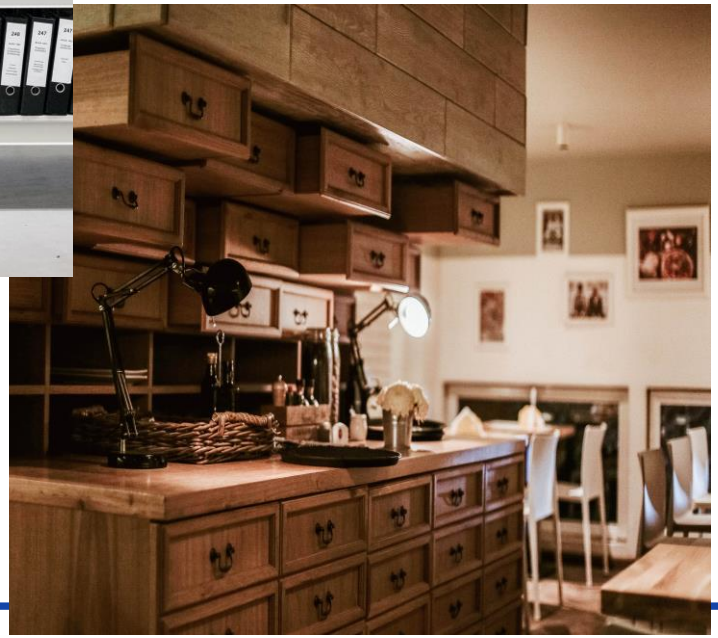
LINKING OPEN STATISTICAL DATA: THE ROLE OF STATISTICAL CLASSIFICATIONS

LOD Workshop, Berlin, 27-28 September 2017

Sangita Dubey, Senior Statistician

Statistics Division, FAO

FOOD AND AGRICULTURE STATISTICS IN SOME COUNTRIES



FOOD AND AGRICULTURE STATISTICS IN SOME COUNTRIES

FINDABLE

- non-shared national data
- difficult to find when available

ACCESSIBILITY

- non-open formats (PDF publications)
- Non-electronic formats

INTEROPERABLE

- Different definitions/vocabularies
- Different classifications/ontologies/hierarchies

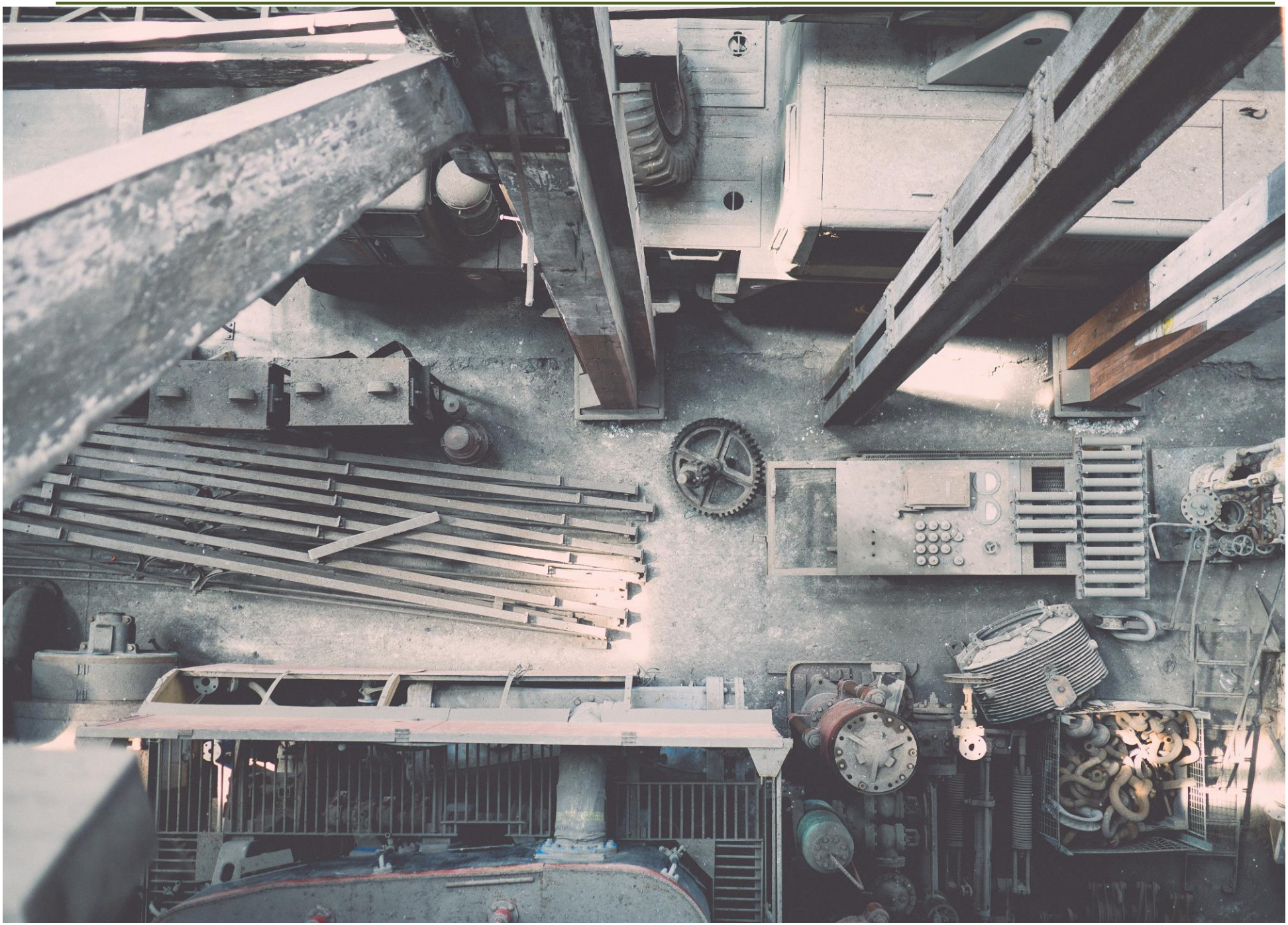
RE-USABLE

- Lack of adoption of Open Data licenses/standards in the statistics world
- Lack of guidelines/metadata

