

THE OPEN DATA INVENTORY

METHODOLOGY



ODIN 2015

125 Countries

20 Data Categories

SOCIAL STATISTICS

Population & Vital Statistics
Education Facilities
Education Outcomes
Health Facilities
Health Outcomes
Reproductive Health
Gender Statistics
Poverty & Income Statistics

ECONOMIC

National Accounts
Labor Statistics
Price Indexes
Government Finance
Money & Banking
International Trade
Balance of Payments

ENVIRONMENT

Land Use
Resource Use
Energy Use
Pollution
Built Environment



10 Elements

COVERAGE

Indicator Coverage
Data Available Last 5 Years
Data Available Last 10 Years
First Administrative Level
Second Administrative Level

OPENNESS

Machine Readable
Non-proprietary
Download Options
Metadata Available
Free/ unrestricted use and reuse

ODIN ONLINE

odin.opendatawatch.com

ODIN concepts and methodology

Data Sources

The Open Data Inventory assesses the coverage and openness of statistics available from websites maintained by national statistical offices. Websites maintained by private or non-governmental agencies or international agencies are not included in the assessment. Websites maintained by other units of the national government or by sub-national governmental units are included if and only if they can be reached from the national statistical office website.

For example, if the national accounts are maintained by the central bank, then data would be included in the ODIN assessment only if the NSO's website provides a link to the appropriate page on the central bank's website or if the NSO reproduces the data on its own website. ODIN is premised on the belief that NSOs can and should take responsibility for providing access to all official statistics.

Data Categories

The Open Data Inventory assesses macrodata. By this we mean data that have been aggregated above the unit record level. We focus on these data because they are the final product released by the NSO or other official agencies. They are used most frequently for policy making and for tracking policy outcomes. Microdata from censuses and surveys are very important, but require a different approach to assessing their openness.

Twenty categories of data are included in the ODIN assessment. [Table A3-1](#) lists the data categories and the sentinel indicators and recommended disaggregations in each category. For the construction of summary measures, the data categories are grouped as social statistics, economic statistics, and environmental statistics.

Table 1. ODIN Data Categories

Data category: Social Statistics

1. Population and vital statistics

Sentinel indicators: Population by 5-year age groups; crude birth rate; crude death rate

Recommended disaggregation: Sex; Marital status

2. Education: Facilities

Sentinel indicators: Number of schools and classrooms; teaching staff; annual budget

Recommended disaggregation: Age group; School stage

3. Education: Outcomes

Sentinel indicators: Enrollment and completion rates; literacy rates and/or competency exam results

Recommended disaggregation: Sex; School stage; Age groups

4. Health: Facilities

Sentinel indicators: Core operational statistics of health system (budget, clinics, hospital capacity, doctors, nurses, midwives)

Recommended disaggregation: Facility type

5. Health: Preventive care and morbidity

Sentinel indicators: Immunization rates; incidence and prevalence major communicable diseases

Recommended disaggregation: Sex; age as applicable

6. Health: Reproductive health

Sentinel indicators: Maternal mortality ratio; infant mortality rate; under-5 mortality rate; fertility rate; contraceptive prevalence rate; adolescent birth rate

Recommended disaggregation: Mortality rates disaggregated by sex

7. Gender statistics

Sentinel indicators: Specialized studies of the status and condition of women; violence against women; women in parliament and management

Recommended disaggregation: None

8. Poverty Statistics

Sentinel indicators: Number and percentage of poor at national poverty line; distribution of income

Recommended disaggregation: Median income; income shares by deciles

Data category: Economic Statistics

9. National accounts

Sentinel indicators: Production by industry; expenditure by government and households

Recommended disaggregation: Production by industrial classification; Current and constant prices

10. Labor statistics

Sentinel indicators: Employment; unemployment

Recommended disaggregation: Sex; Major age groups; Employment by industry and occupation

11. Price indexes

Sentinel indicators: Consumer price index; Producers price index

Recommended disaggregation: By major components

12. Central government finance

Sentinel indicators: Actual revenues; actual expenditures

Recommended disaggregation: Revenues by source; Expenditures by major categories

13. Money and banking

Sentinel indicators: Money supply

Recommended disaggregation: M1; M2; and so forth

14. International trade

Sentinel indicators: Exports and imports

Recommended disaggregation: Major categories using international trade classification

