ADDITIONAL DOCUMENTATION

Data download and extraction

Due to the size of the information to be downloaded, a CPD (Data Processing Center) in a virtual machine, and not locally, is chosen, suitable for the flexibility and security it offers in the flow of information. In addition, it offers high remote availability to researchers through single, non-concurrent access.

For data downloading and processing, a virtual machine hosted in Microsoft Azure is used, equipped with 4 processing cores, 16 GB of RAM, a 2 TB disk of storage, up to 1 TB of network traffic, and remote access by VPN (Virtual Private Network), with Bastion security and monitoring system, and with automated on and off capabilities in a 24/7 operational service.

ORCID data

Public information about ORCID data is located mainly at two URLs (Uniform Resource Locator):

- (1) https://github.com/ORCID/ORCID-Source
- (2) https://info.orcid.org/es/

Use of data in the file is governed by the data file's terms of use (spanish version) (https://info.orcid.org/es/pol%C3%ADtica-de-uso-de-archivos-de-datos-p%C3%BAblicos) and the Privacy Policy (https://info.orcid.org/es/Pol%C3%ADtica-de-Privacidad/), whereas the Public Data Archive is released annually under a dedicated public domain CC0 1.0 developed by Creative Commons, with the publication of recommended community standards for its use, without restrictions or conditions (including those contained in the Terms and Conditions of Use and Membership Agreement) added.

ORCID data structure and hierarchy

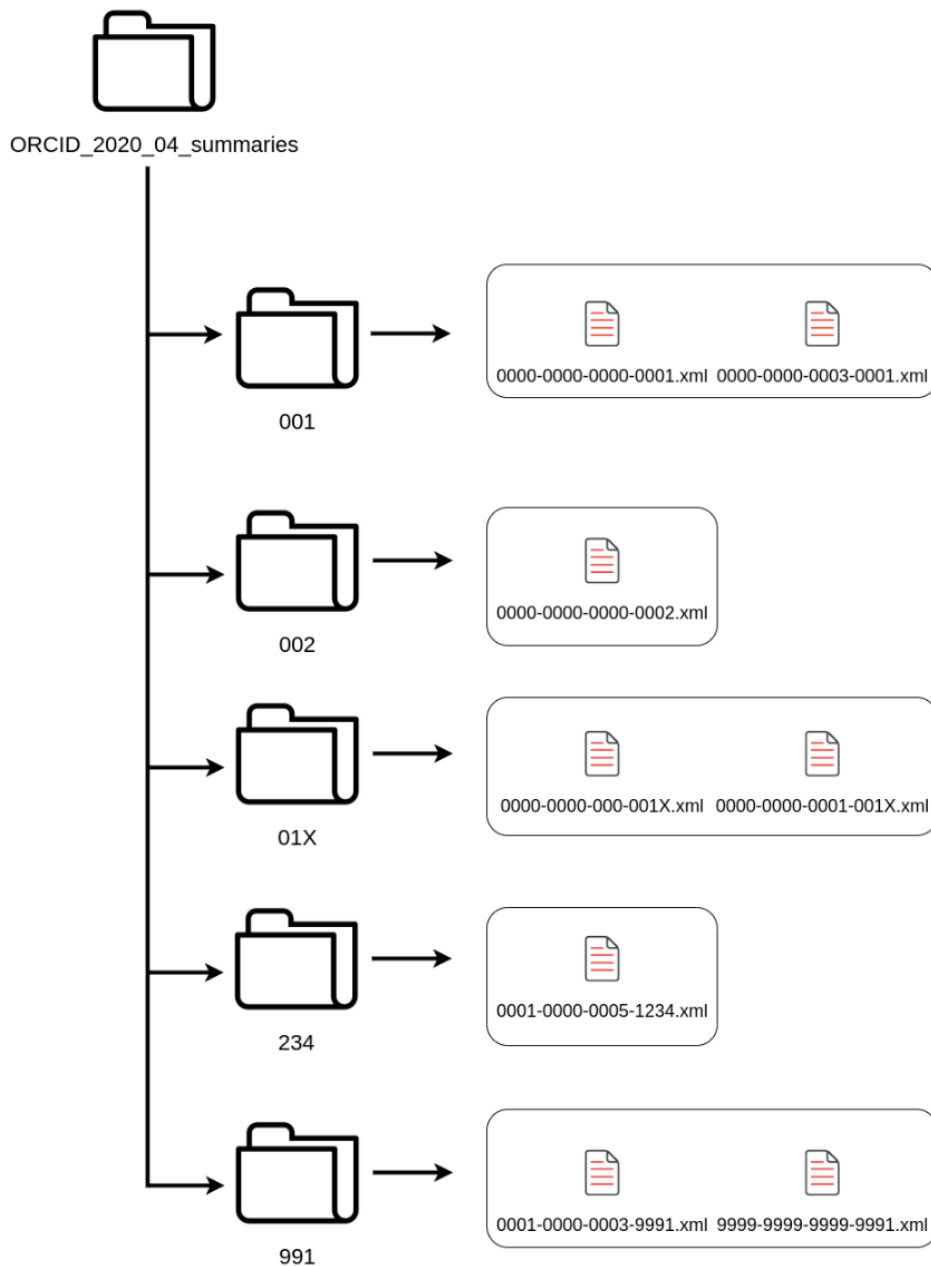Below is an outline of the Activities file hierarchy.

