

# Intellectual Output 4 (Report)

# Job Knowledge Base (JKB)

This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/ or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

**Disclaimer:** The European Commission support for the production of this publication does not constitute an endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.







The DISKOW project has received funding from the Erasmus plus programme under grant agreement no. 2018-1-DE02-KA202-005215

Grant Agreement Number	2018-1-DE02-KA202	-005215 Acronym	DISKOW
Full Title	Discovering Job Knowledge through Web Analytics towards facilitated mobility of European Professionals and Refugees Career Integration		
Topic	KA2 - Cooperation for Innovation and the Exchange of Good Practices  KA202 - Strategic Partnerships for vocational education and training		
Funding scheme	ERASMUS+		
Start Date	01-08-2018	Duration	36
Project website URL	www.diskow.eu		
Project Coordinator	GOTTFRIED WILHELM LEIBNIZ UNIVERSITAET HANNOVER		
Report	IO4 - report		
Intellectual Output	IO4 - Job Knowledge Base (JKB)		
Delivery Month	36	Version	1.0
Actual Delivery Date	31/08/2021		
Lead Beneficiary	ENG		

## **Preface**

The Web is a gold mine for job knowledge discovery. Linked open data, social media sources, job search engines, forums, wikis, data streams and interlinked information are few examples of such readily available job-related sources of knowledge. Improving job knowledge, competencies and skills of job seekers is essential to increase their employability and reduce the risk of poverty and social exclusion. To gain insights into labor demands or supply, researchers and policymakers have traditionally relied on interviews, trade publications, surveys, and vacancies. Although such traditional data sources have some clear advantages, they are also characterized by limitations that can be addressed by using web-based data and Semantic Web technologies instead. The scope of DISKOW project is to create an open "job knowledge base" prototype that can be used by employers, employees, job seekers, labor market experts and policy makers. Such a knowledge base could contain different types of information such as responsibilities and roles, required competences (described using existing standards, such as European e-Competence Framework (eCF), European Qualifications Framework (EQF), etc.), that could be used to develop trainings and identify priorities, wage information, geographical and demographic trends, cultural issues, demands of the job markets in different domains, job announcement information and rates, job popularity and other useful statistics. Engineering Ingegneria Informatica will be responsible for the creation of the Job Knowledge Base (JKB) based on the Knowage open source data platform. The JKB will hosts jobs data collected from the web provided by the Job Knowledge Analysis Engine and Visualization APIs resulting from the work done in collaboration with the project partners, Petanux Gmbh, Gottfried Wilhelm Leibniz Universitaet Hannover. Ekonomicky Ustav Slovenskei Akademie Vied).

## Who We Are

	Participant Name	Short Name	Countr y Code	Logo
1	ENGINEERING – INGEGNERIA INFORMATICA SPA	ENG	IT	ENGINEERING
2	GOTTFRIED WILHELM LEIBNIZ UNIVERSITAET HANNOVER	LUH	DE	(35)
3	EKONOMICKY USTAV SLOVENSKEJ AKADEMIE VIED	SAS BIER	SK	<b>EUSAV</b> Ekonomický ústav SAV Institute of Economic Research SAS
4	PETANUX GMBH	PNX	DE	Petanux

## Partners profile in a nutshell

**LUH** is part of TU9 (the alliance of leading Institutes of Technology in Germany) and it is a leading institute for Web Science. Work at L3S focuses on innovative and cutting-edge methods and technologies for three key enablers for the European Information Society: Knowledge, Information and Learning, projects investigate innovative scenarios and solutions in the areas of Technology Enhanced Learning (TEL) and Web Science. The L3S team builds upon approaches from various areas such as information retrieval, machine learning, natural language processing, semantic web technologies, data management, Distributed and Peer-to-Peer Information Systems, as well as privacy/security related legal issues and sociology.

IER SAS is an economic research institute with lively connections to other research fields covered by the Slovak Academy of Sciences (e.g. educational research, experimental psychology, sociology). IER SAS is delivering policy relevant research to national stakeholders but also internationally (e.g. European Commission, CEDEFOP). A micro-team focusing on skill needs analysis and anticipation operates at IER SAS since 2010. This team has been involved in national as well as European projects, delivering analysis and expertise on skills anticipation and forecasting. IER SAS brings experience with various methodologies of skills anticipation and forecasting, including big data analysis in graduate tracking and content analysis of online job advertisements. Of special relevance is the methodology on assessing skill requirements based on employers' behaviour in searching online CVs of potential employees.

**ENG** is the first largest private ICT group in Italy and among the top 10 ICT groups in Europe, has over 12.000 employees and 40 branch offices in Italy, Belgium, Spain, Germany, Serbia, Latin America and USA with a revenue of over 1 Billion Euro in 2020. The group produces IT innovation, consultancy and training to more than 1.000 large clients, with a complete offer combining system and business integration, outsourcing, cloud services, consulting, and proprietary solutions. ENG Academy "Enrico della Valle" organise more than 300 residential and online courses on technologies, methodologies and soft skills for company and customers employees, cooperating with experts, universities and technology vendors. Moreover, ENG operates 7 different business units: Finance, Central Government, Local Government and Healthcare, Energy & Utilities, Industry and Telecoms. Additionally to vast expertise of ENG in different ICT domains, ENG has a strong experience in developing HR and payroll systems. For instance, one of the controlled companies in ENG group, WebResults srl is specialised on HR services and applications.

**PNX** is a German research and development company in the data science and smart analytics and will contribute to DISKOW with its expertise in labour market data retrieval and analysis as well as dissemination and exploitation of the project results. In particular, PNX members have notable experience in initiating and coordinating European projects especially Transfer of Innovation and Erasmus+ calls, as stated in the PNX profile.

## **Contents**

1.	. Intr	oduction	11
	1.1	Purpose of the document	11
	1.2	Positioning within the Project	
	1.3	Report Structure	11
2	Tasl	cs constituting IO4 – Job Knowledge Base (JKB)	12
3	JKB	Architecture	
	3.1	General Description of the system	
	3.2	JKB Architectural component interaction	
	3.2.1	'	
	3.2.2		
	3.2.3	1 3	
	3.2.4	Data use	14
4	JKB	infrastructure and software requirements	15
	4.1	JKB Cloud infrastructure	15
	4.2	JKB Software requirements	16
5	JKB	Data base	17
	5.1	KNOWAGE suite tables	
	5.1	JKB tables	
	5.3	JKB datasets customised queries and views	
	5.4	GDPR Requirements for collected data	
	5.5	JKB DB Integration with external project data	
	5.6	JKB DB data exploitation	
6	Kno	wage, open source business intelligence platform suite, customisation	
	6.1	JKB application design and refinements	
	6.2	JKB Cockpit main functionalities	
	6.2.1	•	
	6.3	JKB Data Analysis	
	6.3.1	JKB Data Analysis in detail	30
	6.4	JKB ESCO categories	33
	6.5	JKB ESCO categories usage	34
A	ppendi	х	35
F	igure	s	
Fi	igure 1	- JKB Prototype Conceptual Architecture	13

Figure 2 - JKB Virtul Machine (Contabo)	15
Figure 3 - Knowage suite functional tables	17
Figure 4 - JOBPOSTING table structure	18
Figure 5 - JKB Data sets list	20
Figure 6 - Login form	23
Figure 7 - Dashboard Cockpit page	24
Figure 8 - Multi-choice filter	25
Figure 9 - European job positions filtered table	25
Figure 10 - European job positions filtering all columns	26
Figure 11 - European job positions filtering specific column	26
Figure 12 – "Basic and Estimated Salary" graph in detail	27
Figure 13 – "Basic and Estimated Salary" - Zoom view	27
Figure 14 – "Estimated Salary By Contract type" 3D graph	28
Figure 15 - JKB Data Analysis dashboard	29
Figure 16 – "Country job offers by sector" graph	30
Figure 17 – "Country job offers by sector" radar graph	30
Figure 18 – "European job offers by sector" pie graph	31
Figure 19 – "European job offers by sector" wordcloud graph	31
Figure 20 – "European basic salary by sector" graph	32
Figure 21 – "European open positions by country" graph	33
Figure 22 – JKB ESCO categories web page	33
Figure 23 - ESCO categories - Skill table filtering	34
Figure 24 - JKB Cockpit versions	35
Figure 25 - Table widget selection functionality	36
Figure 26 - Table widget multy filtering on columns	36
Figure 27 - Table widget filtering	36
Figure 28 - Refresh/Reset filters icon	36
Figure 29 - Table widget customization (columns to visualize)	37
Figure 30- Table widget - conditions on table data	37
Figure 31 - Table widget configuration	38

Figure 32 - Chart widget dynamic style choice	39
Figure 33 - Chart widget measures configuration	40
Figure 34 - Chart widget structure/measures configuration	40
Figure 35 - Chart widget data filter functionality	41
Figure 36 - Widget datasets selection	41
Figure 37 - Chart widget styles	42
Figure 38 - Selection widget configuration	42
Figure 39 - Selection filter data filtering	43
Figure 40 – Job Posting Data set definition	44
Figure 41 – Job Posting Data set metadata customization	44
Figure 42 - Skills data set definition	45
Figure 43 - Skills data set metadata customization	45
Figure 44 - Occupation data set definition	45
Figure 45 - Occupation data set metadata customization	46
Figure 46 - ISCO group data set definition	46
Figure 47 - ISCO group metadata customization	46
Figure 48 - Custom dataset for graph	47
Figure 49 - Custom data set for graph	47
Figure 50 - Custom data set for graph	47
Figure 51 - Custom data set for graph	48
Figure 52 - Skills table	48
Figure 53 - Occupation table	49
Tables	
Table 1 - List of abbreviations and acronyms	X
Table 2 - Software and Hardware requirements	16

# **Executive Summary**

This document provides information about the project activities performed within IO4 and the relationships that interconnect it with other project IOs. An architectural overview of the Job Knowledge Base is given, defining all the phases that make up the whole platform from a technical and functional point of view. All the aspects related to the IT infrastructure hosting the solution are also mentioned, as well as the main technological and implementation aspects, giving detailed evidence of the main back-end and front-end features. The document also illustrates the web interface of the JKB prototype realised through the customisation of Engineering Group's open source business platform named Knowage, explaining the single functionalities realised and their usage.