15. Balance of payments

Sentinel indicators: Exports and imports of goods and services; foreign investment; foreign exchange rates

Recommended disaggregation: Goods and services disaggregated by principal industry groupings

Data category: Environment Statistics

16. Land use

Sentinel indicators: Land area

Recommended disaggregation: Urban; rural; cropping

17. Resource use

Sentinel indicators: Fishery harvests; forests coverage and deforestation; major mining activities including gas/petroleum; water supply & use

Recommended disaggregation: Data in physical units; Location as appropriate

18. Energy use

Sentinel indicators: Consumption of electricity, coal, oil, and renewables

Recommended disaggregation: Industry; households; in physical units

19. Pollution

Sentinel indicators: Emissions of air and water pollutants; CO2 and other GHG; toxic substances

Recommended disaggregation: In physical units

20. Built environment

Sentinel indicators: Access to drinking water; access to sanitation; housing quality (from census)

Recommended disaggregation: In appropriate units

Elements of Data Coverage and Openness

The data categories are assessed against ten elements of coverage and openness shown in [Table A3-2](#). Each element has a possible score of 1, 0.5, or 0, indicating that the data in a category satisfy the criteria for that element, partially satisfy them, or fail to satisfy them or the data are entirely missing. Thus a country has a maximum potential score of 200: 100 for data coverage and 100 for data openness. The scoring

scheme is deliberately coarse. A finer scoring grid (say from 1 to 10) would inevitably invite greater subjectivity on the part of assessors and create problems when comparing results produced by different assessors or at different times. The scoring guidelines for each element are summarized in [Table 3](#).

Table A3-2. Elements of Data Coverage and Openness

Elements of Data Coverage	
Indicator coverage and disaggregation	Representative indicators and disaggregations available
Time coverage	<ul style="list-style-type: none"> • Data available in last 5 years • Data available last 10 years
Geographic	<ul style="list-style-type: none"> • First admin level • Second admin level
Elements of Data Openness	
Download format	<ul style="list-style-type: none"> • Machine readable • Non-proprietary • User selectable/API or bulk download
Metadata	Metadata available
Licensing terms	Terms of use stated/ CC BY 4.0 or similar

ODIN Scoring Guidelines

Element 1: Coverage and Disaggregation

The first element requires assessors to locate representative indicators within each data category and determine whether important topical disaggregations are available. Guidelines for scoring each data category are shown in [Table A3-3](#). The representative indicators and disaggregations are listed in [Table A3-1](#) above. In the event that the score

for element 1 is less than 1, the remaining four elements of data coverage cannot exceed the score of element 1. However, the elements of data openness (elements six through 10) are scored on the basis of available data, which may receive a full score for openness if they satisfy the guidelines for those elements. If no data are available for a category, all elements are scored 0.

Table 3. Scoring Guidelines for Element 1: Indicator Coverage and Disaggregation

Social Statistics Scoring	
Population and vital statistics	If population data not available by at least 5-year age groups, score no more than 1/2 point; if sex missing, subtract 1/2 point. Birth and death rates are not disaggregated by age
Education: Facilities	Breakdown by school stage (primary, lower secondary, secondary, tertiary) score 1/2 point; additional detail including age groups and/or school types (technical training; apprenticeship programs, and so forth) gets an additional 1/2 points
Education: Outcomes	Score 1/2 point for enrollment and completion rates by school stage or type; score 1/2 point for exam results; If not disaggregated by sex, subtract 1/2 point
Health: Facilities	Score 1/2 point if at least 3 representative indicators present; score 1/2 point more if disaggregated by facility type
Health: Preventative care and morbidity	Score 1/2 point for immunization data; score 1/2 point for disease incidence or prevalence. Subtract 1/2 point if not disaggregated by sex
Health: Reproductive health	Score 1/2 point for mortality rates; score 1/2 point for fertility, contraceptive prevalence, and adolescent birth rate; subtract 1/2 point if infant and under-5 mortality rates not disaggregated by sex
Gender statistics	Score 1/2 point for data on violence against women; score 1/2 point for data on women in management or political office; special studies that include similar information score 1 point. Disaggregation optional
Poverty and income statistics	Score 1/2 point for poverty headcount; 1/2 point for income distribution by deciles or finer. Disaggregation optional

