

DFID project data Validation report

Jiri Skuhrovec, Datlab s.r.o., 19.12.2017

Background

This document summarizes work on validation of data on foreign aid delivered by Datlab. It consists of detailed report of availability of each delivered variable – out of roughly 500 which have been collected and processed within previous deliverables. The data have been extensively manually checked and repetitively processed to deliver maximum quality and coverage within limited time. Overall conclusion of the process is, that the quality of the data follows from the quality of the source data, while the tailored software does not lose significant portion in its processing.

Data and validation report structure

The structure of the data (and hence all contained variables) is thoroughly described here:

<https://docs.google.com/spreadsheets/d/1TQEQTJKIsUQ2GCq83qjeUI4BIOBfP6P0aqLr86VKM0c/edit#gid=625767421>

Note that most of source datasets have less extensive coverage than this general structure, which necessarily causes many variables to appear empty in reports below.

The quality of collected data itself is presented by detailed reports for following tables:

1. ec_beneficiaries¹
2. ec_procurements
3. ec_ted_procurements
4. idb_awarded_contracts_goods_and_works
5. idb_procurement_notices
6. idb_project_details
7. wb_project_details
8. wb_project_documents
9. wb_project_financing_summaries
10. wb_tender_notices

Methodological notes

The report itself describes overall availability and most frequently observed values per each variable and source dataset combination. Apparently in each dataset there is large portion of missing variables, which either follows from poor quality of source data, or just plain fact that such variable is not expected to be contained in given dataset (for example there is 0 % of final prices in wb_tender_notices, which is only natural as tender notices are published before final price is known).

¹ Individual datasets are named as agreed with the research team, the prefixes EC, IDB and WB correspond to EuropeAid, Inter-American Development Bank and World Bank data.

Consequently statistics per each data source should be rather interpreted in combination with others – as there are 3 different sources of idb data, generally a good availability of a variable in at least one of those three is considered satisfactory as long as the datasets can be combined together (which mostly holds for newer IDB and WB data).

Additionally, the report on most frequent values gives better insight into what is actually contained in the data – for instance ec_beneficiaries report does not only confirm reasonable shape of some variables (supply_type indeed contains works/services/supplies) but also provides simple but useful insights into actual practice (services are largely dominant in supply_type, Belgium is by far the most frequent seat of beneficiary etc.)

Stats for dataset: ec_beneficiaries

records count : 17285

variable	average		most frequent values
	availability	length	
title	100%	83	EuropeAid/123116/D/SUP/EG: Tender for the Supply of Medical and Non-Medical Furn.. (0.1%) General Equipment for VET Schools/centres, Bosnia and Herzegovina (0.1%) Suministros para beneficiarios del Programa regional de apoyo a la calidad y a l.. (0.1%)
bidder_country	99.5%	9	Belgium (23.3%) United Kingdom (10.8%) France (8.8%)
final_price	99.7%	6	10000 (1.1%) 5000 (0.8%) 3000 (0.3%)
estimated_duration_in_month	70.1%	2	12 (7.2%) 6 (5.8%) 3 (4.6%)
purpose	96.4%	5	0 (6.5%) 99810 (6.3%) 15110 (5.3%)
supply_type	99.8%	8	SERVICES (81.5%) SUPPLIES (11.9%) WORKS (6.3%)
address_of_implementation__country	99.4%	12	All Countries (5.2%) Serbia (2.4%) Bosnia and Herzegovina (2%)
address_of_implementation	92.6%	39	Brussels (1.8%) Bruxelles (1.6%) Kosovo (1.3%)
bidder_name	100%	29	MOORE STEPHENS LLP (4%) IBF INTERNATIONAL CONSULTING SA (2.7%) ERNST AND YOUNG BEDRIJFSREVISOREN CVBA (2.3%)
contract_number	100%	7	133339 (0%) 135737 (0%)

variable	availability	average length	most frequent values	136131 (0%)
publication_date	100%	11	2010-01-01 (20.8%)	2009-01-01 (19.9%) 2008-01-01 (19.3%)

