

Frequently Asked Questions (FAQs)



I GENERAL FAQs

- What is OAMDI?
- What types of micro data are available?
- Am I eligible to access the data?
- Are there any constraints on using ERF micro data?
- How can I access the data for my analysis?

II FAQs on the ERF Micro Data Catalogue (NADA)

- How do I register to use the micro data on NADA?
- What should I do if I do not receive the confirmation email upon registration or I received an empty email?
- Where do I send questions about the data, procedures, technical problems, etc.?
- How should I cite the ERF data?
- How do I obtain data?
- How many times can I download the data if I am granted an approval?
- Is there a preferred statistical package for using the ERF data? /
What format are the data in?
- Why can't I open the data file?
- From where can I download the survey documentation materials?

III FAQs on Harmonized Household Surveys

- What is Data Harmonization?
- How the Data Harmonization Standards were developed?
- Which Types of surveys are included in the ERF Data Harmonization Project?
- What is a Data Dictionary?
- Why should I check the data dictionary?
- What is the importance of the Expenditure aggregates documents published with HHIES?
- What documentation materials are provided with the dataset?
- What are the classifications used to harmonize expenditures, occupations and economic activities?
- Can I study multiple countries? multiple years?
- Can the data be used for regional analysis?
- Can the data be used for comparing multiple periods of time?
- What is the harmonized data files structure?

- How a record is uniquely identified in the data file?
- What does the data file name refer to?
- In the HHIES, what is the difference between "V1" and "V2"?
- In the HHIES and HLFS, what are "Standardized" and "Detailed" versions of variables?
- In the HHIES and HLFS, what does "universe" mean in the data dictionary?
- Why can't I find some variables in the data file though they exist in the data dictionary?
- There were questions asked in the questionnaires for which I don't see variables in ERF Harmonized dataset. Where are the data?
- Can I get the original data?
- Where can I find country-specific information and documentation?

I GENERAL FAQs

➤ What is OAMDI?

The Open Access Micro Data Initiative (OAMDI) is a dynamic and unique project led by ERF, providing researchers with free access to several types of micro data.

OAMDI serves as the sole archive of a variety of micro data for countries of the ERF region, among which some are already cleaned and harmonized by ERF to ensure comparability across countries and over time. This online automated data portal is set up to encourage responsible use of those data sets while respecting the ERF's and National Statistical Offices' (NSOs) confidentiality and copyrights.

OAMDI was first launched in April 2013 to make micro data available to the data users, covering 17 Harmonized Household Income and Expenditure Surveys from Egypt, Jordan, and Palestine in addition to 4 Labor Market Panel Surveys from Egypt and Jordan, and 7 data sets on Micro and Small Enterprises from Egypt, Lebanon, Morocco, and Turkey.

Starting February 2014, the ERF data team embarked in a revamp of the OAMDI activities including upgrading available surveys and disseminating new ones. A change in the citation requirement has been also implemented and our users are encouraged to use the relevant citation, available with each survey, when acknowledging the use of the data set.

ERF continuously works on making more new data sets publicly available as more memorandum of understanding are signed with the NSOs.

➤ What types of micro data are available?

The following types of micro data are currently published on the ERF data portal:

- Harmonized Household Income and Expenditure Surveys (HHIES).
- Harmonized Labor Force Surveys (HLFS)
- Harmonized Household Health Surveys (HHHS)
- Labor Market Panel Surveys (LMPS)
- Higher Education Graduates Survey (HEGS)
- Survey of Young People in Egypt (SYPE)
- Micro and Small Enterprises Surveys (MSEs)
- Exports and Imports Data (EID)
- Constraints facing the Development of Micro and Small Enterprises (CDMSE)
- UNIDO Industrial Statistics Database at the 2-digit level of ISIC Rev. 3 (INDSTAT2)
- UNIDO Industrial Statistics Database at the 4-digit level of ISIC Rev. 3 & 4 (INDSTAT4)

- UNIDO Industrial Demand-Supply Balance Database at the 4-digit level of ISIC Rev. 3 & 4 (IDSB)
- Mining and Utilities Statistics Database at the 2- and 3-digit level of ISIC Rev. 3 & 4 (MINSTAT)

➤ **Am I eligible to access the data?**

ERF micro data are made available, for free, for all researchers upon a quick and simple registration process, aiming at controlling the access to the micro data while protecting its confidentiality.

➤ **Are there any constraints on using ERF micro data?**

ERF micro data should ONLY be used for scholarly research, or educational purposes. Users are prohibited from using data acquired from the Economic Research Forum in the pursuit of any commercial or private ventures.

➤ **How can I access the data for my analysis?**

To access the micro data, researchers are required to register on the ERF data portal (micro data catalogue) and comply with the data access agreement.

