

Thematic Evaluation Series

Evaluation of FAO's statistical work

Annex 2. Analysis of FAO databases

1. Introduction

1. This document summarises the analysis of 14 databases managed by the Food and Agriculture Organization of the United Nations (FAO), some of which are joint operations with other organizations. The databases are very important dissemination means for FAO statistics. In some cases, they complement paper publications (books/brochures); others are purely electronic publications. Databases are also used to support users with extensive meta-information that can be rapidly updated. The list of databases can be found in the table on the next page.
2. The method of analysis was extensive desk review, comparing data with other published sources where possible. The structure of the analysis follows the structure of the FAO Statistics Quality Assurance Framework (2014).

The comments in italics are statements drawn from the FAO Statistics Quality Assurance Framework that summarise the section heading quality criterion.

2. Relevance

Relevance is the degree to which FAO statistical databases meet current and potential user needs.

Stated purpose of the databases

- Any public database must have a reason for its existence; statistics have no intrinsic value but must have a purpose to be useful. Most of the FAO databases had reasonably clear objectives, although there were two partial exceptions. The website of Domestic Animal Diversity Information System (DAD-IS) focuses on its content rather than purpose, which is to enable the user 'to make informed decisions on the management of animal genetic resources'. The implicit purpose of FishStatJ is to provide a complete, primary source of sector information. Taken as a group, there is no commonality among purposes of the FAO databases reviewed. Some are highly specific; others are generalist and multi-purpose. The two databases mentioned above are both aimed at a relatively small user group of professionals working in a group of connected fields.

Who are the intended users?

- It is not necessary for the presentation of FAO databases to specify the intended user group, although this can be useful in some circumstances. Again, there is no commonality across the databases, whose target audiences range from small groups of professionals to generalist users of data to policy influencers with little technical knowledge.

Databases examined

Database	Website	Achievement of purpose
FAO's Global Information System on Water and Agriculture (AQUASTAT)	http://www.fao.org/aquastat/en/	Yes
FAO/WHO GIFT Global Individual Food consumption data Tool	http://www.fao.org/gift-individual-food-consumption/en/	Work in progress
Global Livestock Environmental Assessment Model (GLEAM)	http://www.fao.org/gleam/en/	Yes
AMIS Outlook – Agricultural Market Model Information System	http://www.amis-outlook.org/	Yes
FAO Yearbook of Fishery and Aquaculture Statistics	http://www.fao.org/fishery/statistics/yearbook/en	
FishStatJ	http://www.fao.org/fishery/statistics/software/fishstatj/en	Not fully
FAOSTAT	http://www.fao.org/faostat/en/#home	Not fully
Global Forest Resources Assessments (FRA)	http://www.fao.org/forest-resources-assessment/en/	Not fully

Database	Website	Achievement of purpose
Agriculture Stress Index System (ASIS) Database and Tool	http://www.fao.org/resilience/news-events/detail/en/c/296089/	Yes
DAD-IS (The Domestic Animal Diversity Information System)	http://www.fao.org/dad-is/en/	Not fully
EMPRES-i (Global Animal Disease Information System) -incomplete	http://empres-i.fao.org/eipws3g/	Not fully
Food Outlook	http://www.fao.org/gIEWS/reports/food-outlook/en/	Yes
FPMA (Food Price Monitoring and Analysis)	http://www.fao.org/gIEWS/food-prices/home/en/	Yes
GLW (Livestock Systems)	http://www.fao.org/livestock-systems/en/	Yes
INFOODS (International Network of Food Data Systems)	http://www.fao.org/infoods/infoods/en/	Not fully

5. An example of the former type is the International Network of Food Data Systems (INFOODS), which targets organisations and individuals working on food composition issues. FAOSTAT evidently attracts a wide range of data users who generally focus on international comparisons of data. The fact that a database has a specialised purpose does not necessarily restrict its user group to a narrow range of professionals. As a result of the high public profile of the issues it covers, the Global Forest Resources Assessment (FRA) is of interest to policy makers, influencers, media and general public; commercial interests globally related to forestry, forest lands and their inhabitants.
6. Alternatively, Livestock Systems (Gridded Livestock of the World [GLW]) has two discrete sets of users:
 - i. a general, policy-oriented group, to which the section introductions and links to publications and policy briefs appear aligned;
 - ii. a technical group who require the data as inputs to their geographical information systems.
7. Most of the information is aimed at the second group, with the generalist website acting as a wrapper for an extensive technical dataset. The generalist content is, however, non-zero and useful.
8. When a database is restructured or its presentation updated, its user groups need to be clearly identified, together with the database's purpose and potential uses.

What uses could the statistical database potentially fulfil?

9. It is not necessarily the case that a FAO database should fulfil all the data needs of its target audience. The main focus of FAO statistics is on inter-country comparisons, since its data is largely based on national data sources and sub-national comparisons are generally

outside FAO's remit. Geographically disaggregated data below national level is therefore relatively scarce. However, Livestock Systems (Gridded Livestock of the World [GLW]) exists to provide a global geo-referenced dataset that is as localised as possible and, as a by-product, facilitates sub-national data comparisons.

