

# 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.



Data Catalogue - an open data repository which manages and publishes all data in the Pacific Data Hub. It is the central component which links to PDH.stat and the Microdata Library; Most spatial datasets in the Data Catalogue can be visualised with PacificMap



**PacificMap** - a geospatial data exploration tool providing easyto-use map-based visualisation of spatial data;



PDH.stat - indicator database explorer which contains the 132
Pacific Sustainable Development Indicators (SDGs) as well as a range of economic, health, demographic and environmental datasets (replaces National Minimum Development Indicator Database - NMDI)



**Microdata Library** - online census and survey documentation and archiving application which also provides access to microdata for some collections.



### Improving the statistical business process

#### **DESIGN**

#### **BUILD**

#### **COLLECT**

#### **PROCESS**

#### **DISSEMINATE**

- Data products are specified explicitly using standard metadata (SDMX)
- Data products are designed as multidimensional data cubes i.e. large multipurpose tables rather than one-sizefits-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

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
- → Re-usable building blocks are domain independent

Data is stored in a reference database

- Derived indicators
   can be calculated
   automatically
- Revision history

   and update dates
   are recorded
   automatically

→ Single version of the truth principle

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



### **Designed for a wide range of users**





### **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.

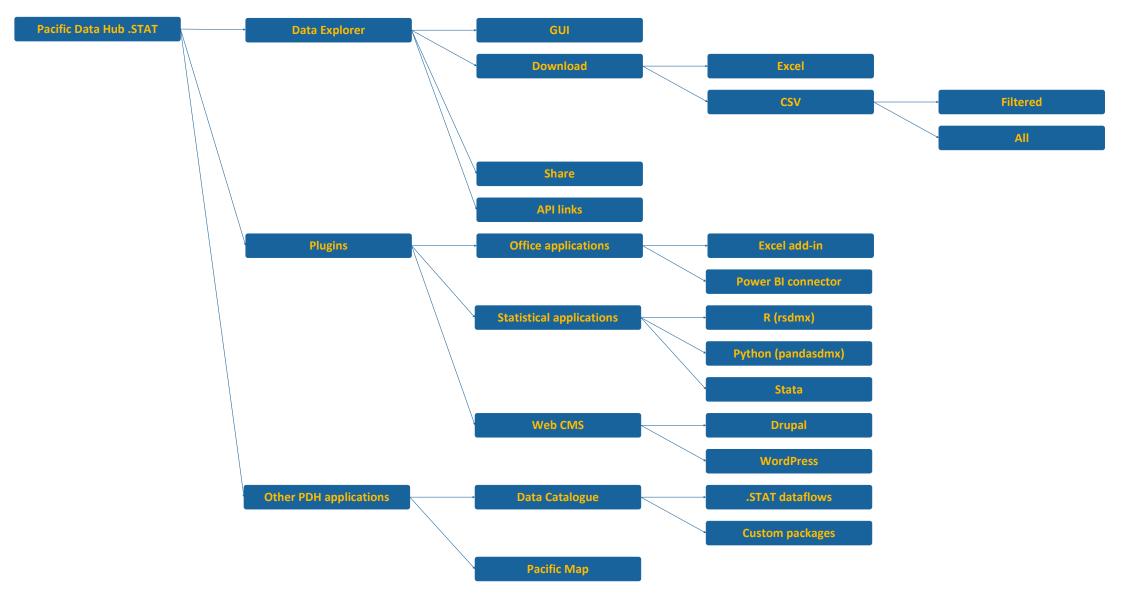


#### SDMX .stat API

A Statistical Data and Metadata eXchange (SDMX) API for programmatically accessing and retrieving the full range of published data and indicators.



