Data Profiling

Alessandro Mosca
KRDB Research Centre for Knowledge and Data
Smart Data Factory
Free University of Bozen-Bolzano, Italy

"Data Curation"
Master in Computational Data Science
Faculty of Computer Science

A.Y. 2022-2023

unibz

KRDB

SMART DATA FACTORY

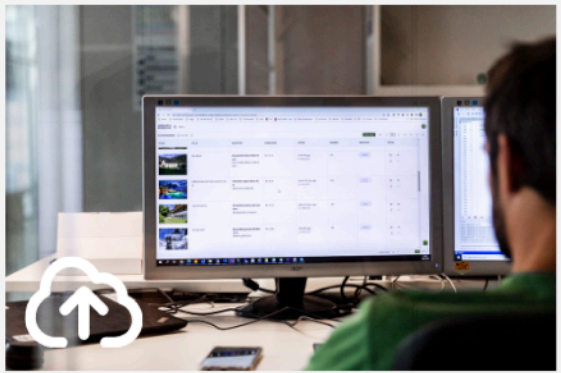
Open Data Hub

https://opendatahub.com/data-access/

OPEN DATA HUB | Services | Datasets | Use Cases | Web Components | Quickstart | Community | Events | Contact us

Data Access: understand now how to build on open data.

If you develop software applications we provide access to open data, through a **machine-readable**, documented and **stable channel**, to updated and **real-time data**, released under an Open Data licence.



Accessing data as a registered data consumer, you will get

- single machine-readable data
- documented and stable channel
- updated and real-time data
- visibility as a data consumer in the Open Data Hub community
- mailing with regular communications
- early communication in case of breaking changes of the datasets
- direct contact to data providers upon request

Join the community →

Datasets

Build your next service accessing **a growing number of datasets**. Get a quick overview on the data we provide. Datasets mostly fall in either Mobility and Tourism domains. Some data are available on request only.

All Datasets | Open Data

Articles →

Tourism | Open Data

Bike Charger →

Mobility | Open Data

Bike Counter →

Mobility | Open Data

Bikesharing →

Mobility | Open Data

Bluetooth Traffic Sensors →

Traffic | Open Data

Carpooling →

Mobility | Open Data

Carsharing →

Mobility | Open Data

Creative Industries →

Others | Open Data

District & Municipalities →

Echarging station →

Events NOI & EURAC →

Locations →

https://docs.opendatahub.com/en/latest/datasets.html

Domains and Datasets

Changed in version 2022.03: Move Domain description in this section

What is a Domain?

A Domain is a category that contains entities that are closely related. In the Open Data Hub, each domain roughly identifies one social or economical category; the domains intended as sources for data served by the Open Data Hub are depicted at the bottom of Figure 1.

Currently, the domains that can be accessed through the Open Data Hub are:

1. Mobility, this domain contains data about public transportation, parkings, charging station, and so on.
2. Tourism: data about events, accomodations, points of interest, and so on.
3. Other domain: a special category that encompasses domains that do not fall in any of the above category.

Each domain is composed by datasets, each of which contains data that provide useful information for the domain.

It is important to note that there is no clear separation between two domains. For example, data about public transportation belong to the Mobility domain, but are also useful for the Tourism domain.

The goal of the Open Data Hub project is to make available datasets containing data about the South Tyrolean ecosystem, to allow third parties to develop novel applications on top of them, consuming the exposed data. These applications may range from a

https://docs.opendatahub.com/en/latest/howto.html

List of HOWTOs

This page contains the list of available howtos, divided into areas. The list of howtos, together with a short description is available here:

Mobility

1. How to Access Mobility Data With API v2 Getting started with the new API v2.
2. How to Access Analytics Data in the Mobility Domain This howto guides you in browse and query data produces from the sensors used in the mobility domain.
3. How to access e-Charging Stations Data? This howto is deprecated and has been removed, please refer to How to Access Mobility Data With API v2 if you are interested in accessing the mobility dataset.

Tourism

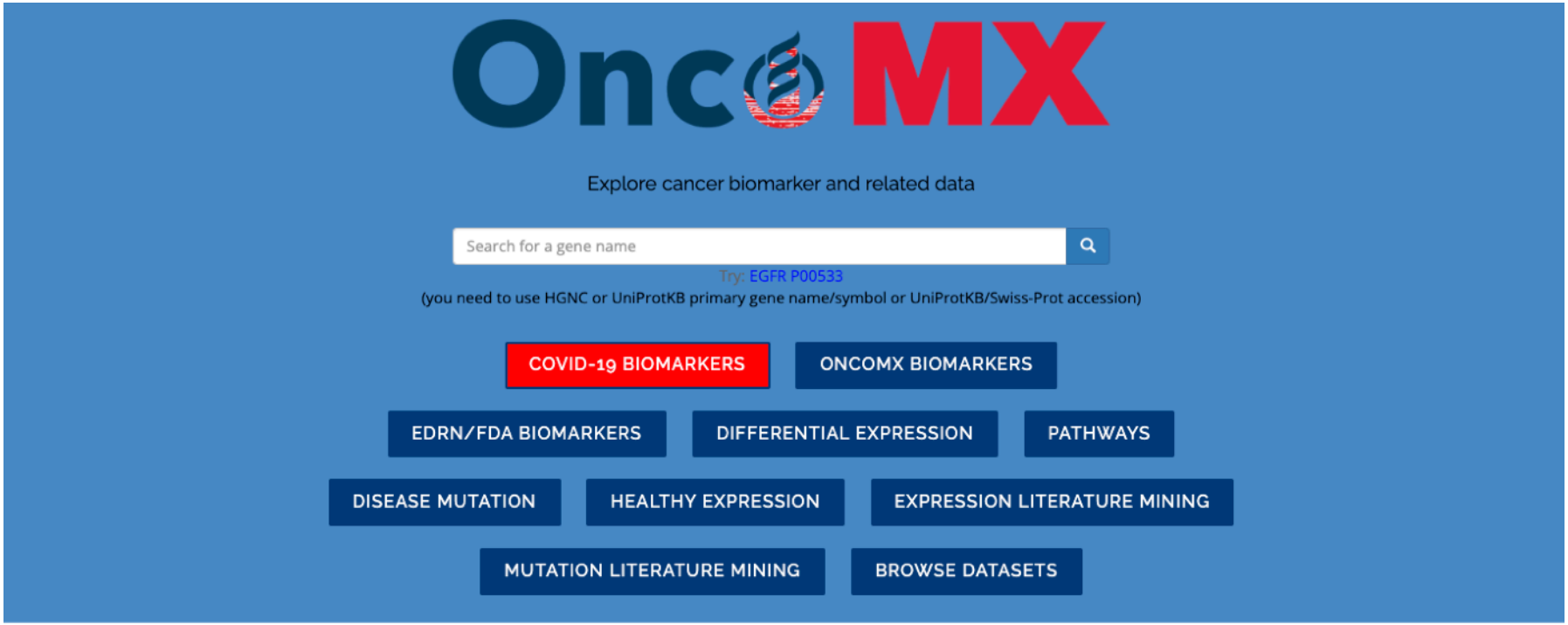
1. How to access Tourism Data? Description of how to access and manipulate data, the input data used, and the various filters available in the Tourism domain.
2. How to use the Open Data Hub's Tourism Data Browser? Access to the open data provided within the Tourism domain.
3. Quick and (not-so) Dirty Tips for Tourism (AKA Mini-howtos) Mini howtos, tricks&tips, and use cases for data in the Tourism domain.
4. How to access Open Data Hub AlpineBits Server as a client access AlpineBits data

Generic HOWTOs

1. How to use authentication? Access to data that are not yet open (mostly for internal use).
2. How to set up your local Development Environment? Set up your workstation to develop for Open Data Hub–This tutorial is still in development and not so useful at the moment!
3. GITHUB Quick Howto get started with GitHub workflow
4. How to set up Postman (API Development Environment)? Setup of Postman, a popular API development environment, to access data from the Open Data Hub.
5. How to insert and modify NOI Events? Create and modify NOI events directly from the Open Data Hub portal.
6. How to Publish your Web Component on the Open Data Hub Store share your Web Components with Open Data Hub team
7. How to Access Open Data Hub Data Using SPARQL query Open Data Hub datasets using W3C's Query Language.
8. How to Access Open Data Hub Data With R and SPARQL query Open Data Hub datasets using R and its SPARQL library.

