ICTD/UNU-WIDER GOVERNMENT REVENUE DATASET (GRD): USER GUIDE, COMPARISONS & FAQs

Table of Contents

This short guide introduces the ICTD/UNU-WIDER Government Revenue Dataset (GRD), providing key information and variable definitions, as well as a brief comparison of the merits of the GRD in light of two other recently released datasets: the IMF's WoRLD and OECD GRSD.

Introduction:

The GRD was developed due to the weaknesses in terms of transparency, coverage and accuracy of existing cross-country data on government revenues. It combines data from a number of different international sources under a standard classification system, leading to demonstrable gains in both coverage and accuracy compared to any other single source. The GRD is constructed and presented in a transparent manner that allows users to track each figure back to its original source. This guide aims to assist users to understand the construction of the dataset and how best to make use of it. More detailed information on the construction of the dataset and its update procedure can be found in McNabb (2017)[1] and Prichard et al. (2014).[2]

The GRD is updated on an annual basis. In keeping with the aims of transparency and openness, users are encouraged to contact UNU-WIDER with feedback and suggestions where they have questions surrounding the data contained in the GRD. Please get in touch at GRD@wider.unu.edu.

About the GRD:

The GRD combines government revenue data from the following sources:[3]

- OECD *Revenue Statistics*: OECD Member Countries
- OECD *Revenue Statistics:* African Countries
- OECD *Revenue Statistics:* Asian and Pacific Economies
- OECD *Revenue Statistics:* Latin American Countries
- IMF *Government Finance Statistics* (GFS)
- IMF Article IV Staff Reports
- CEPALSTAT *Revenue Statistics in Latin America*

For a large number of countries, data is available from more than one source. One of the key strengths of the GRD is that the authors have examined data from each underlying source and hand-picked the best source based on a number of criteria.[4] In general, data that…

- combines a long and consistent time series and high levels of disaggregation of sub-categories of revenue;
- accounts for natural resource revenues (whether classed as *tax* or *non-tax*) as distinct from other types of tax revenue
- reports social contributions separately from tax

---

[1] McNabb, K. (2017) '*Toward Closer Cohesion Of International Tax Statistics: The ICTD/UNU-WIDER GRD 2017*'. WIDER Working Paper 2017/184. Helsinki: UNU-WIDER.
[2] Prichard, W., Goodall, A. and Cobham, A. (2014) '*The ICTD Government Revenue Dataset*', ICTD Working Paper 19, Brighton: International Centre for Tax and Development.
[3] In a very limited number of cases, data from individual countries' public finance records are used as a supplementary source. Where relevant, these sources are noted in the GRD. A variety of additional sources were also consulted in constructing the data, but are not employed in the final dataset.
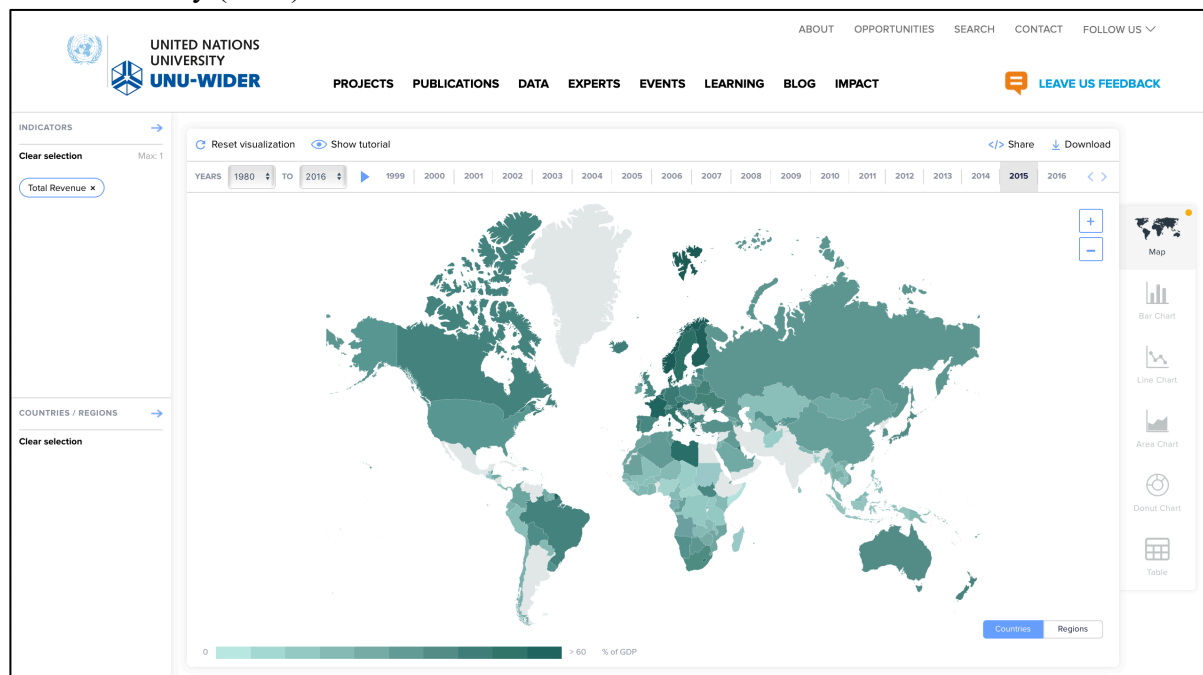[4] Data from all of those underlying sources – as well as noted on data choices where relevant - is available on request.