## 1. LIST OF ABBREVIATIONS AND ACRONYMS

No.	Abbreviation	Description
1	API	Application Programming Interface
2	ESCO	European Skills, Competences, Qualifications, and Occupations
3	ISCO	International Standard Classification of Occupations
4	JK	Job knowledge
5	JKB	Job knowledge base
6	JSON	JavaScript Object Notation
7	XML	Extensible Markup Language
8	SQL	Structured Query Language
9	IO	Intellectual output
10	DB	Data base

Table 1 - List of abbreviations and acronyms

## 1. Introduction

The main objective of the DISKOW project is to create a Job Knowledge Base (JKB) prototype, based on the existing open source Business Intelligence platform named Knowage suite<sup>1</sup>. As it is well known, the web is a gold mine of data and information especially regarding the world of work. Endless information about the world of work can be easily found on the web but all too often this information does not follow standards that help the end user to recognise it unambiguously. In this respect, the DISKOW project aims not only to collect job offers from the web portals identified in IO2 but also to harmonise them thanks to specific micro services developed in IO3. Data harmonisation follows specific schema defined in the IO1 and adopts standards at European level to categorise the job offers. Harmonised data are finally available through a web and mobile ready application developed in IO4 which is the subject of this report and of which all the necessary details will be given in this document. At the same time the interconnection between all the project IOs is a fundamental factor for the correct achievement of the final project goals and although this report will purely illustrate the IO4 activities, the interactions with the other project IOs will be quickly explained.

## 1.1 Purpose of the document

The purpose of this document is to illustrate all the activities carried out in relation to IO4 without losing sight of the interconnections with the other IOs. The conceptual architecture of the project will be described, as well as the individual tasks and the solution will be explained from a technical and functional point of view. Some detailed information will be better listed and explained in the Appendix section of the document.

## 1.2 Positioning within the Project

The purpose of this report is to explain the activities carried out in relation to the project IO4 "The Job Knowledge Base, JKB" and to highlight the interconnections with the other IOs.

## 1.3 Report Structure

The report is divided in 6 main sections as following:

- Section 1 reports a general introduction to the document
- Section 2 reports the tasks that compose the IO4
- Section 3 describes the JKB architecture
- Section 4 describes the project infrastructure and software requirements
- Section 5 reports details of the JKB data base
- Section 6 reports the customization activities of the Knowage suite

IO4 report: Job Knowledge Base (JKB)

<sup>&</sup>lt;sup>1</sup> Knowage suite, https://www.knowage-suite.com/site/

# 2 Tasks constituting IO4 – Job Knowledge Base (JKB)

The main activities addressed within IO4 are defined and described in the five tasks as per the project proposal and grant agreement. These five tasks, listed below, represent the different moments of development of the JKB starting from the design and customisation of the solution, continuing with the various releases and tests, to end with the final version of the prototype.

#### IO4 tasks:

- O4/A1 Customisation of Knowage instance
- O4/A2 App design and implementation
- O4/A3 Alpha release
- O4/A4 Beta release
- O4/A5 Final release

## 3 JKB Architecture

The Job Knowledge Base prototype architecture is based on three different layer better explained into the next sections. The main aspects of this architecture are related to the definition three layer, the Data provider layer, the Data governance layer and the Data user layer that will exploit the different project phases: *Data acquisition, Data preparation and processing, Data publishing* and *Data use*.

## 3.1 General Description of the system

The DISKOW Job Knowledge Base will simultaneously collect, analyse and transform and made available via a specific web user interface a series of job posting available on European web portals addressing needs from a very large user community including job seekers, policy maker, job centers, job analyst, and anyone who takes up employment.

To build this platform a series of blocks representing different stages of the work done on the data have been implemented and are defined in the next picture.

- Data acquisition (To collect data)
- Data preparation/curation (To prepare the data under a common data model/ontology)
- **Data publishing** (To send data to the JKB data base)
- Data use (To provide data via a web user interface)

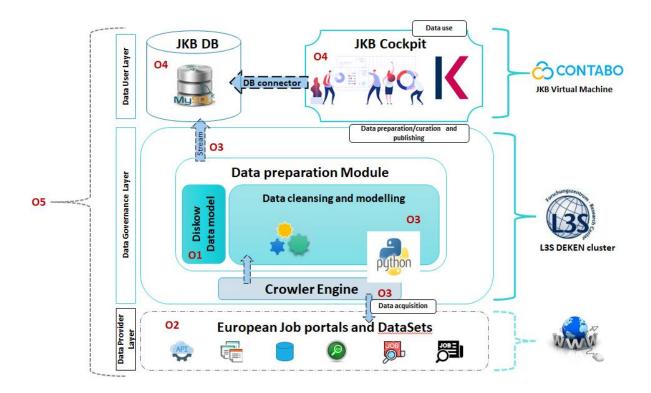


Figure 1 - JKB Prototype Conceptual Architecture

IO4 report: Job Knowledge Base (JKB)

## 3.2 JKB Architectural component interaction

This section of the report describes the main interactions between the different components that build the architecture of the JKB prototype. Their interactions will be explained in order to have a global vision of how the result of each output is in turn the input for another output and how the global result is the object of O5 assessment and reporting.

## 3.2.1 Data acquisition

The data acquisition phase takes care of collecting valid job offers from European job portals. A specific module, the "Crawler engine", collects all available information coming from different data sources identified within the IO2 tasks such as html pages, comma separated files, APIs and sends it to the module for cleaning, harmonisation, conversion and final formatting of the data under the common data model identified within the IO1 tasks.

## 3.2.2 Data preparation/curation

The data preparation/curation phase is formed from a series of micro-services, I.e. optimal data preprocessing pipelines to ensure restructuring, reformatting of fields, outlier detection, inconsistencies handling, noise reduction and data structure reformatting. In this phase all the data gathered from the web are managed and kept up-to-date in a suitable way to be stored into the JKB data base during the data publishing phase. All these process are explained in more detail in the IO3 report.

## 3.2.3 Data publishing

The data publishing phase is built from a service that updates the *JobPosting* table with the new data crawled and processed during the Data preparation/curation phase. This process is explained in more detail in the IO3 report.

### 3.2.4 Data use

The data usage phase is the customisation of the business analysis suite developed by ENG. Within this phase, specific data sets are defined with appropriate conditions of research, aggregation and pre-processing by drawing from the database populated with data coming from the data publication phase and from some basic data pre-loaded in the DB. This data will be made available to the end user via JKB's web user interface. Thanks to this data, the end user will be able to carry out his own analysis such as job search, analysis of the labour market, analysis of skills for specific job sectors and other analyses related to the available data.

## 4 JKB infrastructure and software requirements

In this section the software and infrastructure requirements to build the JKB are mentioned.

## 4.1 JKB Cloud infrastructure

After a thorough analysis of possible cloud providers by the project partners, it was decided to rent a virtual machine provided by Contabo hosting. The management of the VM was directly followed by the Petanux partner and the access credentials to the machine answering to IP address: 173.249.3.120 on the port 22 were provided to all project partners for their activities. The Diskow software runs on top of the aforementioned Contabo VM and the direct contact with the cloud provider is managed by Petanux. All the information about the aforementioned hosting provider can be found on the official website<sup>2</sup>.

```
## NonbardoQ-md12446 -

| Welcome: | Welcome
```

Figure 2 - JKB Virtul Machine (Contabo)

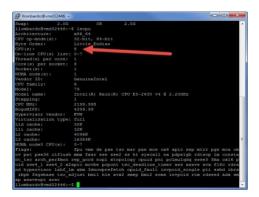
Below are the characteristics of the VM used for the DISKOW project:

#### VM's characteristic:

Number of Virtual CPU cores: 8;

Memory: 30 Gb;Storage: 730 Gb;OS: Ubuntu 18.04.1





<sup>&</sup>lt;sup>2</sup> Contabo, https://contabo.com/en/

## 4.2 JKB Software requirements

In order to work properly, the Knowage suite requires certain basic hardware and software requirements to be satisfied. Regarding this, the following table reports the compatible environment from the Knowage official installation online manual<sup>3</sup> and the ones adopted for the Diskow project installation.

Knowage compatible environment		Diskow environment (Contabo)	
Operating system	Version	Operating System	Version
Ubuntu	18LST	Ubuntu	18.04.1
Java	Version	Java	Version
JDK	1.8	JDK	1.8.0_191-b12
Application server	Version	Application server	Version
Tomcat	8.5	Tomcat	8.5
Database schema	Version	Database schema	Version
MySql	5.7,8	MySqI	5.7,24
Disk usage minimum free space		Disk usage available free space	
2GB		730 GB	

Table 2 - Software and Hardware requirements

-

 $<sup>^3</sup>$  Knowage online manual, https://knowage-suite.readthedocs.io/en/7.2/installation-guide/index.html

## 5 JKB Data base

Based on the results of IO1 and IO2 activities related to the design of the DISKOW schema and data model, the usage of a relational data base to hold the JKB data has been identified. Among the various open source and free databases that were evaluated such as PostgreeSQL<sup>4</sup>, MariaDB<sup>5</sup>, MySQL<sup>6</sup>, the choice fell on the latter. In particular, MySQL DB is also adopted by the Knowage suite for the configuration of many of its parameters and for the correct functioning of the software. This led the consortium to this choice and allowed us to manage both DBs under a single installation, reducing maintenance and monitoring times. All the information about the installation steps and related configuration are available at the official MySQL web site<sup>7</sup>. Dedicated schemas have been created within the DB to host both Knowage configuration and JKB tables. The main DB schemas required for the application to work properly are explained below.

## 5.1 KNOWAGE suite tables

In order to save information relating to the configurations and customisations of the platform, Knowage suite uses some dedicated tables that are fed and used by its software components. Figure 3 shows some of the configuration tables.