OncoMX

https://oncomx.org/



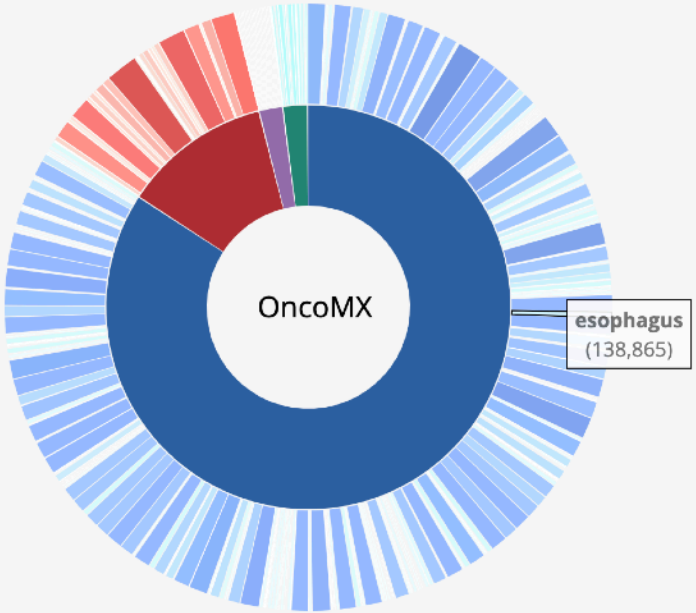
View OncoMX data across all data sources. This includes data from EDRN, Bgee, BioXpress, Reactome, and BioMuta.

Perspectives

- Biomarkers
- Evolutionary Context
- Literature Mining
- Biomarkers within Pathways

Data Sources

- Bgee
- BioMuta
- BioXpress
- DEXTER
- DiMeX
- EDRN
- Reactome



Statistics at a Glance

- Cancer Terms
- Bgee Stats
- Biomarker Stats
- BioMuta Stats
- BioXpress Stats
- Proteins
- Reactome Stats

News

- 2020 ITCR PI Meeting Virtual Poster #NCIITCR20
- OncoMX: A Knowledgebase for Exploring Cancer Biomarkers in the Context of Related Cancer and Healthy Data

OncoMX

OncoMX is an integrated cancer mutation and expression resource for exploring cancer biomarkers alongside related experimental data and functional information. OncoMX is a collaboration between the George Washington University (GW), NASA's Jet Propulsion Laboratory (JPL), the Swiss Institute of Bioinformatics (SIB), and the University of Delaware (UD). The core knowledgebase of OncoMX is derived from integrated cancer mutation (BioMuta) and expression (BioXpress) knowledgebases developed at GW. This collection of cancer data is mapped to biomarkers from EDRN and other sources, and augmented by normal expression data from Bgee and custom literature mining for mutation (DiMeX) and expression (DEXTER) software developed at UD. Combining this information is expected to improve functional interpretation of the reported variants, expression profiles, and other evidence of cancer involvement. Resulting data are wrapped into the OncoMX database, mapped to additional functional information and made available through the web portal. Here, a user can explore cancer data from various perspectives: for example, the biomarker vizualization enables a a user to see all attributes related to any biomarker annotations for a gene of interest. All datasets used by OncoMX, as well as additional custom datasets, are made available through the parallel data portal at www.data.oncomx.org.

External Downloads

Advanced API coming soon...

BIOMUTA | BIOXPRESS | EDRN | REACTOME | BGEE | SCRNA

FDA APPROVED BIOMARKERS | BGEE MOUSE | HCC N-GLYCANS

*Click on data.oncomx.org to see additional datasets for FDA approved biomarkers, neoepitope data and other custom datasets.

https://data.oncomx.org/

Dataset Collection

Webs

Search term

Search

You can search by BCO ID, name or contributor.

Total of 76 datasets (2 passed filter)

Filter by species
- [ ] Homo sapiens (71)
- [x] Mus musculus (3)
- [ ] Other sp... (2)

Filter by disease_status
- [ ] cancer/normal (2)
- [x] normal (5)
- [ ] cancer (64)
- [x] covid19 (1)
- [ ] thrombosis (2)
- [ ] Other di... (2)

Filter by scope
- [x] internal (74)
- [x] Other sc... (2)

Filter by file_format
- [x] csv (74)
- [x] Other fi... (2)

Filter by dataset_status
- [x] reviewed (74)
- [x] Other da... (2)

Filter by study_type
- [x] expression (24)
- [x] clinical data (3)
- [x] map file (1)
- [x] neoepitope (1)
- [x] biomarker (7)
- [x] literature mining (2)
- [x] mutation (4)
- [x] biomarkers (26)
- [x] glycans (1)
- [x] functional elements (1)
- [x] glycosyltransferases (2)
- [x] microrna (1)
- [x] mutations (1)
- [x] Other st... (2)

Genes normally expressed in mouse tissues (Bgee)

gene_symbol...	ensembl_gene...
Gnai3	ENSMUSG00000000001...
Gnai3	ENSMUSG00000000001...

List of mouse [taxid:10090] genes with normal RNA-Seq and Affymetrix expression data in Bgee; addit ...

view details

Genes normally expressed in mouse tissues (Customized)

gene_symbol...	human_orthol...
Gnai3	GNAI3...
Gnai3	GNAI3...

List of mouse [taxid:10090] genes with normal RNA-Seq and Affymetrix expression data from Bgee for a ...

view details

Genes normally expressed in mouse tissues (Bgee)

List of mouse [taxid:10090] genes with normal RNA-Seq and Affymetrix expression data in Bgee; additional documentation available at https://github.com/BgeeDB/bgee_pipeline/tree/develop/pipeline/collaboration/oncoMX#information-about-the-files-generated-for-oncomx