Stats for dataset: ec_procurements

records count : 39531

variable	availability	length	most frequent values
first_ca	49.2%	11	2009-01-21 (0.2%) 2004-12-23 (0.2%) 2013-02-12 (0.2%)
url	100%	251	https://webgate.ec.europa.eu/europeaid/online-services /index.cfm?ADSSChck=149575.. (0%) https://webgate.ec.europa.eu/europeaid/online-services /index.cfm?ADSSChck=149575.. (0%)
tender_id	100%	26	EuropeAid/110006/C/SER/Multi (0%) EuropeAid/110010/C/SER/EG (0%) EuropeAid/110031/C/SER/AM (0%)
title	100%	99	Programme d'appui au développement socio-économique local dans le Nord-Est de .. (0.1%) Supply of equipment for inspection of fishery products, IT equipment and vehicle.. (0.1%) Information and Communication Programme (0%)
status	97.3%	9	FINISHED (93.5%) CANCELLED (3.8%)
bids_receipient	0%	0	
project_name	100%	30	PreAccession Countries / New Member States (45.3%) Neighbourhood (18.8%) Africa, Caribbean and Pacific (17.6%)
publication_date	100%	11	2003-12-23 (0.3%) 1999-03-01 (0.3%) 2003-03-07 (0.2%)
bid_deadline	91.7%	20	2007-01-01T00:00:00 (0.4%) 2006-04-28T00:00:00 (0.4%) 2004-05-01T00:00:00 (0.3%)
address_of_implementation	94.8%	10	Poland (9.3%) Romania (6.5%) Turkey (4.9%)
estimated_price	46%	7	1000000 (2.8%) 2000000 (2.5%) 1500000 (2%)
variable	availability	length	most frequent values

currency	46%	4	EUR (45.8%)
estimated_price_in_eurs	45.8%	7	1000000 (2.8%) 2000000 (2.5%) 1500000 (2%)
supply_type	100%	8	SERVICES (43.4%) SUPPLIES (27.9%) WORKS (16.8%)
firs_cft	61.5%	11	2004-04-30 (0.4%) 2006-12-28 (0.3%) 2004-04-29 (0.2%)
last_cft	61.5%	11	2004-04-30 (0.4%) 2006-12-28 (0.3%) 2004-04-28 (0.2%)
cft_count	100%	2	1 (58.3%) 0 (38.5%) 2 (2.7%)
last_ca	49.2%	11	2009-01-21 (0.2%) 2004-12-23 (0.2%) 2013-02-12 (0.2%)
ca_count	100%	2	0 (50.8%) 1 (44.1%) 2 (2.8%)
first_cc	14.9%	11	2002-07-16 (0.1%) 2005-09-20 (0.1%) 2007-06-06 (0%)
last_cc	14.9%	11	2002-07-16 (0.1%) 2005-09-20 (0.1%) 2007-06-12 (0%)
cc_count	100%	2	0 (85.2%) 1 (13.1%) 2 (1.5%)
last_update	99.2%	11	2016-09-30 (79.2%) 2013-01-10 (2.6%) 2016-07-19 (0.3%)

Stats for dataset: ec_ted_procurements

records count : 7662

variable	availability	length	average		most frequent values
			most frequent values	average	
variable	availability	length	most frequent values	average	length
publication_url	100%	75	http://ted.europa.eu /udl?uri=TED%3ANOTICE%3A34308 2013%3ATEXT%3AEN%3AHTML (0.2%) http://ted.europa.eu /udl?uri=TED%3ANOTICE%3A45984 2015%3ATEXT%3AEN%3AHTML (0.2%) http://ted.europa.eu /udl?uri=TED%3ANOTICE%3A04059 2011%3ATEXT%3AEN%3AHTML (0.2%)		
addres_of_implementation_country	95.6%	31	Contract forecast notice (17.7%) Contract award notice (13.4%) Service procurement notice (6.5%)		
are_variants_accepted	46.8%	6	False (46.8%)		
award_decision_date	6.7%	11	2014-06-02 (0.2%) 2010-11-04 (0.2%) 2013-10-10 (0.2%)		
bid_deadline	42.5%	20	2013-11-29T16:00:00 (0.3%) 2015-01-27T12:00:00 (0.2%) 2011-04-18T12:00:00 (0.2%)		
buyer_city	100%	8	Ankara (26.5%) Brussels (8.1%) Algiers (2.4%)		
buyer_country	100%	3	TR (26.5%) BE (8.1%) DZ (2.4%)		
buyer_address	0.5%	44	265, avenue du Gouverneur Jules Brévi PO Box 13854 (0.1%) Ehlibeyt Mah., Ceyhun Atif Kansu Cd., 1271 Sk., No: 13 (0.1%) American Drywall Building, Vide Boutielle (0.1%)		
cpvs	100%	22	71356200 (11.2%) 75130000 (1.4%) 71520000 (1.3%)		

