4th International Conference on the Science of Science and Innovation, June 16-18, 2025 Copenhagen Business School, Copenhagen, Denmark

Enhancing the Accessibility of ORCID Public Data, now additionally hosted on Google BigQuery

Simon Porter¹, Hélène Draux², Michele Pasin¹, Jared Watts¹, Julie Petro², and Tom Demeranville²

¹Digital Science ²ORCiD

March 2025

Keywords: ORCiD, Scientometric Infrastructure, BigQuery

Extended Abstract

Background

ORCID is committed to openness, exemplified by the annual release of its Public Data File since 2012. This dataset, encompassing all public ORCID records, has been downloaded over 190,000 times and serves as a resource for analyzing research community dynamics, scientific migrations, collaboration networks, and ORCID adoption trends. However, the file's substantial size poses challenges for users lacking advanced data management skills, hindering exploratory analyses [1].

Objective

To improve accessibility and facilitate data exploration, ORCID partnered with Digital Science to host the 2024 Public Data File on Google BigQuery, a cloud-based data analytics platform optimized for large datasets [1].

Methods

By leveraging Google BigQuery, users can now perform exploratory analyses directly on the cloud without the need to download and locally process the entire dataset. This approach reduces technical barriers and enables more efficient data interaction [1].

Results

The beta version of this service is now available, allowing the research community to develop innovative use cases for ORCID data, such as reporting on peer review practices or linking ORCID data with other open external bigQuery datasets like those from the World Bank, or commercial datasets such as Dimensions [2][3]. While the dataset remains freely accessible, users must establish their own Google BigQuery accounts. Google offers a free usage tier, with fees applicable beyond certain usage levels. Additionally, Digital Science provides sample queries to assist users in efficiently querying different parts of the ORCID dataset [4].

Conclusion

This collaboration between ORCID and Digital Science enhances the accessibility and usability of the ORCID Public Data File, empowering researchers and stakeholders to conduct comprehensive analyses without extensive data management overhead. This initiative aligns with ORCID's commitment to openness and supports the research community in deriving valuable insights from the data [1].

References

- [1] ORCID, Digital "ORCID Partners with Science Make Opento Available: https://info.orcid.org/ ness More Accessible." Even orcid-partners-with-digital-science-to-make-openness-even-more-accessible/
- [2] Draux H and Porter S, "Blooming research profiles: cultivating identity with ORCiD and Dimensions" Available: https://researchmusings.substack.com/p/ blooming-research-profiles-cultivating
- [3] Digital Science, "Dimensions on Google BigQuery" Available: https://docs. dimensions.ai/bigquery/
- [4] Porter S, Exploring the ORCiD dataset on Google Bigquery" Available: https:// bigquery-lab.dimensions.ai/tutorials/09-orcid/