II FAQs on the ERF Micro Data Catalogue (NADA)

➤ How do I register to use the micro data?

To register for access to the *ERF* micro data via *ERF data portal*, please follow these steps:

Step 1: Complete and submit the ERF data portal user Registration Form

Step 2: Activate your account through the user ID and password received via e-mail

Step 3: Fill in a Micro Data Request Form for each of the surveys you are interested in

Step 4: Micro Data Requests are reviewed within 5 working days. Subject to the information provided, approval to download the data will be granted or denied. In some cases, more information is required

➤ What should I do if I do not receive the confirmation email upon registration or I received an empty email?

If you do not receive a confirmation email after the registration, check your junk mail folder. If the email is missing or you find an empty email from the ERF data portal, you will receive an email from the ERF data team, within 2 working days, indicating your account details.

➤ Where do I send questions about the data, procedures, technical problems, etc.?

Please send all queries to the ERF data team on erfdataportal@erf.org.eg.

➤ How should I cite the ERF data?

Within each dataset, the ERF provides a suggested citation on the data portal under "Access Policy" that the users are requested to use.

An example on Egypt Household Income, Expenditure, and Consumption Survey, HIECS 2012/2013 citation is as follows:

"OAMDI, 2014. Harmonized Household Income and Expenditure Surveys (HHIES), <http://www.erf.org.eg/cms.php?id=erfdataportal>. Version 2.0 of Licensed Data Files; HIECS 2012/2013 - Central Agency for Public Mobilization and Statistics (CAPMAS). Egypt: Economic Research Forum (ERF)."

➤ How do I obtain data?

When the user is granted an approval to download the data, an email is sent to the user providing the links that can be used to download the required data.

➤ **How many times can I download the data if I am granted an approval?**

ERF usually allows for 3 attempts to download the data.

If extra downloads trials are required, the data user is kindly requested to send an email to erfdataportal@erf.org.eg, stating the reason for which he/she is asking for an extension.

➤ **Is there a preferred statistical package for using the ERF data?**

➤ **What format are the data in?**

ERF micro data are available in SPSS and/or STATA formats.

➤ **Why can't I open the data file?**

There are two likely explanations:

1) Due to slow internet connections, the data file is in some cases not fully downloaded and hence is corrupted. In those cases, the data file should be re-downloaded using a high speed downloading connection.

The user is also encouraged to compare the size of the file downloaded to that identified on the data portal to make sure he/she has a full correct download.

2) The version of the software used by the researcher whether SPSS or STATA may mismatch the version uploaded by ERF. In this case, the user needs to either re-save the file in a matching version or convert the data file to a usable software version.

➤ **From where can I download the survey documentation materials?**

All documentation materials are published under the "related materials" section for each survey link on the data portal.

III Harmonized Household Surveys FAQs

➤ What is Data Harmonization?

Data Harmonization, as defined by ERF, is a process aiming at producing a clean and comparable version of national Household Income and Expenditure Surveys, and Labor Force Surveys, collected by the Statistical Offices in the Arab countries. These datasets are then acquired by ERF, upon signing confidentiality agreements with the designated authorities.

The ERF team of statisticians work on cleaning and harmonizing these datasets, to a list of predefined variables, that are standardized across all micro data subject to harmonization.

The result is a set of cleaned and comparable data files across all countries and years.

➤ How the Data Harmonization Standards were developed?

The data harmonization standards were developed by the ERF team of statisticians after carefully and individually reviewing the surveys micro data and documentation materials provided by the Statistical offices, and following the best practices of other organizations working in the same field.

A list of standardized/harmonized variables was then compiled to include the variables' name, label, definition, values, standard classification, ...etc. This list constitutes the data dictionary considered to be an essential material for the harmonization process.

The data dictionary is customized for each data set in a certain country/year to indicate the original variables versus the harmonized ones, and how the variables provided in the raw data were standardized and harmonized to create a list of comparable variables across all datasets.

➤ Which Types of surveys are included in the ERF Data Harmonization Project?

The data Harmonization project mainly covers the micro data of the Household Income and Expenditure Surveys and Labor Force surveys in Arab countries, Iran and Turkey.

Recently, ERF started to acquire, harmonize, document and disseminate Health surveys as well. However, health surveys are only harmonized on the country level to create an inter-year's comparable version of all available surveys in a certain country.

Harmonization at this stage only included unifying variables' names, labels and some definitions. Thus, the health data are only harmonized on the year level in the same country, but are not comparable across countries.

➤ **What is a Data Dictionary?**

The data dictionary is a documentation table produced by ERF for each data set. It presents the ERF set of harmonized variables and harmonization standards on the household as well as the individual levels.