10. Some FAO statistics, such as the Global Forest Resources Assessment (FRA), are the only global data source available in that specific domain. In such a case, all relevant uses need to be considered, although whether they are feasible and affordable to fulfil is another question.
11. In other cases, FAO databases are not the only source of comparable information. This does not make them redundant. For example, the Agriculture Stress Index System 2, based on remote sensing, is one of eight global and regional early warning systems oriented to food security. A comparative study found that "the systems are tailored to meet the needs of different customers and hence they differ in the importance they place on inputs to the system as well as how they disseminate their results" (Fritz. S *et al*, 2019).
12. FAOSTAT, as a general purpose database, is evidently aimed at fulfilling a wide variety of uses, although the results of the evaluation questionnaire for users suggests that the range of applications is not as wide as it could be.
13. The potential uses are not always obvious to the builders of a database, especially as they change over time. This knowledge requires close coordination with a range of users.

Given the methodology, can the stated purpose be attained?

14. The question is intended to identify any databases that require rethinking, rather than solely technological redevelopment.
15. Comments on the International Network of Food Data Systems (INFOODS), the Domestic Animal Diversity Information System (DAD-IS) and FishStatJ have at least partial difficulties in fulfilling their objectives. This is generally because the aims are not clear or else the data collection methods have inherent difficulties in fulfilling the objective which appear to be difficult to resolve by improved data collection.
16. The more demanding question is whether the database actually fulfils its mandate. This can be for a number of reasons, the most pressing of which is when the data collection methods are insufficient. This is the case both with FAOSTAT and the Global Forest Resources Assessment (FRA). Given the wide range of data within FAOSTAT and data sources, completely fulfilling its mandate is a very demanding objective. Other reasons include accuracy and coverage, which are both issues with the Global Animal Disease Information System (EMPRES-i). With the FAO/WHO Global Individual Food consumption data Tool (GIFT), it appears that the database is still in the process of being populated, so that judgement cannot yet be made.
17. The question as to whether a database completely fulfils its mandate is to an extent a subjective one. The column 'Achievement of purpose' in the 'Databases examined' table should be considered in this light.

3. Accuracy and reliability

FAO statistics accurately and reliably portray reality.

What does accuracy entail in the context of this database?

18. The objective and content of each database determine the appropriate measure of accuracy. Analysing database accuracy generally rests on comparison of data internally, with source data and against other FAO datasets. For this reason, some of the analyses of accuracy were the same as those for coherence and comparability.

What are the actions that are intended to be taken to ensure accuracy?

19. In general, data quality management is required to ensure accuracy of data and metadata.

What documentation is planned or provided to ensure accuracy?

20. Metadata at the database and individual statistic level were sought but not always found. Some of the metadata found was too general to lead to conclusions at the level of the statistic. The Domestic Animal Diversity Information System (DAD-IS) was one database that featured very little systematic information about data sources and methods. Moreover, FishStatJ data is not fully internally comparable or consistent, perhaps due to incomplete updating of earlier period data or else use of different data sources.
21. There are documented reliability issues with some FAOSTAT data, including levels of estimation and imputation, as well as acceptance of national data that external studies have criticised. The degree of FAO's responsibility, as contrasted with the national data sources, could be clarified by improved and more systematic metadata.¹
22. In general, systematic implementation of FAO's existing standard approach to data quality and metadata reporting would lead to accuracy improvements.

¹ An example of FAOSTAT metadata is at:
<http://www.fao.org/faostat/en/#data/PP/metadata>

4. Timeliness and punctuality

What are the timeliness objectives for this database, relative both to the reference period and the completion of the source data, if applicable?

23. Timeliness and punctuality are again determined by the nature of the database – reporting periods range from tens of hours to years.
24. The generally accepted norm is that annual data should not be published more than 12 months after the end of the reference year; similar norms apply for other periods. There are cases where this rule is not adhered to and estimates or imputations are used instead. For example FAOSTAT annual data is often only available more than one year after the end of the reference period. Data for many developing and middle-income countries is not available even then, so that the last one or two years' data in FAOSTAT are commonly estimated or imputed.
25. Databases that are updated only after long periods also suffer from data delays, presumably because the data collection exercise is only carried out a single time and any data that is not available is excluded. The Global Forest Resources Assessment (FRA) is an example of this type of issue.
26. Timeliness and punctuality issues can be addressed by publishing a forward calendar of data releases.

5. Coherence and Comparability

FAO statistics are consistent internally, and over time. FAO statistics are internationally comparable, and can be used in conjunction with statistics provided by other statistical organizations.