…is preferred to data that does not.

Furthermore, in the '*Merged*' dataset, General Government data is preferred to Central or Budgetary Central Government, all else being equal; there is, however, **no strict preference ordering that can be applied to all countries**. Each country has been considered in isolation, in order to ensure that the most accurate data is included. The inconsistent nature of revenue data across sources means that applying a one-size-fits-all rule or preference order would lead to a dataset with significant gaps and inconsistencies.

Accessing the GRD

The GRD can be accessed in a number of ways. In 2019, UNU-WIDER launched the *GRD Explorer*. This presents the GRD in a user-friendly, intuitive online tool that allows users to visualize or download custom selections of the data. The *Explorer* tool is the quickest and easiest way to access the data and allows users to download the data in both % of GDP and Local Currency (LCU).

For users wishing to access to the entire dataset, this can be downloaded in Excel or Stata format from within the *Explorer* tool, or from the GRD Homepage.

More advanced users, wishing to delve deeper into the underlying Central and General government data from the GRD can find these files on the Resources page. These files, in some places, contain additional detail on source choices or calculations made using original underlying data.

Also available on this page is a folder of 'Full GRD' files: these excel sheets contain all of the original formulae for converting the data from LCU into % of GDP. **NB.** These files are considerably larger and will run slowly on some computers.

Comparison with other similar datasets

This section briefly considers the ICTD/UNU-WIDER GRD in light of two alternative datasets, The OECD's Global Revenue Statistics Database (GRSD) and The IMF's World Revenue Longitudinal Dataset (WoRLD). This high-level summary attempts to guide would-be users by highlighting the differences as well as some of the respective strengths and weaknesses of each.

Table 1: Comparison of GRD, GRSD & WoRLD

	GRD 2019	**GRSD 2018**	**WoRLD 2019**
# countries covered	196	80	189
Years covered	1980 – 2017	1990 - 2017[5]	1990 - 2017
Total Obsv. (Years * Countries)	7,448	2,160	5,191
Observations as % of potential 1990-2017 (at least 1 indicator)	87.88%	97.20%	88.87%
Data available as	% of GDP - LCU (from GRD Explorer)	% of GDP % of Total Tax - LCU - USD	% of GDP
Data formats available	- Online tool - Stata - Excel	- Limited online tool - Excel	- Online tool - Stata
Sources included:	OECD IMF GFS CEPALSTAT IMF Article IV Reports National Sources	OECD	IMF WEO IMF GFS OECD
Levels of Gov't	'Merged'* Central General (where available)	General Central Subnational levels	Merged (General preferred, otherwise Central)
Updated	Annually	Annually	Unknown.

