

This publication is a Technical report by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication.

JRC Science Hub

<https://ec.europa.eu/jrc>

JRC104721

EUR 28339 EN

PDF ISBN 978-92-79-64594-5 ISSN 1831-9424 doi:10.2788/226897

Luxembourg: Publications Office of the European Union, 2016

© European Union, 2016

The reuse of the document is authorised, provided the source is acknowledged and the original meaning or message of the texts are not distorted. The European Commission shall not be held liable for any consequences stemming from the reuse.

How to cite this report: Gammieri V; Alvarez Alvarez M; Barbas T. *Data Catalogue Knowledge Centre on Migration and Demography*, EUR 28339 EN, doi:10.2788/226897

All images © European Union 2016

Contents

- Acknowledgements & Abstract2
- 1 Introduction.....3
- 2 Methodology4
 - 2.1 Collection of datasets4
 - 2.1.1 Classification4
- 3 Implementation.....5
 - 3.1 The CKAN software5
 - 3.1.1 Features used5
 - 3.2 Apache Solr5
 - 3.3 Metadata5
- 4 Tool6
 - 4.1 Datasets.....6
 - 4.2 Publishers.....8
 - 4.3 Domains.....9
- Conclusions 11
- References 12
- Annexes 13
 - Annex 1. Classification into domains 13
 - Domains* 13
 - Keywords* 16
 - Annex 2. Metadata..... 17

Acknowledgements

We would like to thank the various DGs and services of the Commission that contributed in reviewing the tool, and sharing their feedback with the authors. Their contribution has been very valuable in implementing a number of improvements.

We would also like to thank the colleagues working for the JRC Data Catalogue who helped us with the software. Finally, we would like to thank the many colleagues of our team who helped us to review the metadata of the datasets.

Abstract

The Migration Data Catalogue is an inventory of migration relevant datasets made available by international organisations, European institutions, in including in particular the European Commission and EU agencies as well as several EU Member State administrations. This thematic catalogue classifies and organises the datasets in domains addressed by the European Agenda on Migration reflecting current Commission priorities, including topics like legal migration and integration, asylum and refugees, irregular migration and returns and unaccompanied minors. In addition to finding data for each of these predefined domains, the catalogue can be searched using keywords.

For each dataset the catalogue gives key information e.g. on its content, methodology of collection, geographical and temporal coverage and frequency of update. By giving this insight into the usability of the datasets for analysis, the catalogue provides guidance on available migration related data not only to EU policymakers but also to other interested parties.

The purpose of this document is to describe the catalogue, its implementation and its functionalities.

1 Introduction

The **KCMD Data Catalogue** is a product of the Knowledge Centre on Migration and Demography. It is a thematic collection of datasets organised in different policy areas that take into account the current EU priorities in migration. Information on contents, methodology, geographic and temporal coverage, as well as on frequency of update for the different sources is there to provide insights into what data is available, including its potential for supporting different types of analysis. The catalogue is a live product and as such it is expected to evolve and be kept up-to-date over time. Note that it does not provide direct access to the data for analysis, but rather links to the relevant data sources.

The target audience of this data catalogue is primarily policy makers and practitioners of migration. The catalogue is intended to answer the question "what data exists already that can be of use to study migration and develop migration policies". Academics and other experts are encouraged to use the links to the data sources to find more detailed and updated information directly from the data providers themselves.

The catalogue contains both official, international statistics at the EU and the global level, and selected operational data, including at the Member State level. Key cross-cutting datasets for the study of migration are all intended to be included. Users are warned to exercise care before using short-term operational data, for instance, to provide evidence. Inclusion of such data is there to help with completing the overview, which is the primary purpose of this knowledge management product. KCMD welcomes suggestions as to additional data sources to be included in this product or about modifications of the descriptions. In its web site it suggests use of the functional email kcnd@jrc.ec.europa.eu with "KCMD data catalogue" in the subject.

2 Methodology

2.1 Collection of datasets

Data sets were identified and analyzed in the following way:

1. Exploring relevant websites; (international datasets)
2. Harvesting Open data portals (European, national government datasets);
3. Reviewing the extracted metadata.

In Table 1, the main Open data portals that were considered in the second step.