description	58.2%	636	Under the contract the successful tenderer shall supply, deliver, install, confi.. (0.2%) The subject of the contract shall be the supply, delivery, unloading, installati.. (0.1%) Supply, delivery, unloading, installation commissioning of leak detection equip. (0.1%)
requirements	42.2%	70	See the criteria used in section 12.2 of the 'Instructions to tenderers'. (0.1%) The following selection criteria will applied to tenderers; where the tender (0%) Tenders will be selected on the basis of an administrative compliance grid and a (0%)
estimated_duration_in_months	0%	0	average
variable	availability	length	most frequent values
estimated_duration_in_days	0%	0	
estimated_invitation_date	14.9%	11	2014-08-01 (0.1%) 2015-06-01 (0.1%) 2016-02-01 (0.1%)
estimated_start_date	15.1%	11	2013-10-01 (0.2%) 2014-12-01 (0.2%) 2016-01-01 (0.2%)
publication_publication_date	100%	11	2013-02-12 (0.7%) 2016-01-05 (0.5%) 2015-12-29 (0.5%)
estimated_price	37.7%	7	2000000 (1.8%) 1000000 (1.6%) 1500000 (1.5%)
final_price	5.6%	7	4.2 (0.1%) 2.2 (0%) 4.1 (0%)
fundings_programme	77.3%	63	Financing agreement. (22.1%) IPA budget item 13.05.02 of the general budget of the EU and national contributi.. (6.8%) Budget line. (2.8%)
fundings_source	100%	43	External aid and European Development Fund (100%)

legal_basis	98.1%	182	Council Regulation (EC) No 1085/2006 of 17.7.2006 establishing an Instrument for.. (15%) ACP-EC Partnership Agreement signed at Cotonou on 23.6.2000 as amended. (9.5%) 9th or 10th EDF: ACP-EC Partnership Agreement signed in Cotonou on 23.6.2000, as.. (5.4%)	
procedure_type	93.4%	8	RESTRICTED (48.4%) OPEN (44.8%) NEGOTIATED_WITHOUT_PUBLICATION (0.2%)	
publication_source_id	100%	17	2013/S 199-343084 (0.2%) 2015/S 251-459843 (0.2%) 2011/S 26-040591 (0.2%)	
publication_source_tender_id	99.8%	33	EuropeAid/133510/D/SUP/GH. (0.4%) EuropeAid/130197/D/SUP/TR. (0.4%) EuropeAid/137540/IH/WKS/NG. (0.4%)	
publication_dispatch_date	100%	11	2015-04-13 (0.5%) 2013-01-30 (0.4%) 2015-02-27 (0.4%)	
publication_source_form_type	100%	22	Contract notice (42.5%) Prior Information Notice (22.1%) Contract award (15.9%)	
variable	availability	average	length	most frequent values
publication_form_type	100%		18	CONTRACT_NOTICE (42.5%) PRIOR_INFORMATION_NOTICE (33.9%) CONTRACT_AWARD (23.6%)
selection_method	34.1%		14	LOWEST_PRICE (29.8%) QUALITY_AND_COST_BASED_SELECTION (3.2%) MEAT (1.1%)
supply_type	100%		8	SERVICES (51.4%) SUPPLIES (40.8%) WORKS (7.8%)
title	100%		113	EDF — Supply of laboratory and metrology equipment for testing and calibration.. (0.4%) EDF — Construction and renovation of cold stores and health facilities in EU-S.. (0.4%) EDF — Supply of tools and equipment for technical and vocational education and.. (0.3%)
related_publications	15.6%		35	2011/S 196-318527 (0.2%) 2011/S 77-125201 (0.2%) 2010/S 118-177003;2011/S 26-040591 (0.2%)

bidder_name	19.3%	63	Globe Corporation bv, Galvaniweg 4, 6101 XH Echt, NETHERLANDS. (0.2%) CEDI Europe bv. (0.2%) GFA Consulting Group GmbH. (0.2%)
bidder_country	13.5%	9	Turkey. (2.6%) France. (1.4%) Italy. (1.4%)
bidder_address	17.5%	116	Straatweg 62, 3621 BR, Breukelen, NETHERLANDS. (0.2%) Hindenburgring 18, 61348 Bad Homburg, GERMANY. (0.2%) Lothringerstraße 16, 1030 Vienna, AUSTRIA. (0.2%)
lot_bid_count	23.5%	2	4 (3.7%) 5 (3.5%) 3 (3.3%)
lot_title	54.4%	49	1 lot only. (6.7%) No, 1 lot only. (4.1%) Not applicable. (3.9%)