*Important to note that the GRD and WoRLD datasets 'merge' data from different levels of government in different ways. See note below.

---

[5] Data is available back to 1965 for OECD countries in the underlying *Revenue Statistics* dataset.

Summary of strengths and weaknesses of GRD and similar datasets

ICTD/UNU-WIDER GRD:

Strengths:
- Coverage back to 1980, where available.
- Superior coverage of tax subcomponents, especially for low and middle-income countries, than GRSD & WoRLD.
- Includes most of the data present in GRSD but synthesises with data from other sources.
- Taxes are presented inclusive and exclusive of social contributions and natural resource revenues are separated out from total revenues.
- Data can be accessed/downloaded in both % of GDP and Local Currency on the *Explorer* tool
- ICTD and UNU-WIDER experts have analysed data from numerous sources for each observation, ensuring that the most accurate data is included. As part of this process, experts have drawn up detailed notes and user guidance regarding the data.

Weaknesses:
- At times coverage might be sacrificed for the availability of a data source that includes natural resource revenues (or vice versa).
- Data from Article IV Staff Reports might have been classified into a certain category on the basis of assumptions – although where this is the case, it will have been detailed in the user notes.

OECD GRSD:

Strengths:
- Long coverage for OECD countries (available to 1965 from underlying *Revenue Statistics*), allowing for potentially very fruitful time series analysis albeit for a limited number of countries.
- All revenues are classified according to the OECD *Interpretive Guide* and thus users can make cross-country comparisons with confidence.
- Detailed metadata accessible through the underlying *Revenue Statistics* portals, although not downloadable.
- Data can be accessed and downloaded either as % of GDP, as % of Total tax revenue, in local currency or in US Dollars.

Weaknesses:
- Limited country coverage
- Short time coverage for many developing countries
- Reports only the subcomponents of taxation (including social contributions) for most countries. Total Government Revenue (and by extension grants, nontax) is available only for African and Asian & Pacific economies.
- All tax figures only expressed inclusive of social security contributions.

- Doesn't systematically report natural resource revenues (although the underlying country pages do, where available).

IMF WoRLD:

Strengths:
- Offers a long time series for the Revenue and Tax variables for a large cross-section of countries.

Weaknesses:
- No metadata or guidance provided.
- Frequent gaps in the series for subcomponents
- Can switch from Central to General (& vice versa) within one country over time. Where all taxes from the relevant indicator are collected at the Central level this is not problematic, but in other cases can induce large jumps and inconsistencies, rendering the data unsuitable for time-series analysis.
- Often merges data from two or three sources for a single country-year observation. This can introduce problems whereby the subcomponents are not commensurate with the totals. This leaves many countries with a residual that is not allocated to any one category. The dataset includes a number of 'residual' columns highlighting this.
  - Similarly, there is no Nontax revenue figure or 'Other Tax' figure. The WoRLD is thus not useful for gaining a complete breakdown of government revenues.

Ultimately, different users will have different needs. Where users are unsure of where or how to access and employ cross-country data, they are encouraged to get in touch with the experts at UNU-WIDER, who are happy to provide guidance. GRD@wider.unu.edu

APPENDIX A: Frequently Asked Questions

Should analysis employ data from the merged, general government or central government dataset?

There is no universal answer to this question, as it will depend on the nature of the research question. That said, for studying total revenue collection most users are likely to use the *merged* dataset, which employs general government data where available, while inserting central government data if, and only if, there is evidence that subnational revenue collection is limited. By contrast, relying exclusively on central government data will dramatically underestimate total revenue collection in fiscally decentralized states, while general government data is only available for a limited group of countries. While relying on the merged dataset implies a modest underestimation of revenue collection in countries for which general government data is not available, this will generally be consistent over time, and preferable to the major risks introduced by relying exclusively on central or general government data.