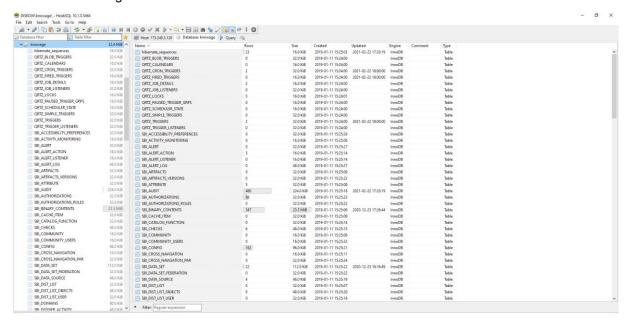


Figure 3 - Knowage suite functional tables

## 5.2 JKB tables

During the solution design phase the JKB tables have been defined considering all the functionalities to be implemented. In this regard have been defined tables for saving basic data such as the countries

<sup>&</sup>lt;sup>4</sup> PostgreeSQL, https://www.postgresql.org/

<sup>&</sup>lt;sup>5</sup> MariaDB, https://mariadb.com/

<sup>&</sup>lt;sup>6</sup> MySQL, https://www.mysql.com/

MySQL documentation, https://dev.mysql.com/doc/

belonging to the European community and its characteristics (useful for possible advanced statistics), the tables that map the ESCO definitions such as skills, occupations, groups and so on to be able to retrieve detailed information on each single skill required, and the table for storing the dynamic data such as the job offers as result of the IO3 services. Data stored in these tables and mainly those concerning job vacancies properly harmonized and standardized can be a valuable source for future implementations and research projects in the same scope of the Diskow project. Figure 4 below shows the "jobposting" table structure as an example, the remaining tables and details are included in the Appendix section.

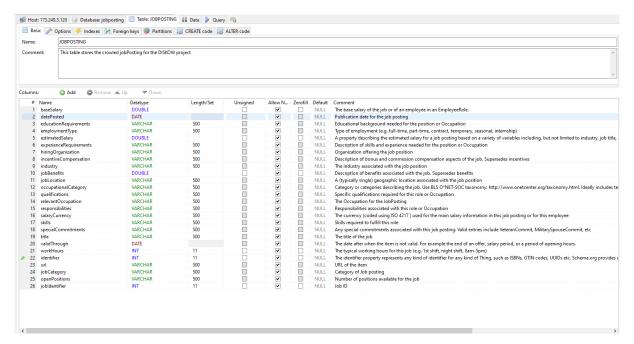


Figure 4 - JOBPOSTING table structure

#### Below some details about the JKB tables:

- Jobposting: this table contains the job offers crawled from the web as result of IO3 activities. The information contained in this table, appropriately interpreted, can provide a general picture of the European world of work both for those seeking a job, those offering job opportunities, and those who have to legislate in this context. By way of example, the main information that are collected within it are:
  - Title: The title of the job
  - o Skills: Skills required to fulfill this role (using the ESCO definition)
  - Industry: The industry sector associated with the job position (based on the project choosen classification as from IO3 activity)
  - o JobLocation: The European country (as from the country table definitions)
  - EducationRequirement: Educational background needed for the position or Occupation (based on a standard education definition)
  - EmploymentType: Type of employment (based on a standard definition e.g. full-time, part-time, contract, temporary, seasonal, internship)
  - WorkHours: The typical working hours for this job expressed as a number (eg: 4 or 8 etc)

- BaseSalary: The base salary of the job or of an employee in an Employee Role
- estimatedSalary: A property describing the estimated salary for a job posting based on a variety of variables including, but not limited to industry, job title, and location. The estimated salary is usually computed by outside organizations and therefore the hiring organization is not bound to this estimated salary
- **Country**: this table contains the European countries. The data is used as reference for the jobPosting tables and some of the information collected within this table can be used to execute an analysis based for example on the number of citizens and open positions for each country. The main information that are collected in this table are the following:

Name: The name of the country (eg: Italy, Germany, France, Slovakia)

Code: The code of each country (eg: IT, D, F, SK)

Capital: The country capital

Province: The country province

Area: The country area

Population: The population size

- **Skills**: this table contains the European skills as from the ESCO official classification to connect people to the job and made by the European Commission. Skills defined into this table are the base for the "skill" column of the jobPosting table. The main information that are collected in this table are the following:
  - conceptType: The concept of the kind of information contained in the row
  - conceptUry: The uri to retrieve the complete definition of the specific skill and all related information
  - o skillType: The type of the skill
  - PreferredLabel: The label that synthetize the skill concept
  - AltLabel: The label that describe more in detail the skill
  - Description: The skill description
- Occupation: this table contains the European occupation definition as from the ESCO official
  classification language to connect people to the job and made by the European Commission.
  Occupation defined into this table are the base for the "relevantOccupation" column of the
  jobPosting table. The main information that is collected in this table is the following:
  - conceptType: The concept of the kind of information contained in the row
  - conceptUry: The uri to retrieve the complete definition of the specific occupation and all related information
  - o iscoGroup: The id of the related iscoGroup
  - PreferredLabel: The label that synthetizes the occupation concept
  - AltLabel: The label that describes in more detail the occupation
  - Description: The occupation description

## 5.3 JKB datasets customised queries and views

Figure 5 shows some specific datasets that were defined and tested by querying the JKB data base in order to feed the JKB UI and perform analysis on top of these.

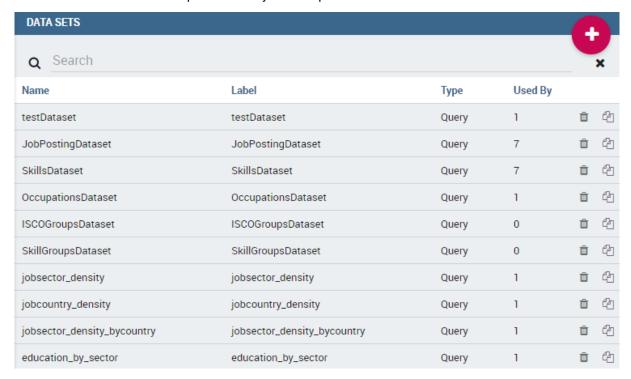


Figure 5 - JKB Data sets list

Below are details of the data sets and related usage:

- *JobPostingDataset*: Provide data about the available Job Posting crawled from the web, this data set is used mainly in the Dashboard page to show the job posting in detail.
- *SkillsDataset*: Provides data of the skill categories defined by ESCO, it is mainly used in the ESCO categories dashboard.
- OccupationDataset: Provides data of the occupation categories defined by ESCO, it is mainly used in the ESCO categories dashboard.
- *JobSectorDensity*: Provides the density of each job sector in terms of number of open job position in Europe based on the data stored into the JKB data base.
- *JobCountryDensity*: Provides the density of open job position for each European country based on the data stored into the JKB data base.
- JobSector\_density\_bycountry: Provides the density of each job sector in terms of number of open job position for each European country based on the data stored into the JKB data base.
- *Education\_by\_sector*: Provides the education level required by each job sector based on the data stored in the JKB data base.

More details about the aforementioned data sets are presented in the Appendix at the end of this

## 5.4 GDPR Requirements for collected data

The data shown in the JKB app are the result of data extraction activities relating to data sources such as portals or social channels identified in IO2. In this regard, what is defined in the IO2 report regarding the data privacy is set out below. For instance, according to the terms and conditions of one job portal - "You can use the material available on this site for any non-commercial purpose. Usage of material for a commercial purpose is not allowed." This type of declaration is common nowadays to ensure data privacy and protect their proprietary data. Most job portals charge a fee, according to their service policy, to provide the customized service. Customized services can be of two types: data of job seekers to job providers and job provider postings to job seekers. Hence, if we plan to use the data from commercial job portals, then we might have to make an official (or legal) agreements with these job portals for using their data to develop the JKB. To address the promised tasks of IO2, we have extracted data in the following manner. Initially, by browsing the websites of existing job portals, we have extracted the freely available data from these paid websites. As we mentioned earlier, most of these job portals are paid portals for customized service. Hence, we have also identified data sources where job portals allow to legally access their data. In addition, this report also offers some insights regarding alternative free data sources such as Twitter."

## 5.5 JKB DB Integration with external project data

The JKB data base population as previously mentioned is carried out through specific scripts better explained in detail in the IO3 report taking care of websites changes. In case the JKB is being updated with data from possible new data crawlers wishing to provide their contribution with data from other sources, several solutions can be adopted. For example, those who wish to enrich the JKB DB will be able to perform simple configurations/scripts in their crawling software, these configurations/scripts are related to the connection to the project DB using specific credentials that will be provided upon request and subsequent authorization to applicants. These are information relating to Ip, port and username and password of the JKB DB. It is understood that the data entry scripts must be compatible with the project schema and provide the same backup rules as already defined in the IO3 report. Another solution to integrate new data into the JKB is to define your own externally accessible DB with a schema compatible with the data defined in the DISKOW data model. This new DB will then be configured at the level of data source and tables in the JKB application as explained in the appendix. At this point, having the connection with the new external DB, it will be possible to create views of these tables that must reflect the project tables structure and merge them with the project tables to obtain the sum of the available data from both DBs. A further solution for less expert users who are familiar with Microsoft XLS is to use an xls plugin named "MySQL for Excel" that provide an interface which allows to connect to the project DB (subject to authorisation and obtaining credentials) and insert/update the data of the project tables with the new crawled data in the form of an xls file compatible with the project schema. Details on how to use this plugin are provided in the appendix.

## 5.6 JKB DB data exploitation

The JKB data can certainly be a good source of information for other similar applications that want to make use of it. Sharing the data in accordance with the GDPR rules is one of the aspects of the project to be taken into account. Different ways of sharing can be considered in this respect. The datasets could be periodically shared in a special section of the project site in the form of exports in the most common formats such as CSV, Sql file, XLS. Access to the data could be provided through the creation of read-only users and information on access configuration to be provided with whoever requests it. Another interesting way to exploit the project data is the usage of a tool for less experienced users in programming named "MySQL for XLS" plugin already mentioned in section 5.5 and described in detail in the appendix of the document. This plugin/tool give the possibility to easily download the tables content as XLS file and work with it by integrating it into your own applications.

