..., United Kingdom.
 - The temporal coverage of CLC 2012 is the time interval 2011-2012.
 - CLC is an acronym for “CORINE land cover”.
 - CORINE land cover 2012 is available on the following URL: <https://land.copernicus.eu/pan-european/corine-land-cover/clc-2012>
- **Geospatial and temporal knowledge**
 - Chania is a city in Crete which is an island of Greece.
 - The WKT encoding of the geometry of Chania is “MULTIPOLYGON(...)”.
 - 2011 is included in 2011-2012.
- **Knowledge about publications**
 - The publication “The lifecycle of big linked EO data” uses the dataset CLC 2012.

Open Problems

- **Extensions of schema.org** for annotating EO datasets
- **EO KG construction and maintenance**
- **Question answering** over EO KGs



The ExtremeEarth project (<http://earthanalytics.eu/>)



- The ExtremeEarth project started on January 1, 2019 and will last for 3 years. Main emphasis:
 - Distributed deep learning for satellite images
 - Big linked geospatial data
 - Two use cases (Food Security and Polar) in the context of the two respective ESA Thematic Exploitation Platforms.





The ExtremeEarth Partners

- National and Kapodistrian University of Athens (co-ordinator)
- VISTA
- The Arctic University of Norway
- University of Trento
- KTH
- NCSR Demokritos
- DLR
- Polar View
- Norwegian Meteorological Institute
- Logical Clocks
- British Antarctic Survey



food security

tep



The Food Security Use Case

- The objective of the Food Security use case is to provide **water availability maps** for selected agricultural areas, allowing field level irrigation support.
- Information is based on the catchment wide assessment of the water, including seasonal storage as snow.
- The water availability maps will be made available to **farmers and decision makers** in agriculture, using the Food Security TEP.
- Big EO data processing, crop type information derivation using deep learning and water-to-plant modelling are applied.
- The focus will be the **Danube and Duero river catchments**.

The Polar Use Case

- To produce **high resolution ice charts** in a semi-automatic fashion from massive volumes of heterogeneous Copernicus data.
- The charts will be made available as linked data and will be combined with other information such as sea surface temperature and wind information for **informing maritime users and Polar TEP users.**



Main Research Results Up to Now



- The **ExtremeEarth platform**: a platform for distributed deep learning for satellite images and big linked geospatial data based on **Hops** (<https://www.logicalclocks.com/>).
- **New large training datasets**:
 - 2,632,230 pixels of Sentinel-2 images labelled with a single label chosen from a set of 17 crop types.
 - 263,541 patches of Sentinel-1 images labelled as “sea” or “ice”.
- **New deep neural network architectures**:
 - A multi-layer LSTM architecture for crop type classification at pixel level.
 - Two CNN architectures for sea/ice classification.
 - A two-step sea/ice classification architecture based on a variational autoencoder for feature learning, and a supervised method for classification.
- New **highly scalable** versions of the **linked geospatial data tools** GeoTriples, JedAI, Strabon and SemaGrow.

AI4EU (<https://www.ai4eu.eu/>)

- AI4EU started in January 1, 2019 and will last for 3 years. It brings together **81 partners from 21 European countries**, and has a budget of 20M Euros.
- AI4EU is creating an **AI ecosystem** in Europe, it is developing a **European AI-on-demand platform**, it is preparing a **strategic research and innovation agenda for AI** and more!
- Copernicus data will be accessed through the AI4EU platform by building interfaces to **Mundi** and **WEkEO DIAS**.
- The **AI4Agriculture** pilot is using Copernicus data.



Conclusions

- **AI can benefit Copernicus and Copernicus can benefit AI.**
- Some AI technologies are already **mature** enough to be deployed in the Copernicus Ground Segment (ontologies, linked geospatial data, semantic annotation).
- **More European research and development** in AI technologies is needed so that Copernicus can benefit even more.

Thanks! Questions?

- Thanks to all my colleagues for their contributions.
- For more, see the web page of my group <http://ai.di.uoa.gr> and my personal web page <http://www.di.uoa.gr/~koubarak> .
- Follow us on Twitter!



[@ExtremeEarth_EU](https://twitter.com/ExtremeEarth_EU)

[@mkoubarakis](https://twitter.com/mkoubarakis)



[@AI4EU](https://twitter.com/AI4EU)



Manolis Koubarakis

Professor

National and Kapodistrian
University of Athens