Should analysis include or exclusive social contributions from total revenue and total tax revenue?

Conceptually, the answer is ambiguous and highly dependent on the research question: social contributions are not the same as other taxes, as they are contributions toward a specific area of public spending, but they nonetheless represent an important area of government involvement. Pragmatically, however, most users are advised to rely on revenue and tax figures exclusive of social contributions, owing to problems of completeness and comparability for social contributions figures. Specifically, many countries do not consistently report social contributions, leading to more extensive missing data, whilst what is reported in the 'social contributions' category does not appear to be entirely consistent across countries, particularly for developing countries and when relying on IMF Article IV reports, thus raising concerns that including social contributions in the analysis could produce misleading results.[6]

If I am interested in resource revenues, should I use the "total resource revenue" variable, or an alternative strategy to measure resource dependence?

We would ideally be able to produce a figure for "total resource revenue" for every country, which would capture all resource revenues, whether those revenues are recorded as tax or non-tax revenue in official government accounts. However, in practice, this is simply not possible: Relatively few countries report a figure for "total resource revenue", with some additional countries identifying the resource components of total tax revenue or of total non-tax revenue, but not both. Current international initiatives to strengthen the collection of resource revenue data are correspondingly extremely valuable.

---

[6] The issue of how social contributions are reported in different sources is discussed in further detail in McNabb (2017).

In the absence of such data, users may nonetheless wish to include a proxy for resource wealth in econometric analyses. In these cases the best option available to most users is likely to be to calculate "total non tax revenue" as the difference between "total revenue" and "total non-resource tax revenue".   This variable will thus include all types of non-tax revenue, including resource revenue, but acts as a useful proxy for resource revenue because resource revenue explains the vast majority of the variation in "total non tax revenue" across countries. See discussion in Prichard et al. (2014).[7]

Why do the subcomponents of tax revenue not sum to the respective totals?

(i)   Total tax revenue is generally equal to the sum of the sub-components of tax revenue. However, the total tax figure exceeds the sum of the sub-components in cases where some revenue was not allocated to any specific category in the underlying source, which may occur for a variety of reasons. Thus, for example, Total Taxes on Income, Profits and Capital Gains is sometimes larger than the sum of CIT + PIT, as some sources allocate some revenue to "other income taxes", which cannot be allocated to either personal or corporate income taxes.

(ii)   Very occasionally, the subcomponents of revenues are slightly greater than or smaller than the reported totals in the underlying source. The GRD, as a rule, does not modify data from underlying sources, thus retaining these minor discrepancies in order to ensure transparency.[8] If users wish to correct such data for their own analyses, they are free to do so, though changes of this kind to the data should be carefully recorded for any analysis.

(iii)   For most EU countries, the data in the GRD comes from the OECD Revenue Statistics. The total tax figure in such cases is *inclusive* of 'Customs duties collected for the EU', reflecting the 'Total Tax Revenue' figure in the OECD. However, this item is not allocated to any subcomponent, meaning that the sum of subcomponents will often be marginally lower than the total tax figure.

Why does revenue data in the GRD not match that in other sources?

(i)   The GRD expresses all data in terms of a common underlying GDP figure, taken from the World Economic Outlook (WEO).[9] This helps to ensure comparability of data across different sources, but might be higher, or lower, than tax ratios expressed in other sources, based on different GDP data.

(ii)   The underlying GDP figures taken from the most recent version of the WEO will reflect the rebasing of GDP or recent adoption of more modern systems of national

---

[7] Prichard, W., Salardi, P., and Segal, P., (2014), Taxation, Non-Tax Revenue and Democracy: New Evidence Using New Cross-Country Data, *ICTD Working Paper 23*, Brighton: Institute of Development Studies.
[8] The only exception being where the surrounding data all sums and there is a glaringly obvious error. This would be noted in the notes column of the underlying file.
[9] In cases where no data is available in the WEO, GDP figures are taken from the underlying source of the tax data, e.g. Article IV Staff Reports.

accounting. In some cases, this leads large changes in GDP, which might cause users to find lower tax ratios than an older source that does not express tax figures in terms of up-to-date GDP figures.