# 6 Knowage, open source business intelligence platform suite, customisation

As per project proposal the prototype of the JKB build within the IO4 is based on the customisation of open source business analytics suite called Knowage made by Engineering Ingegneria Informatica. The suite is released under two different versions, the Enterprise edition with a commercial offering and the Community Edition that is the one used for the DISKOW project, with the whole set of analytical capabilities, open source and entirely free under AGPL v3<sup>8</sup> license. All the improvements made to the suite are regularly released in the various releases of the software and are available to users who want to use it. The adoption of this suite makes the JKB prototype even more interesting, especially in view of exploitation of the project results. Knowage users benefit from free online documentation<sup>9</sup> that guides them through the installation of the suite and provides all the necessary information for both the end user and the administrator as well as a community support the suite adopters helping to solve issues and sharing solutions. The Knowage suite source code repository is available on the official GitHub repository<sup>10</sup>.

## 6.1 JKB application design and refinements

During the different project's phases, the JKB underwent several versions and customizations on the basis of the scenario analyses carried out, evolution of the crawled data from the web and questionnaires assessment provided by the associated partners. These JKB versions have represented respectively the first app design and test within the O4/A2 task, the Alpha release with services and components tests and the collection of the usability feedback via specific questionnaire made by Ekonomicky Ustav Slovenskej Akademie Vied partner within the O4/A3 task, the Beta release to solve bugs, improve services and usability within the O4/A4 task and the final version within the O4/A5 that is available online that will be shown below and whose various functionalities will be explained in detail. The JKB cockpit is reachable at the following <a href="link11">link11</a> with username and password: biuser / biuser . Each user who uses the prototype can use the username mentioned above or request a specific one in the case of special profiling.



Figure 6 - Login form

<sup>&</sup>lt;sup>8</sup> AGPL v3 license, https://www.gnu.org/licenses/agpl-3.0.html

 $<sup>^9</sup>$  Knowage online documentation, https://knowage-suite.readthedocs.io/en/7.2/index.html  $\,$ 

<sup>&</sup>lt;sup>10</sup> Github Knowage, https://github.com/KnowageLabs/Knowage-Server

<sup>&</sup>lt;sup>11</sup> JKB online, http://173.249.3.120:8080/knowage/servlet/AdapterHTTP?PAGE=LoginPage&NEW\_SESSION=TRUE

## 6.2 JKB Cockpit main functionalities

Figure 7 shows the Cockpit of the JKB application home page. This page contains several information about the JKB available data thanks several graphs and a table with the active job posting. The main functionalities provided by the home page are related to:

- The number of European job positions stored in the JKB data base with all the necessary information such as: job title, country, job sector, type of employment, required skills, basic salary, level of education required and the direct link to the relative job provider. All this information is collected in a navigable table, which can be paginated, sorted by columns and filtered through a single or multiple column search. It is also possible to apply a graphical effect to each individual column, such as different colours to emphasize the values of certain data on the basis of certain thresholds such as average salary rather than other specific values.
- A graph that reports for each European country the difference between the basic average salary versus the Estimated salary on the basis of the JKB available data grouped by country.
- A graph that reports the average salary based on the contract type (*freelancing, full-time, internship, part-time, zero hours*).
- A graph that reports the average estimated salary based on the contract type (*freelancing*, *full-time*, *internship*, *part-time*, *zero hours*).

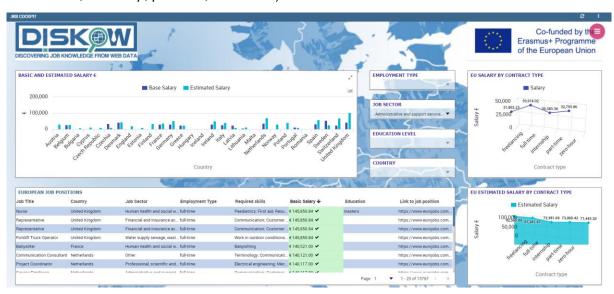


Figure 7 - Dashboard Cockpit page

## 6.2.1 JKB Cockpit in detail

The JKB Cockpit is totally dynamic and the values shown on it are cross-referenced and filterable through a series of filters in the form of a multi-choice menu. The main filters that can be applied are 'Employment type', 'Job Sector', 'Education Level' and 'Country'. The appropriate combination of these filters varies the results in the graphs and in the table of the job offers. At the same time, the filters are linked to each other so they apply an additional level of filtering to what may or may not be selected in

the next filter. These mechanisms are the result of an appropriate customization of the application to benefit the end user in the most modular filtering possible.

Figure 8 shows how a multiple selection is performed on the filter related to the "Job sector" and at the same time shows the search area (red arrow) in case the user want to perform a quick search of the available job sectors directly by search key.



Figure 8 - Multi-choice filter

The values of each filters are not fixed but are retrieved directly from the data base and are applied to all the UI widgets (linked to the same data set of the filter widget). It means that when the user apply a specific filter both graphs and table, are affected and show the new values. Figure 9 shows how the table data is changed based on the applied filter and in particular shows a graphical behavior that change the colors of the *Basic Salary* columns based on some thresholds configured at administrator level better mentioned into the Appendix section.



Figure 9 - European job positions filtered table

Figure 9 shows the table data filtered as from the multiple selection filter. Figure 10 shows how it is possible also to apply free text filter on the table data using the lens-shaped icon on the upper right corner of the table. By using this filter capability it is possible to search a specific word or phrase inside a set of table's columns, these columns can be selected or removed to have a more efficient search. In the following example the user that could be for example a VET agency is looking for all the job offers related to the "data scientist" word. The user in this case is applying the filter to the columns "Country, Job Sector, Employment type, required skills, Education, Link to job position".

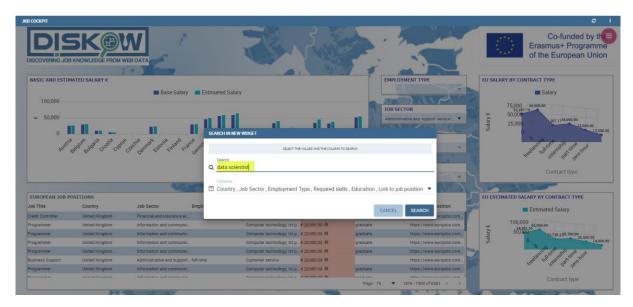


Figure 10 - European job positions filtering all columns

Figure 11 shows the same use of the search functionality however the user in this case prefers to search a specific word "team" into the "Job title" column. The result will report all the job position that contains this specific word. As from image it is possible to note that all the results contain the work "teams" like "Team Manager, Team leader, Team supervisor" and so on.

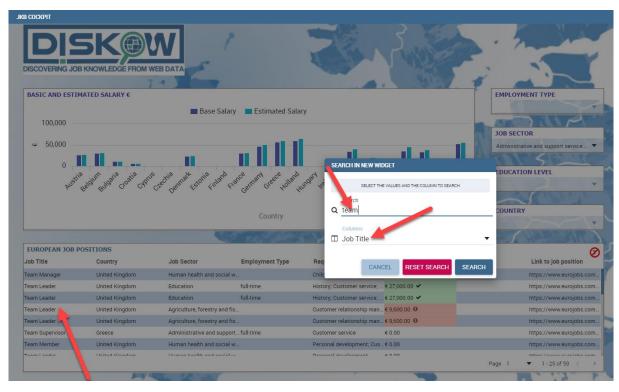


Figure 11 - European job positions filtering specific column

Figure 12 shows another detail of the dashboard that is the graph related to a comparison between the "Basic and Estimated salary" at European country level. These data are related to the job position offers stored inside the JKB data base that contain this information. By this graph the user can have an idea about this remuneration aspect in an easy way and do their own analysis.

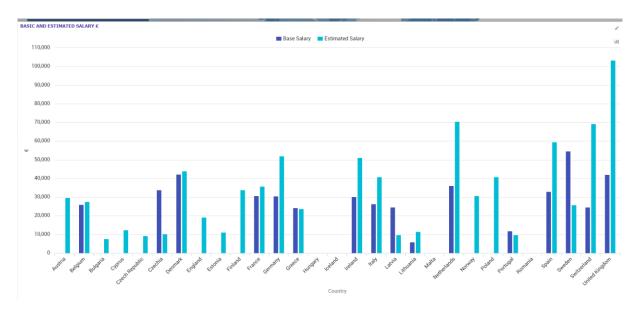


Figure 12 - "Basic and Estimated Salary" graph in detail

It is also possible to select a portion of the graph to obtain a zoom in and better visualize for example data of some specific countries, Figure 13 shows the zoom and the "Reset room" button to restore the previous visualization.

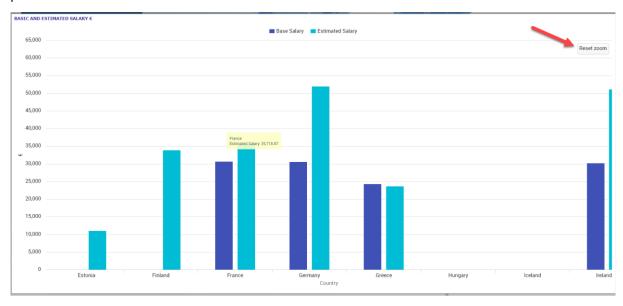


Figure 13 – "Basic and Estimated Salary" - Zoom view

Other two graphs present into the dashboard are the ones related to the "Basic salary and Estimated salary by contract type". Figure 14 shows the one that reports the data of the *Estimated salary*, the same graph is available for the *Basic salary*.

