Pacific Data Hub
Store once, publish many

SDMX GLOBAL CONFERENCE

October 2023
The Pacific Data Hub programme

The Pacific Data Hub, generally abbreviated « PDH », is a program aiming to improve access to Pacific data. It includes a series of deliverables related to document and data accessibility and also some activities in the area of outreach and capacity-building.
Improving the statistical business process

- Data products are specified explicitly using standard metadata (SDMX)
- Data products are designed as multi-dimensional data cubes i.e. large multipurpose tables rather than one-size-fits-all products

From metadata, a data collection template (Excel), database tables (PostgreSQL) and output products on .STAT are generated automatically (metadata-driven approach)

- Separation of concerns between data collection and data dissemination
- Re-usable building blocks are domain independent

Templates offer a series of common functions out-of-the box:
- Validate against data structure
- Compare with previous version
- Publish to internal or public version of .STAT

- Single version of the truth principle

Data is stored in a reference database
- Derived indicators can be calculated automatically
- Revision history and update dates are recorded automatically

Data is directly available on a wide range of dissemination channels: .STAT data Explorer, PDH Data Catalogue, websites and dashboards, Excel, Power BI, statistical packages, ...

- Collect once, re-use many times principle
Graphical User Interface (GUI) - .STAT Data Explorer

Basic user
A powerful search engine allows finding quickly most common statistical tables

Advanced user
User-friendly interface to extract tailored information from voluminous data tables

Application Programming Interface (API) – SDMX Web-service

Statistician and Data Scientist
.Stat data is accessible directly from most popular data analysis software (Excel, Power BI, R, Python, Stata)

Web-developer and IT specialist
.Stat data can be included in any IT application using SDMX web services
Integrating .Stat Suite in a broader data experience

Data Explorer browser interface
You can view, query and download data directly from your internet browser.

Plugins
Create a live connection between PDH.stat and your Excel Workbooks, Power BI reports, STATA, R or Python applications, etc.

SDMX .stat API
A Statistical Data and Metadata eXchange (SDMX) API for programatically accessing and retrieving the full range of published data and indicators.
Store once, re-use many times: Overview of data access channels

- Pacific Data Hub .STAT
- Data Explorer
  - GUI
    - Download
      - Excel
      - CSV
      - Filtered
      - All
    - Share
    - API links
  - Plugins
    - Office applications
      - Excel add-in
      - Power BI connector
    - Statistical applications
      - R (rdsdmx)
      - Python (pandasdmx)
      - Stata
    - Web CMS
      - Drupal
      - WordPress
  - Other PDH applications
    - Data Catalogue
      - .STAT dataflows
      - Custom packages
  - Pacific Map
Example: SPC’s Statistics for Development Division website integration with Drupal plugins

https://sdd.spc.int/
Example: COVID-19 Pacific Community Updates Web pages with Drupal plugins

https://www.spc.int/updates/blog/2023/02/covid-19-Pacific-community-updates#CurrentStatus
https://php.spc.int/covid-19
Example: Thematic dashboards with Power BI

https://pacificdata.org/dashboards
Example: Country Statistical Snapshots with Python/pandaSDMX
Example: .STAT content on the Pacific Data Hub CKAN catalogue

https://pacificdata.org/
Benefits of having an SDMX-driven statistical business process

When a new statistical process is created for the Pacific Data Hub, the first step is to design the structure of the new statistical product and specify it explicitly using standard SDMX metadata.

Efficient processes

✓ Data products are designed as multi-dimensional data cubes by customising common SDMX data structure templates and by re-using as far as possible common concepts and code lists.
✓ Data collection templates, database tables and output products are generated to a large extent by domain-independent application building blocks (metadata-driven approach).
✓ Data is integrated from various sources and disseminated with SDMX as backbone model supporting a reference database implementing the “single version of the truth” principle.

Interoperable data

✓ SDMX standard syntaxes and information model provide syntactic interoperability.
✓ Common data structure templates, concepts and code lists shared among PDH datasets further provide semantic interoperability.
✓ Datasets from PDH.STAT are consistently modelled which make them easy to link together.
Thank you!

Denis Grofils
Statistics Advisor Process Modernisation
Statistics for Development Division
Pacific Community
denisg@spc.int