(iii)   Tax statistics in other sources may be inflated by revenues earned from natural resources. The GRD, where possible, presents both total revenue and total tax figures both inclusive and exclusive of resource revenues.

Is it possible to estimate total subnational revenue by subtracting Central Government revenue from General Government revenue?

Generally speaking, the difference between General and Central Government revenue should be equal to total subnational revenue. However, any such calculations should be undertaken with extreme caution, and checked carefully, as definitional differences, or other unusual accounting practices, between central and general government sources could lead to over or under-estimating subnational revenue.

Data that I am interested in using has been flagged as potentially problematic.

(i)   The authors of the GRD have striven to alert users to cases where data might be potentially unreliable or misleading, through a serious of 'flags'. These are defined in each data file, and often notes relating to the need for the individual flags are included. Users are advised to carefully consider the inclusion of any data that is flagged as *Caution 1 – Caution 4* from their econometric or cross - country analysis, depending on the variables required. In cases where results do not hold following the exclusion of such data, users are recommended to
    a.   communicate this to their audience and/or
    b.   exclude the data all together.

(ii)   More generally, because of well-known imperfections with the collection of government revenue data – and with GDP data used in some countries to calculate tax revenues – best practice for econometric analysis should always be to pursue extensive robustness testing, including sensitivity to the inclusion of data from individual, or small groups, of countries. Caution should also always be employed when comparing tax ratios across countries, as high, or rapidly rising, tax ratios in individual countries may reflect the underestimation and irregular rebasing of GDP.

What are some of the limitations to the GRD?

(i)   There are widely recognised concerns over the quality of government revenue data in existing sources. The GRD strives to present the 'best' available data consistently and to flag any noticeably problematic or questionable data. This already represents an improvement over any other source. However, the broader accuracy of the numbers contained in the GRD is only as good as the underlying data obtained from each source.

(ii)   In particular, data on natural resource revenues have not historically been collected consistently, or according to strictly enforced definitions. As such, the disaggregation between non-resource taxes and resource taxes is likely to be imperfect at the margin – though still better than the alternative of not making any distinction. There is potential for more precise data in the future, as more systematic data collection efforts are put in place – something toward which both the IMF and OECD are working. In a similar vein, no data on resource revenues is generally available for countries with small levels of resource revenue (generally less than 1% of GDP) and, as such, all revenue is treated as non-resource tax revenue in these cases.

(iii)   There are also potential inconsistencies across countries, or over time, in the disaggregation of tax revenue between taxes on international trade and taxes on goods and services. This reflects potential inconsistencies in the classification of taxes on goods and services collected at the border, by customs officials. While every effort has been made to ensure consistency, users should be aware of this potential issue in individual countries when working with these disaggregated categories.

(iv)   There may be cases where the data in the GRD appears slightly less complete (in terms of disaggregation) than that in, for example, the GFS. This would be as a result of using IMF Article IV data that, whilst not as detailed, allows the resource component of tax to be isolated.

UNU-WIDER-welcomes all feedback from users of the GRD. Queries should be directed to: GRD@wider.unu.edu