Stats for dataset:

idb_awarded_contracts_goods_and_works

records count : 103481

variable	availability	average length	most frequent values
project_approval_date	0%	0	
project_borrower_financing	0%	0	
project_country	0%	0	
variable	availability	average length	most frequent values
project_description	0%	0	
project_donor_financing	0%	0	
project_final_cost	0%	0	
project_signature_dates	0%	0	
project_id	0%	0	
major_sectors	100%	5	IN (9.5%) AS (8.1%) TR (6.7%)
sectors	0%	0	
tender_project_name	0%	0	
tender_project_country	0%	0	
tender_project_operations	100%	10	950/SF-BO (1.5%) 1118/OC-AR (1.3%) 955/OC-BR (1.2%)
tender_title	0%	0	
tender_donor_financing	100%	6	1000 (2.3%) 0 (2%) 2000 (1.9%)
bidder_name	99.9%	19	VARIOS PROVEED. MENORES (0.3%) COPUI (0.2%) MITSUBISHI CORPORATION (0.2%)
bidder_city	91.1%	10	SANTIAGO (3.5%) SAN SALVADOR (2.9%) BUENOS AIRES (2.6%)

bidder_country	100%	10	UNITED STATES OF AMERICA (13.2%) BRAZIL (12.8%) ARGENTINA (11.4%)
bidder_address	0%	0	
buyer_assigned_id	0%	0	
description	0%	0	
final_price	100%	6	1000 (2.4%) 2000 (1.9%) 3000 (1.4%)
fiscal_year	0%	0	
addres_of_implementation	0%	0	
supply_type variable	100%	6 availability	WORKS (100%) average length most frequent values
contract_signature_date	100%	11	1977-01-01 (3.8%) 1976-01-01 (3.7%) 1975-01-01 (3.6%)
estimated_completion_date	0%	0	
project_operations	0%	0	
tender_bid_deadline	0%	0	
tender_procedure_type	100%	4	ICB (43.5%) OTHER (39.2%) NCB (10.3%)
tender_procurement_type	100%	29	GENERAL BUILDING CONTRACTORS (9.1%) HIGHWAY AND STREET CONSTRUCTION (6.7%) INDUSTRIAL MACHINERY AND EQUIPMENT (6.5%)
buyer_name	54.8%	32	MINISTERIO DE EDUCACION (3.3%) MINISTERIO DE TRANSPORTES Y COMUNICA (2%) FONDO INVERSION SOCIAL DE EL SALVADO (1.7%)
buyer_city	0%	0	
buyer_country	100%	8	BRAZIL (14.9%) ARGENTINA (11.9%) CHILE (8.1%)

buyer_address	0%	0
publication_source_id	100%	8
publication_date	0%	0
publication_source_form_type	0%	0
publication_form_type	0%	0

Stats for dataset: idb_procurement_notices

records count : 15690

		average		
variable		availability	length	most frequent values
project_id		91.4%	8	BR-L1165 (1.2%) BR-L1284 (0.8%) BR-L1241 (0.7%)
project_approval_date		0%	0	
project_borrower_financing		0%	0	
project_country		0%	0	
project_description	0%	0		
				average
variable		availability	length	most frequent values
project_donor_financing		0%	0	
project_final_cost		0%	0	
project_signature_dates		0%	0	
major_sectors		0%	0	
sectors		0%	0	
tender_project_name		91.4%	51	PROFISCO Pernambuco (1.2%) Pro-Energy RS Distribution (0.8%) Serra do Mar and Atlantic Forest Mosaics System Socioenvironmental Recovery (0.7%)
tender_project_country		0%	0	
tender_project_operations		0%	0	
tender_title		99.9%	63	Consulting Services (4.2%) General Procurement Notice (1.8%) NOTIFICACIÓN DE ADJUDICACIÓN DE CONTRATO (0.6%)
tender_donor_financing		0%	0	
bidder_name		0%	0	
bidder_city		0%	0	
bidder_country		0%	0	
bidder_address		0%	0	