FOOD AND AGRICULTURE STATISTICS IN SOME COUNTRIES

Photo by MILKOVÍ on Unsplash

Food and Agriculture Statistics in other countries



BENEFITS OF OPEN LINKED STATISTICAL DATA

- Greater degree of comparability, integration and coherence
- Better availability of hidden data
- Better data quality, improved access and re-use
- Potential for greater detail/disaggegration
- Reduced duplication of efforts/resource use from open classifications and concordances

➢ *More efficient and effective evidence base for decision making*

➢ *Improved FAIR-ness of agriculture statistics*

NEEDS IN LINKING MULTI-SOURCE STATISTICAL DATA

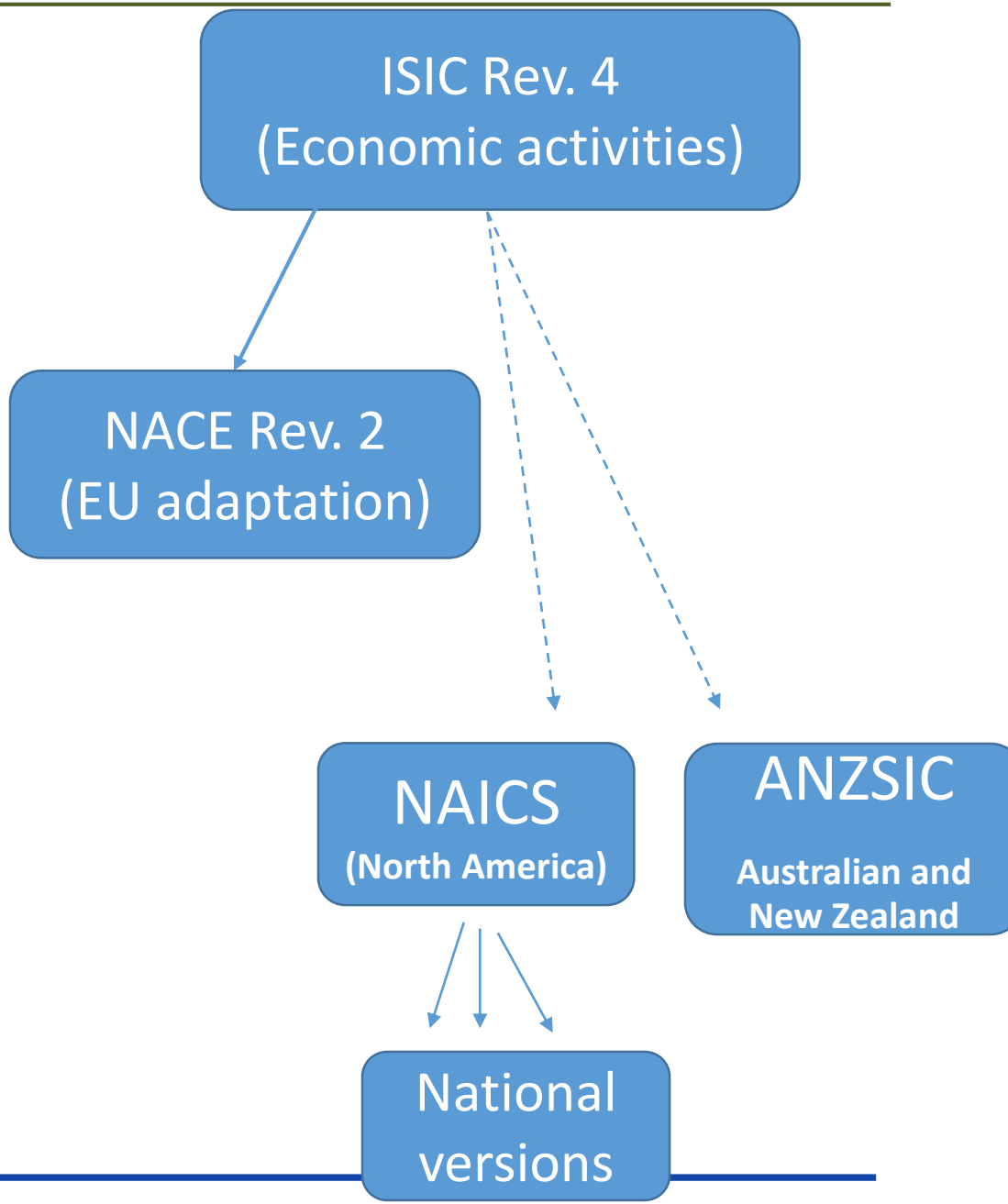
- **A unique "key" or identifier** to **link micro data across source**s for the same statistical unit (i.e. person, family, enterprise, farm, holding, product, etc.)
  - Example, linking tax/administrative and census and survey data
    - For persons: name, tax record, passport number, address, other?
    - For enterprises: name, address, tax number, other?
    - For commodities: name, unit of measure (kg, ton, litre), other?
    - What if a single "key" does not exist? – probabilistic matching

- **Consistent definitions/vocabularies** of items to be measured:
  - What is a farm? A family? A household? A business?
  - What are agriculture and food commodities? What is a potato, livestock, peppers, etc.?