APPENDIX B. Variable definitions for sheet ICTDWIDERGRD_2019.xlsx

Sheet Descriptions	
Merged	The final 'merged' dataset. Central or General data denoted by '0' or '1' in the MergeCode column respectively
General	All general government revenue data.
Central	All central government revenue data.
GDP Series	The underlying GDP figures used to express the data in % of GDP. Data is in millions of LCU and almost always comes from the IMF's WEO (most recent edition), apart from those observations in red, where the data comes from Article IV consultations. This is most often due to a case where a country has adopted a new currency, but the existing revenue figures are still expressed in the old currency.
Classifier & GDP info	Contains further info (e.g. System of National Accounts, base year used to compile statistics).
Metadata Variable Descriptions	
Identifier	A unique identifier used to identify country-year observations for performing data operations.
MergeCode	= 1 if General Gov't data used; 0 if Central Gov't data.
Source	Original source for the data, where made explicit
Country	Full Country name, following the World Bank naming convention.
Reg	Geographical region, following the World Bank classification as of 08/2019 (http://data.worldbank.org/about/country-and-lending-groups). Key: 1 = East Asia & Pacific. 2 = Europe & Central Asia. 3 = Latin America & The Caribbean. 4 = Middle East & North Africa. 5 = North America. 6 = South Asia. 7 = Sub-Saharan Africa
Inc	Income group of country following World Bank classification as of 08/2019 (http://data.worldbank.org/about/country-and-lending-groups). Key: 1 = Low Income. 2 = Lower-Middle Income. 3 = Upper-Middle Income. 4 = High Income
Year	Nearest calendar year, to which data is attributed.
ISO	Three letter country code following World Bank classification
GDP	Gross domestic product (in millions of local currency units)
Full Variable List	
Variable	*Description*
General Notes	Notes on the observation that are deemed useful for users but do not explicitly fall under one of the 'Caution' categories.
Caution 1: Accuracy, Quality or	=1 if there are concerns about the quality, accuracy or comparability of these observations. This suggests that the data

Comparability of data is questionable.	differs from surrounding years, or from other countries. Users are urged to consult the accompanying notes and, in many cases, should ideally exclude such data from cross country or econometric analysis
Caution 2: Resource Revenues / taxes are significant but cannot be isolated from total revenues / taxes	=1 if it has been impossible to exclude resource revenues from one or more of the sub-components of total revenue. Users should consult the notes and exclude from x-country or econometric analysis which relies on revenue categories being net of resource revenue.
Caution 3: Un-excluded Resource Revenue are Marginal, but Non-Negligible	=1 if users should treat this data with caution, as resource revenues are small, but not entirely negligible, and are not excluded from tax or total revenue because they are not reported separately in underlying sources
Caution 4: Inconsistencies with Social Contributions	=1 if there are specific circumstances in this country that suggest the treatment of social contributions is not entirely consistent. For example: two sources report social contributions at different levels of the general government, or if the country does not collect much in the way of social contributions, instead funding social security through taxes, or private sector contributions - making x-country comparisons difficult.
Revenue including social contributions	Total government revenue including taxes, non-tax revenue, grants and social contributions
Revenue excluding social contributions	Total government revenue, excluding social contributions
Revenue excluding Grants (including social contributions)	Total government revenue, excluding grants
Revenue excluding grants and social contributions	Total government revenue, excluding grants and social contributions. **This is the suggested total revenue variable for econometric analysis, as it is most consistent and complete across countries.**
Total Resource Revenue	Total natural resource revenues, including natural resource revenues reported as "tax revenue" or "non-tax revenue". Natural resources are here defined as natural resources that include a significant component of economic rent, primarily from oil and mining activities.
Total Non-Resource Revenue including social security	Total non-resource revenues from both tax and non-tax sources, including social contributions
Taxes including social contributions	Total tax revenue, including social contributions
Taxes excluding social contributions	Total tax revenue, excluding social contributions
Resource taxes	Component of reported tax revenue that is from natural resource sources, most often corporate taxation of resource firms