Fig. 1. Activity File Hierarchy
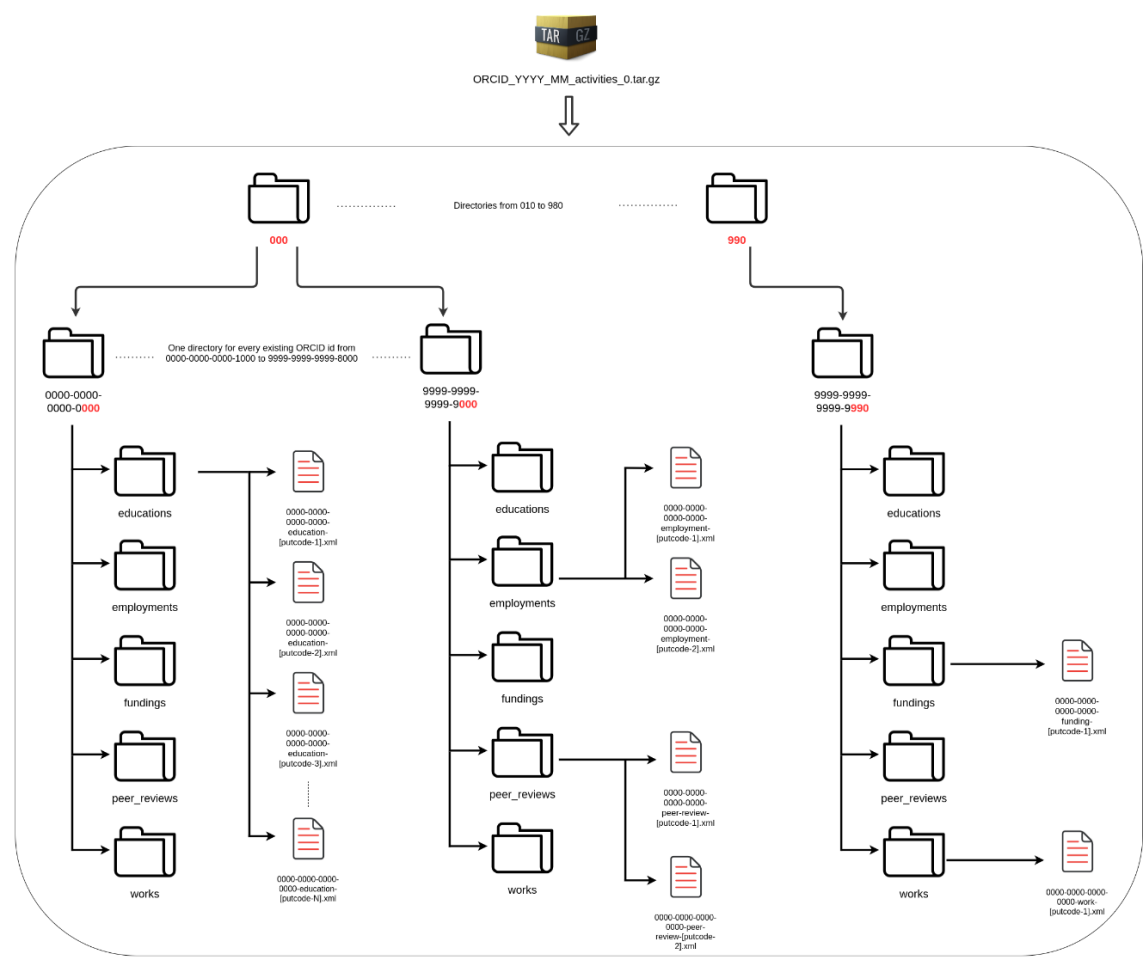
Source: ORCID