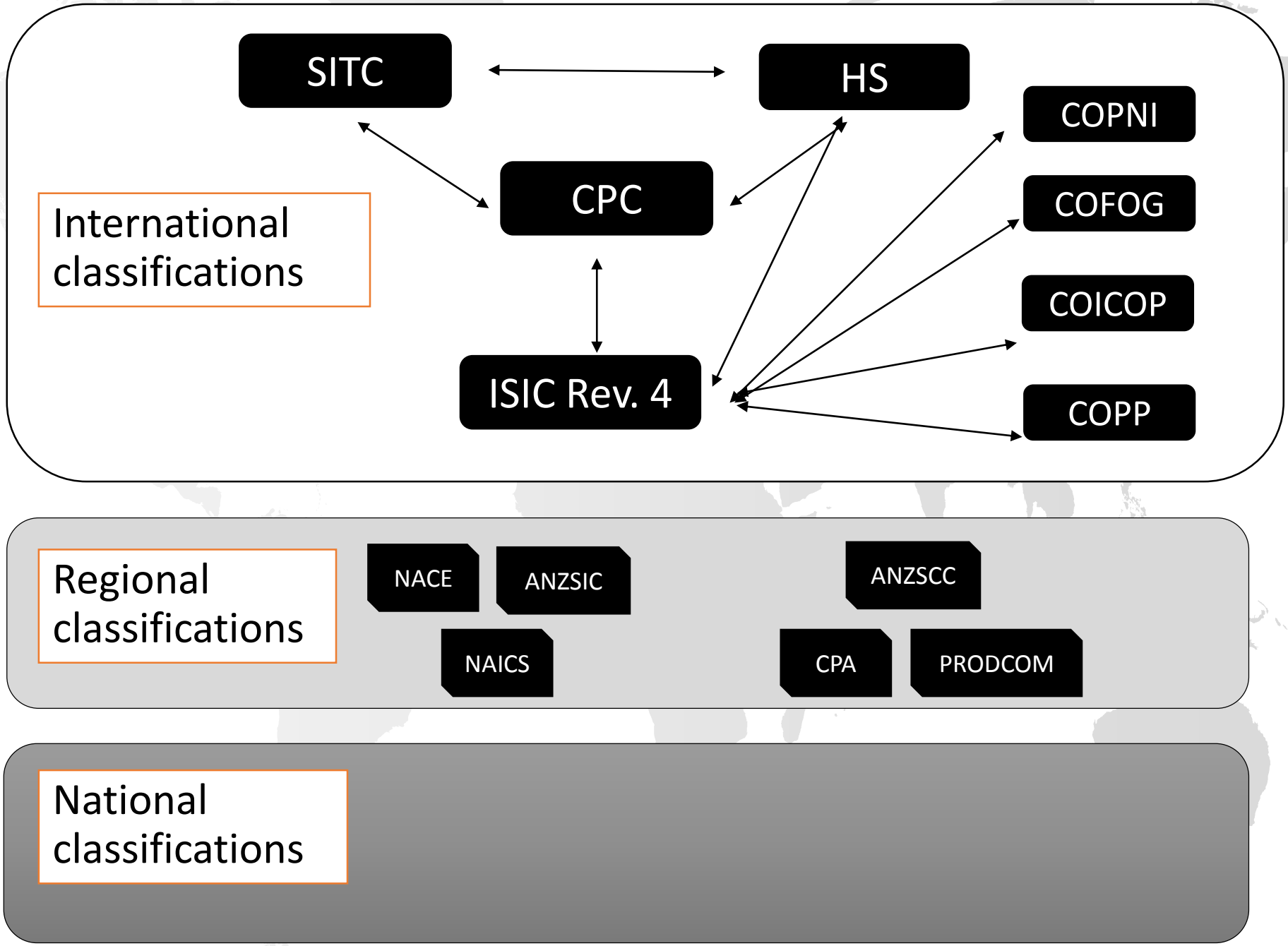
- **Consistent classifications/ontologies** to compare and link aggregated data

STATISTICAL CLASSIFICATIONS FOR LINKING DATA

- **Reference classification**
- **Derived classification**
  - Adopt (hierarchical) structure
    - consistency of aggregates
    - additional detail where necessary
  - Tailored for use at national or multinational level
- **Related classifications**
  - Only partial concordance
  - Differ in structure and/or building blocks from reference classifications
  - ***Linked mainly through correspondence or concordance tables***
  - many links with partial correspondences



THE SYSTEM OF NATIONAL ACCOUNTS (SNA)



THE INTERNATIONAL FAMILY OF STATISTICAL CLASSIFICATIONS

Group of international classifications

- Product of international agreement, approved by UN Statistical Commission (or other international board) and recommended as reference classification
- Provides framework against which to collect, organize, compile, analyze, disseminate and archive data, from both statistical and administrative collections, and to report and compare statistics at international level
- Facilitates standardization of statistical information, enables aggregate and disaggregate data sets in a meaningful way, and helps support policy and decision-making
- Improves cooperation, coordination and data quality at national and international levels

Requirements for Reference Classifications

- Have to be general enough to
  - be applicable to all countries
  - provide room for national adaptation and specifications
- Have to be specific enough to
  - allow for useful data collection
  - provide reasonable guidelines for creation of national classifications
- Must fulfill basic principles and essential criteria of international statistical classifications; be reviewed and approved by the UNSC (comprised of member states and experts) or similar competent body