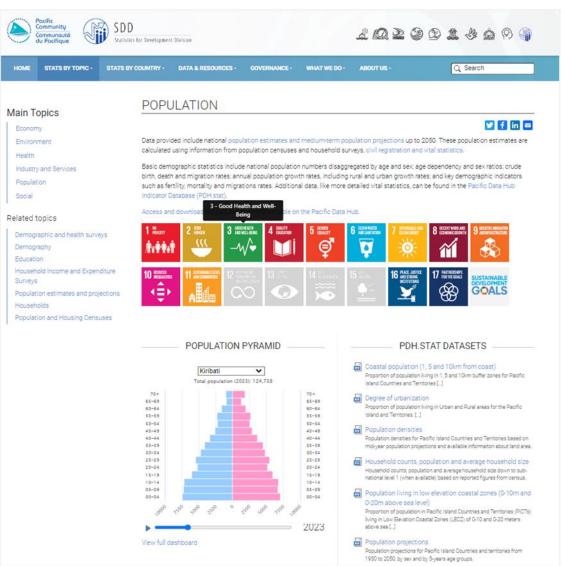
### Store once, re-use many times: Overview of data access channels





### **Example: SPC's Statistics for Development Division website integration with Drupal plugins**



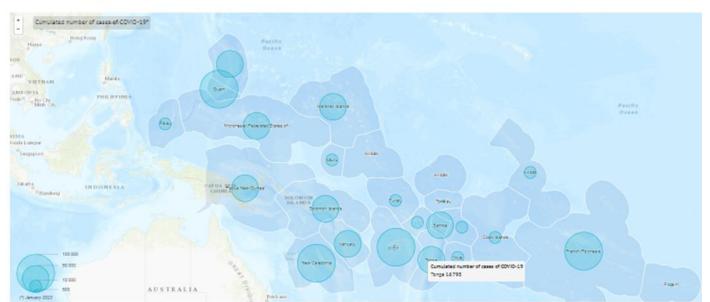


https://sdd.spc.int/

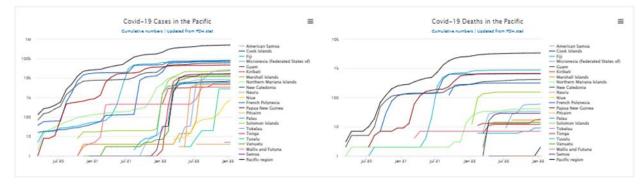


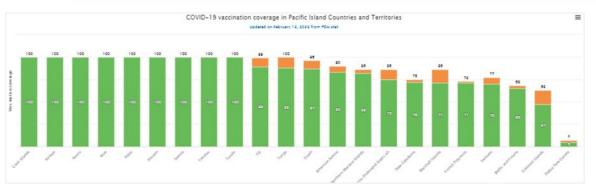
### **Example: COVID-19 Pacific Community Updates Web pages with Drupal plugins**