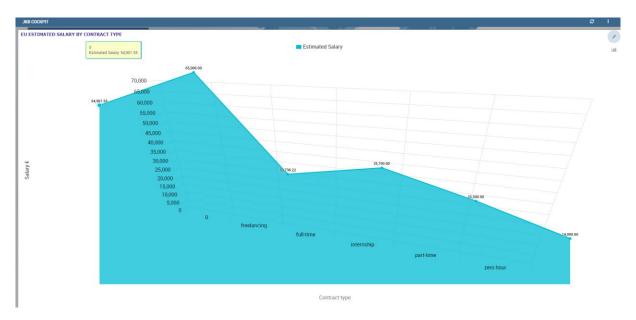


Figure 14 – "Estimated Salary By Contract type" 3D graph

## 6.3 JKB Data Analysis

As is known the main target group and beneficiaries of the DISKOW project and its JKB are job seekers, employees, policy makers, employers, VET providers and labour market experts. As the JKB contains different types of information such as job advertisements, job roles and responsibilities, skills and competences required, salary information, geographical information etc., all of which are available from the initial dashboard, it was considered useful to enrich the functionality of the platform by adding through the Job Knowledge Analysis Engine some dynamic statistics that could facilitate and at the same time address the various stakeholders for their analysis.

Figure 15 shows the JKB Data Analysis page with relative graphs, each graph will be better explained in the next sections.

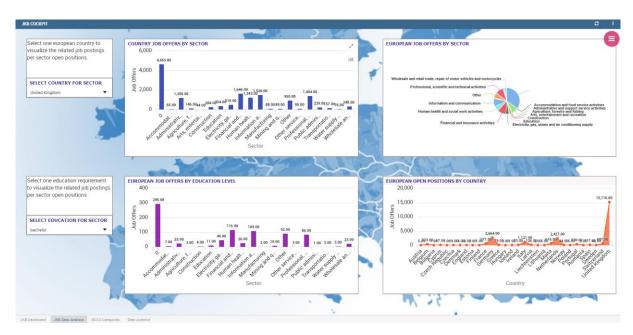


Figure 15 - JKB Data Analysis dashboard

## 6.3.1 JKB Data Analysis in detail

The following dynamic graph in Figure 16 shows the number of vacancies at individual European country level by sector/industry area. The aim of this graph is facilitating the final user of the JKB in identifying for each specific country of the European Union what are the major areas of work and at the same time mature a choice of sectors on which to invest resources for training and productive activities. The graph is the result of an appropriate aggregation of the historical values in the JKB and can be displayed either as a histogram (default version) or as a radar graph Figure 17 (see the appendix to know how to change the chart type). It is also possible to zoom in on a particular portion of the graph to highlight its details. This graph can be filtered by choosing a specific European country through a special selection menu.

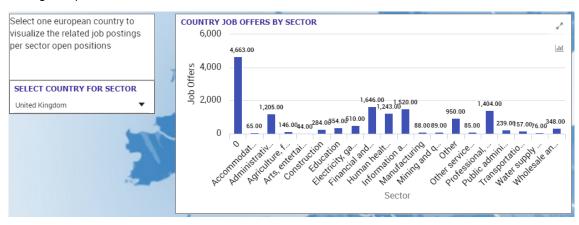


Figure 16 - "Country job offers by sector" graph

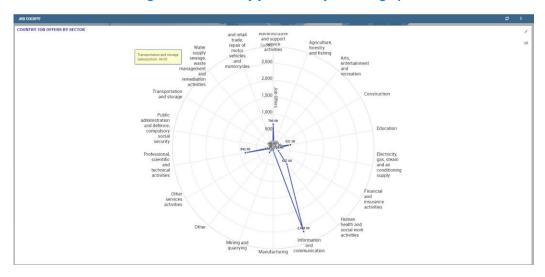


Figure 17 - "Country job offers by sector" radar graph

The following graph in Figure 18 shows the percentage of vacancies at European level broken down by sector/industry area. This graph has been constructed with the aim of facilitating the end-user of JKB to have an overview at European level of the sectors and industrial areas with the highest number of vacancies. The target group of possible beneficiaries is as in the previous chart made up of job seekers, politicians who have to discuss and approve policies related to the world of work and companies offering job training courses. The graph is the result of an appropriate aggregation of the values historicized in JKB and for its peculiarity it can be visualized both as a pie (default version) and

as a wordcloud graph Figure 19 or histograms and so on (see the Appendix to know how to change the chart type).

#### **EUROPEAN JOB OFFERS BY SECTOR**

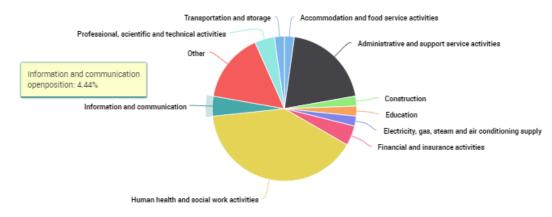


Figure 18 - "European job offers by sector" pie graph

#### **EUROPEAN JOB OFFERS BY SECTOR**



Figure 19 – "European job offers by sector" wordcloud graph

The following graph in Figure 20 shows the number of job vacancies at European level broken down by the level of education required for each industry sector. The target group of possible beneficiaries is, as in the previous graphs, made up of jobseekers who want to understand what studies to follow, politicians who have to discuss and approve policies related to the world of work and academics who want to have a clear idea of the occupational outlets related to the training provided. The graph is the result of an appropriate aggregation of the historical values in the JKB that takes into account all the jobs for which it was possible to determine the required level of education during the data collation phase. Because of the peculiarities of the data, not all job vacancies report this value, so the graph may be smaller than the number of jobs actually in the JKB. In any case, this graph shows the scaled value of the required educational level per sector.

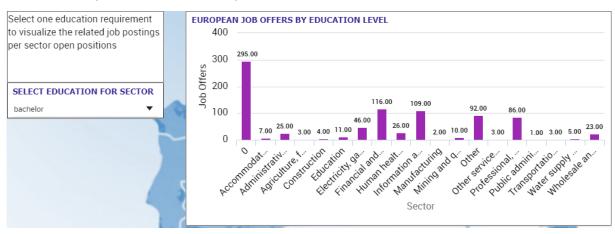


Figure 20 - "European basic salary by sector" graph

The following graph in Figure 21 shows the number of job vacancies at individual European country level. The figure shown is based on an appropriate aggregation on the data in the JKB database. This graph has been constructed with the intention of providing an immediate snapshot of the active job offers for each individual country and to give a possible job seeker the opportunity to see which country might have the most job opportunities. This graph can be assisted by the use of the other graphs in the analytical section of the JKB to identify a specific sector on which to focus their attention. Once the user has made the appropriate evaluations, he or she can then find a job offer through the project dashboard, which contains all the details of each individual job offer saved in JKB. At the same time, the graph can be of help to all those bodies or persons from the world of politics who want to make enquiries regarding the countries offering the most job opportunities and determine their specific social/employment analyses and policies.



Figure 21 – "European open positions by country" graph

## 6.4 JKB ESCO categories

The ESCO categories page in Figure 22 reports the definition, description and relater URL of the available Skills and Occupation definition at European level. This specific pages has been created to give the opportunity to the user to search in an easy way a possible skill or occupation definition and via the URL go to the related web pages that describes all the details of the specific definition. Figure 22 and Figure 23 show the tables that contains the standardized skills and occupation definition and how to search for example a specific definition.

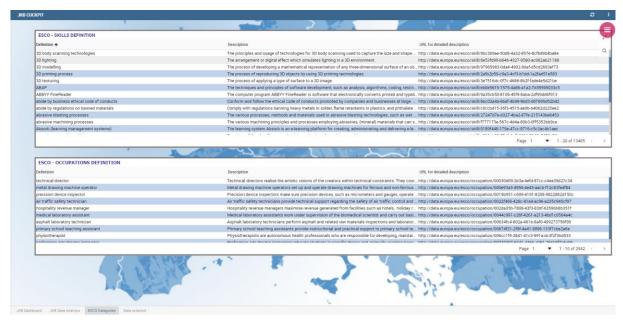


Figure 22 – JKB ESCO categories web page

# 6.5 JKB ESCO categories usage

Figure 23 shows an example how to search the "python" skill from the ESCO –SKILLS DEFINITION table. Opening the corresponding URL the user can visualize the details of this specific skills. The skills definitions reported in this table are the same used into the job position table of the dashboard.

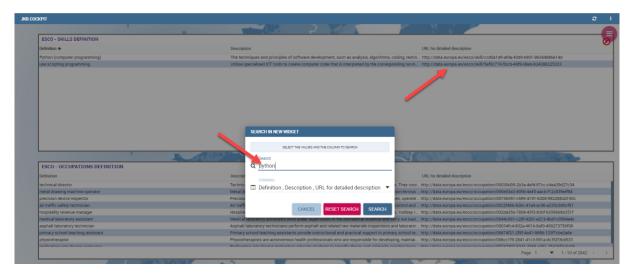


Figure 23 - ESCO categories - Skill table filtering

# **Appendix**

The development cycle of the JKB web UI has seen the creation of several versions, starting with the test versions and continuing with the alpha, beta and final versions. Figure 24 shows the interface used by the Administrator through which it is possible to view the various Cockpit versions, modify, delete or run these.



Figure 24 - JKB Cockpit versions

One of the widgets that has been used for the creation of the JKB is the table widget, this has been modified, customised and refined during the life of the project. Here are some images that show how it is possible to apply filters on the data in the table.

Figure 25 shows how to launch a filter.

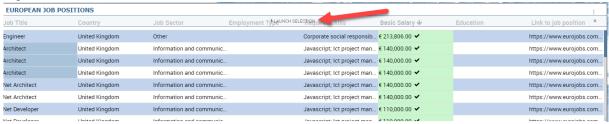


Figure 25 - Table widget selection functionality

Figure 26 shows with the red arrows that it is possible click on the labels of each column to use the values of the labels as filter. The result is shown in Figure 27.

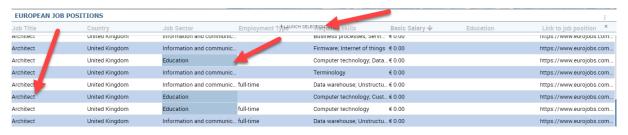


Figure 26 - Table widget multy filtering on columns

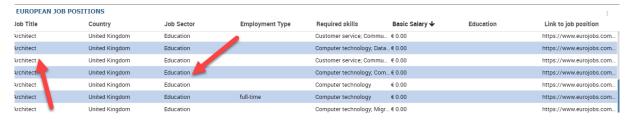


Figure 27 - Table widget filtering

Figure 28 shows the icon to reset all the filters in one click.