buyer_assigned_id	0%	0	
description	0%	0	
final_price	0%	0	
fiscal_year	0%	0	
addres_of_implementation	0%	0	
supply_type	0%	0	
contract_signature_date	0%	0	
estimated_completion_date	0%	0	
project_operations	0%	0	
tender_bid_deadline	72.4%	20	2001-01-01T00:00:00 (0.9%) 2011-01-01T00:00:00 (0.2%) 2016-07-25T00:00:00 (0.1%)
tender_procedure_type	0%	0	
tender_procurement_type	0%	0	average
variable		availability	length
buyer_name	0%	0	most frequent values
buyer_city	0%	0	
buyer_country	100%	8	Brazil (20.8%) Argentina (8.8%) Nicaragua (6%)
buyer_address	0%	0	
publication_source_id	0%	0	
publication_date	100%	11	2001-01-01 (28%) 2014-09-28 (0.5%) 2013-12-16 (0.5%)
publication_source_form_type	99.6%	8	SPECIFIC (69.4%) AWARD (16.9%) GENERAL (8.6%)
publication_form_type	99.6%	15	CONTRACT_NOTICE (81.1%) CONTRACT_AWARD (18.5%)

Stats for dataset: idb_project_details

records count : 21114

variable	availability	length	most frequent values
project_id	100%	9	TC9905067 (0.1%) BO0180 (0%) AR0127 (0%)
project_approval_date	93.2%	11	2006-12-20 (0.2%) 1997-12-17 (0.2%) 2013-12-12 (0.2%)
project_borrower_financing	25.1%	7	100000 (0.8%) 50000 (0.6%) 150000 (0.5%)
project_country	100%	8	Regional (22%) Brazil (5.9%) Colombia (5.1%)
project_description flexible and	45.6%	339	The operation will allow for responding in a dynamic, expedient man.. (0.2%) The operation includes the following components: tourism product strategy, touri.. (0%) To strengthen airport security through the reform and adjustment of policies and.. (0%)
variable	availability	length	most frequent values
project_donor_financing	93.2%	7	150000 (7.1%) 10000 (5.4%) 30000 (5.1%)
project_final_cost	99.5%	7	150000 (3.4%) 30000 (2.9%) 500000 (2.8%)
project_status	100%	10	COMPLETED (77.3%) IMPLEMENTATION (11.1%) APPROVED (8.3%)
project_signature_dates	75.3%	18	2016-01-22 (0.1%) 1979-09-15 (0.1%) 2015-01-07 (0.1%)
major_sectors	89.4%	23	REFORM / MODERNIZATION OF THE STATE (13.7%) AGRICULTURE AND RURAL DEVELOPMENT (10.2%) PRIVATE FIRMS AND SME DEVELOPMENT (9.8%)

sectors	89.4%	43	PRIVATE FIRMS AND SME DEVELOPMENT- MICROENTERPRISE DEVELOPMENT (4%) REFORM / MODERNIZATION OF THE STATE-REFORM AND PUBLIC SECTOR SUPPORT (3.9%) REFORM / MODERNIZATION OF THE STATE (3.8%)
tender_project_name	0%	0	
tender_project_country	0%	0	
tender_project_operations	0%	0	
tender_title	0%	0	
tender_donor_financing	0%	0	
bidder_name	0%	0	
bidder_city	0%	0	
bidder_country	0%	0	
bidder_address	0%	0	
buyer_assigned_id	0%	0	
description	45.6%	406	The operation will allow for responding in a dynamic, flexible and expedient man.. (0.2%) The operation includes the following components: tourism product strategy, touri.. (0%) To strengthen airport security through the reform and adjustment of policies and.. (0%)
final_price	0%	0	
fiscal_year	0%	0 average	
variable	availability	length	most frequent values
addres_of_implementation	0%	0	
supply_type	0%	0	
contract_signature_date	0%	0	
estimated_completion_date	0%	0	