Table 1. Main Open data portals that were considered

Portal	URL link	description	Result of search (# items) ⁽¹⁾
EU open data portal (EU ODP)	http://data.europa.eu/euodp/en/data	data and reports from Commission DGs and services or EU Agencies	"Migration" (165 items)
European Data Portal	http://www.europeandataportal.eu/	Data from MS national data portals.	"Migration" (1,882 items)

⁽¹⁾ searching with keyword='migration'.

In the European Data Portal, using "migration" as the keyword, we found various governmental portals with more than 20 datasets each: Netherlands, Germany, United Kingdom, Italy, France, Spain, Finland, Austria, Ireland.

Through these portals we found many entries for the same source (e.g. all the quarterly issues of the same data report). In these cases, we considered them as if they were one. For example in EU ODP, there are 25 datasets corresponding to the "FRAN Quarterly Reports" published by Frontex, while in the KCMD Data Catalogue metadata about it is entered only once.

2.1.1 Classification

Two main ways of classification are used, namely, through predefined domains (policy areas, high-level topics) and through free keywords.

Classification was based on an interim KCMD thesaurus-like classification (see annex 1).

3 Implementation

3.1 The CKAN software

We selected the CKAN tool as it is one of the most used open-source tools worldwide for publishing data portals and catalogues. CKAN is currently used by governments and organizations to power both official and community data portals. Some examples are: EU Open data portal; JRC data catalogue¹; European Data Portal;

CKAN² provides for a search capability which allows for quick 'Google-style' keyword search as well as faceting by tags for drill-downs and also browsing between related datasets. Users can quickly see what datasets are available, straight from the search results. All dataset fields are searchable.

CKAN has several features including fuzzy-matching; faceted search (drill-down); search via API. CKAN is highly customizable, allows one to choose which features to use. There are over 60 different extensions and a modular architecture that allows extensions to be developed to provide additional features such as harvesting or data upload.

3.1.1 Features used

Fuzzy-matching – option to search for closely matching terms instead of exact matches.

Faceted search – drill-down via facets – for example domains, tags, provider. *Faceted search* it is the ability to consecutively narrow the search by further facets allowing users to limit their search to datasets for a specific domain or with certain tags after they see the search results.

Customization by extension - A custom extension was built in order to provide the theme and additional functionalities like custom sort, version management, new faceted search and so on.

3.2 Apache Solr

Apache Solr³ is an open source software search platform. Solr powers the search and navigation features. Our search function is entirely based on metadata cached in Apache Solr. Two search solutions are combined together to allow for an efficient search mechanism to take place:

- Text search, which combines full text search with faceted search on filters
- Keywords search, based on predefined Keywords

3.3 Metadata

The vocabulary of the metadata is using the Data Catalogue Vocabulary (DCAT-AP⁴) standard defined at EU level and designed to facilitate interoperability between data catalogues published on the web.

The main attributes are: domains; keywords; title (or dataset name); description; methodology; publisher; mode; data type; temporal coverage; spatial coverage; identifier at publisher; link to the data; links to documentation.

More details are given in annex 2.

¹ <http://data.jrc.ec.europa.eu/>

² <http://ckan.org/>

³ <http://lucene.apache.org/solr/>

⁴ https://joinup.ec.europa.eu/asset/dcat_application_profile/description

4 The tool

The tool allows searching the catalogue by domain, by publisher or through keywords.

Link: <https://bluehub.jrc.ec.europa.eu/catalogue/>

In Figure 1 the main page, that offers direct search in the dedicated box (in the middle of the page) and search by domain, clicking on one of the twelve domain icons (in the bottom of the page).

There are 5 tabs accessible in the navigation bar: home; datasets; publishers; domains; about.

Figure 1. KCMD Data Catalogue main page (home).

