STATISTICAL DISCLOSURE CONTROL FOR CENSUS TABLES WITH REDATAM

Build improved disclosure control solutions and algorithms into REDATAM

CELADE - Population Division of ECLAC, United Nations
Software tool developed by CELADE since 1985 for processing and disseminating population and household census micro data. With the support of the Saint Lucia and Chile Statistical Offices, who provided CELADE with their census microdata to be used as examples for all the testing.

Handles large volumes of micro data in a hierarchical structure and facilitates data processing for small geographical areas.

Use of the software has expanded and it is currently used for processing all types of data (micro or aggregated).

User-defined study areas up the smallest geographical division.

And 35 years after that, the PCs are bigger, faster, and we are still using Redatam! Amazing!
User friendly and Free
Reduces original database size
Store efficiently each variable in a separate binary files
Produce tabulations rapidly and efficiently (1 million record per second)
Customize table, graph and thematic map styles according to users' requirements
Command language for aggregated data generation
Made for actual humans, not just data scientists
Census/Stats and role of REDATAM

- Data check
- Hierarchical structure validation
- Indicators System for Quality Control

Field enumeration

- Efficient data storage
- Imputation validation
- Tabulation plan
- Sociodemographic profiles

Data processing

- Websites widely accessed
- Democratization of access to microdata
- Training in the use, processing and analysis of data

Dissemination

- EIM & EIF Demographic Modules
- Population Analysis and Indicators for CdeM, SDG
- Redatam modules for UBN, overcrowding, poverty hotspots, humanitarian aide

Analysis

- Data check
- Hierarchical structure validation
- Indicators System for Quality Control

Virtual Expert Group Meeting on Statistical Disclosure Control for Caribbean Census Tables

November, 10th 2020
During these years Redatam evolved, moved one of its branches to the internet, providing online access to the NSO’s databases.
Special uses of Redatam Platform

Redatam Command Language
- Expand the use of Redatam databases
- Not common language, training is the key
- Outputs only in Excel format

Dynamic Graph Visualization
- JavaScript modules
- Adaptable by other developers

TOD Project: Using Redatam with Python (pyRedatam)
- Laboratorio de Gobierno de la Ciudad de Buenos Aires (Subsecretaria de Transporte)
- [http://gcba.github.io/tod/](http://gcba.github.io/tod/)
### Main Goals (RedatamX incoming version)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anonymization</td>
<td>Preventing disclosure of the tabular data at the lowest geography levels</td>
</tr>
<tr>
<td>Data encryption</td>
<td>Protecting the microdata and new internal data formats</td>
</tr>
<tr>
<td>Program accessibility</td>
<td>Programmers will be able to develop plug-ins and independent modules to customize access to Redatam databases</td>
</tr>
</tbody>
</table>
Redatam technology ecosystem

**RedatamX**
- Multiplatform (X stands for Cross platform)
- Json output format
- Open source for components
- Integration with stats packages

**Map Integration**
- Geojson map format
- Queries to REDATAM databases from GIS applications

**API’s**
- Interoperability with Redatam databases
- Use of Redatam Command Language in multiple environments and analytics applications
- Replicable, easily adaptable by other developers