Version

v-1.0.36 06/13/2022

BCO JSON | README | DOWNLOAD

gene_symbol	ensembl_gene_id	uberon_anatomical_id	uberon_anatomical_name	uberon_developmental_id	uberon_developmental_name	expression_level_gene_relative	expression_level_anatomical_relative	call_quality	expression_rank_
Gnai3	ENSMUSG00000000001	UBERON:0002371	bone marrow	MmusDv:0000061	early adult stage	HIGH	HIGH	GOLD	856
Gnai3	ENSMUSG00000000001	UBERON:0000998	seminal vesicle	MmusDv:0000061	early adult stage	HIGH	HIGH	SILVER	857
Gnai3	ENSMUSG00000000001	UBERON:0002369	adrenal gland	MmusDv:0000061	early adult stage	HIGH	HIGH	GOLD	871
Gnai3	ENSMUSG00000000001	UBERON:0002370	thymus	MmusDv:0000061	early adult stage	HIGH	HIGH	GOLD	934
Gnai3	ENSMUSG00000000001	UBERON:0001068	skin of back	MmusDv:0000050	6 weeks (mouse)	HIGH	HIGH	SILVER	971
Gnai3	ENSMUSG00000000001	UBERON:0002116	ileum	UBERON:0000113	post-juvenile adult stage	HIGH	HIGH	GOLD	1.00e3
Gnai3	ENSMUSG00000000001	UBERON:0001068	skin of back	MmusDv:0000051	7 weeks (mouse)	HIGH	HIGH	SILVER	1.02e3
Gnai3	ENSMUSG00000000001	UBERON:0002115	jejunum	MmusDv:0000076	aged adult stage (mouse)	HIGH	HIGH	SILVER	1.02e3
Gnai3	ENSMUSG00000000001	UBERON:0002370	thymus	MmusDv:0000052	8 weeks (mouse)	HIGH	HIGH	SILVER	1.04e3
Gnai3	ENSMUSG00000000001	UBERON:0001043	esophagus	MmusDv:0000063	3 month-old stage (mouse)	HIGH	HIGH	SILVER	1.09e3
Gnai3	ENSMUSG00000000001	UBERON:0001155	colon	MmusDv:0000063	3 month-old stage (mouse)	HIGH	HIGH	GOLD	1.12e3
Gnai3	ENSMUSG00000000001	UBERON:0002115	jejunum	UBERON:0000113	post-juvenile adult stage	HIGH	HIGH	GOLD	1.12e3
Gnai3	ENSMUSG00000000001	UBERON:0001068	skin of back	MmusDv:0000052	8 weeks (mouse)	HIGH	HIGH	SILVER	1.13e3
Gnai3	ENSMUSG00000000001	UBERON:0000002	uterine cervix	MmusDv:0000050	6 weeks (mouse)	HIGH	HIGH	SILVER	1.14e3
Gnai3	ENSMUSG00000000001	UBERON:0001155	colon	MmusDv:0000053	9 weeks (mouse)	HIGH	HIGH	SILVER	1.14e3
Gnai3	ENSMUSG00000000001	UBERON:0002369	adrenal gland	MmusDv:0000050	6 weeks (mouse)	HIGH	HIGH	SILVER	1.15e3
Gnai3	ENSMUSG00000000001	UBERON:0001155	colon	MmusDv:0000051	7 weeks (mouse)	HIGH	HIGH	SILVER	1.16e3
Gnai3	ENSMUSG00000000001	UBERON:0002370	thymus	MmusDv:0000063	3 month-old stage (mouse)	HIGH	HIGH	SILVER	1.16e3
Gnai3	ENSMUSG00000000001	UBERON:0000995	uterus	MmusDv:0000061	early adult stage	HIGH	HIGH	GOLD	1.16e3
Gnai3	ENSMUSG00000000001	UBERON:0001155	colon	MmusDv:0000061	early adult stage	HIGH	HIGH	GOLD	1.18e3
Gnai3	ENSMUSG00000000001	CL:0002138	endothelial cell of lymphatic vessel	MmusDv:0000052	8 weeks (mouse)	HIGH	HIGH	SILVER	1.18e3
Gnai3	ENSMUSG00000000001	UBERON:0000995	uterus	UBERON:0000113	post-juvenile adult stage	HIGH	HIGH	GOLD	1.28e3
Gnai3	ENSMUSG00000000001	UBERON:0001043	esophagus	UBERON:0000113	post-juvenile adult stage	HIGH	HIGH	GOLD	1.28e3
Gnai3	ENSMUSG00000000001	UBERON:0002369	adrenal gland	MmusDv:0000052	8 weeks (mouse)	HIGH	HIGH	BRONZE	1.29e3
Gnai3	ENSMUSG00000000001	UBERON:0004645	urinary bladder urothelium	MmusDv:0000064	4 month-old stage (mouse)	HIGH	HIGH	SILVER	1.30e3
Gnai3	ENSMUSG00000000001	UBERON:0001347	white adipose tissue	MmusDv:0000062	2 month-old stage (mouse)	HIGH	HIGH	SILVER	1.30e3
Gnai3	ENSMUSG00000000001	UBERON:0000995	uterus	MmusDv:0000050	6 weeks (mouse)	HIGH	HIGH	GOLD	1.30e3
Gnai3	ENSMUSG00000000001	UBERON:0000006	islet of Langerhans	MmusDv:0000061	early adult stage	HIGH	HIGH	GOLD	1.39e3
Gnai3	ENSMUSG00000000001	UBERON:0000992	female gonad	MmusDv:0000064	4 month-old stage (mouse)	HIGH	HIGH	SILVER	1.42e3
Gnai3	ENSMUSG00000000001	UBERON:0005200	thoracic mammary gland	MmusDv:0000064	4 month-old stage (mouse)	HIGH	HIGH	BRONZE	1.43e3
Gnai3	ENSMUSG00000000001	UBERON:0000992	female gonad	MmusDv:0000061	early adult stage	HIGH	HIGH	GOLD	1.43e3
Gnai3	ENSMUSG00000000001	UBERON:0002114	duodenum	MmusDv:0000051	7 weeks (mouse)	HIGH	HIGH	SILVER	1.43e3
Gnai3	ENSMUSG00000000001	UBERON:0004645	urinary bladder urothelium	MmusDv:0000097	late adult stage	HIGH	HIGH	SILVER	1.44e3
Gnai3	ENSMUSG00000000001	UBERON:0000945	stomach	MmusDv:0000062	2 month-old stage (mouse)	HIGH	HIGH	SILVER	1.45e3

CRISP

https://crisp.inf.unibz.it/

CRISP VKG

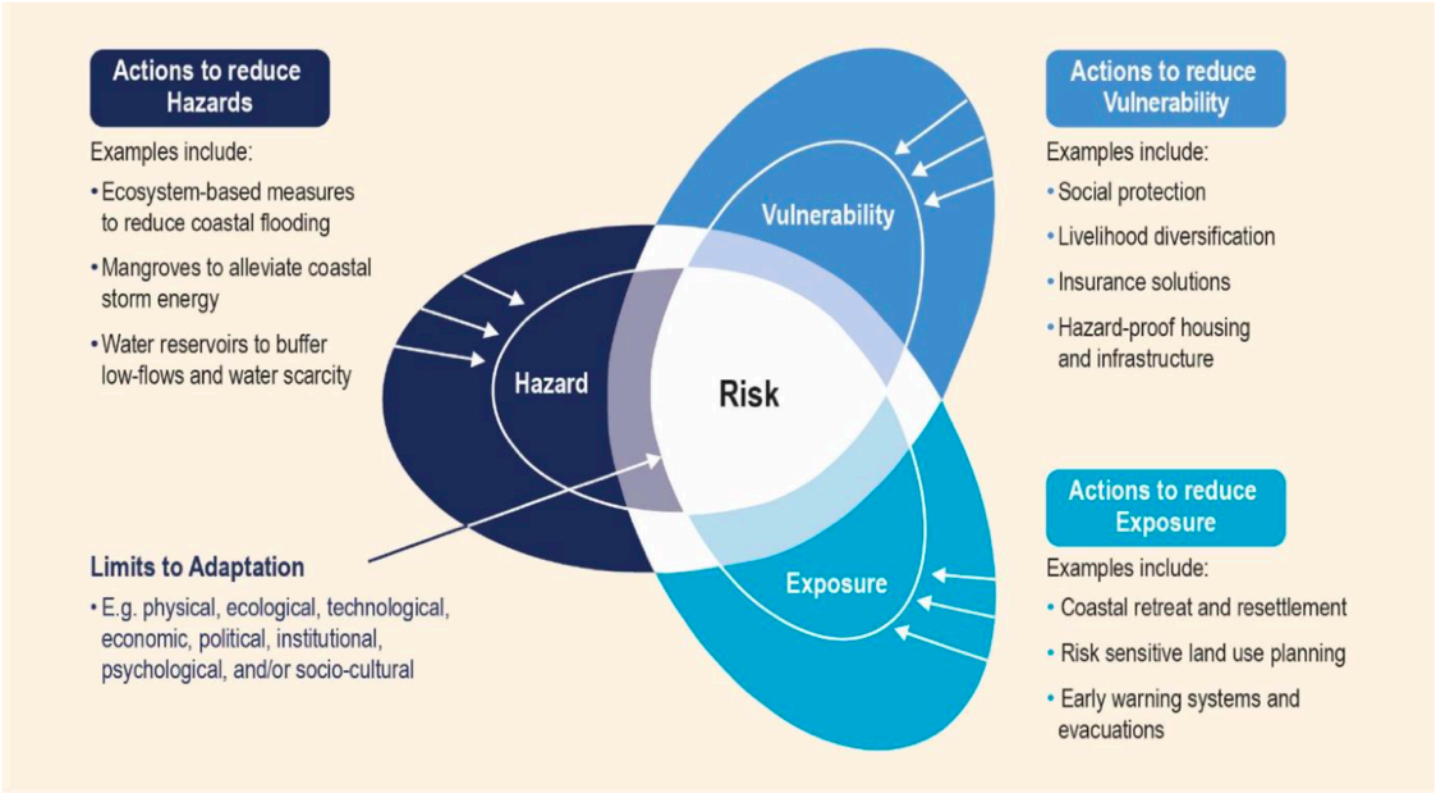
Climate Risk Planning & Managing Tool for Development Programmes in Agri-food Systems

The CRISP Virtual Knowledge Graph (VKG), powered by Ontop, provides an integrated RDF/OWL ontological view with SPARQL querying support of all data related to the CRISP project. Running from August 2021 to April 2023, CRISP is a project funded by GIZ (German Development Cooperation) that aims at building a freely available web-based tool for climate risk planning & managing, targeting agricultural & rural development project managers and grounded on the impact chain method developed by the Eurac Climate and Disaster Risk research group

This page provides access to the CRISP VKG *knowledge base*, to available *resources* and to related *links*.
Contact the project team for any enquiry.