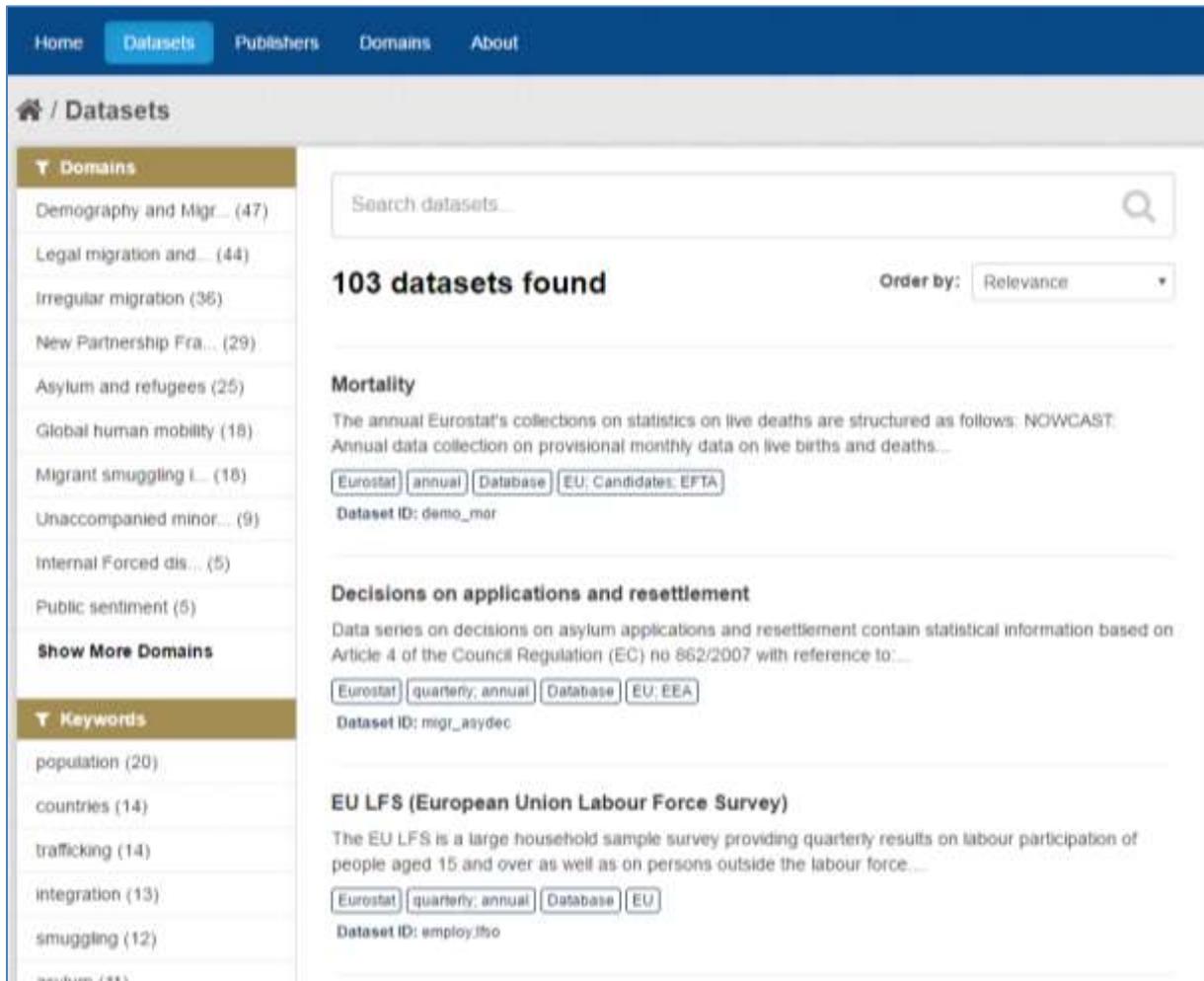


4.1 Datasets

The main datasets page (figure 2) provides the access to the datasets, with a direct search using a keyword on the top box. This page presents a summary for each selected dataset: title, short description, publisher, frequency, mode, spatial coverage, dataset ID.

It is possible to restrict the selection using the search facets available on the left-hand side, filtering by domain, keyword, mode, spatial coverage, data frequency or publisher.

Figure 2. Datasets search page



Selecting a single dataset will allow to see more details (figure 3): long description and methodology, list of domains the dataset belongs to, direct links to the publisher: data and documentation (metadata if available), keywords, dataset ID, data frequency, temporal coverage, status, spatial coverage, mode, data type, access on, category.

Figure 3. Single dataset description page

Home / Publishers / Eurostat / Residence permits

Publisher



Eurostat
statistical office of the European Union [read more](#)

i Catalogue Record

Added on: 2016-12-02

Updated on: 2016-12-02

Social

Google+

Twitter

Facebook

Residence permits

Residence permits issued for third country nationals in the EU and EFTA countries.