THE INTERNATIONAL FAMILY OF STATISTICAL CLASSIFICATIONS

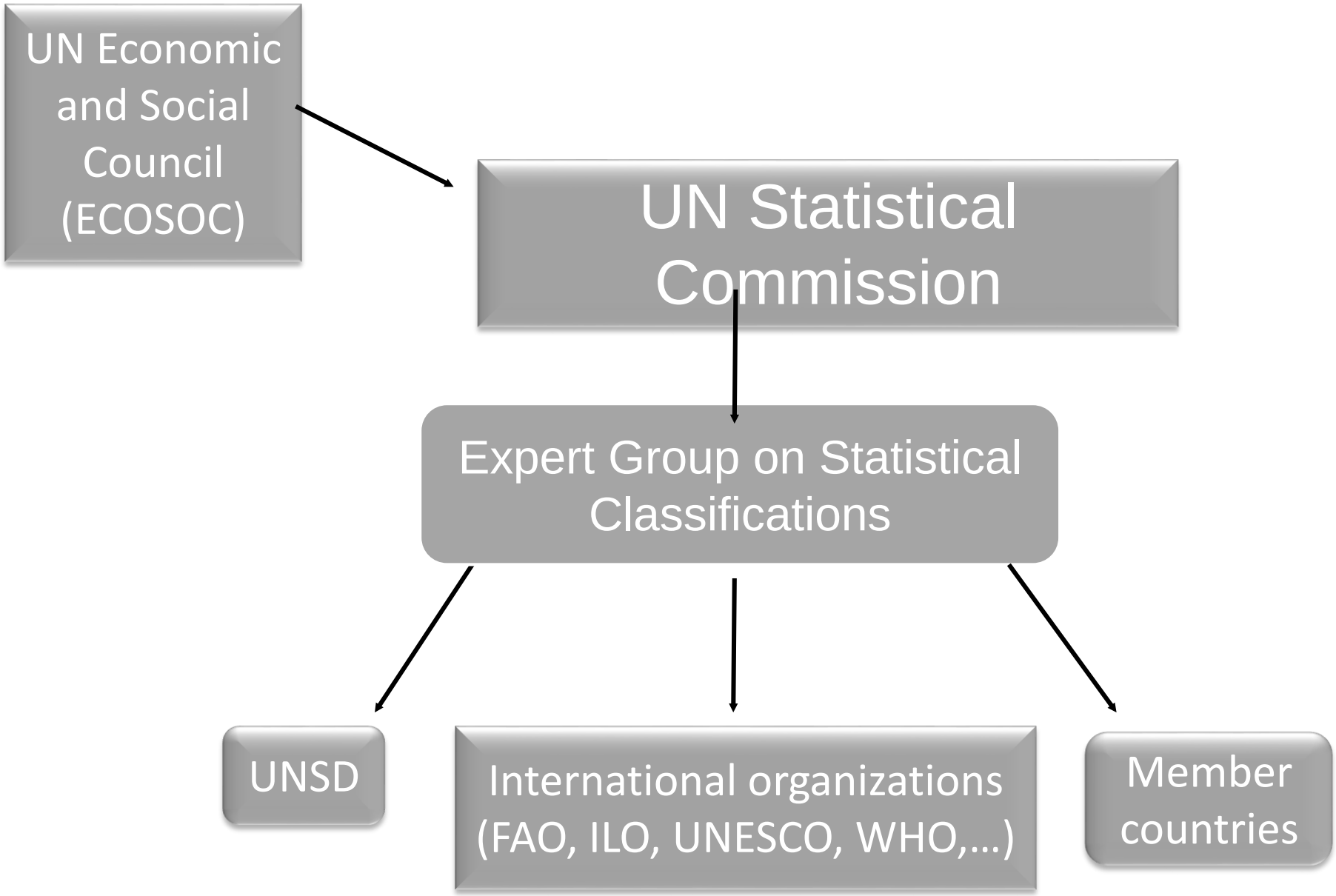
Covers international classifications from all statistical domains:

- Classifications officially approved as standard classifications by the UNSC (or similar)
- Other classifications that are internationally used and have become de-facto standards in their fields, even if originally created for regional or national purposes
- International & (multi-)national

Economic activities	Products	Health	Justice and crime
Trade	Environment	Population and migration	Geographical
Employment	Expenditures	Education	Agriculture & Food

https://unstats.un.org/unsd/classifications/Family

THE INTERNATIONAL INSTITUTIONAL SETTING



INTERLINKAGES BETWEEN CLASSIFICATIONS

- Within subject matter areas (e.g. Products)
  - CPC ↔ HS (Central Product Classification & Harmonized System)

- Between subject matter areas (e.g. Products, Activities)
  - ISIC ↔ HS ; ISIC ↔ CPC (International Standard Industrial Classification)

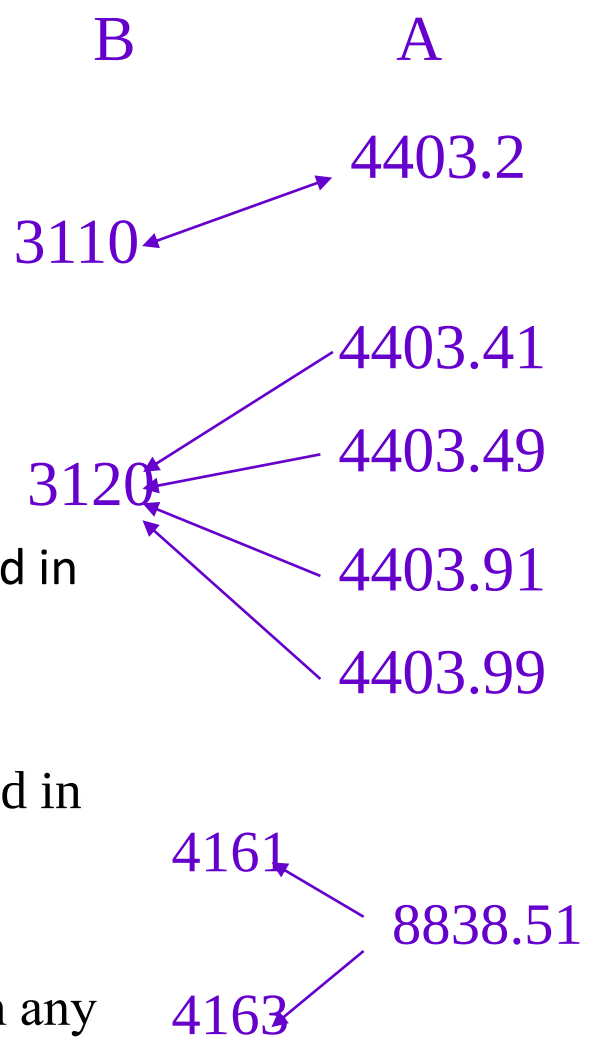
- Between types of classifications only (e.g. Activities)
  - ISIC ↔ NACE ↔ NAICS (EU and North American version of ISIC)