The table's fields indicate the variables' names and labels, as appearing in the data file, a description of the content of each variable, general and country specific comments, and the formula used to generate the variable if computed by ERF.

In addition, the data dictionary provides a variable by variable mapping between the original and the standardized values to indicate how each variable was generated, followed by a column illustrating the variable universe and a key column denoting the availability of the variable in the harmonized data set.

➤ **Why should I check the data dictionary?**

The data user is encouraged to download and check the data dictionary before requesting the data for download, to make sure it matches the purpose of his/her research.

He/she is also encouraged to study well the dictionary before further use of the data for analysis.

➤ **What is the importance of the Expenditure aggregates documents published with HHIES?**

The expenditure aggregates is a documentation table produced by ERF for each data set. It presents a list of products' (expenditure items) codes/labels categorized under each major group of the Classification of Individual Consumption According to Purpose (COICOP). It also illustrates all general comments on different expenditure groups and formulas used to compute the expenditure aggregates.

➤ **What documentation materials are provided with the dataset?**

A set of documentation materials is published with each survey in PDF/XLS format, subject to availability, in all available languages. This set includes all documentation available for the original survey provided by the Statistical Agency (questionnaires, codebooks, methodology, user guide, etc...), and for the harmonized data sets produced by ERF (List of harmonized variables, data dictionary, expenditure aggregates for HHIES and list of available harmonized variables for HLFS). In addition to a copy of all international classifications of expenditures, occupations and economic activities used during the harmonization process.

➤ **What are the classifications used to harmonize expenditures, occupations and economic activities?**

The Classification of Individual Consumption According to Purpose (COICOP) is used to compose expenditure aggregates in all HHIES. The occupation classification follows the 1988 or 2008 International Standard Classification of Occupations (ISCO), in most HHIES and HLFS surveys. As for the economic activity classification, it follows the Rev3.0, Rev3.1 or Rev.4 of the International Standard Industrial Classification of All Economic Activities (ISIC).

Relevant documentation of various classifications standards are made available whenever needed under the related materials for each survey link on the data portal.

➤ **Can I study multiple countries? multiple years?**

➤ **Can the data be used for regional analysis?**

ERF dataset is designed to facilitate cross-national and/or cross-temporal research.

ERF acquired, in most cases, datasets that are available at multiple points in time– in some cases spanning decades. The data team has then harmonized the data acquired from the statistical Offices, after cleaning it, into a common cross-national template, pre-defined by ERF, and guided by best practices pursued by other organizations working in the same field. However, the HHHS can only be used to produce across years' analysis in a certain country; but the datasets are not harmonized across countries.

➤ **Can the data be used for comparing multiple periods of time?**

All HHIES and HHHS micro data are cross-sectional (not panel data). Therefore, it cannot be used for household- or individual-level longitudinal analysis, because there are no identifiers that link households or persons across waves of data.

Some of the HLFS are cross-sectional and others are panel. A variable available in each dataset (DTYPE) identifies the type of data being cross-sectional or panel. However, data users are always encouraged to carefully check the survey documentation and related materials before any attempt of using the data for analysis.

➤ **What is the harmonized data files structure?**

Harmonized Household Income and Expenditure Surveys (HHIES)

The HHIES dataset is produced in 2 data files, the first contains the generated household-level harmonized variables and the second presents the individual-level harmonized variables. The household data file covers the following categories, subject to availability:

- File identification
- Geographical characteristics
- Household composition
- Dwelling characteristics
- Ownership of durables, livestock and land
- Head's and spouse's characteristics
- Annual household expenditure
- Annual household income

The individual data file envelopes the following categories, subject to availability:

- File identification
- Geographical characteristics
- Demographics
- Father's and mother's characteristics
- Nationality and immigration
- Education
- Health
- Labor
- Annual individual income

Harmonized Labor Force Surveys (HLFS)

The HLFS dataset is produced in 2 data files, the first contains the generated household-level harmonized variables and the second presents the individual-level harmonized variables.

The household data file covers the following categories, subject to availability:

- File identification and information.
- Geographical characteristics.
- Household composition.
- LFS classification of household members.
- Head's and spouse's characteristics.
- Dwelling characteristics.
- Ownership of durables.

The individual data file envelopes the following categories, subject to availability:

- File identification and information.
- Demographics.
- Nationality and immigration.
- Education.
- Current labor status.
- Employment in main job.
- Wages and incomes.
- Employment in secondary job.
- Last held job characteristics.
- Unemployment characteristics.
- Inactivity reason(s).

Harmonized Household Health Surveys (HHHS)

The household health surveys are only harmonized across the survey years available for a certain country. Thus, the datasets are not comparable across countries.

Therefore, the datafiles structure has a country-specific format, which depends on the original raw micro data provided.