Likewise, below is a diagram of the hierarchy of the Summaries file.

Fig. 2. Summary Archive Hierarchy

ORCID_2020_04_summaries

001 → 0000-0000-0000-0001.xml 0000-0000-0003-0001.xml

002 → 0000-0000-0000-0002.xml

01X → 0000-0000-000-001X.xml 0000-0000-0001-001X.xml

234 → 0001-0000-0005-1234.xml

991 → 0001-0000-0003-9991.xml 9999-9999-9999-9991.xml

Source: ORCID

Queries and data analysis

PowerBI is used to generate subsample query reports in the virtual machine, which is a unified and scalable data analysis tool from Microsoft that is aimed at interactive data visualization. It has a simple and intuitive interface for any type of end user to create their own reports and dashboards. These types of digital tools are used in the so-called business intelligence (BI), which tries to transform basic information (data) into knowledge to optimize the decision-making process in a company.

The following image shows one of the generated previews.

Fig. 3. Map and table of previous results
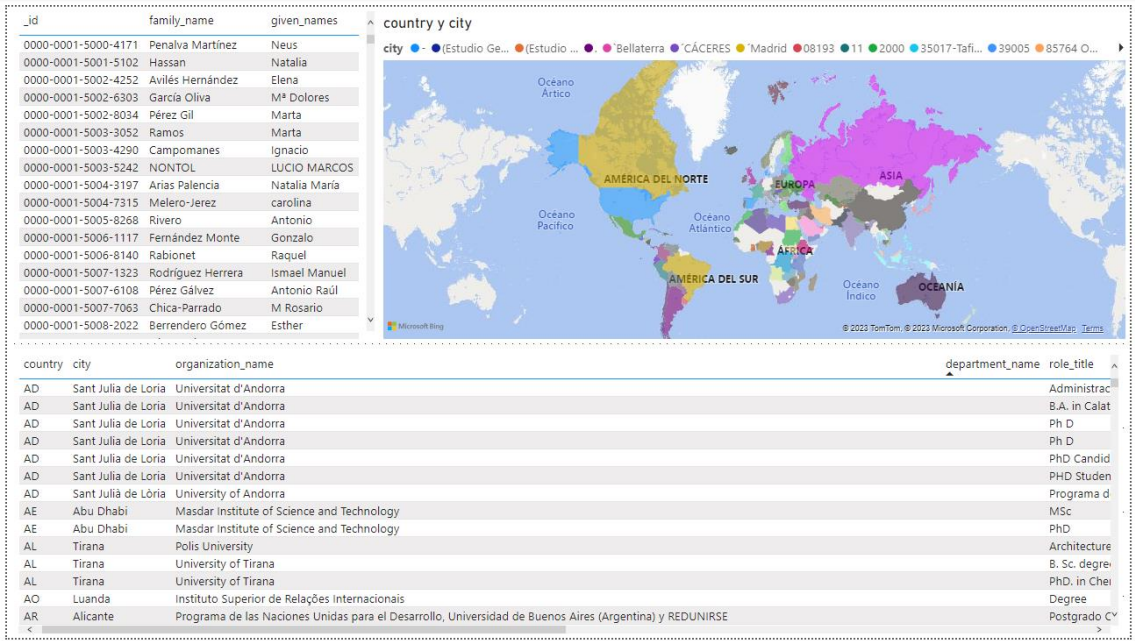
Source: own elaboration based on ORCID data