- Between types of classifications and subject matter areas (e.g. Activities, Products)
  - ISIC ↔ COFOG ↔ CRS Purposes (Classification of Functions of Government, and the Creditor Reporting System)

INTERLINKAGES BETWEEN CLASSIFICATIONS

Can generally be of the form (from A to B):

- One-to-one
  - Results can be transformed in both ways



- Many-to-one
  - Results of A can be compared in terms of B
- One-to-many
  - Results of B can be compared in terms of A
- Many-to-many
  - Results can't be compared in any direction

ROLES OF CUSTODIANS

- Maintain, update and revise “their” classification

- Follow “family” obligations
  - Collaborate with other custodians
  - Support national implementation of international classifications
  - Exchange information and work transparently
  - Contribute experience to advance Reference Classification (guidelines, definitions, interpretations,…)
  - Coordinate timing of changes/updates
  - Develop guidance and training materials

- UN Statistics Division (UNSD) - main custodian of international statistical classifications; sector-specific international classifications owned by other agencies, such as FAO, the WHO, the ILO, UNESCO and the ISO.

NATIONAL IMPLEMENTATION

- Depending on applicability, statistical classifications may be: national, regional (or supranational) or international.
- To ensure data harmonization and comparability, international standards should be taken as points of reference when developing national and supranational schemes, and adapted to suit individual countries' statistical needs.
    - International classifications are "living tools" subject to continuous improvement.
    - Countries play a crucial role.

When implementing international classifications, it is not always possible or appropriate to use them in their original formats and structures. Different solutions may be envisaged to take into account specific requirements and country needs.

NATIONAL IMPLEMENTATION

- When international classifications are used in their original state, the highest degree of comparability is ensured; though national adaptations may be developed to meet user needs while maintaining compatibility with the international classification of reference. Two options are recommended:

  • Case 1 (preferred): high-level categories maintained but further detail added at lower level for regional or national purposes.

  • Case 2: lower-level items are retained, but different aggregates defined as the sum or split of the international items. Allows for a comparability as aggregates may be reconstructed as in the ICs, while lower levels are linked on a one-to-one basis.

Case study – Practical example of revision of an international classification : FAO switchover to CPC

On how flexible enough classification frameworks can help serve the needs and on how end-users can contribute to the improvement and evolution of international classifications

PRODUCTION & TRADE STATISTICS AT FAO - COMMODITY LIST

53

years

245

countries, regions, territories

> 1800

commodities

FAO COMMODITY LIST

AGRICULTURE (700)
- **primary products & livestock**
- **processed products**

FISHERIES & AQUACULTURE (1000)
- **live aquatic animals & seaweeds**
- **fishery products**

forest products (54)
- **machinery & equipment (16)**
- **fertilizers (26)**
- **pesticides (30)**

MAIN AREAS OF WORK ON CLASSIFICATIONS AT FAO

Harmonization of standards used in FAO and increased use of international classification is a priority:

IN THE ORGANIZATION

- **Developed CPC expanded for agriculture**
- **CPC and HS introduced in main statistical databases (e.g. FAOSTAT)**

WITH INTERNATIONAL ORGANIZATIONS

review international classifications for FAO needs

- **most recently: HS, CPC and COICOP**

WITH MEMEBER COUNTRIES

promote the use of international classification to integrate agricultural statistics into national statistical system

- **developed guidelines on classifications**

NEW FAO (COMMODITY) CLASSIFICATION SYSTEM

Today

CPC agricultural production SUA	HS Agricultural trade

before

FAOSTAT commodity list
reference classification in FAO for production, trade & SUA

REVIEW PROCESS

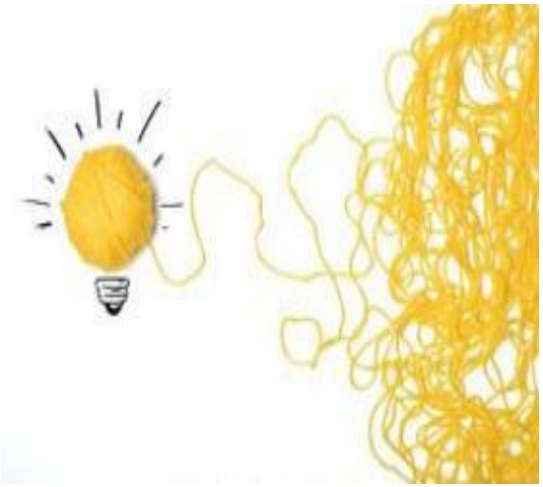
Over the past ten years FAO contributed to **CPC & HS review to make them suitable for agricultural statistics**

HS 2012 & 2017	CPC Ver.2 & Ver.2.1
8 years	8 years
23 meetings	5 meetings
373 new sub-heading	271 new sub-classes

REVIEW PROCESS & RESULT

problem

- ❑ CPC is a general scope classification
- ❑ FAO's classification was a sector-specific organization classification
- ❑ CPC needed more detail needed on food & agriculture products



solution

CPC EXPANDED
for agriculture & food
(7 digits)

now an official annex to CPC Ver.2.1
105 additional codes

CPC EXPANDED CODE

One more level to accommodate FAO detail

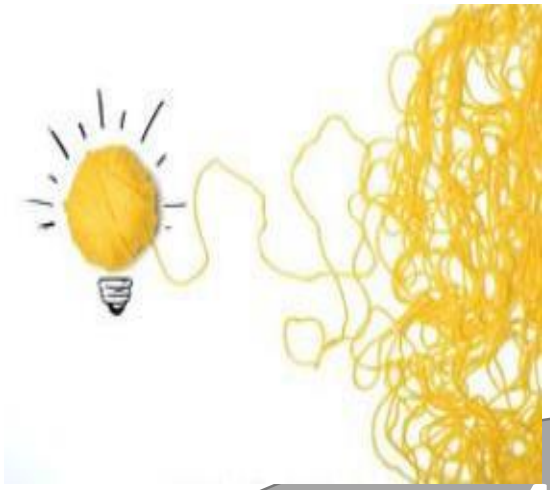
0	0	0	0	0	.	0	0
Section							
Division							
Group							
Class							
Sub-class							
FAOSTAT detail							