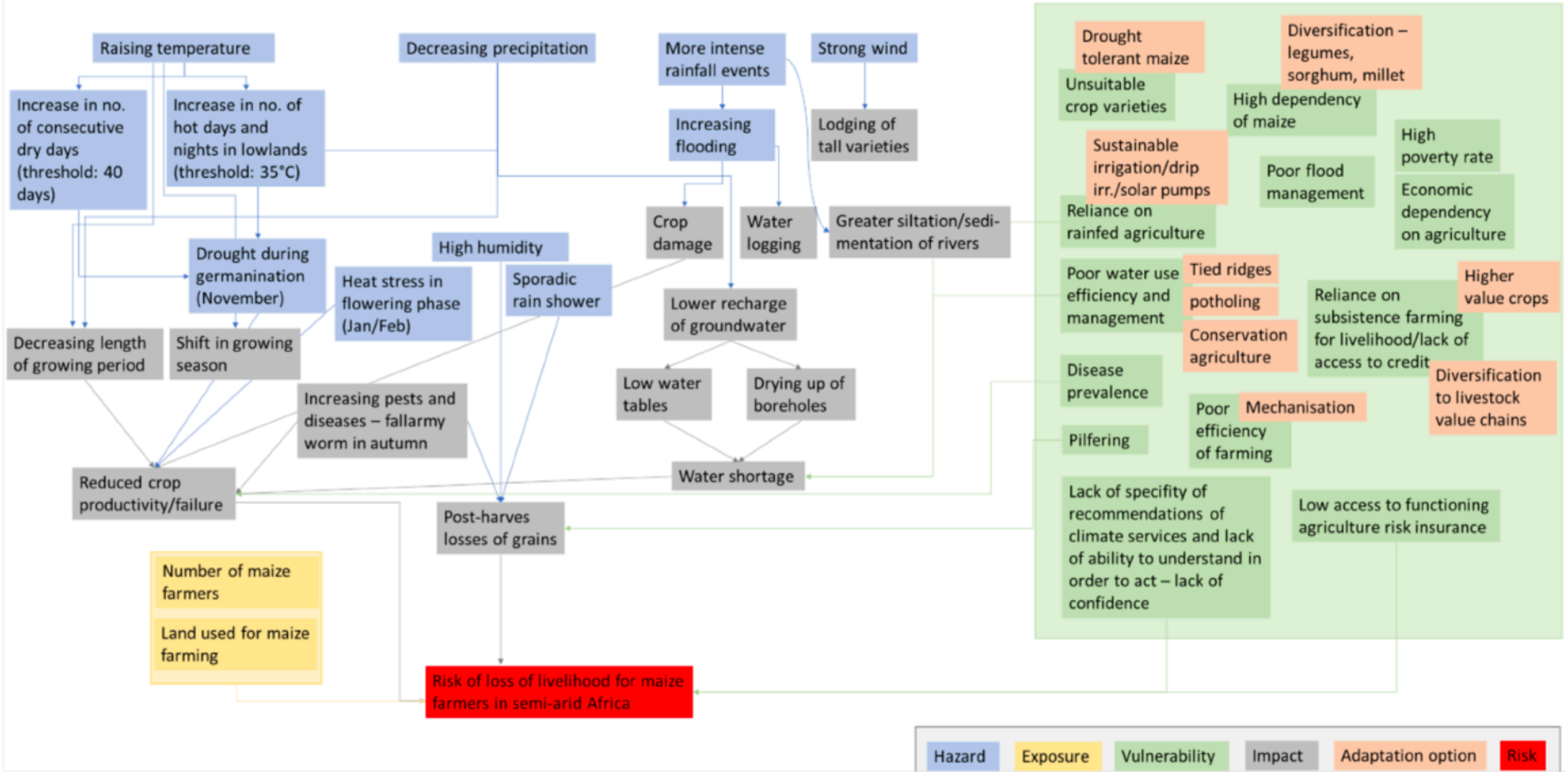
Explore the data | Example entities | Resources | Links

- Risk results from the interaction of **vulnerability, exposure, and hazard**



Adaptation can lower the risk by decreasing vulnerability and exposure (and hazard?)

Impact Chain Model instance



CONNECTIONS

	From	Link type	To	Description	Bibliographic resource	Tags
HAZARD TO BIOPHYSICAL IMPACT	Increase in drought events	Leads to	Reduced crop productivity/failure	Drought is expected to have an adverse impact on agricultural production and food security in the irrigated drylands of Central and South America. For example, in Chile, drought is projected to have a significant impact on the production of fruit farmers. In Mexico, farmers highlighted drought as the main cause for agricultural losses, with Baja California, Sonora, and Sinaloa being badly affected. In Central-West Argentina, drought (2010/2011- 2014/2015) resulted in water shortages, reducing grape and seasonal crop production as well as the loss of livestock, negatively impacting regional food security.	(Prager et al., 2020); (Hagen et al., 2022); (Ochoa-Noriega et al., 2020); (Rivera et al., 2021)	Rainfall related; Biophysical
	Increase in drought events	Leads to	Reduced livestock productivity	Drought is expected to have an adverse impact on livestock productivity in the irrigated drylands of Central and South America. For example, in Central-West Argentina, drought (2010/2011- 2014/2015) resulted in water shortages and livestock losses. In Mexico, drought in the arid and semi-arid regions has resulted in significant losses in the livestock sectors.	(Prager et al., 2020); (Pontifes et al., 2018)	Rainfall related; Biophysical
	Decrease in average precipitation	Leads to	Water scarcity	Decreasing rainfall contributes to water scarcity issues in the arid and semi-arid areas of the irrigated agricultural system.	(Lambrano et al., 2020); (Peña-Guerrero et al., 2020)	Rainfall related; Biophysical
	Increase in average temperatures	Leads to	Water scarcity	Higher temperatures contribute to water scarcity issues in the arid and semi-arid areas of the irrigated agricultural system. For example, in Chile, increasing temperatures are expected to intensify aridity in northern parts of the country as well as enhance the risk of water scarcity in central and southern areas.	(Lambrano et al., 2020); (Peña-Guerrero et al., 2020)	Temperature related; Biophysical
	Decrease in average precipitation	Leads to	Increase in soil water deficit	Decreases in average precipitation contribute to soil water deficits limiting the water available for plant growth.	(Lambrano et al., 2020)	Rainfall related; Biophysical
	Increase in drought events	Leads to	Increase in soil water deficit	Drought leads to soil water deficits limiting the water available for plant growth.	(Lambrano et al., 2020)	Rainfall related; Biophysical
	Increase in average temperatures	Leads to	Increase in soil water deficit	Increases in average temperatures increase evaporation and decrease soil water viscosity thereby restricting the water available for plant growth.	(Lambrano et al., 2020)	Temperature related; Biophysical
	Increase in average temperatures	Leads to	Reduced produce quality	With increases in temperature and solar radiation, plant respiration and photosynthesis rates decrease causing plants to use their carbon reserves, this can affect the sweetness of fruit and vegetables.	(Lambrano et al., 2020); (Prager et al., 2020)	Temperature related; Biophysical
	Increase in extreme rainfall events	Leads to	Increase in pests	In Peru and Chile the increasing frequency of extreme rainfall due to El Niño/La Niña events causes agricultural pest outbreaks.	(Prager et al., 2020); (Reyer et al., 2015)	Rainfall related; Biophysical; Pests and diseases
	Increase in average temperatures	Leads to	Increase in pests	Higher temperatures favour an increase in insect population densities in Mexico.	(Servin and Mendoza, 2014)	Temperature related; Biophysical; Pests and diseases
	Increase in extreme rainfall events	Leads to	Increased flooding	Increases in extreme rainfall events in the region will lead to flooding. For example, in northern Mexico, monsoon thunderstorms are responsible for locally severe weather and flooding. In Argentina, flooding due to extreme rainfall events has increased in recent years.	(Prager et al., 2020); (Reyer et al., 2015); (Hagen et al., 2022); (World Bank, 2021)	Rainfall related; Biophysical
	Increase in average temperatures	Leads to	Reduced crop productivity/failure	Higher temperatures can reduce fruit production and shelf-life, for example, in Peru, increasing temperature is causing a decline in banana production.	(Prager et al., 2020)	Temperature related; Biophysical
	Increase in extreme heat (no. of hot days)	Leads to	Reduced crop productivity/failure	High temperatures limit pollen production leading to reduced fruit set. Heat stress in fruit and vegetables typically includes scorching of leaves and stems, sunburn on fruits and stems, enhanced leaf senescence, rapid leaf drop, and reduced plant growth.	(Hagen et al., 2022); (World Bank, 2021)	Temperature related; Biophysical
	Increase in extreme heat (no. of hot days)	Leads to	Increase in heat stress	Cattle production is efficient within a thermoneutral zone of 20 °C (varies between 10 to 26 °C). When temperatures exceed 27 °C, especially if relative humidity is higher than 40%, cattle struggle to self-regulate, this is known as heat stress.	(Zazueta-Gutiérrez et al., 2021)	Temperature related; Biophysical
	Rising sea levels	Leads to	Increased flooding	Rising sea levels can cause flooding of coastal areas. For example, in Chile, flooding is expected to increase due to an increase in ice-sheet melting contributing to rising sea levels.	(Hagen et al., 2022); (Feld and Galiani, 2015)	Maritime related: Biophysical
	Rising sea levels	Leads to	Saline intrusion	Seawater intrusion contaminates coastal aquifers. For example, the Maneadero aquifer in Baja California of Mexico is severely contaminated, as well as the soil irrigated by the water, leading to the abandonment of farmlands.	(Bohn et al., 2018); (Mendoza-Espinosa and Daessle, 2018)	Maritime related: Biophysical
	Increase in extreme rainfall events	Leads to	Increase in diseases	In Peru and Chile the increasing frequency of El Niño/La Niña events due to climate change is expected to lead to an increase in agricultural disease outbreaks.	(Prager et al., 2020)	Rainfall related; Biophysical; Pests and diseases
	Increase in average temperatures	Leads to	Increase in diseases	Higher temperatures tend to increase the prevalence and range of livestock diseases, which negatively affect productivity and increase mortality rates.	(Gianella et al., 2019)	Temperature related; Biophysical; Pests and diseases