However, and since the data dictionary is not provided in the cases of HHHS, the ERF has created and published a "Variables Mapping & Availability Matrix" documentation material. This document indicates the list of variables subject to harmonization; the final names and labels under which those variables were published, the level of harmonization of those variables across different survey years (if fully or partially harmonized) in addition to anymore comments that need to be illustrated on specific variables.

The fully harmonized variables are those having the same definition, universe, and categories, while those partially harmonized are those having different universe, or categories.

➤ How a record is uniquely identified in the data file?

Harmonized Household Income and Expenditure Surveys (HHIES)

Harmonized Household Health Surveys (HHHS)

In the household data file: A record is uniquely identified by the household identifier (the variable "CASESER").

In the individual data file: A record is uniquely identified by both the household and the individual identifiers (the variables: "CASESER" and "PNUM").

Harmonized Labor Force Surveys (HLFS)

In the household data file: In case of cross-sectional datasets, a record is uniquely identified by the household identifier (the variable "CASESER"). In case of panel datasets, paneled households can be tracked using the variable "HPNLID".

In the individual data file: A record is uniquely identified by both the household and the individual identifiers (the variables: "CASESER" and "PNUM"), in case the dataset is cross-sectional. While in panel datasets, both the variables "HPNLID" and "PPNLID" are used to identify a paneled individual.

➤ What does the data file name refer to?

Harmonized Household Income and Expenditure Surveys (HHIES)

The data file name consists of 4 components. It starts by the country name followed by a 2-digits survey year, an indication on the data file unit of analysis whether household (HH) or Individual (IND), and finally the version number of the data file disseminated.

A description of the version published is available for each data set on the ERF data portal.

10

Harmonized Labor Force Surveys (HLFS)

The data file name consists of 5 components. It starts by the country name followed by a 2-digits survey year, the abbreviated survey title (LFS), an indication on the data file unit of analysis whether household (HH) or Individual (IND), and finally the version number of the data file disseminated.

A description of the version published is available for each data set on the ERF data portal.

Harmonized Household Health Surveys (HHHS)

The data file name consists of 5 components. It starts by the country name followed by a 4-digits survey year, the abbreviated survey title (HHS), an indication on the data file unit of analysis whether household (HH) or Individual (IND), and finally the version number of the data file disseminated.

A description of the version published is available for each data set on the ERF data portal.

➤ In the HHIES, what is the difference between "V1" and "V2"?

V2 includes all variables previously published in V1 in addition to a number of new/detailed-composite coded version of the variables considered essential on the household as well as the individual level, produced by the Economic Research Forum for dissemination.

➤ **In the HHIES and HLFS, what are "Standardized" and "Detailed" versions of variables?**

Variables considered essential in the ERF list of harmonized variables have a composite coding structure, where the first digit (and in some cases the first 2 digits) is comparable across all samples, while subsequent digits provide progressively more details available in some samples and not others. The original variable in the dataset is usually recognized by the term "[Standardized version]" in the variable label while the composite coding structure is formally recognized by the term "[Detailed version]".

➤ **In the HHIES and HLFS, what does "universe" mean in the data dictionary?**

The "universe" indicates the part of the sample that should respond to the variable/question. In other words, it represents the sub-sample of households or individuals to whom the survey questions were asked, as reflected on the questionnaire/data file.

For example, the marital status question in most surveys is collected for individuals aged 12 years or older, hence this age group represents the universe of the marital status variable.

Cases that are outside the universe of the variable are coded system missing.

To avoid the loss of cases available in the original data, the variables universe was not unified across surveys, yet it is clearly identified for each variable in the Data Dictionary of every survey disseminated by ERF.

➤ **Why can't I find some variables in the data file though they exist in the data dictionary?**

In the data dictionary, the column titled "key" indicates the availability of the variables in each dataset.

Whereas the letter [A] denotes a variable that is available in/can be generated from the survey original data, while [/] indicates a variable that is not available in/cannot be generated from the survey original data.

Consequently, all variables in front of which an [A] is indicated are then generated and thus will be available in the data file, while other variables for which [/] is marked, will not exist in the final harmonized data set.

- **There were questions asked in the questionnaires for which I don't see variables in ERF Harmonized dataset. Where are the data?**

Some countries do not supply variables for every question available in the survey's instrument. And in other cases some items of the questionnaire are not relevant to the ERF list of harmonized variables.

- **Can I get the original data?**

In complying with ERF's agreements with the Statistical Offices, ERF ONLY produces and releases the harmonized versions of the Household Income and Expenditure Surveys and Harmonized Labor Force Surveys, in both SPSS and STATA formats.

- **Where can I find country-specific information and documentation?**

The country-specific information is available in the Data Dictionary produced by ERF for each data set. All documentation materials are published on the data portal under related materials for each survey link.