Countries can use the CPC expanded to identify local products while ensuring data comparability at the global level

AID DATA - PROBLEM, REVIEW PROCESS & RESULT

problem

- Gaps in OECD's CRS classification of purposes related to food and agriculture (i.e. Food Safety, One Health)
- Inability to monitor flows to SDG targets

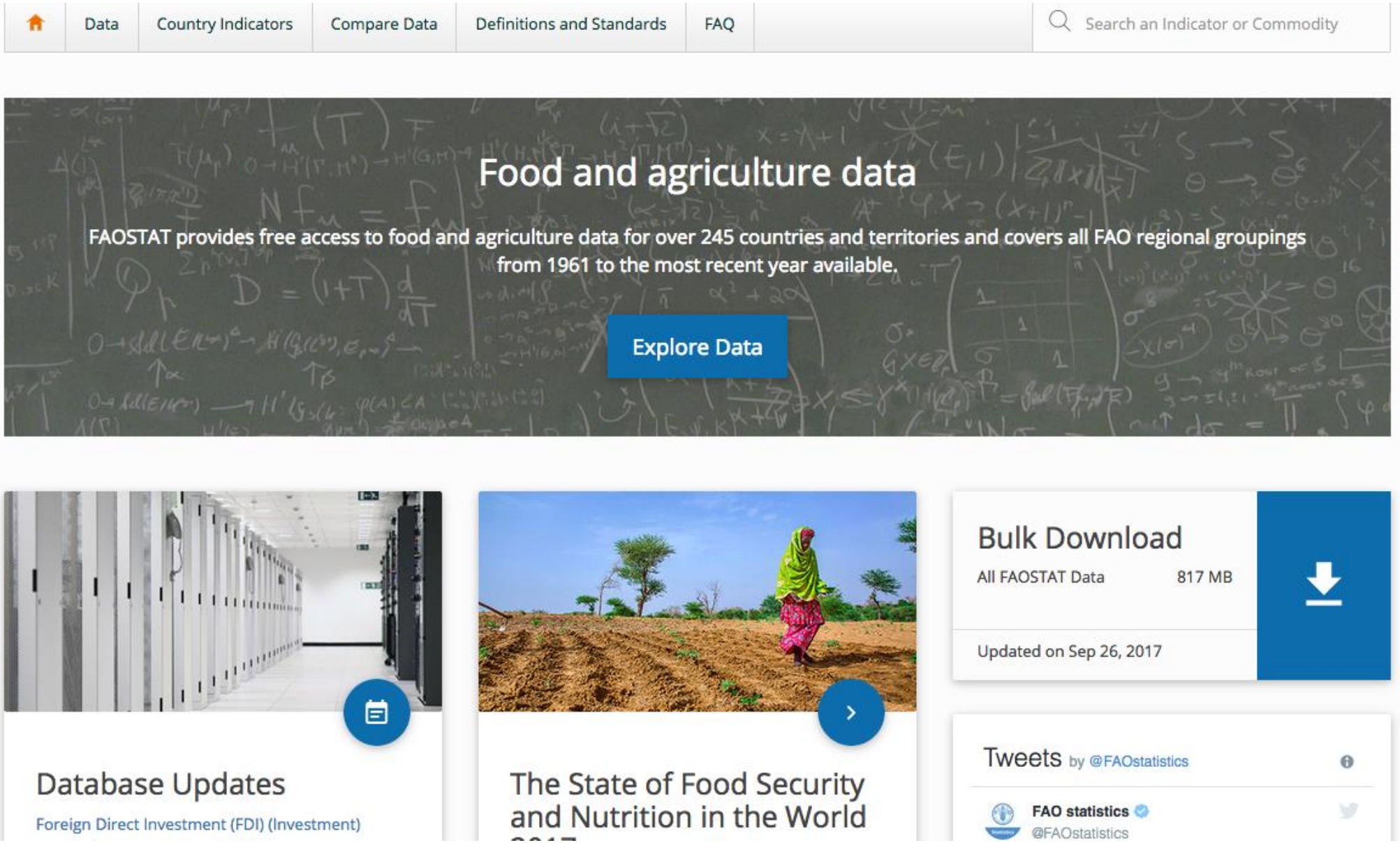


solution

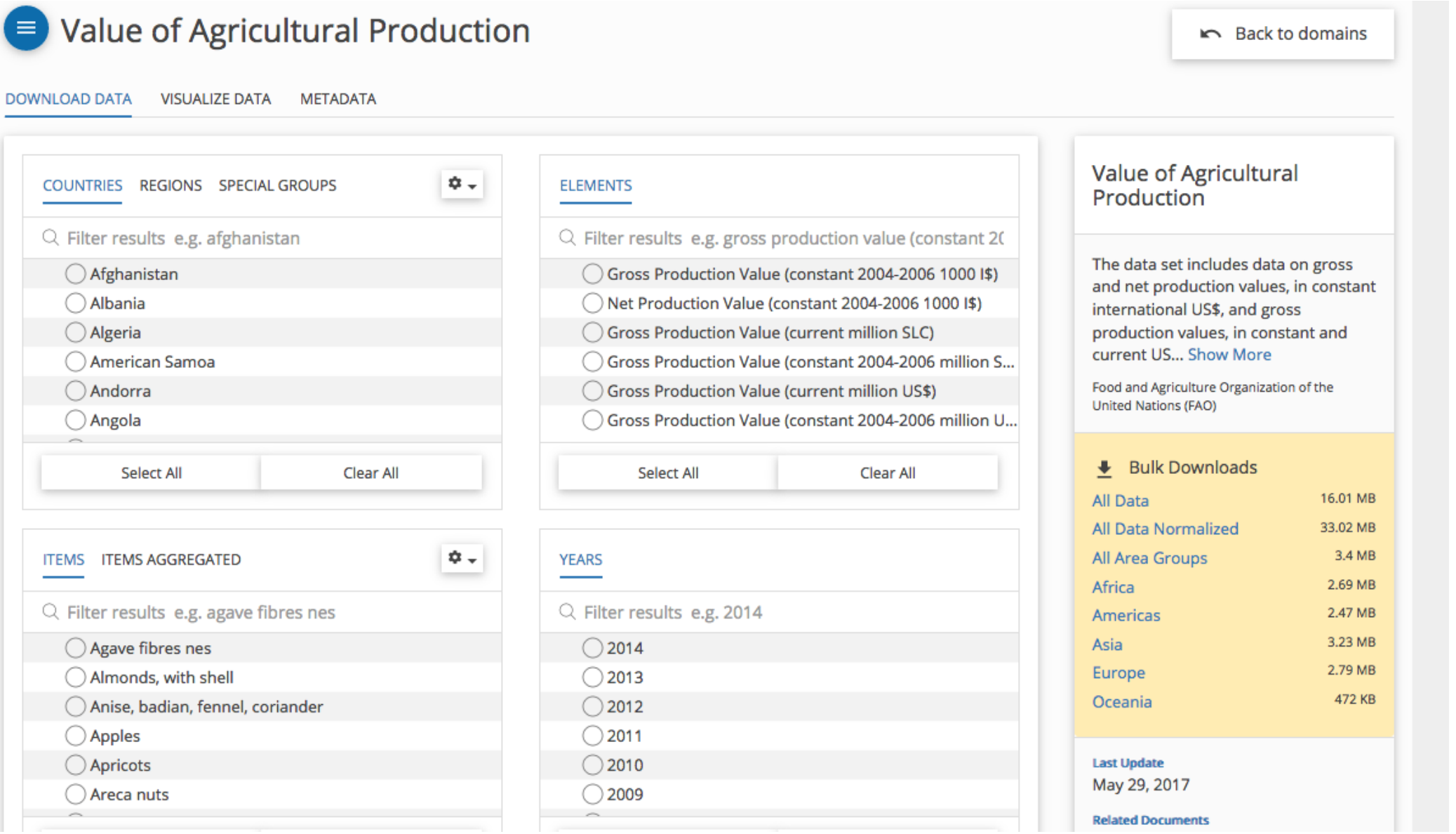
Consultation-based proposal underway

4 new purposes being proposed,
and new sector on food security & nutrition

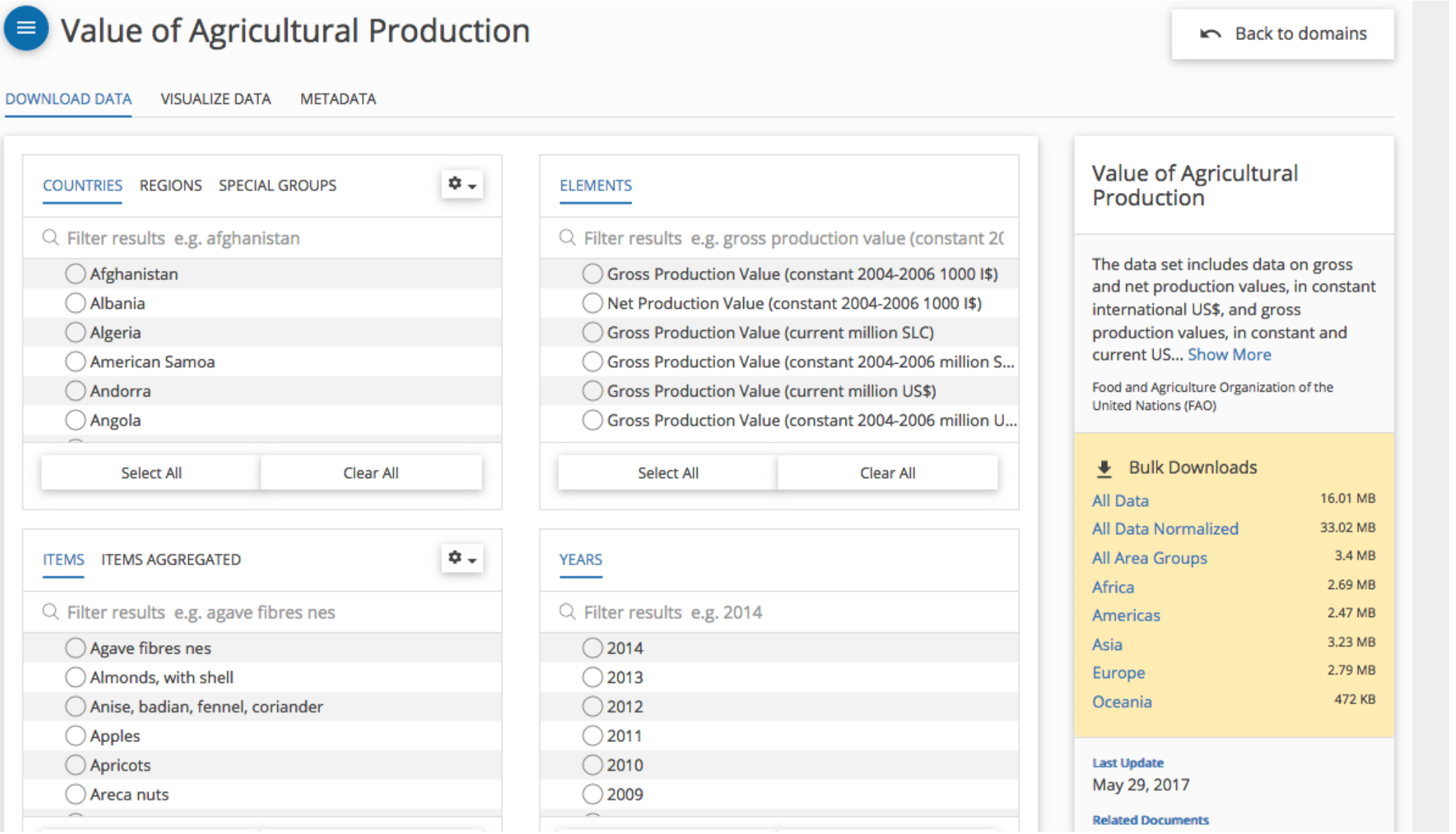
FAOSTAT


Data
Country Indicators
Compare Data
Definitions and Standards
FAQ
Search an Indicator or Commodity
Food and agriculture data
FAOSTAT provides free access to food and agriculture data for over 245 countries and territories and covers all FAO regional groupings from 1961 to the most recent year available.
Explore Data
Database Updates
Foreign Direct Investment (FDI) (Investment)
The State of Food Security and Nutrition in the World
Bulk Download
All FAOSTAT Data
817 MB
Updated on Sep 26, 2017
Tweets by @FAOstatistics
FAO statistics
@FAOstatistics

FAOSTAT


Value of Agricultural Production
Back to domains
DOWNLOAD DATA
VISUALIZE DATA
METADATA
COUNTRIES
REGIONS
SPECIAL GROUPS
Filter results e.g. afghanistan
Afghanistan
Albania
Algeria
American Samoa
Andorra
Angola
Select All
Clear All
ELEMENTS
Filter results e.g. gross production value (constant 2(
Gross Production Value (constant 2004-2006 1000 I$)
Net Production Value (constant 2004-2006 1000 I$)
Gross Production Value (current million SLC)
Gross Production Value (constant 2004-2006 million S...
Gross Production Value (current million US$)
Gross Production Value (constant 2004-2006 million U...