Contextual information: Agricultural system

Name of agri system: Irrigated
Overview of agri system (250-300 words)
The irrigated agricultural system covers large areas of arid land in Northern and Central Mexico and coastal and inland valley areas of Peru, Chile and Western Argentina. The total land area is close to 200 million ha (Dixon et al., 2001). As the name suggests, the agricultural system is irrigated, which allows for a relatively high degree of intensification. The system is commercially oriented involving the production of fruit, rice, cotton, horticulture and vines. Poverty levels in the irrigated agricultural system are generally low to moderate (Dixon et al., 2001). In Chile, approximately 52% of the coast is irrigated, with irrigation efficiency low at around 35–40% (World Bank et al., 2015). The main irrigation system used is gravitational flooding, which has a low irrigation efficiency (30%). The second-most adopted irrigation type is micro-irrigation (23%), followed by sprinkler irrigation (5%) (Donosa, 2021). The main irrigated crops in Central Chile are maize, rice, blueberries, raspberries, and perennials, such as apple and cherry trees. Fruit production, such as grape and avocado, are typically exported (Peña-Guerrero et al., 2020). Irrigated agriculture exerts considerable pressure on the country's water resources, especially in the coastal areas. In the northern areas of the Peru coastline, irrigated agricultural production includes the cultivation of asparagus, artichokes, peppers, avocados, pineapples, mangos and blueberries (Portugal, 2019). These products, destined for export markets, have resulted in agricultural expansion and an increased demand for irrigation. Investments in modern irrigation infrastructure by large-scale farms have increased groundwater abstraction (Salmoral et al., 2020). There is a clear difference in the level of modernisation between agro-export farms, which typically use automated high-efficiency drip irrigation systems, and small-scale farms that rely on low energy, lower efficiency surface irrigation systems. Small-scale farmers tend to obtain their water from surface water sources as they lack access to finance to drill boreholes or increase the depths of existing boreholes/wells to draw upon the falling water table (Salmoral et al., 2020). As in Peru and Chile, in Argentina, the irrigated farming system has allowed for the intensification and commercialisation of agricultural production, such as grape production (World Bank et al., 2014). In Mexico, the key fruits and vegetables grown for export markets include bell pepper, onion, spinach, cucumber, mango, papaya, banana and avocado. Increased demand for irrigation in conjunction with climate change has led to aquifer overexploitation and declines in regional groundwater levels.

		Bibliographic resource
Countries within the agri system	Mexico; Chile; Peru; Argentina	(Dixon et al., 2001)
Total area (m ha)	200	(Dixon et al., 2001)

Contextual information: Farming system

FARMING SYSTEM																		
Farming system	Landscape	Agro-ecological zone	Total area (m ha)	Cultivated area (m ha)	Cattle population (m head)	Irrigated area (m ha)	Total population (m)	Agricultural population (m)	Farm size	Dominant livelihood source	Irrigated/rainfed	Location: Macro region	Location: Macro Region (M49-code)	Location: Region	Location: Region (M49-code)	Location: Subregion	Location: Subregion (M49-code)	Location: Countries	Location: Countries (M49-code)
Irrigated (LAC)		Arid and semi-arid	200							Commercial horticulture, fruit and cattle	Irrigated	Latin America and Caribbean	419					Mexico; Chile; Peru; Argentina	484; 152; 604; 032

Contextual information: Commodity

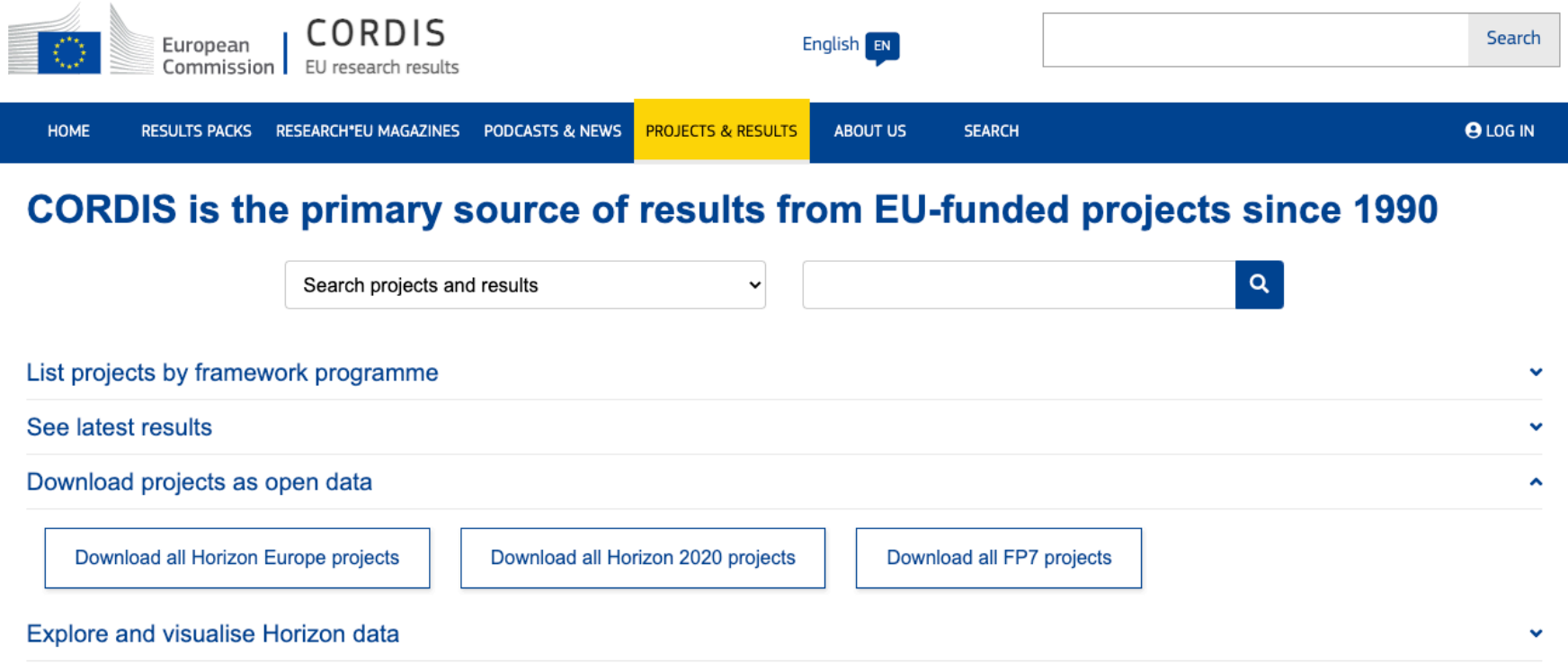
COMMODITY												
Name	NCBITaxonName	NCBITaxonID	Soil type	Max THI	Min temperature (°C)	Max temperature (°C)	Average temperature (°C)	Min precipitation (mm)	Max precipitation (mm)	Average precipitation (mm)	Min elevation (masl)	Max elevation (masl)	Bibliographic resource
Blueberries			Various		-3	30		250	1.000		100	625	(Banados, 2009); (Retamales et al., 2014); (Lyrene and Munoz, 2008)
Bananas	Musa x paradisiaca	89151	Loamy		16	38		1.200	2.200				(FAO, 2022)
Grapes	Vitaceae	3602	Sandy loams		14	29	21	75	100				(Yzarra et al., 2017); (Covarrubias and Thach, 2015)
Avocado	Persea americana	3435	Clay loams; Silty loams; Sandy loams		16		18-26			1,200-1,800			(Gruter et al., 2022)
Asparagus	Asparagus	4685	Sandy loams		15	26		60	100				(Walker, 2009); (CABI, 2021)
Cattle	Bos taurus	9913		>74									(Diaz et al., 2020)