project_operations	99.8%	15	1075/SF-BO;1075/SF-BO-10;1075/SF-BO-4;1075/SF-BO-5;1075/SF-BO-6;1075/SF-BO-7;107.. (0%) 1192/OC-AR-1;1192/OC-AR-2;1192/OC-AR-3;1192/OC-AR-4;1192/OC-AR-5;1192/OC-AR-6 (0%) ADM/MP-0008-RG;ADM/MP-0007-RG;ADM/MP-0001-RG;ADM/MP-0002-RG;ADM/MP-0003-RG;ADM/M.. (0%)
tender_bid_deadline	0%	0	
tender_procedure_type	0%	0	
tender_procurement_type	0%	0	
buyer_name	0%	0	
buyer_city	0%	0	
buyer_country	0%	0	
buyer_address	0%	0	
publication_source_id	0%	0	
publication_date	0%	0	
publication_source_form_type	0%	0	
publication_form_type	0%	0	
financing_summary_disbursed	85.5%	6	150000 (1.5%) 30000 (1.2%) 200000 (0.6%)
financing_summary_repayments	14.9%	7	5000000 (0.1%) 2500000 (0.1%) 1000000 (0.1%)
financing_summary_cancelled_amount	56.8%	5 1 (0.3%)	0 (0.5%) 150000 (0.2%)

Stats for dataset: wb_project_details

records count : 16047

variable	availability	length	most frequent values	average
variable	availability	length	most frequent values	average
project_approval_date	100%	11	2010-04-06 (0.1%) 2001-06-26 (0.1%) 2007-06-28 (0.1%)	
project_borrower_name	67.7%	20	GOVERNMENT (3.1%) MINISTRY OF FINANCE (1.7%) GOI (1.2%)	
project_closing_date	81%	11	2017-12-31 (1%) 2017-06-30 (1%) 2013-06-30 (0.9%)	
project_country	99.4%	9	India (4.2%) Indonesia (3.2%) China (3.2%)	
project_description	8.9%	1015	Manpowlight of the prevailing country practices and conditions and with the obje.. (0%) Manpowective of the Additional Financing for the Community Development and Safet.. (0%) Manpowective of the additional financing for the Municipal Development Project f.. (0%)	
donor_financing	100%	7	0 (19.4%) 100000000 (2.6%) 50000000 (2.5%)	
environmental_and_social_category	64.8%	2	B (28.3%) C (26.8%) A (5%)	
final_costs	100%	8	10000000 (1.9%) 50000000 (1.8%) 30000000 (1.8%)	
financing_1	65.4%	8	100000000 (2.1%) 5000000 (1.8%) 50000000 (1.7%)	

financing_2	30%	8	10000000 (0.8%) 5000000 (0.7%) 2000000 (0.6%)
financing_3	11.8%	8	10000000 (0.3%) 2000000 (0.3%) 5000000 (0.3%)
financing_rest	5.3%	20	10000000 (0.1%) 1000000 (0.1%) 30000000 (0%)
financier_1	99.3%	32	No data available. (33.9%) International Development Association (Ida) (19.5%) International Bank For Reconstruction And Development (17.3%) average
variable	availability	length	most frequent values
financier_2	30%	28	Borrower (11.5%) International Development Association (Ida) (3.9%) International Bank For Reconstruction And Development (1.5%)
financier_3	11.8%	28	Borrower (4.2%) International Development Association (Ida) (0.9%) International Bank For Reconstruction And Development (0.4%)
financier_rest	5.3%	78	Borrower (1.2%) Bilateral Agencies (Unidentified) (0.1%) Local Communities;Borrower (0.1%)
grant_amount	99.3%	10	{ } (81.9%) { 'amountWithVat': 5000000, 'currency': 'USD' } (0.5%) { 'amountWithVat': 3000000, 'currency': 'USD' } (0.5%)
implementing_agency	66.6%	28	MINISTRY OF FINANCE (3.9%) MINISTRY OF HEALTH (1.5%) TBD (1.3%)
lending_instrument	98%	26	SPECIFIC_INVESTMENT_LOAN (47.6%) TECHNICAL_ASSISTANCE_LOAN (10.3%) INVESTMENT_PROJECT_FINANCING (8%)
name	25.3%	51	Pacific Resilience Program (0%) Skills Development Project (0%) MDTF KP Emergency Roads Recovery Project (0%)