Table 3. Scoring Guidelines for Element 1: Indicator Coverage and Disaggregation

Economic statistics Scoring	
National accounts	Score 1/2 point for production by industry; score 1/2 point for expenditure data. Subtract 1/2 point if industrial production (value added) not disaggregated by major industry groups: agriculture (including forestry and fishing), industry, and services
Labor statistics	Score 1/2 point for employment; score 1/2 point for unemployment; subtract 1/2 point if not disaggregated by sex; subtract 1/2 point if no age group data
Price indexes	Score 1/2 point for CPI; score 1/2 point for PPI. Disaggregation optional
Central government finance	Score 1/2 point for budget disaggregated by budget categories; score 1/2 point for actual revenues and expenditures by major categories. No points if only totals given
Money and banking	Score 1/2 point monetary aggregates; score 1/2 point for data on the banking system such as total credit to private sector or public sector
Trade	Score 1 point if exports and imports of goods disaggregated by major product categories
Balance of payments	Score 1/2 point for export and imports and current account balance; 1/2 point for foreign direct investment or international investment position
Environment statistics Scoring	
Land use	Score 1/2 point if disaggregated by urban/rural or environmental zones; score 1/2 point if disaggregated by agricultural uses (forest, arable, cropping)
Resource use	Score 1/2 point for any two categories; score 1 point for all
Energy use	Score 1/2 point for any two categories; score 1 point for three of four. Subtract 1/2 point if electricity not disaggregated by industry and household consumption
Pollution	Score 1/2 point for CO ₂ and other greenhouse gases; score 1/2 point for other emissions and pollutants if source identified
Built environment	Score 1/2 point for access to water and sanitation; disaggregation by facility type optional; score 1/2 point for housing quality information with disaggregation by characteristics such as housing type, construction material, or number of rooms

Elements 2 through 5: Other Elements of Data Coverage

Scoring guidelines for the data coverage elements 2 through 5 are summarized in Table 4. Elements 2 and 3 assess the availability of annual data within each category over the 10-year period, 2006 – 2015. Although many countries now provide quarterly data for economic indicators, scoring is based only on annual values. Elements 4 and 5 score the availability of subnational data at the level of first and second

administrative units. Assessors are instructed to determine the administrative levels from official sources. Certain categories of economic statistics are not expected to be available for first or second administrative levels; no scores are recorded for those categories.

Table A3-4. Scoring Guidelines for Elements of Data Coverage

1.	Indicator coverage and disaggregation – see Table 3
2.	Data coverage for the last 5 years a. 1 point if data are available for 3 of the last 5 years b. 0.5 points if data are available for 1-2 of the last 5 years c. 0 points if data are unavailable for last 5 years
3.	Data coverage for the last 10 years a. 1 point if data are available for 6 of the last 10 years b. 0.5 points if data are available for 3-5 of the last 10 years c. 0 points if data are unavailable for 2 or fewer of last 10 years
4.	First administrative level a. 1 point if data available at first subnational level (state, province, and so forth) b. 0.5 if some data available at first subnational level c. 0 points if data only available at national level
5.	Second administrative level a. 1 point if data available at two levels of subnational level (municipality or other similar division) b. 0.5 if some data available at second subnational level c. 0 points if no data available at this level

Elements 6 through 10: Data Openness

Elements 6 through 10 assess the openness of data in a category using criteria derived from the Open Definition. (See <http://opendefinition.org/>.) Scores for coverage and openness were considered independently. If only one indicator for a certain category was published but that indicator was published in a fully open, it was given full points for openness. Scores for openness could, therefore, exceed the scores for coverage in the same category, but in practice this rarely happens. The scoring guidelines for the elements of openness are shown in [Table 5](#).

Elements 6 and 7 assess whether data are downloadable in machine readable, non-proprietary

formats. Open data should be available to anyone in convenient and readily modifiable form. Element 8 asks whether users can select the data they are interested in and whether they are able to establish an API connection to the data, which would allow data to be linked to other applications. The alternative is often that data are only available in predetermined tables. The availability of metadata (element 9) is of importance in providing users with information on how the data were collected and compiled. Clear licensing terms (element 10) state what users may do with the data and permit for reuse of data with some restrictions; fully open data may be used and reused without restriction other than providing attribution to the original source.