Bibliographic resources

BIBLIOGRAPHIC RESOURCES					
Authors	**Title**	**Publication date**	**URL**	**Resource type**	**Citation**
Steven Prager; Ana R. Rios; Benjamin Schiek; Juliana S. Almeida; Carlos E. Gonzalez	Vulnerability to climate change and economic impacts in the agriculture sector in Latin America and the Caribbean	2020	https://publications.iadb.org/publications/english/document/Vulnerability-to-Climate-Change-and-Economic-Impacts-in-the-Agriculture-Sector-in-Latin-America-and-the-Caribbean.pdf	Journal article	(Prager et al., 2020)
Brian Feld; Sebastian Galiani	Climate change in Latin America and the Caribbean: policy options and research priorities	2015	https://latinaer.springeropen.com/articles/10.1007/s40503-015-0028-4	Journal article	(Feld and Galiani, 2015)
Christopher P.O. Reyer; Sophie Adams; Torsten Albrecht; Florent Baarsch; Alice Boit; Nella Canales Trujillo; Matti Cartsburg; Dim Coumou; Alexander Eden; Erick Fernandes; Fanny Langerwisch; Rachel Marcus; Matthias Mengel; Daniel Mira-Salama; Mahe' Perette; Paola Pereznieto; Anja Rammig; Julia Reinhardt; Alexander Robinson; Marcia Rocha; Boris Sakschewski; Michiel Schaeffer; Carl-Friedrich Schleussner; Olivia Serdeczny; Kirsten Thonicke	Climate change impacts in Latin America and the Caribbean and their implications for development	2015	https://link.springer.com/article/10.1007/s10113-015-0854-6	Journal article	(Reyer et al., 2015)
I Hagen; C Huggel; L Ramajo; N Chacón; J P Ometto; J C Postigo and E J Castellanos	Climate change-related risks and adaptation potential in Central and South America during the 21st century	2022	https://ui.adsabs.harvard.edu/abs/2022ERL....17c3002H/abstract	Journal article	(Hagen et al., 2022);
Onil Banerjee; Martin Cicowiez; Ana R. Rios; Cicero Z. de Lima	Climate Change Impacts on Agriculture in Latin America and the Caribbean: An Application of the Integrated Economic-Environmental Modeling (IEEM) Platform	2021	https://publications.iadb.org/publications/english/document/Climate-Change-Impacts-on-Agriculture-in-Latin-America-and-the-Caribbean-An-Application-of-the-Integrated-Economic-Environmental-Modeling-IEEM-Platform.pdf	Report	(Banerjee et al., 2021)
Deissy Martínez Barón and Bram Govaerts	AgriLAC Resiliente: Resilient Agrifood Innovation Systems Driving Food Security, Inclusive Growth, and Reduced Out-Migration in Latin America and the Caribbean (LAC)	2021	chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://storage.googleapis.com/cgiarorg/2021/12/AgriLAC-Resiliente.pdf	Report	(Baron and Govaerts, 2021)
Sharon Hutchinson; Charmaine Gomes; Dillon Alleyne; Willard Phillips	An assessment of the economic and social impacts of climate change on the agriculture sector in the Caribbean	2013	chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.cepal.org/sites/default/files/publication/files/38278/LCCARL398_en.pdf	Report	(Hutchinson et al., 2013)
World Bank; CIAT; CATIE	Climate-Smart Agriculture in Mexico. CSA Country Profiles for Latin America Series	2014	chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://assets.publishing.service.gov.uk/media/57a089dde5274a31e00002da/CSA-in-Mexico.pdf	Report	(World Bank et al., 2014)
Francisco J. Meza; Daniel S. Wilks; Luis Gurovich; and Nicolás Bambach	Impacts of Climate Change on Irrigated Agriculture in the Maipo Basin, Chile: Reliability of Water Rights and Changes in the Demand for Irrigation	2012	chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://cpb-us-e1.wpmucdn.com/blogs.cornell.edu/dist/	Journal article	(Meza et al., 2012)

CORDIS EU Research Results

https://cordis.europa.eu/projects/en



https://data.europa.eu/data/datasets/cordis-eu-research-projects-under-horizon-europe-2021-2027?locale=en

Dataset

CORDIS - EU research projects under HORIZON EUROPE (2021-2027)

Publications Office

Dataset | Categories | Quality | Similar datasets

Dataset feed | Linked data | Cite

This dataset contains information about projects and their results funded by the European Union under the Horizon Europe framework programme for research and innovation from 2021 to 2027.

The dataset is composed of six (6) different sub-set (in different formats):

- HORIZON projects – which includes participating organisations, legal basis information, topic information, project URLs and classification with the European Science Vocabulary (EuroSciVoc)
- H2020 project IPRs (Intellectual Property Rights) - NOT YET AVAILABLE
- H2020 project deliverables (meta-data and links to deliverables) - NOT YET AVAILABLE
- H2020 project publications (meta-data and links to publications) - NOT YET AVAILABLE
- H2020 report summaries (periodic or final publishable summaries) - NOT YET AVAILABLE
- Principal Investigators in Horizon 2020 ERC projects - NOT YET AVAILABLE

Reference data (programmes, topics, topic keywords funding schemes (types of action), organisation types and countries) can be found in this dataset: https://data.europa.eu/euodp/en/data/dataset/cordisref-data

EuroSciVoc is available here: https://data.europa.eu/data/datasets/euroscivoc-the-european-science-vocabulary

CORDIS datasets are produced monthly. Therefore, inconsistencies may occur between what is presented on the CORDIS live website and the datasets.

Landing Page	https://cordis.europa.eu/
Updated	25.07.2022
Created	25.07.2022
Languages	English
Publisher	Name: Publications Office of the European Union Homepage: https://op.europa.eu
Contact Points	E-Mail: cordis@publications.europa.eu Telephone: +352292942210 URL: https://cordis.europa.eu/about/
Other Identifiers	Identifier: 10.2906/112117098108/20 Scheme: http://purl.org/spar/datacite/doi

Show more

https://cordis.europa.eu/project/id/863410

HORIZON 2020

INODE - Intelligent Open Data Exploration

Fact Sheet | Reporting | Results

Project description

DE EN ES FR IT PL

Helping humans talk to their databases

Countless sets of data in different formats and about almost everything imaginable are continuously being generated and collected. What is more, open data repositories and the internet ensures data are available to all. One important question now is about how humans are interacting with all the data available. The EU-funded INODE project will provide a set of agile, fit-for-purpose and sustainable services. It will ensure open data sets help users, who are invited to explore data and discover new insights through visualisation. The project will respond to the scientific community. The goal of this project is to make communication with databases more humanlike.