product_line	99.3%	12	IBRD/IDA (80.8%) Recipient Executed Activities (11.9%) Global Environment Project (2.7%)
project_id	100%	8	
region	100%	17	AFRICA (31.1%) LATIN_AMERICA_AND_CARIBBEAN (18.5%) EAST_ASIA_AND_PACIFIC (16.2%)
status	100%	7	CLOSED (83.9%) ACTIVE (16.1%)
team_leader	72.2%	19	PS Dummy for Conversion (0.9%) George Soraya (0.1%) Garó J. Batmanian (0.1%)
sectors	100%	74	No data available., (33.3%) Central government administration,100.0 (2.1%) (Historic)Social assistance,100.0 (0.9%)
themes	100%	98	No data available., (33.9%) Education for all,100.0 (0.9%) Climate change,100.0 (0.9%)
variable	average availability	length	most frequent values
number_of_cft	15.3%	2	1 (6.3%) 2 (2%) 3 (1.6%)
number_of_ca	34.5%	2	1 (3.2%) 2 (2.2%) 3 (2%)
number_of_documents	0%	0	
newest_cft_date	15.3%	11	2012-08-19 (0.1%) 2015-07-23 (0%) 2017-02-22 (0%)
newest_ca_date	8.9%	11	2017-02-07 (0%) 2017-02-27 (0%) 2017-02-14 (0%)
newest_document_date	0%	0	
oldest_cft_date	15.3%	11	2011-07-22 (0.1%) 2012-08-19 (0.1%) 2011-02-08 (0%)

oldest_ca_date	8.9%	11	2011-07-22 (0%) 2009-12-16 (0%) 2011-07-07 (0%)	
oldest_document_date	0%	0		
evaluation_fiscal_year	61%	5	2015 (2.6%) 2014 (2.5%) 2013 (2.1%)	
evaluation_data	61%	11	2095-06-30 (0.5%) 2090-06-29 (0.4%) 2095-12-29 (0.4%)	
evaluation_type	61%	3	ES (27.6%) PAR (20.7%) PCR (9.3%)	
evaluation_err_ex_ante	24%	2	17 (1.3%) 16 (1.3%) 15 (1.2%)	
evaluation_err_ex_post	20.8%	2	12 (1.1%)	-5 (1%)
			13 (0.9%)	
evaluation_project_outcome	61%	16	Satisfactory (29.2%) Moderately Satisfactory (12.5%) Unsatisfactory (10.8%)	
		average		
variable		availability	length	most frequent values
evaluation_project_impact		43.7%	9	MODEST (13.8%) NOT RATED (12.8%) SUBSTANTIAL (10.9%)
evaluation_sustainability_rating		43.7%	9	LIKELY (15.9%) NOT RATED (12.3%) UNCERTAIN (6.7%)
evaluation_risk_to_development		17.2%	10	MODERATE (7%) SIGNIFICANT (5.8%) NEGLIGIBLE TO LOW (2.1%)
evaluation_icr_quality		60.9%	11	NOT RATED (31.9%) SATISFACTORY (24.4%) UNSATISFACTORY (2.4%)

evaluation_me_quality	16.4%	9	MODEST (8.8%) SUBSTANTIAL (4.7%) NEGLIGIBLE (2.3%)
borrower_overall_performance	60.9%	14	NOT RATED (25.7%) SATISFACTORY (16.4%) MODERATELY SATISFACTORY (7.9%)
borrower_government_performance	60.7%	14	SATISFACTORY (20.6%) NOT RATED (19.4%) UNSATISFACTORY (7.4%)
borrower_implementing_agency_performance	60.7%	14	SATISFACTORY (20.7%) NOT RATED (19.1%) UNSATISFACTORY (8.3%)
borrower_quality_ex_ante	36.8%	11	NOT RATED (19.4%) SATISFACTORY (12.2%) UNSATISFACTORY (3.6%)
donor_supervision_quality	60.7%	14	SATISFACTORY (26.7%) NOT RATED (16.1%) UNSATISFACTORY (6.8%)
donor_overall_performance	60.9%	14	NOT RATED (25.7%) SATISFACTORY (17.1%) MODERATELY SATISFACTORY (7.9%)
donor_quality_ex_ante	60.7%	14	NOT RATED (20.5%) SATISFACTORY (18.9%) UNSATISFACTORY (8.4%)

Stats for dataset: wb_project_documents

records count : 37008

variable	availability	average length	most frequent values
variable	availability	average length	most frequent values
project_id	100%	8	P099047 (0.6%) P070291 (0.2%) P089566 (0.2%)
document_source_type	100%	16	Procurement Plan (25.4%) Loan Agreement (16.6%) Agreement (15.5%)
document_type	100%	15	Loan agreement (74.6%) Procurement plans (25.4%)
publication_date	99.2% 2011-01-01 (0.4%)	11 2012-01-01 (0.4%)	2010-01-01 (0.4%)