Residence permit is defined as any authorisation valid for at least 3 months issued by the authorities of a Member State allowing a third country national to stay legally on its territory. First permit means the residence permit issued to a person for the first time. A residence permit is considered as a first permit also if the time gap between expiry of the old permit and the start of validity of the new permit issued for the same reason is at least 6 months, irrespective of the year of issuance of the permit.

The statistics are provided by the national responsible authorities, mainly Ministries of Interior or Immigration Agencies of the EU MS and EFTA countries. Data are based entirely on relevant administrative sources. These data are supplied by MS as part of the annual Residence Permits Data Collection conducted by Eurostat according to the provisions of Article 6 of Regulation 862/2007 on statistics on migration and international protection. Data can be disaggregated by reason, length of validity, age, sex and citizenship.


Asylum and refugees


Global human mobility


Legal migration and integration


New Partnership Framework

Data and Resources

-  **Residence permits**
-  **Reference Metadata in Euro SDMX Metadata Structure (ESMS) HTML**
-  **Residence permits statistics explained on Eurostat Website**

asylum
residence permits

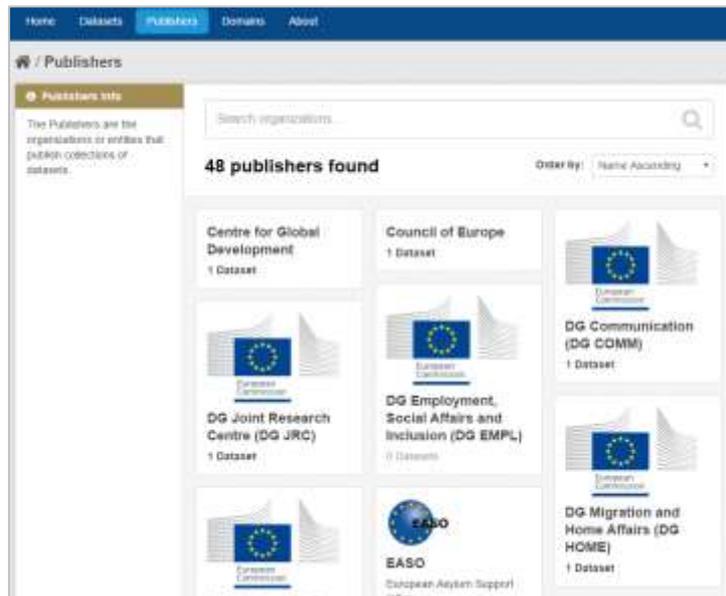
Additional Info

Field	Value
Dataset ID	migr_res
Data Frequency	annual
Temporal Coverage Start	2008
Status	Active
Spatial Coverage	EU, EFTA
Mode	Database
Data Type	statistical
Accessed On	11/24/16
Category	official statistics

4.2 Publishers

The Publishers are the organisations or entities that publish collections of datasets. At the moment of the launch, we had 48 different publishers in our catalogue. Although some of them still do not have an associated dataset inserted, they are in the list as important organizations in the field of migration.

Figure 4. Main Publishers page



4.3 Domains

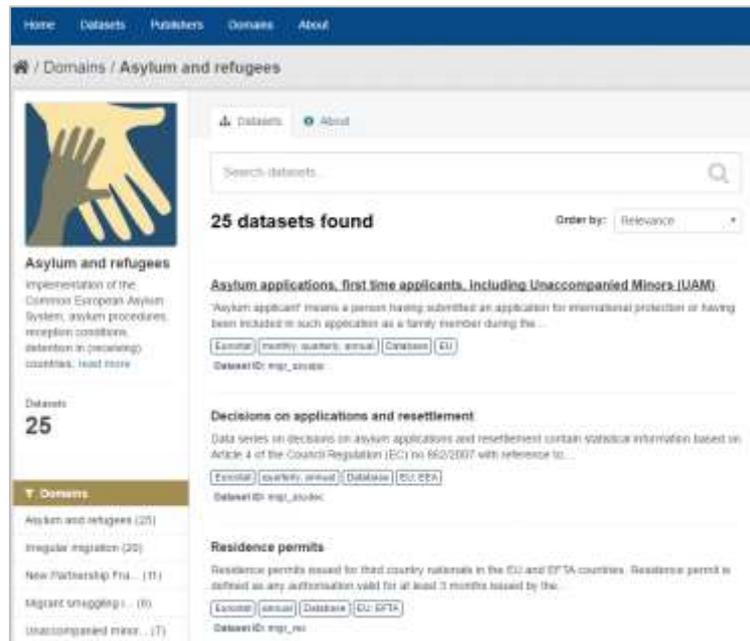
Domains are used to classify by subject the various datasets in the catalogue. They offer an alternative way – in addition to searching for specific terms - to find datasets that are relevant for specific topics or areas. These domains include, for example, irregular migration, asylum and refugees, migrant smuggling, development and migration, legal migration, migrant integration, forced displacement, and other topics.