Table A3-5. Scoring Guidelines for the Elements of Openness

- | | | |
|----|---|--|
| 6. | Machine readable format | <ul style="list-style-type: none"> a. 1 point if data are downloadable in a machine-readable format (such as XLS, CSV, Stata, SAS, and so forth) b. 0.5 point if some but not all the data are downloadable in machine-readable format c. 0 points if data are not available in machine-readable format (such as HTML, JPEG, PDF) |
| 7. | Non-proprietary format | <ul style="list-style-type: none"> a. 1 point if data downloads are in non-proprietary format (such as CSV) b. 0.5 point if some but not all data are available non-proprietary format c. 0 points if data are not available in non-proprietary format (such as XLS, Stata, SAS, PDF, JPEG) |
| 8. | User selection/ API or bulk download | <ul style="list-style-type: none"> a. 0.5 points if user can select specific indicators from a dashboard for download; 0 otherwise b. Add 0.5 points if an Application Program Interface (API) or other mechanism is available that allows for bulk download |
| 9. | Metadata available | <ul style="list-style-type: none"> a. 1 point if metadata are present that provide specific details about the definition of the indicator or the method of data collection and compilation for that indicator b. 0.5 points if metadata are provided about a large survey or group of data of which the indicator is part. It may require a search of a different section of the website than where the data are to find such metadata c. 0 points if no metadata are available |

Table 5. Scoring Guidelines for the Elements of Openness

10.	Licensing terms
a.	1 point if terms of use are consistent with the Creative Commons Attribution 4.0 (CC BY 4.0) license or similar terms. This means that data must be licensed to permit free use and reuse for commercial and noncommercial use with, at most, an obligation to attribute data to the original source
b.	0.5 points if terms of use for the data are clearly stated on the website (a copyright symbol at the bottom of the page is not sufficient) and allow for data use with some restrictions
c.	0 points if terms of use for data are not found or do not allow for use or reuse of data

Aggregate ODIN Scores

ODIN scores are summarized along both dimensions of the ODIN assessment: by categories and by elements. In addition, subscores are computed for the combined categories of social statistics, economic statistics, and environmental statistics and for the combined elements of coverage and openness. The overall score aggregates all scores across both dimensions. For convenience, all aggregate scores are standardized by rescaling them to a range of 0 to 100.

Weighting

Because the three principal topical groupings (social, economic, and environmental) contain different numbers of data categories, aggregates computed over these categories would be implicitly weighted by the number of categories in each grouping. To neutralize this effect, the data categories are reweighted so that each group has equal weight in aggregates computed over all categories. The reweighting does not affect

aggregates computed within each grouping. All elements have equal weights in all aggregates.

ODIN Online has an option for downloading both the raw and weighted scores for further analysis. In the future an option for user-specified weights for both categories and elements will be included in the online version of ODIN.

Standardized scores

The aggregate scores shown in ODIN tables and charts have been standardized. Scores are standardized by dividing by the maximum score achievable and multiplying by 100. For most subscores, the maximum score is the product of the number of data categories and the number of elements included. However, some of the elements of geographic disaggregation have been excluded a priori from the economic categories. Specifically, it is assumed that the national accounts and government finance statistics will not be available at the second administrative level and that money and banking,

international trade, and balance of payments statistics will not be available at the first or second administrative levels. Therefore, the maximum, unweighted score for five data coverage across all seven economic categories is 27 not 35 and the maximum achievable score over all data categories and elements is 192 not 200. Standardized scores involving any of these categories are reweighted to give them full weight. Because of this discrepancy, subscores over data categories or across elements involving economic statistics will not “add up” consistently, but the treatment of each subscore is internally consistent.

Accessing ODIN Online

The ODIN website is located at: <http://odin.opendatawatch.com/>. The website should be easy to navigate without additional instructions, but here is a short guide to what you will find.