Show the project objective

Fields of science

natural sciences › physical sciences › astronomy › **astrophysics**
engineering and technology › electrical engineering, electronic engineering, information engineering › electronic engineering › **sensors**
medical and health sciences › clinical medicine › **oncology**

Project Information

INODE
Grant agreement ID: 863410

DOI
10.3030/863410

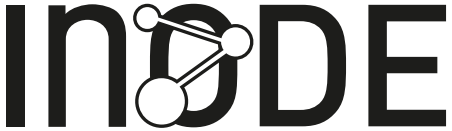
Start date 1 November 2019
End date 30 April 2023

Funded under
EXCELLENT SCIENCE - Research Infrastructures

Total cost
€ 5 732 000

EU contribution
€ 5 732 000

Coordinated by
ZURCHER HOCHSCHULE FUR ANGEWANDTE WISSENSCHAFTEN
Switzerland



Participants (8)

Sort alphabetically ⇕ Sort by Net EU contribution ⇕ Expand all

ATHINA-EREVNITIKO KENTRO KAINOTOMIAS STIS TECHNOLOGIES TIS PLIROFORIAS, TON EPIKOINONION KAI TIS GNOSIS Greece	Net EU contribution € 798 000,00	▾
MAX-PLANCK-GESELLSCHAFT ZUR FORDERUNG DER WISSENSCHAFTEN EV Germany	Net EU contribution € 322 500,00	▾
FRAUNHOFER GESELLSCHAFT ZUR FORDERUNG DER ANGEWANDTEN FORSCHUNG EV Germany	Net EU contribution € 545 500,00	▾
SIRIS ACADEMIC SL Spain	Net EU contribution € 519 250,00	▾
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE CNRS France	Net EU contribution € 558 500,00	▾
LIBERA UNIVERSITA DI BOLZANO Italy	Net EU contribution € 765 500,00	▾
SIB SWISS INSTITUTE OF BIOINFORMATICS Switzerland	Net EU contribution € 475 500,00	▾
INFILI TECHNOLOGIES SOCIETE ANONYME Greece	Net EU contribution € 558 000,00	▾

INODE

Objective

Data growth and availability as well as data democratization have radically changed data exploration in the last 10 years. Many different data sets, generated by users, systems and sensors, are continuously being collected. These data sets contain information about scientific experiments, health, energy, education etc., and they are highly heterogeneous in nature, ranging from highly structured data in tabular form to unstructured text, images or videos. Furthermore, especially online content, is no longer the purview of large organizations. Open data repositories are made public and can benefit more types of users, from analysts exploring data sets for insight, scientists looking for patterns, to dashboard interactors and consumers looking for information. As a result, the benefit of data exploration becomes increasingly more prominent. However, the volume and complexity of data make it difficult for most users to access data in an easy way.

In this project we propose INODE – Intelligent Open Data Exploration. The core principle of INODE is that users should interact with data in a more dialectic and intuitive way similar to a dialog with a human. To achieve this principle, INODE will offer a suite of agile, fit-for-purpose and sustainable services for exploration of open data sets that help users (a) link and leverage multiple datasets, (b) access and search data using natural language, using examples and using analytics (c) get guidance from the system in understanding the data and formulating the right queries, and (d) explore data and discover new insights through visualizations.

Our service offering is formed by and will initially respond to the needs of large and diverse scientific communities brought by our three use case providers: (a) Cancer Biomarker Research - SIB Swiss Institute of Bioinformatics, Switzerland, (b) Research and Innovation Policy Making - SIRIS, Spain, and (c) Astrophysics - Max Planck Institute for Extraterrestrial Physics, Germany.

Fields of science

natural sciences › physical sciences › astronomy › **astrophysics**
engineering and technology › electrical engineering, electronic engineering, information engineering › electronic engineering › **sensors**
medical and health sciences › clinical medicine › **oncology**

https://webgate.ec.europa.eu/dashboard/sense/app/98dcd94d-ca66-4ce0-865b-48ffe7f19f35/sheet/KVdtQ/state/analysis

Key Figures

Programme | Country Name | Country Group | Organisation Type | Partner Type | Pillar | Thematic Priority | Topic | ...

Net EU Contribution
13,35B 8,46% of Programme (All)

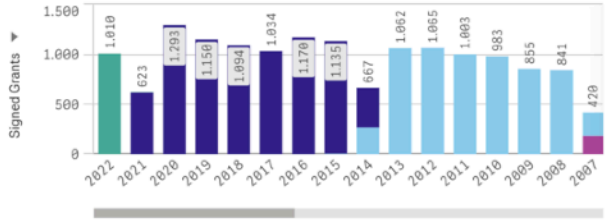
Total Cost
18,75B 5,78% of Programme (All)

Signed Grants
30.765 25,97% of Programme (All)

Participation
58.592 9,37% of Programme (All)

Unique Participants
5.598 9,10% of Programme (All)

Evolution over time

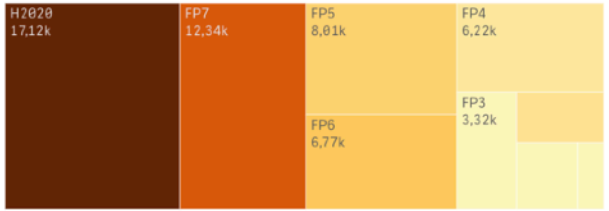


Net EU Contribution by Country / Region

Country
NUTS 1 Name
Area layer

© Eurogeographics © UN-FAO © TURKSTAT - The designations employed and the presentation of material on the map do not imply the expression of any opinion whatsoever on

Overview by programme/pillar/thematic priority/keyword



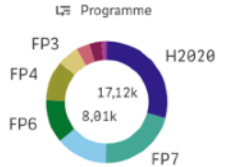
Top participating countries - regions

[NUTS 1 Name]	Programme	Country Gro
Totals		
CENTRO (IT)	HORIZON EUROPE	Member Sta
NORD-OVEST	HORIZON EUROPE	Member Sta
NORD-EST	HORIZON EUROPE	Member Sta
SUD	HORIZON EUROPE	Member Sta
ISOLE	HORIZON EUROPE	Member Sta
-	HORIZON EUROPE	Member Sta

Top participants

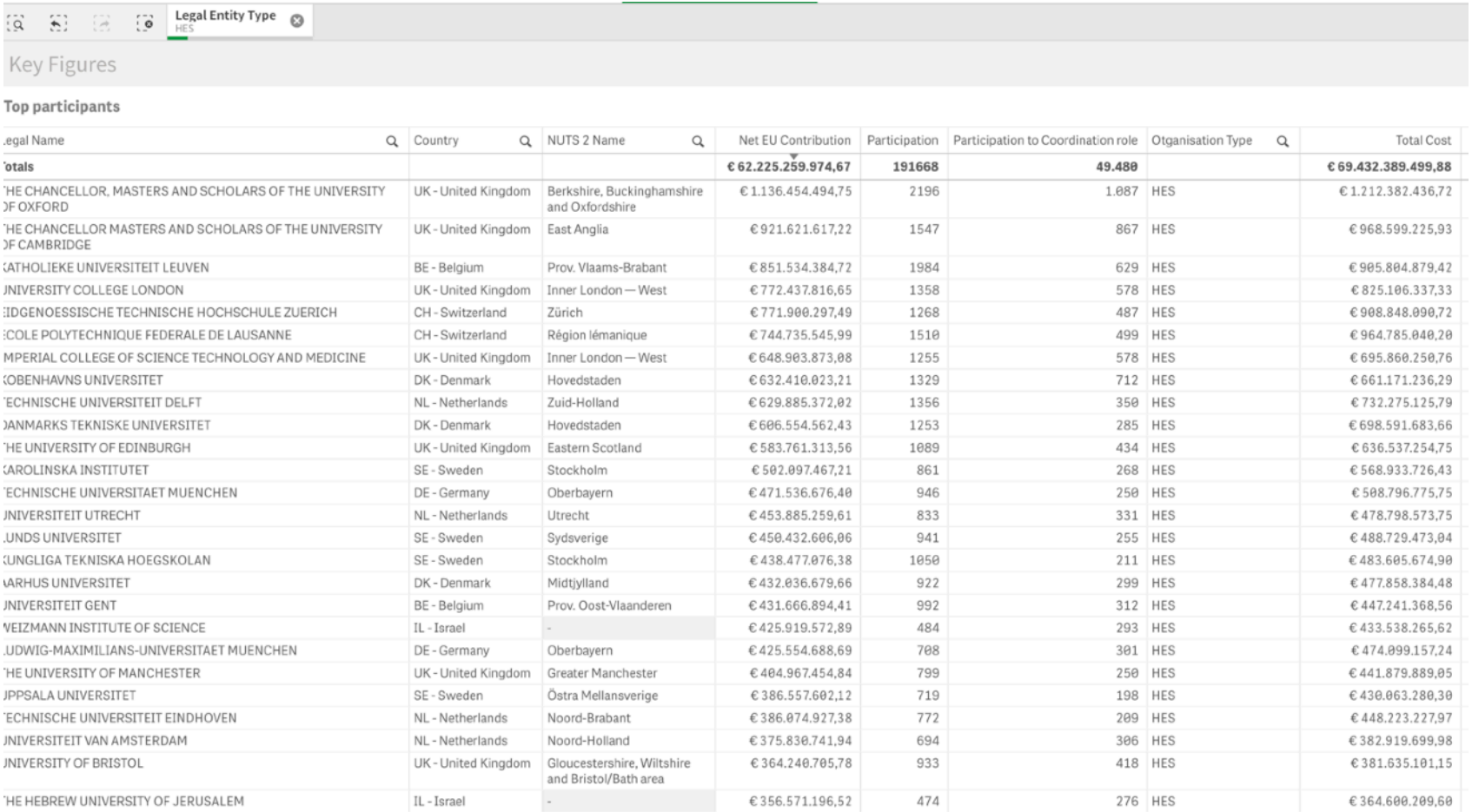
Legal Name	Country	NUTS 2 Name
Totals		
CONSIGLIO NAZIONALE DELLE RICERCHE	IT - Italy	Lazio
POLITECNICO DI MILANO	IT - Italy	Lombardia
ALMA MATER STUDIORUM - UNIVERSITA DI BOLOGNA	IT - Italy	Emilia-Romagna
POLITECNICO DI TORINO	IT - Italy	Piemonte
FONDAZIONE ISTITUTO	IT - Italy	Liguria