Select All
Clear All
ITEMS
ITEMS AGGREGATED
Filter results e.g. agave fibres nes
Agave fibres nes
Almonds, with shell
Anise, badian, fennel, coriander
Apples
Apricots
Areca nuts
YEARS
Filter results e.g. 2014
2014
2013
2012
2011
2010
2009
Value of Agricultural Production
The data set includes data on gross and net production values, in constant international US$, and gross production values, in constant and current US... Show More
Food and Agriculture Organization of the United Nations (FAO)
Bulk Downloads
All Data 16.01 MB
All Data Normalized 33.02 MB
All Area Groups 3.4 MB
Africa 2.69 MB
Americas 2.47 MB
Asia 3.23 MB
Europe 2.79 MB
Oceania 472 KB
Last Update
May 29, 2017
Related Documents

FAOSTAT


Value of Agricultural Production
Back to domains
DOWNLOAD DATA
VISUALIZE DATA
METADATA
COUNTRIES
REGIONS
SPECIAL GROUPS
Filter results e.g. afghanistan
Afghanistan
Albania
Algeria
American Samoa
Andorra
Angola
Select All
Clear All
ELEMENTS
Filter results e.g. gross production value (constant 2(
Gross Production Value (constant 2004-2006 1000 I$)
Net Production Value (constant 2004-2006 1000 I$)
Gross Production Value (current million SLC)
Gross Production Value (constant 2004-2006 million S...
Gross Production Value (current million US$)
Gross Production Value (constant 2004-2006 million U...
Select All
Clear All
ITEMS
ITEMS AGGREGATED
Filter results e.g. agave fibres nes
Agave fibres nes
Almonds, with shell
Anise, badian, fennel, coriander
Apples
Apricots
Areca nuts
YEARS
Filter results e.g. 2014
2014
2013
2012
2011
2010
2009
Value of Agricultural Production
The data set includes data on gross and net production values, in constant international US$, and gross production values, in constant and current US... Show More