Non-Resource tax including social contributions	Total non-resource tax revenue, including social contributions. Calculated as "Taxes including social contributions" minus "resource taxes"
Non-resource tax excluding social contributions	Total non-resource tax revenue, excluding social contributions. Calculated as "Taxes excluding social contributions" minus "resource taxes".
Direct taxes including social contributions and resource revenue	Total direct taxes, including social contributions and resource taxes. Includes taxes on income, profits and capitals gains, taxes on payroll and workforce and taxes on property, as well as social contributions. The total values of direct taxes may sometimes exceed the sum of these sub-components, owing to revenue that is unclassified among these sub-components
Direct taxes including social contributions, excluding resource revenue	Total direct taxes, including social contributions but excluding resource taxes. Includes non-resource taxes on income, profits and capitals gains, taxes on payroll and workforce and taxes on property, as well as social contributions. The total values of direct taxes may sometimes exceed the sum of these sub-components, owing to revenue that is unclassified among these sub-components
Direct taxes excluding social security, including Resource revenue	Total direct taxes, excluding social contributions but including resource taxes. Includes taxes on income, profits and capitals gains, taxes on payroll and workforce and taxes on property. The total value of direct taxes may sometimes exceed the sum of these sub-components, owing to revenue that is unclassified among these sub-components
Direct taxes excluding social security and resource revenue	Total direct taxes, excluding social contributions and resource taxes. Includes non-resource taxes on income, profits and capitals gains, taxes on payroll and workforce and taxes on property. The total value of direct taxes may sometimes exceed the sum of these sub-components, owing to revenue that is unclassified among these sub-components.
Taxes on income, profits, and capital gains	Total taxes on income, profits and capital gains, including taxes on natural resource firms. This figure is always exclusive of social contributions. The total value of Taxes on Income, Profits and Capital Gains may sometimes exceed the sum of Individuals and Corporations, due to revenues that are unallocated between the two.
Resource Component of Taxes on income, profits, and capital gains	Component of reported taxes on income, profits and capital gains that is from natural resource sources, most often corporate taxation of resource firms
Non-resource component of taxes on income, profits, and capital gains	Total non-resource taxes on income, profits and capital gains.

Individuals	Total income, capital gains and profit taxes on individuals. This figure is always exclusive of resource revenues in available sources.
Corporations and other enterprises	Total income and profit taxes on corporations, including taxes on resource firms.
Resource Component of Corporations & other enterprises	Component of reported income and profit taxes on corporations that is from the taxation of resource firms.
Non-resource component of Corporations & other enterprises	Total non-resource corporate taxes on income and profits
Taxes on payroll and workforce	Total taxes on payroll and workforce. This variable is entirely distinct from social contributions, though in underlying sources social contributions are very occasionally reported as payroll taxes.
Taxes on property	Total taxes on property.
Indirect	Total Indirect Taxes, including resource revenues. Includes taxes on goods and services, taxes on international trade and other taxes. Indirect may exceed the sum of Taxes on Goods and Services, Taxes on International Trade and Transactions and Other Taxes due to unallocated revenue not classified in any of these categories
Resource Component of Indirect	Component of indirect taxes from natural resources, for examples through export taxes. Note that this value is only non-zero is a small number of cases.
Non-Resource Component of Indirect Tax	Total non-resource indirect taxes.
Taxes on goods and services	Total taxes on goods and services, which includes (but it not necessarily always equal to) sales taxes and excise taxes.
General taxes on goods and services	Previously called 'Sales', this includes VAT, Sales Tax, Turnover taxes and Taxes on Financial and Capital Transactions (previously part of Property and still reported as such in some sources).
VAT	Value-added tax, including VAT on domestic and imported goods and services.
Excises	Total excise taxes.
Taxes on int'l. trade and transactions	Total taxes on international trade, including both import and export taxes. In some cases, this figure may also include VAT collected at the border, where countries inconsistently report revenue in this way. The "Problem 4" flag is included in the very infrequent cases in which this figure includes resource revenues.
Import	Total taxes on imports
Export	Total taxes on exports
Other taxes	Total other taxes

Consolidated Non-Tax Revenue	Total non-tax revenue, comprising data categorized as either "non-tax revenue" or "other revenue" depending on the underlying source. Includes revenue from both resource and non-resource sources.
Resources Component of Non-Tax	Total resource component of non-tax revenue.
Non-Resource Component of Non-Tax	Total non-resource component of non-tax revenue.
Social contributions	Total social contributions.
Grants	Total grants received by the government