Figure 28 - Refresh/Reset filters icon

Figure 29 and Figure 30 show how the administrator can apply colour effects on a specific column of the table defining specific conditions on the data.

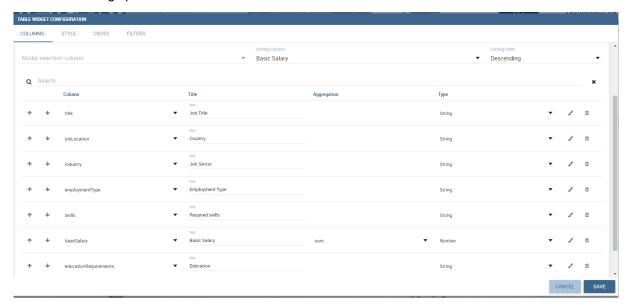


Figure 29 - Table widget customization (columns to visualize)

Figure 30 shows how to define the Threshold.

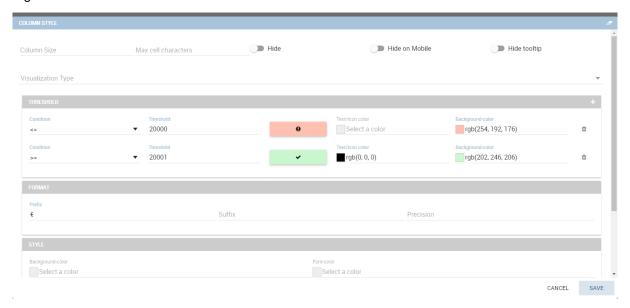


Figure 30- Table widget - conditions on table data

Figure 31 shows how to customize the table widget style.

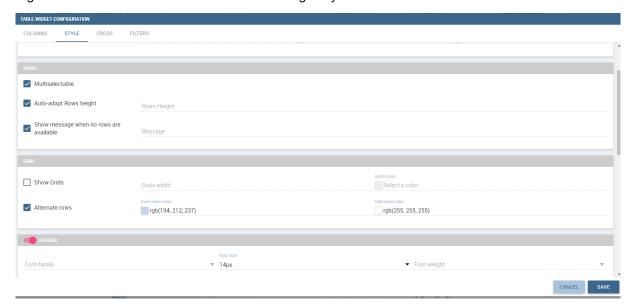


Figure 31 - Table widget configuration

Figure 32 shows how clicking on a specific icon " of the graph widget it is possible to choice a different graph's style from the available ones.



Figure 32 - Chart widget dynamic style choice

Another important widget that has been used in the creation of the JKB web UI is the chart. This widget has been customised in its components in order to better display the data contained in the JKB. The following images show how the Administrator can configure and customize it.

Figure 33 shows how to define the axis values from the available ones of the dataset used for the graph.

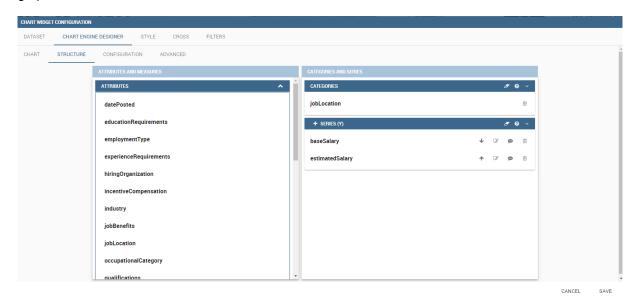


Figure 33 - Chart widget measures configuration

Figure 34, as for Figure 33 but with a smaller dataset.

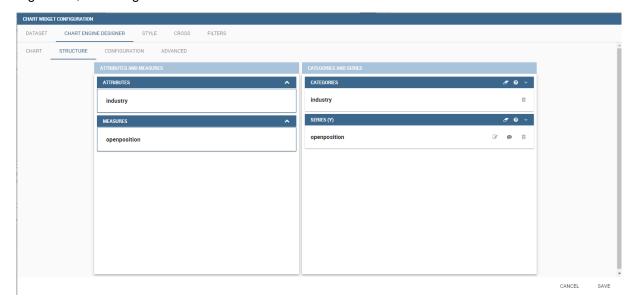


Figure 34 - Chart widget structure/measures configuration

Figure 35 shows where the Administrator can apply a specific filter via a SQL command.

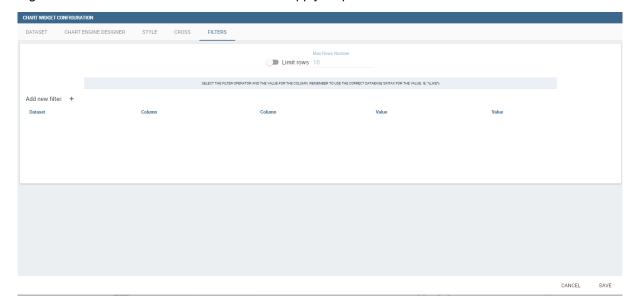


Figure 35 - Chart widget data filter functionality

Figure 36 shows how the Administrator can choice between the available datasets already defined via the specific functionality.

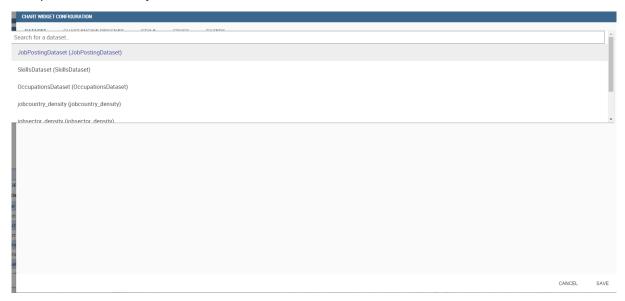


Figure 36 - Widget datasets selection

Figure 37 shows the available chart types that can be used according with the type of data to show.

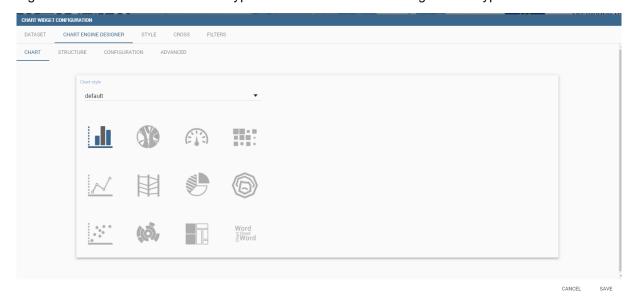


Figure 37 - Chart widget styles

The filter widget is used to apply various forms of selection to data that is displayed in tables and charts. In order to apply filters to these it is important that they are all linked to the same data set so that the filter acts on the data displayed. Filters can be in the form of combo-boxes, multi filters, radio buttons, free text, etc...

Figure 38 and Figure 39 show the configuration of some of the filters used in the JKB web UI.

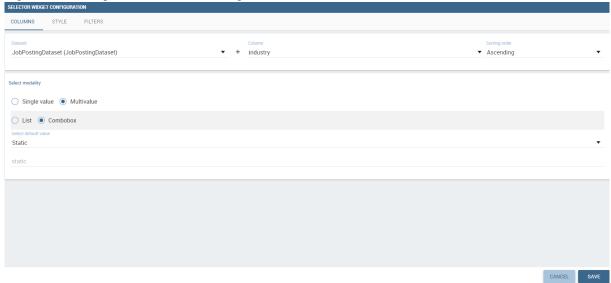


Figure 38 - Selection widget configuration

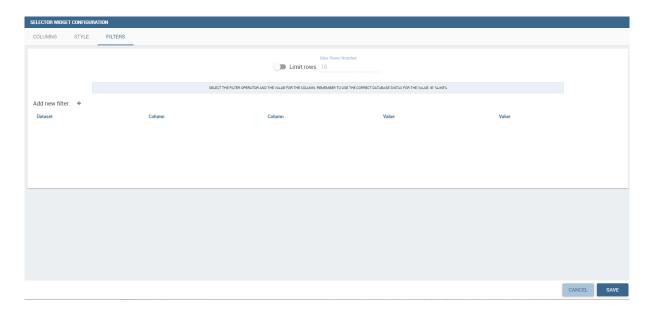


Figure 39 - Selection filter data filtering

Below are some images showing the administration interface through which it is possible to define and create the data sets that will be used within the JKB web UI. These data sets can have various data sources, in the case of JKB the source is the MySql DB that has been already described in this report. The following images show the main data sets used in the project, paying attention to the definition of meta data that are used mainly to define what in the graphics should be considered measures rather than text attributes.

Figures from 40 to 50 show the definition of the main datasets created within the JKB app.