Food and Agriculture Organization of the United Nations (FAO)
Bulk Downloads
All Data 16.01 MB
All Data Normalized 33.02 MB
All Area Groups 3.4 MB
Africa 2.69 MB
Americas 2.47 MB
Asia 3.23 MB
Europe 2.79 MB
Oceania 472 KB
Last Update
May 29, 2017
Related Documents

SUCCESSES AND CHALLENGES AT FAO

- Open-ness
  - FAO Success: FAO a signatory to the International Aid Transparency Initiative (IATI) since April 2016; began publishing its aid project data as of May 2017, and now publishes quarterly data under open data licenses.
  - Challenge: FAOSTAT, the FAO portal for agricultural data, provides free data. Open licenses not yet in place, creating uncertainty on use terms for some

- Linkages (Statistical Classifications)
  - FAO Success: improvements fo CPC, HS and COICOP international statistical classification to improve detail and inter-country comparability of agricultural production, farm-gate prices, trade, and consumption.
  - Challenge: provide multi-lingual open data compliant classifications and concordances (international, regional and national)

NEXT STEPS

- Provide multi-lingual open data compliant international classifications and concordances (international, regional and national)
  - Starting with the OECD's Creditor Reporting System (CRS), used to report aid expenditures to IATI
  - Continue with CPC, HS, etc.

- Provide concordances across versions of a classification, and between classifications

- Develop and publish national-to-international concordances for select partner countries

- Determine platform to host these "vocabularies"/ontologies in collaboration with other experts (e.g. GACS, RDA, GODAN)

CONTACT INFORMATION: Sangita.Dubey@fao.org