Covid Recent cases Map



Pacific	c Island Countries and territories	2022-01	2022-02	2022-03	2022-04	2022-05	2022-06	2022-07	2022-08	2022-09	2022-10	2022-11	2022-12	2023-01
AS	American Samoa	22	15	4,474	1,249	6,192	6,555	7,952	8,168	8,247	8,260	2,265	8,502	2,32
ж	Cook Islands	0	42	1,926	4,990	3,675	5,774	6,042	6,265	6,319	6,405	6,475	6,533	6,91
P2	PD:	62,855	63,930	64,409	64,643	65,031	68,713	67,617	60,171	68,242	69,264	60,373	68,750	60,00
FМ	Micronesia (Federated States of)	0	0	٥	1	50	35	5,214	7,475	13,953	22,208	22,055	22,247	23,6
SU	Goam	82,881	44,344	47,265	48,132	49,078	\$1,090	\$3,208	\$7,947	\$4,577	\$1,133	89,379	60,053	60,7
o	Kirlbati	460	2,927	1,066	3,082	3,114	3,236	3,430	3,430	1,430	4,366	4,066	4,887	5,0
мн	Marshall Islands	4	4	4	14	14	15	64	15,042	15,305	15,382	15,541	15,554	15,5
ΗР	Northern Marlana Islands	4,908	9,550	11,065	11,263	11,366	11,759	12,573	13,120	13,171	13,212	11,227	13,358	13,5
NC	New Caladonia	20,622	54,313	60,413	60,653	62,529	64,337	69,796	73,934	74,288	74,406	76,187	79,877	79,9
NR	Nauru	0	0	٥	3	3	3,394	4,407	4,610	4,610	4,610	4,623	4,625	4,7
NU	Nice	0	0		- 1	9.	15	49	71	80	15	147	439	2
pp	French Polynesia	42,979	67,299	72,166	72,987	73,194	73,601	75,701	76,745	76,046	76,911	77,284	78,104	78,23
PG	Pepus New Guines	37,270	41,302	43,374	43,817	44,551	44,725	44,793	44,887	43,010	45,465	46,182	46,663	46,61
PN	Piscelm	0		0	0	0	0	4	4	4	4	4	4	
PW	Pelau	1,633	3,781	4,024	4,610	5,065	5,199	5,232	5,348	5,460	5,515	5,785	5,971	5,91
58	Solomon Islands	1,416	7,049	11,470	15,677	21,237	21,544	21,544	21,544	21,544	21,544	24,575	25,684	25,61
TK	Tokulasi	0	0	0	0	0	0	0	0	0	0	0	5	
TO	Tonga	2	355	6,546	10,330	11,769	12,346	12,597	16,182	16,182	16,182	16,222	16,437	16,71
TV	Tovals	0	0	0	0	1	3	1	20	20	20	2,779	2,779	2,77
VU	Vancatu	7	12	1,717	7,914	9,230	11,271	11,710	11,825	11,920	11,964	11,962	12,011	12,0
WF	Walife and Fetuna	454	454	414	455	455	455	825	3,218	3,411	3,415	3,416	3,431	3,4
WS	Samoa	32	33	1,889	9,392	13,647	14,964	15,405	13,339	19,941	15,946	15,946	18,975	16,3
Total		211,065	296,086	336,723	363,634	101,909	396,341	420,235	453,845	462,638	473,492	413,213	491,439	496,1





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**



Microsoft Power Bl



### **Example: Country Statistical Snapshots with Python/pandaSDMX**







































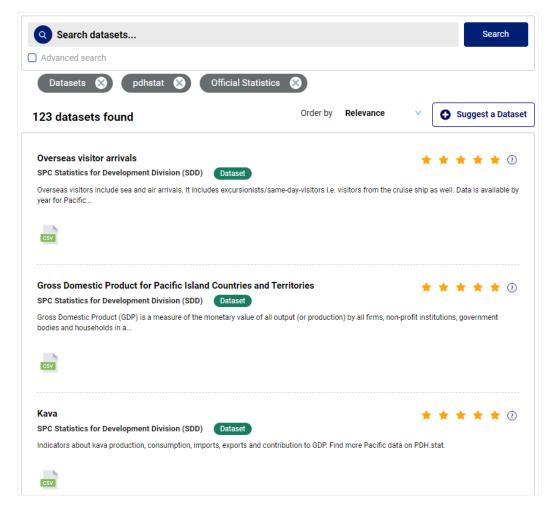




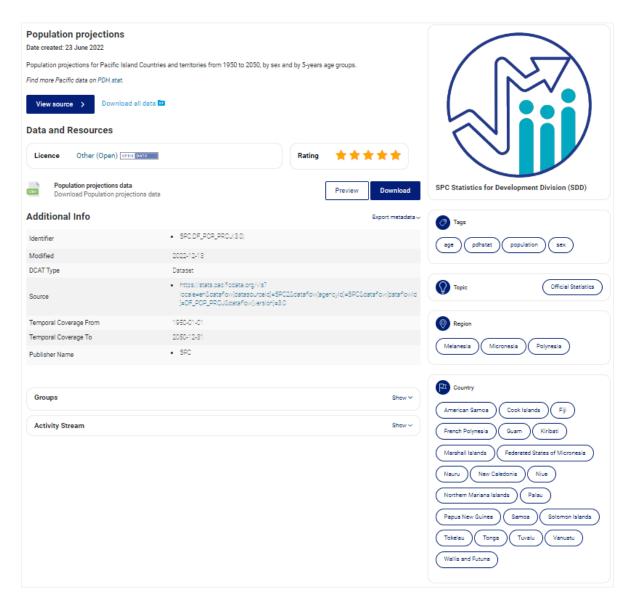




### **Example: .STAT content on the Pacific Data Hub CKAN catalogue**



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