Overview by type of action



Open the Exploration Menu to select an alternative dimension

https://webgate.ec.europa.eu/dashboard/sense/app/98dcd94d-ca66-4ce0-865b-48ffe7f19f35/sheet/KVdtQ/state/analysis



.egal Name	Country	NUTS 2 Name	Net EU Contribution	Participation	Participation to Coordination role	Otganisation Type	Total Cost
'otals			**€ 62.225.259.974,67**	**191668**	**49.480**		**€ 69.432.389.499,88**
'HE CHANCELLOR, MASTERS AND SCHOLARS OF THE UNIVERSITY)F OXFORD	UK - United Kingdom	Berkshire, Buckinghamshire and Oxfordshire	€ 1.136.454.494,75	2196	1.087	HES	€ 1.212.382.436,72
'HE CHANCELLOR MASTERS AND SCHOLARS OF THE UNIVERSITY)F CAMBRIDGE	UK - United Kingdom	East Anglia	€ 921.621.617,22	1547	867	HES	€ 968.599.225,93
:ATHOLIEKE UNIVERSITEIT LEUVEN	BE - Belgium	Prov. Vlaams-Brabant	€ 851.534.384,72	1984	629	HES	€ 905.804.879,42
JNIVERSITY COLLEGE LONDON	UK - United Kingdom	Inner London — West	€ 772.437.816,65	1358	578	HES	€ 825.106.337,33
:IDGENOESSISCHE TECHNISCHE HOCHSCHULE ZUERICH	CH - Switzerland	Zürich	€ 771.900.297,49	1268	487	HES	€ 908.848.090,72
:COLE POLYTECHNIQUE FEDERALE DE LAUSANNE	CH - Switzerland	Région lémanique	€ 744.735.545,99	1510	499	HES	€ 964.785.040,20
MPERIAL COLLEGE OF SCIENCE TECHNOLOGY AND MEDICINE	UK - United Kingdom	Inner London — West	€ 648.903.873,08	1255	578	HES	€ 695.860.250,76
:OBENHAVNS UNIVERSITET	DK - Denmark	Hovedstaden	€ 632.410.023,21	1329	712	HES	€ 661.171.236,29
'ECHNISCHE UNIVERSITEIT DELFT	NL - Netherlands	Zuid-Holland	€ 629.885.372,02	1356	350	HES	€ 732.275.125,79
)ANMARKS TEKNISKE UNIVERSITET	DK - Denmark	Hovedstaden	€ 606.554.562,43	1253	285	HES	€ 698.591.683,66
'HE UNIVERSITY OF EDINBURGH	UK - United Kingdom	Eastern Scotland	€ 583.761.313,56	1089	434	HES	€ 636.537.254,75
:AROLINSKA INSTITUTET	SE - Sweden	Stockholm	€ 502.097.467,21	861	268	HES	€ 568.933.726,43
'ECHNISCHE UNIVERSITAET MUENCHEN	DE - Germany	Oberbayern	€ 471.536.676,40	946	250	HES	€ 508.796.775,75
JNIVERSITEIT UTRECHT	NL - Netherlands	Utrecht	€ 453.885.259,61	833	331	HES	€ 478.798.573,75
.UNDS UNIVERSITET	SE - Sweden	Sydsverige	€ 450.432.606,06	941	255	HES	€ 488.729.473,04
:UNGLIGA TEKNISKA HOEGSKOLAN	SE - Sweden	Stockholm	€ 438.477.076,38	1050	211	HES	€ 483.605.674,90
\ARHUS UNIVERSITET	DK - Denmark	Midtjylland	€ 432.036.679,66	922	299	HES	€ 477.858.384,48
JNIVERSITEIT GENT	BE - Belgium	Prov. Oost-Vlaanderen	€ 431.666.894,41	992	312	HES	€ 447.241.368,56
'VEIZMANN INSTITUTE OF SCIENCE	IL - Israel	-	€ 425.919.572,89	484	293	HES	€ 433.538.265,62
.UDWIG-MAXIMILIANS-UNIVERSITAET MUENCHEN	DE - Germany	Oberbayern	€ 425.554.688,69	708	301	HES	€ 474.099.157,24
'HE UNIVERSITY OF MANCHESTER	UK - United Kingdom	Greater Manchester	€ 404.967.454,84	799	250	HES	€ 441.879.889,05
JPPSALA UNIVERSITET	SE - Sweden	Östra Mellansverige	€ 386.557.602,12	719	198	HES	€ 430.063.280,30
'ECHNISCHE UNIVERSITEIT EINDHOVEN	NL - Netherlands	Noord-Brabant	€ 386.074.927,38	772	209	HES	€ 448.223.227,97
JNIVERSITEIT VAN AMSTERDAM	NL - Netherlands	Noord-Holland	€ 375.830.741,94	694	306	HES	€ 382.919.699,98
JNIVERSITY OF BRISTOL	UK - United Kingdom	Gloucestershire, Wiltshire and Bristol/Bath area	€ 364.240.705,78	933	418	HES	€ 381.635.101,15
'HE HEBREW UNIVERSITY OF JERUSALEM	IL - Israel	-	€ 356.571.196,52	474	276	HES	€ 364.600.209,60

Data profiling final project:
Guidelines

1. Choose a domain of interest among those that have been introduced here and for which you can identify a collection of datasets that need to be integrated.
2. Export the datasets of interest and perform Elementary Data Analysis (EDA) by means of either existing tools and libraries or your own EDA algorithm implementations. Report the results of the EDA and comment on them in the final project (pdf) documentation.
3. Design one or more relational database to structure and store the exported datasets and show how the dependency discovery algorithms (see, UCC, FD, and IND) introduced in the course have been used to extract metadata that supported the database/s designing process.

Note that the choice of having multiple databases, instead of a single one, has an impact on the technological solution that you will need to deployed in the integration part of the project. In particular, having multiple databases to integrate implies the necessity to rely on a federation system (like Teiid or Denodo, for instance) as an intermediate layer between the sources and ontop. This is obviously not needed if you will work with a single database.

The results obtained by profiling the datasets for dependency discovery, and the consequent database schema design choices, must be properly commented in the the final project (pdf) documentation.

Data profiling final project:
Additional notes

1. The project can be developed either alone, or in a group of two. For projects developed by a group of two, only one of the two students should submit the project.
2. Before starting the effective development of the project, present and discuss with the lecturer the domain and the data sources that you intend to use, and the application that you intend to develop.
3. Any doubt that you have should be discussed before taking decisions that then might force you to revise your work.
4. The documentation produced for the project should include a self-contained pdf document, which will be the basis for the discussion of the project during the oral exam. The document should contain:
5. a header with title of the project, name(s) of the student(s) that have developed it, name of the course (i.e., Data Integration), and academic year;
6. a description of the domain of interest;
7. a description of the content and format of the selected data sources, and an indication of how and where these data sources can be accessed;
8. a description of the profiling analysis that have been performed and a critical analysis of the obtained results (e.g., which techniques have been applied and for which purpose, what the results convey about the datasets at hand and the value distributions, significant attribute correlations, how the profiling results have been exploited in schema design and data cleansing, etc.);

9. All documents specified in the previous items, together with those concerning the "Data Preparation and Integration" module of the course, have to be bundled in a single ZIP file and have to be uploaded to OLE (exact storing place will be communicated by Prof. Calvanese and Prof. Mosca before the end of the course).
10. The deadline for submitting the project is at 23:55 two days before the day set for the project discussion.

See you on Tuesday!

http://www.inf.unibz.it/~almosca/