Figure 5. Main Domains page



There is the possibility to search for all datasets classified under a given domain, clicking in the domain icon either from the Home page (in figure 1) or from the domains main page (in Figure 4). When this is done, the domain page appears, like in figure 6 (for domain "Asylum and refugees").

Figure 6. Single domain page



It is possible to click on "About" to see the description of the domain (figure 7). More details on the twelve domain can be found in annex 1.

Figure 7. Single domain page



Conclusions

The KCMD Data Catalogue gained very much from feedback by the competent DGs and Services of the Commission, including by the EU Agencies.

The KCMD Data Catalogue will continue to be developed and is destined to include additional relevant datasets as they become available or known. It will evolve over time to include datasets that are to date still inconsistent or scattered across public or private sources and alternative sources of data like big data.

This exercise is intended to contribute to the overall knowledge management effort related to the KCMD Knowledge Centre, including inventories of activities by DGs and Services, stakeholder organisations and networks, relevant projects and studies.

References

CKAN. <http://ckan.org/>

DCAT-AP. European Commission, ISA Programme. DCAT application profile for data portals in Europe. Online: https://joinup.ec.europa.eu/asset/dcat_application_profile/

Open Data (<https://ec.europa.eu/digital-single-market/en/blog/european-countries-are-reaping-benefits-open-data>)

SOLR. <http://lucene.apache.org/solr/>

Vaccari, L., Perego, A., Friis-Christensen, A. and Dalla Benetta, A., *JRC Data Catalogue*, JRC95305, 2015.

List of abbreviations and definitions

KCMD	Knowledge Centre on Migration and Demography
ODP	Open Data Portal
CKAN	Comprehensive Knowledge Archive Network
API	Application Program Interface
DCAT-AP	Data Catalogue vocabulary – Application Profile

Annexes

Annex 1. Classification into domains

Domains

12 domains were selected to represent the migration and demography landscape, as described below. Note that data sets may be classified into more than one domain, and that there are overlaps between domains.

1-6 of the domains below were indicated as priority by the Steering Group of the KCMD. Note that not all domains are comparable in terms of volumes and scope.

This list is only indicative and by no means exhaustive.

1. Legal migration and integration (code = 1LM)



- Legal migration and visas (high or low skilled labour, students, seasonal workers, family reunification, blue cards etc.);
- Integration of third country nationals (socio-cultural aspects, education and skills, labour force participation, family reunification, residence permits, naturalization, acquisition of citizenship, regularisation, diasporas etc.);
- Cities (for the local impacts of integration);
- Impacts on destination countries related to fiscal, economic, labour, welfare aspects etc.;
- Databases for the management of external border management, asylum and visas;
- Various statistics, surveys or other data and reports on income, labour force participation, living conditions, education, social inclusion, poverty, inequality, international migrant stocks and migration flows, EU statistics on short-stay visas issued, determinants and evolution of international migration.

2. New Partnership Framework (code = 2PF)



New Migration Partnership Framework, reinforced cooperation with third countries to better manage migration.

The external dimension of migration policy:

- Development and Migration (GAMM and the new Partnership Framework, Trust Funds, compacts, EU Neighbourhood policy, EU development policy, human development, development indicators, governance indicators, poverty indicators, remittances, SDGs, follow-up to the 19 September 2016 UN Summit, etc.); development aid by rich countries;
- Root causes and determinants (conflicts, inequality, climate, economy, demographics etc.);
- Country profiles;
- Impacts of migration on origin countries;
- Use of funds (including soft loans);
- Readmissions, reintegration, resettlement; illegal border crossings;
- Job creation compacts, school enrolment;
- International migration flows, determinants and evolution of international migration, migrant aspirations.