ODIN Home

- The Home page displays a map of the world, showing in color the countries that have been included in the 2015 ODIN assessment. Colors indicate the range of their overall ODIN score. Countries in gray were not include in the 2015 ODIN assessments
- Clicking on a country brings up an information box with the country's aggregate scores and rank. Clicking on the country name takes you to the Country Profile page. (See below.)

Rankings

- The Rankings page displays the overall score and aggregate subscores for data coverage and openness for all countries. The display can be sorted by country name, region, or scores by clicking on the table headers.
- The Rankings dataset can be downloaded with the Export button.

Country Profile

- The Country Profile page provides the most detailed information on a country's ODIN scores. Summary scores are shown for the 20 data categories (aggregated over the elements of coverage and openness) and for the 10 elements of coverage and openness (aggregated over the social, economic, and data categories). Graphs provide regional and global comparisons.

Regional Profile

- ODIN countries are grouped by geographic regions and sub-regions defined by the United Nation Statistics Division's M49 Macro Geographical Regions and Sub-Regions Listing (<http://unstats.un.org/unsd/methods/m49/m49regin.htm>). Country codes are three character ISO codes. ODIN also includes the Republic of Kosovo with ISO code XKS, which is not included in the UN list. Three character regional codes were created for use in ODIN and are not part of the M49 listing.
- ODIN countries have also been classified by the World Bank's income groups. On the Regional Profile page you can choose to view countries grouped by geographic region or by income group. First select the type of display, then select the regions and sub-regions.
- Data from the Country Profile page can be downloaded with the Export button.

Country Comparison

- The Country Comparison page allows users to tabulate aggregate scores for one or more countries. The overall score and five scores aggregated over categories and elements are displayed.
- First select the regions or sub-regions from which to select countries; then select some or all of the countries.
- Data from the Country Comparison page can be downloaded with the Export button. The "spark charts" to the right of the table do not download.

Data Download

- The Data Download provides access to the full ODIN dataset at the item level. Three types of scores can be selected: raw, weighted, and standardized. Raw scores are the original scores recorded by the assessors. Weighted scores have been multiplied by a weighting matrix that gives greater weight to the environment and economic data categories in order to compensate for the fewer number of categories in the overall score. Standardized scores are derived from the weighted scores by dividing by the sum of their weights and multiplying by 100. The item level standardized scores differ from the raw scores by a factor of 100. Weighting only has an effect on the aggregate scores.
- First select regions or sub-regions and then select countries. The entire database can be selected by choosing all regions and countries.
- The aggregate subscores for social, economic, and environmental categories and subscores for coverage and openness elements can be selected for downloading. Aggregates or raw scores and weighted scores are simple sums. Aggregates for standardized scores are weighted averages.

Reports

- The Reports page gives access to the ODIN Annual Report, one page country and regional briefs, and other documentation in PDF format.

Acknowledgements

The Open Data Inventory is a team effort. We are pleased to acknowledge the help of all who contributed to our work.

Open Data Watch

Shaïda Badiëe, Misha Belkindas, Eric Swanson, Zach Christensen, Jamison Crowell, Amelia Pittman, Reza Farivari, and Martin Getzendanner

ODIN Assessors

Chandrika Kaul, Amelia Pittman, Jamison Crowell, Maria Vallenilla, Morgan Smith, Tawheeda Wahabzada, Usman Masood, Mandy Badamkhand, Sophia Rozas, Mariya Fedorchuk, Ela Comanescu, Amira Khalil, Maïssa Khattab, Zach Christensen, Marcus Getzdanner, and Erik Champenois.

Peer Reviewers

Tim Herzog (World Bank), Martine Durand (OECD), Jon Clifton (Gallup), Geoffrey Greenwell (PARIS21), Jessica Espey (Sustainable Development Solutions Network), Mor Rubinstein (Open Knowledge), Joel Gurin and Laura Manley (Center for Open Data Enterprise)

Website and publication design

District Design Group

Website development

Akron, Inc.

Funding support provided by the William and Flora Hewlett Foundation

Cover photo "Harvesting Crops" courtesy of the World Bank Photo Collection.

Copyright: Flickr/Curt Carnemark/World Bank

Back cover "Reading books by the Chinggis Monument" courtesy of the World Bank Photo Collection. Copyright: Flickr/Khasar Sandag / World Bank