27. The existence of multiple databases with limited common technical governance risks the provision of data that are not coherent or comparable. An example is in agricultural commodity production data, whose values differ between Food Outlook, FAOSTAT and AMIS, as well as external data sources. While there are some valid reasons for variations in values and AMIS is explicit in its different data source exploitation from FAOSTAT, not all differences can be explained in this way. The following table provides an example.
28. The data for wheat production in 2017 contained in the November 2018 edition of 'Food Outlook' was compared with other data sources from FAO and other producers for a number of major producers. This data point was chosen because FAOSTAT data for 2018 is not yet available. Food Outlook presents data to the nearest 100 000 tonnes; other sources to at least an extra decimal place.

2017 wheat production data	India	Egypt	Argentina	USA	EU
Food Outlook Nov 2018 - estimated	98.5	8.8	18.5	47.3	154.0
FAOStat	98.51	8.80	18.40	47.37	150.25
FAO-AMIS	98.51	8.45	18.50	47.38	151.95
IGC (reported by FAO-AMIS)	98.51	8.60	18.51	47.38	151.36
USDA-PSD (reported by FAO-AMIS)	98.51	8.45	18.50	47.38	151.13
<i>Units: million tonnes</i>					

Note: The Food Outlook data is that reported in November 2018. Other sources are the most recent available, so they may have had the opportunity to make later revisions. For countries where, on the basis of FAOSTAT metadata, the data is unofficial (Egypt) or come from a single government estimate / statistical release (India), the data are identical. It is more difficult to understand the other data variations, especially for the European Union.

6. Accessibility and Clarity

FAO statistics are easily accessible to all users on an impartial basis, are presented in a clear and understandable format, and are accompanied by relevant supporting metadata.

29. Some FAO databases are excellent examples of data presentation. These are generally recently developed for specific purposes. Examples include Livestock Systems (Gridded Livestock of the World [GLW]); Food Price Monitoring and Analysis (FPMA); and Agriculture Stress Index System 2 (ASIS 2).
30. Others are not so well developed. Technical problems are idiosyncratic, since these databases were developed in isolation from other statistical software projects. Examples are FishStatJ and the 'Global Animal Disease Information System' (EMPRES-i).
31. The inability to group data into coherent subheadings is also an issue, again with FishStatJ and also with the Domestic Animal Diversity Information System (DAD-IS).
32. The solution to accessibility issues, in most cases, is to redevelop databases using FAO's standard data dissemination framework, believed to be under development.

7. Conclusions and recommendations

33. Databases that are aimed at a relatively small user group of professionals working in a group of connected fields are often less well conceived and executed than large, generalist databases, especially the newer ones.
34. When a database is restructured or its presentation updated, its user groups need to be clearly identified, together with the database's purpose and potential uses. The potential uses are not always obvious to the builders of a database, especially as they change over time. This knowledge requires close coordination with a range of users.
35. The degree of FAO's responsibility, as contrasted with the national data sources, could be clarified by improved and more systematic metadata. In general, systematic implementation of FAO's existing standard approach to data quality and metadata reporting would lead to accuracy improvements.
36. Timeliness and punctuality issues can be addressed by publishing a calendar of future data releases.
37. In most cases, the solution to accessibility issues is to redevelop databases using FAO's standard data dissemination framework, which is currently under development.

References

FAO. 2014. *The FAO Statistics Quality Assurance Framework*. Rome. (also available at: <http://www.fao.org/3/i3664e/i3664e.pdf>).

S. Fritz et al. 2019. A comparison of global agricultural monitoring systems and current gaps. *Agricultural Systems*. Volume 168, January 2019, Pages 258-272. (also available at: <https://www.sciencedirect.com/science/article/pii/S0308521X17312027>).

Appendix 1. List of databases

Database	Website
FAO's Global Information System on Water and Agriculture (AQUASTAT)	http://www.fao.org/aquastat/en/
FAO/WHO GIFT Global Individual Food consumption data Tool	http://www.fao.org/gift-individual-food-consumption/en/
Global Livestock Environmental Assessment Model (GLEAM)	http://www.fao.org/gleam/en/
AMIS Outlook – Agricultural Market Model Information System	http://www.amis-outlook.org/
FAO Yearbook of Fishery and Aquaculture Statistics	http://www.fao.org/fishery/statistics/yearbook/en
FishStatJ	http://www.fao.org/fishery/statistics/software/fishstatj/en
FAOSTAT	http://www.fao.org/faostat/en/#home
Global Forest Resources Assessments (FRA)	http://www.fao.org/forest-resources-assessment/en/
Agriculture Stress Index System (ASIS) Database and Tool	http://www.fao.org/resilience/news-events/detail/en/c/296089/
DAD-IS (The Domestic Animal Diversity Information System)	http://www.fao.org/dad-is/en/
EMPRES-i (Global Animal Disease Information System) - incomplete	http://empres-i.fao.org/eipws3g/
Food Outlook	http://www.fao.org/giews/reports/food-outlook/en/
FPMA (Food Price Monitoring and Analysis)	http://www.fao.org/giews/food-prices/home/en/
GLW (Livestock Systems)	http://www.fao.org/livestock-systems/en/
INFOODS (International Network of Food Data Systems)	http://www.fao.org/infoods/infoods/en/