3. Unaccompanied minors and vulnerable groups in migration (code = 3UNMIN)



Unaccompanied minors and vulnerable persons in migration:

- Unaccompanied minors in irregular migration or seeking asylum; vulnerable groups;
- Trafficking in human beings (THB);
- Fundamental rights;
- Child labour;
- Modern slavery.

4. Migrant smuggling into the EU (code = 4SMUG)



- Borders and security;
- Illegal border crossings;
- Migrant smuggling;
- Trafficking in human beings;
- Compliance with relevant international instruments or policies (against human trafficking);
- EU statistics on short-stay visas issued;
- Global slavery;
- Forced labour;
- Mixed migration flows;
- Undocumented migrants in the EU.

5. Irregular migration and returns (code=5IM)



- Illegal border crossings;
- Asylum applicants and asylum applicant decisions;
- Mixed migration flows towards Europe, arrivals including irregular migrants and fatalities;
- Country profiles at the origin (push factors);
- Stocks and conditions for undocumented migrants in the EU;
- Refusals of entry to the EU territory;
- Apprehensions of irregular migrants;
- Overstaying of visas;
- Return decisions and effective returns;
- Databases for the management of external border management, asylum and visas.

6. Forced displacement (code=6FD)



Forced (internal) displacement, root causes and populations of concern.

- Forced (internal) displacement of populations;
- Root causes in countries of origin;
- UNHCR populations of concern (for internally displaced persons).

7. Asylum and refugees (code=7ASY)



- Implementation of the Common European Asylum System, asylum procedures, reception conditions, detention in (receiving) countries;
- Asylum applicants including unaccompanied minors, their numbers, trends and relevant indicators, applications for asylum and their status, decisions on asylum applications granting protection statuses at various stages including negative decisions, information relevant for asylum decision makers e.g. country of origin profiles, relocations, resettled persons, residence permits for TCNs by reasons for issuance;
- Mixed migration flows towards the EU, migrant arrivals including irregular migrants and fatalities, UNHCR populations of concern;

8. Demography and Migration (code=8DEMO)



- EU household and labour force surveys;
- EU income and living conditions, labour force estimates, acquisition of citizenship, child labour, global estimates of migrant workers;
- Demographic characteristics of immigrants, international migration statistics (stocks and flows);
- Population statistics;
- Adult competences assessments, international student assessments;
- Global patterns of migration, global migration policy change monitoring.

9. Internal EU mobility (code=9MOBI)



- Intra-EU mobility;
- EU regions and cities;
- Transport statistics.

10. Global human mobility (code=10GMOB)



- Global human mobility;
- Transport statistics;
- International migration statistics;
- Global labour force statistics and estimates;
- Global demographics.

11. Public sentiment (code=11SENT)



- EU public opinion, attitudes, beliefs, behaviours, Eurobarometer;
- Migration aspirations and sentiments (for non-EU prospective migrants).

12. Migration governance (code=12MGOV)



- Multi-level governance to manage migration;
- Comprehensive approaches linking migration, development and security;
- Governance indicators (not only on migration), input for governance indicators;
- Global migration policies monitoring.

Keywords

The keywords shown below constitute the initial set of keywords for the exercise. They are listed in thematic order. The list is not exhaustive, as more keywords were introduced in the course of cataloguing the sources. The authors of the catalogue were allowed to select their keywords freely.

	keywords
(Disciplines):	economics, sociology, demography, policy, politics
(type of work):	model, theory, empirical analysis, spatial analysis
(chars of migrants/migration):	internally displaced, migrant, returnee, migrant stocks, migrant flows, emigration, immigration, asylum seeker, refugee, persons of concern (UNHCR)
(types):	regular, irregular, forced, voluntary, economic, family, study, seasonal, high-skilled, low-skilled
(geo / levels concerned):	countries, communities, individuals, origin, transit, destination, regional, local, urban
(chrono):	causes, impacts, impact on origin, impact on destination, impact on transit, impact on migrants
(type of document):	academic literature, analysis report, policy analysis, communication material
Various	
(external)	Development, remittances, drivers, conflicts, fatalities, climate, trade, diasporas, migrant communities, programmes, funds, compacts, development aid, humanitarian aid, regional cooperation, development cooperation, returns, readmissions, resettlement, country profile
(rights)	Citizenship, ethnicity, national identity, multiculturalism, assimilation, discrimination, gender, language, religion, fundamental rights, women, children, youth, vulnerable groups, disabilities, racism, xenophobia, minorities
(impacts on)	Housing, economic, education, labour, fiscal, welfare, health
(internal)	Inequalities, poverty, diversity, employment, visa, blue card, skills, educational attainment, occupations
	Smuggling, trafficking, visa overstayers
	Globalization, governance, sentiment
(demography)	population, census