Stats for dataset: wb_project_financing_summaries

records count : 15828

variable	availability	average length	most frequent values
project_id	100%	8	P005748 (0.1%) P126974 (0.1%) P004925 (0.1%)
approval_date	59.2%	11	2010-09-30 (0.1%) 1994-12-08 (0.1%) 2001-06-26 (0.1%)
closing_date	59.2%	11	1999-06-30 (1.1%) 2001-12-31 (1.1%) 2000-12-31 (1%)
loan_number	100%	9	IDA46720 (0%)
cancelled_amount	0%	0	
disbursed	93.9%	9	10000000 (0.6%) 25000000 (0.5%) 50000000 (0.5%)
interest_charges_fees	86.7%	8	10000 (0.5%) 20000 (0.5%) 30000 (0.4%)
principal	98.4%	9	10000000 (2.6%) 50000000 (2.6%) 100000000 (2.5%)
repayments	71.2%	8	10000 (0.1%) 2000000 (0.1%) 1120000 (0.1%)

Stats for dataset: wb_tender_notices

records count : 38976

variable	availability	length	average most frequent values
notice_type	100%	19	PRIOR_INFORMATION_NOTICE (38.4%) CONTRACT_NOTICE (31.4%) CONTRACT_AWARD (30.2%)
project_borrower_country	0%	0	
project_country	100%	9	Vietnam (7.2%) Africa (4.4%) India (4.4%)
project_name	100%	32	BD: Health Sector Development Program (2%) VN-Distribution Efficiency Project (0.9%) BD: Primary Education Dev. Program III (0.6%)
project_id	100%	8	P118708 (2%) P125996 (0.9%) P113435 (0.6%)
project_region	0%	0	
project_product_line	0%	0	
bid_deadline	65.1%	20	2011-01-17T16:00:00 (0.1%) 2011-09-06T11:00:00 (0%) 2014-05-30T17:00:00 (0%)
buyer_assigned_id	69%	20	ICB Award (2.3%) Consultant Award (QCBS, QBS, FCS, LCS) (1%) Small Contracts Award (DIR, CQS, SSS) (0.9%)
contract_signature_date	0%	0	
title	100%	96	Small Contracts Award (dir,cqs,indv,sss) (4.8%) General Procurement Notice (4.8%) Small Contracts Award (dir, Cqs, Sss) (1.4%)
description	0%	0	
final_price	0%	0	
fiscal_year	0%	0	
objection_date	0%	0	
procurement_type	0%	0	
last_update	0%	0	

selection_method	88.4%	32	INTERNATIONAL_COMPETITIVE_BIDDING (4%) QUALITY_AND_COST_BASED_SELECTION (2%) SELECTION_BASED_ON_CONSULTANTS_QUALITY (10%)
supply_type	0%	0	average
variable	availability	length	most frequent values
further_information_provider_name	69.8%	39	Ministry of Health (0.6%) FER (0.5%) Ministry of Education (0.5%)
further_information_provider_email	69.7%	21	cmsdstore@dhaka.net (1.4%) procurement@fer.ru (0.6%) omvssphc@omvs.org (0.3%)
further_information_provider_phone	69.8%	13	9129231 (0.9%) +7(495)7923010 (0.3%) 02-9129231 (0.3%)
further_information_provider_city	68.9%	9	Dhaka (2.1%) Hanoi (1.5%) Kabul (1.4%)
further_information_provider_street	69.8%	54	CMSD, Tejgaon (0.2%) 30 SHAHEED TAJUDDIN AHMED SHARANI (0.2%) Luzhnetskaya Naberejnaya str., 2/4, building 16, 2nd floor (0.2%)
further_information_provider_country	64.7%	3	VN (3.4%) IN (3.1%) BD (2.9%)
bidder_name	0%	0	
bidder_country	0%	0	
contract_number	0%	0	
major_sectors	0%	0	
notice_language	100%	7	English (75%) French (14%) Spanish (9.9%)
notice_status	100%	9	Published (85.5%) Revised (14.6%)

notice_publication_date	100%	11	2012-08-19 (0.3%) 2011-07-22 (0.3%) 2014-07-09 (0.2%)
-------------------------	------	----	---

notice_source_id	100%	11	
------------------	------	----	--