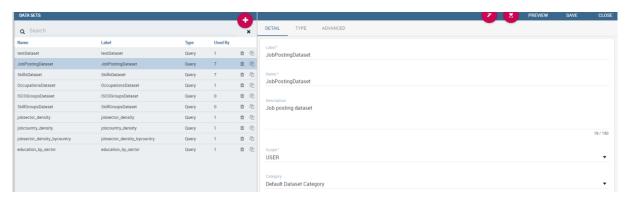


Figure 40 – Job Posting Data set definition

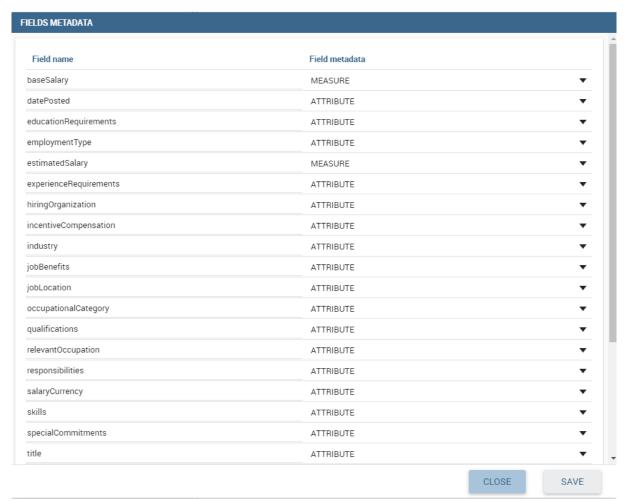


Figure 41 – Job Posting Data set metadata customization

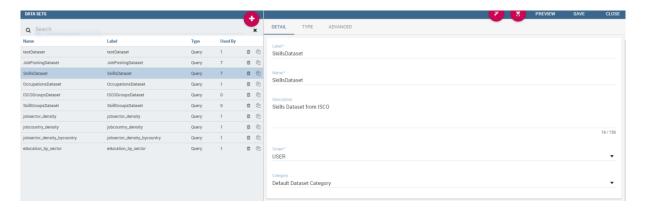


Figure 42 - Skills data set definition

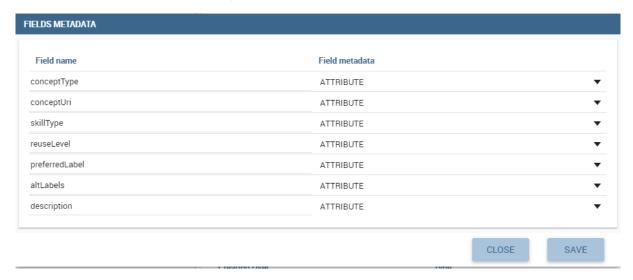


Figure 43 - Skills data set metadata customization

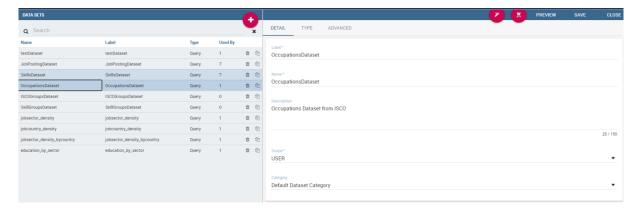


Figure 44 - Occupation data set definition

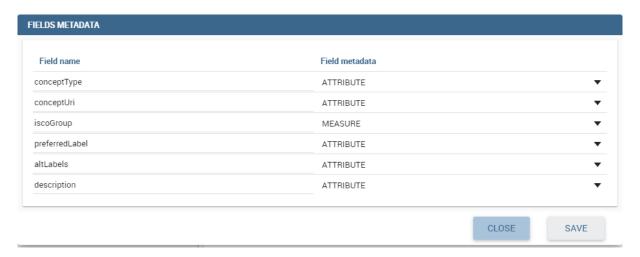


Figure 45 - Occupation data set metadata customization

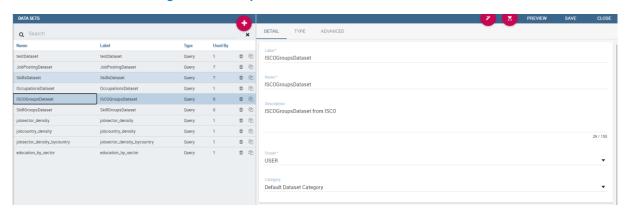


Figure 46 - ISCO group data set definition

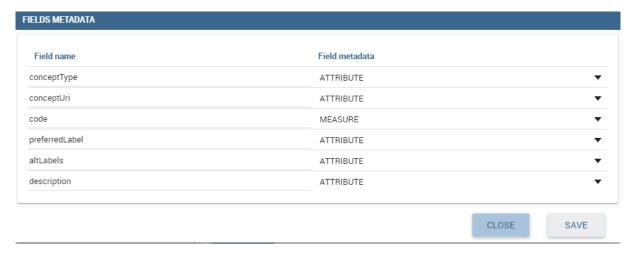


Figure 47 - ISCO group metadata customization

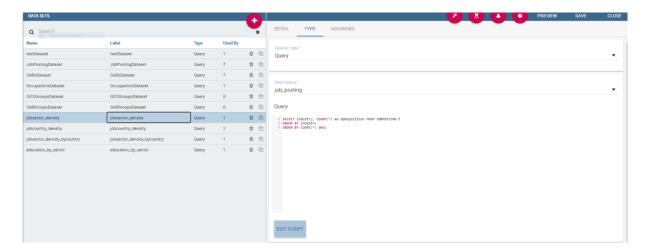


Figure 48 - Custom dataset for graph

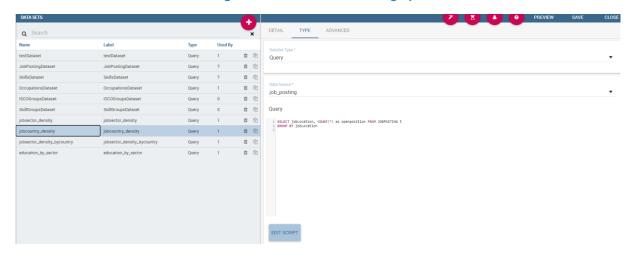


Figure 49 - Custom data set for graph

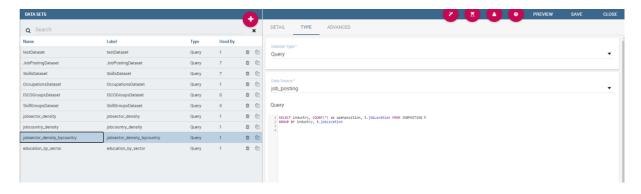


Figure 50 - Custom data set for graph

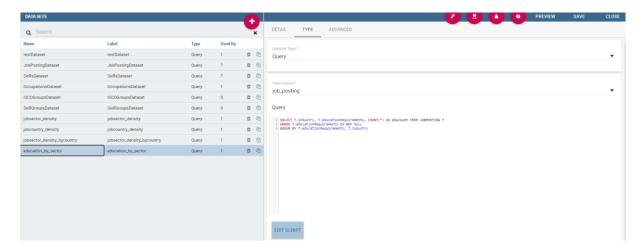


Figure 51 - Custom data set for graph

Below some screenshots related to the main JKB tables visualized throws the free MySQL client HeidiSQL<sup>12</sup>.

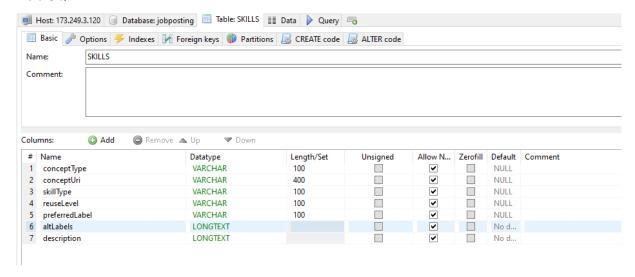


Figure 52 - Skills table

IO4 report: Job Knowledge Base (JKB)

<sup>&</sup>lt;sup>12</sup> HeidiSQL, https://www.heidisql.com/

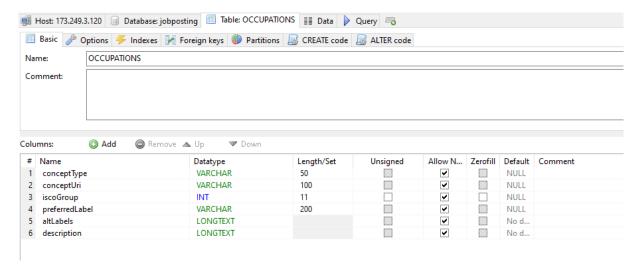


Figure 53 - Occupation table

## MySQL for Excel

MySQL for Excel is a 32-bit add-in for Microsoft Excel, which can be installed with MySQL Installer for Windows. MySQL for Excel does not run on OS X or macOS. MySQL for Excel can interact with MySQL JKB to simplify the management of MySQL connections when both MySQL client tools are installed.

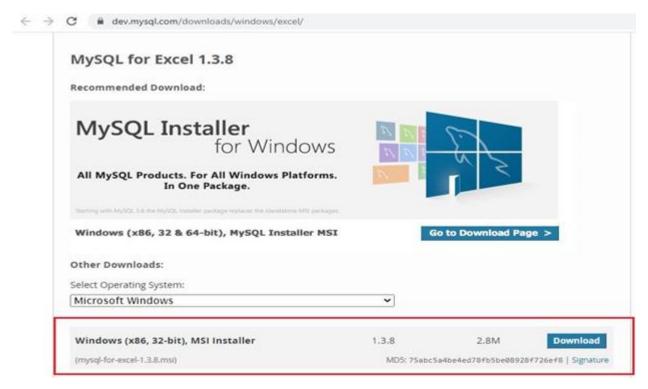
## MySQL for Excel Requirements:

- .NET Framework 4.5.2 (Client or Full Profile).
- Microsoft Office Excel 2007 or later (for Microsoft Windows).
- Visual Studio 2010 Tools for Office Runtime. MySQL Installer may install this for you. (download from <a href="https://www.microsoft.com/en-us/download/details.aspx?id=48217">https://www.microsoft.com/en-us/download/details.aspx?id=48217</a>)

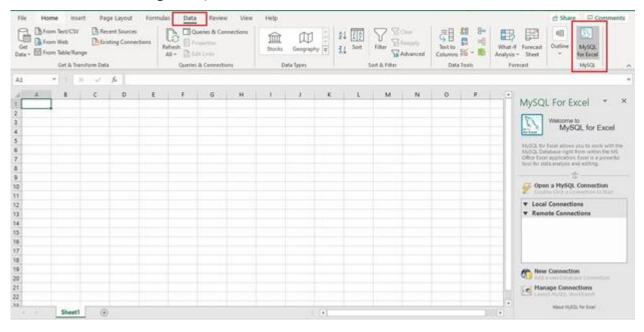
## MySQL for Excel Download and Install

There are two download options available for installing MySQL for Excel:

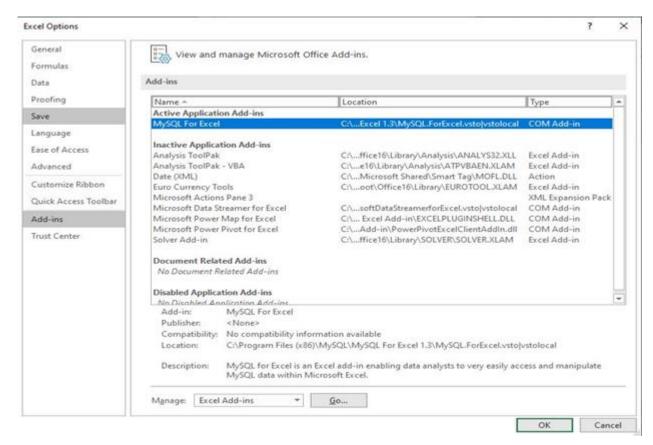
**1. Standalone**: Download and execute the MySQL for Excel <u>standalone MSI file</u>. With this option, it is possible to ensure that all MySQL for Excel requirements are met before executing the MSI file.