Annex 2. Metadata

Fields	Mapping CKAN	values; examples	Notes
id	id		internal dataset identifier
title	title	free text	
short description	notes	free text	
methodology	extra:provenance	free text	
publisher	organization	Ex. EUROSTAT; IOM; UNHCR	
domain	group	12 domains defined in the KCMD thesaurus	a single dataset can be associated to more than one domain (use ;) use ids from domains tab: 1LM;2PF;3UNMIN;4SMUG;5IM;.....
keywords	tags	Ex. Labour market; income; fatalities	a single dataset can be associated to more keywords (use ;) use keywords from thesaurus
potential for analysis		x	when the dataset is VERY RELEVANT for "potential for analysis"
coverage (spatial)	spatial	international; ASEAN; OECD; EU;Mediterranean;EFTA; Schengen;Central Asia; national;Western Balkans; Italy; Spain; Germany; ...	national datasets from MS will be indicated here with the country publishing those as well as 'national' ex. National;Italy
mode		database; index; report; project	often index can be found in the dataset title or description. Even if more appropriate in the following attribute, we leave it here database (also if the datasets are provided in excel tables) only regular published reports , when those include data in form of tables and it is the only way this type of data it is published (Ex. From FRONTEX)
type_of_dataset		survey; statistical; estimations; projections; indicators; index; empirical	often this can be found in the dataset title or description in few cases (specially reports) there is the need to specify to types, use ; Ex: statistical; projections

data frequency		regular; occasional; daily; weekly;biweekly; monthly; quarterly; annual; biannual;5-years;10-years	periodicity (frequency of data update) Example of ' occasional ': incidents from missing migrants project
publication (frequency of publication)		as above	Only in few cases the frequency of publication is not the same as of the data update
temporal coverage: start	extra:temporal_start	Ex. 2008; 2008M3 or 2008-03; 01/10/2015	
temporal coverage: end	extra:temporal_end	as above	often, when the dataset it is not historical, this will be empty only for annual datasets, the last available year is going to be specified
status		active;	to indicate if the dataset is historical (no, therefore discontinued) or still being updated (active)
Dataset ID	extra: alternate_identifier	Ex. migr_asyappctzm	internal identifier by publisher, not always available at the provider website (try to put name of file)
link	extra:publisher_url		link to the data at publisher website
documentation_links		list of urls separated by newlines, ex: url1 url2	Ex. metadata link (to metadata description at publisher website) links to explanatory data statistics
documentation_descriptions		list of descriptions separated by newlines, ex: description1 description2	Each line of documentation link must match with a line on documentation description. Ex: 'metadata description' 'statistics on website'
suggested_by			name of person or organization suggesting dataset
accessed on		25/11/2016	date of last access to original website
category		official statistics; operational data; international organization; research data	

***Europe Direct is a service to help you find answers
to your questions about the European Union.***

Freephone number (*):

00 800 6 7 8 9 10 11

(*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

More information on the European Union is available on the internet (<http://europa.eu>).

HOW TO OBTAIN EU PUBLICATIONS

Free publications:

- one copy:
via EU Bookshop (<http://bookshop.europa.eu>);
- more than one copy or posters/maps:
from the European Union's representations (http://ec.europa.eu/represent_en.htm);
from the delegations in non-EU countries (http://eeas.europa.eu/delegations/index_en.htm);
by contacting the Europe Direct service (http://europa.eu/eurodirect/index_en.htm) or
calling 00 800 6 7 8 9 10 11 (freephone number from anywhere in the EU) (*).

(*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

Priced publications:

- via EU Bookshop (<http://bookshop.europa.eu>).

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.



EU Science Hub
ec.europa.eu/jrc



@EU_ScienceHub



EU Science Hub - Joint Research Centre



Joint Research Centre



EU Science Hub