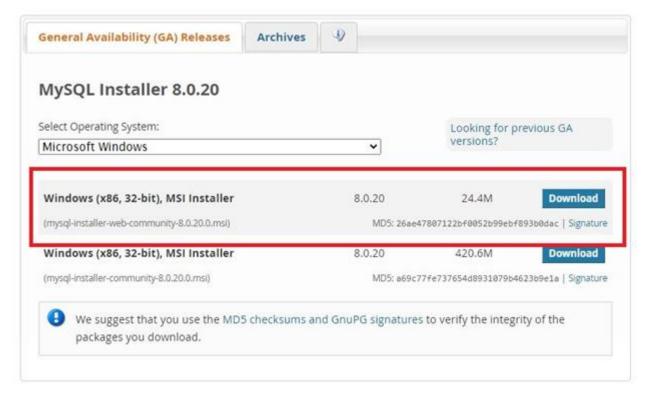
After installation, MySQL for Excel is loaded and executed by selecting the **Data** menu tab in Excel, and then clicking the **MySQL for Excel** button from the ribbon.



Note: If the MySQL for Excel plugin is installed, but can't be find it in Microsoft Excel, it could be necessary to check the Excel Add-Ins list. This is found by clicking the File (top left corner), click Options from the Menu, then select Add-Ins. MySQL for Excel should be listed as a COM Add-in.

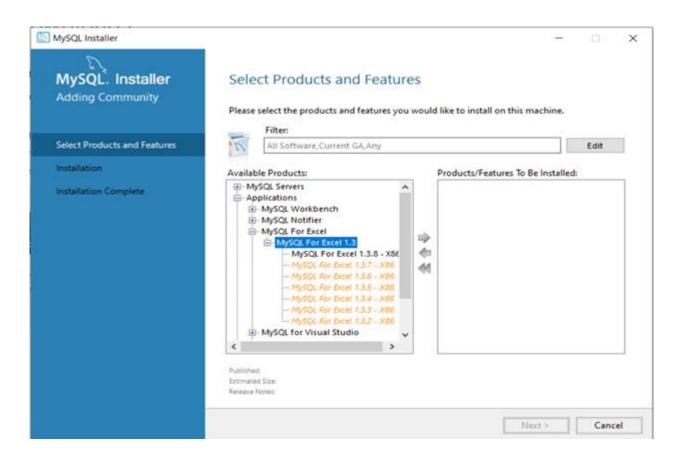


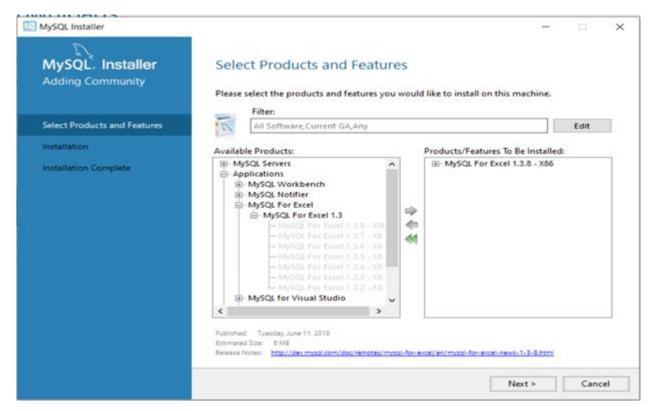
**2. MySQL Installer**: Download and execute the MySQL Installer MSI file. This option is recommended because, in addition to managing all MySQL client applications for Windows, MySQL Installer evaluates the installation requirements on the computer for the user.



When MySQL Installer is already installed: Use the Add operation to install MySQL for Excel.

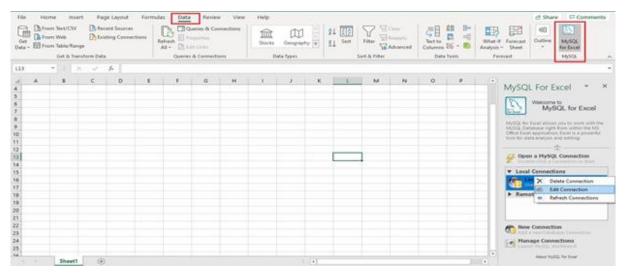




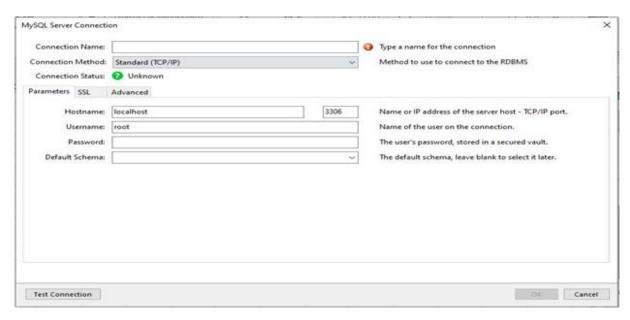


MySQL Connections in Excel

MySQL for Excel provides several options to create and manage MySQL connections. You must open a connection to a MySQL server before you can configure global options, add a new schema, or perform operations that move data between Excel worksheets and MySQL tables.

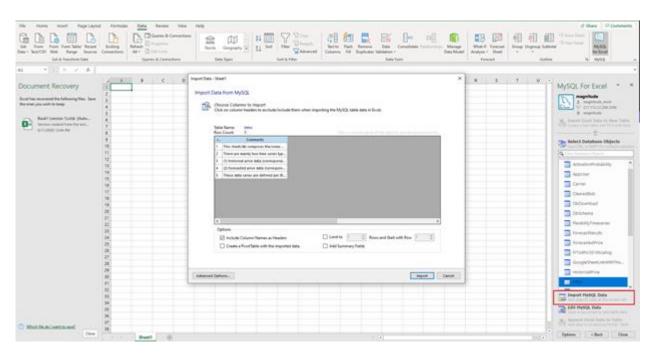


**Adding New MySQL Connections**: To add a new connection, click **New Connection** in the MySQL for Excel task pane to open the MySQL Server Connection dialog. Connection names must be unique. An alert icon (!) indicates that an option value is required. The figure that follows shows the connection dialog with the **Parameters** tab selected.



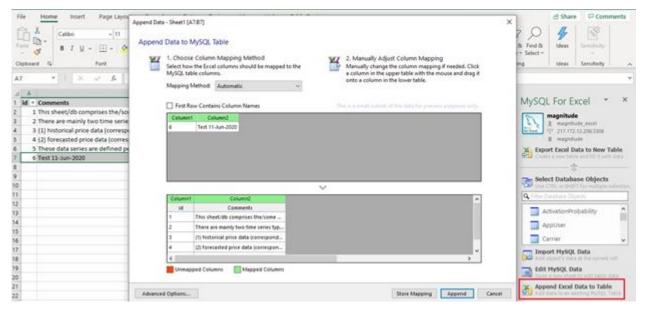
MySQL for Excel can be used to edit existing MySQL connections. To edit a MySQL connection, right-click the connection you want to modify and select **Edit Connection** from the context menu.

To import MySQL data into an Excel worksheet you need to choose table, for example into (click on it), then click on Import MySQL Data, click on Import.



Other option is to right-click on the table you choose and then choose **Import selected and related tables**. This option should not be used due to Excel limitations, such as not supporting composite keys, so relationships between tables cannot be created properly (check: *Unsupported database features in the Excel Data Model*).

**To add a new row**: Add the row you want in the sheet, select that row and then click **Append Excel Data To Table** in the sidebar on the right. In the opened modal, map the columns and then click **Append**. And the rows are appended.

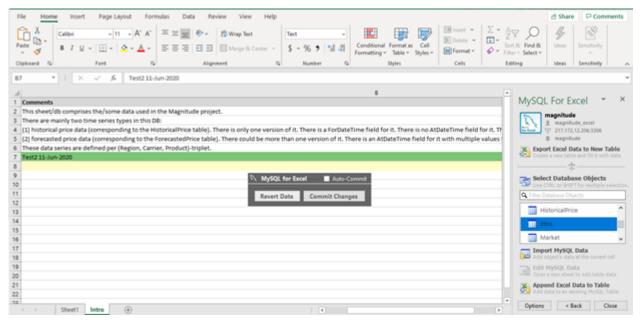


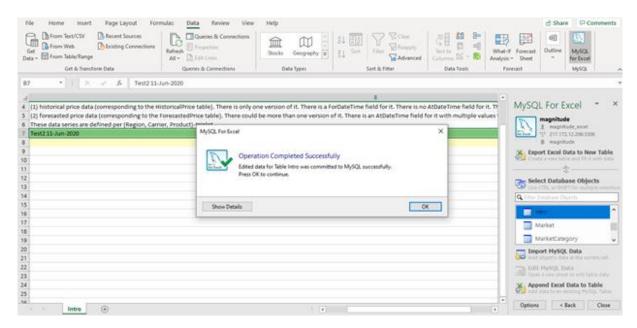
The color of the header field for each column defines the current mapping status of the column. The colors include:

- Green: A source column is mapped to a target column.
- Red: A target column is not mapped.
- Gray: A source column is not mapped.

Mapping can be done automatically, stored (using a *dbname.tablename* naming scheme) or manually (by dragging a column from the upper source grid and dropping it into the lower target column MySQL table grid).

To change the data in the table: Choose table first then click on Edit MySQL Data in the sidebar on the right. You will see the preview and click OK. Make the changes you want in the sheet, save the sheet, and click on the Commit changes and changes will be done in the MySQL database.





**Delete data from table**: If you want to delete a whole row from the MySQL table you need to delete the row as you would in Excel, i.e. by right-clicking the row heading and selecting from the pop-up context menu the **Delete** option and click on the **Commit changes** to save the changes to the database.

