ORCID Connecting Research and Researchers

DIGITAL science

Enhancing the Accessibility of ORCID Public Data via Google BigQuery

Simon Porter 0000-0002-6151-8423, **Hélène Draux** 0000-0001-8837-168X, **Michele Pasin** 0000-0001-8909-7766,
Jared Watts 0000-0002-3315-1572, **Julie Petro** 0000-0003-4967-747X, **Tom Demeranville** 0000-0003-0902-4386

4th International Conference on the Science of Science and Innovation, June 2025

Copenhagen Business School

Introduction

ORCID's Public Data File has been released annually since 2012, providing open access to millions of public researcher records.

While the data is rich in value for scientometric analysis, the dataset's size and format have posed accessibility challenges for non-technical users.

The Partnership

To lower these barriers, ORCID has partnered with Digital Science to host the 2024 Public Data File on Google Big-Query. This collaboration aims to make ORCID's open data more usable, supporting researchers, policy makers, and infrastructure developers.

Impact and Future

- Reduces entry barriers for open data use
- Encourages cross-platform research
- Enables FAIR-aligned scientometric analysis
- More updates and tutorials coming in 2025

What's New?

Before	With BigQuery
Large downloads	Cloud-hosted SQL access
Local storage needed	Serverless query interface
Complex data setup	Ready-to-use dataset
Static analysis	Join with live data sets

Getting Started

1. Sign up at cloud.google.com
2. Access BigQuery and locate 'orcid.public'
3. Use SQL or provided sample queries
4. Visualize results or join with other datasets

Dataset Access

Platform: Google BigQuery

Dataset: 'ds-open-datasets.orcid'

Access: Free with Google Cloud account (1TB/month free tier)

Tutorial:
https://bigquery-lab.dimensions.ai/tutorials/09-orcid/

Links and Resources

Project Blog
https://info.orcid.org/orcid-partners-with-digital-science-to-make-openness-even-more-accessible/

BigQuery Dataset
https://bigquery-lab.dimensions.ai/tutorials/09-orcid/

Dimensions Docs
https://docs.dimensions.ai/bigquery/open-datasets.html

Contact

Digital Science
Simon Porter
Hélène Draux

ORCID
Julie Petro
Tom Demeranville

Real-World Examples from the ORCID BigQuery Tutorial

Active ORCID iDs

QUERY

Most frequent keywords across ORCID records

SQL

```
SELECT
COUNT(orcid_identifier)
FROM
ds-open-datasets.orcid.summaries_2024
WHERE history.deactivation_date IS NULL;
```

ORCID Registrations by Year

QUERY

Number of ORCID iDs created each year

SQL

```
SELECT EXTRACT(YEAR FROM
TIMESTAMP(history.submission_date)) AS year,
COUNT(orcid_identifier)
FROM
ds-open-datasets.orcid.summaries_2024
WHERE history.deactivation_date IS NULL
GROUP BY year
ORDER BY year;
```

Top Domains in Researcher URLs

QUERY

Most common domains in researcher URLs

SQL

```
SELECT SUBSTR(url.url, 1, 25) AS
url_beginning,
COUNT(orcid_identifier) AS orcid_count
FROM
ds-open-datasets.orcid.summaries_2024,
UNNEST(person.researcher_urls.urls) AS url
GROUP BY url_beginning
ORDER BY orcid_count DESC
```

Top Keywords in ORCID Records

QUERY

Most frequent keywords across ORCID records

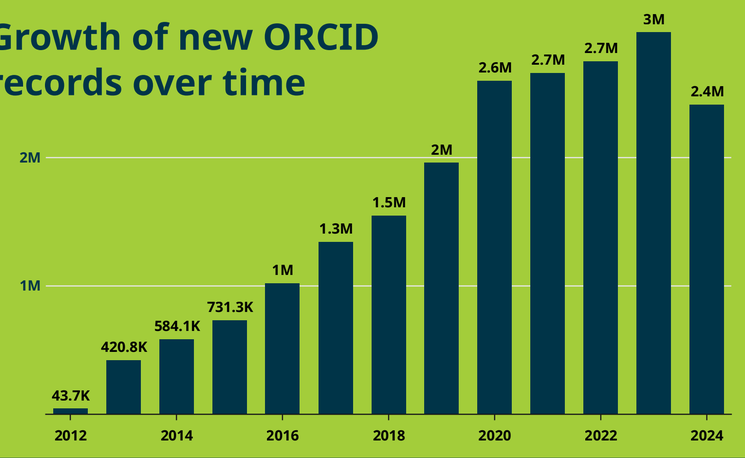
SQL

```
SELECT LOWER(keyword.content) AS keyword,
COUNT(orcid_identifier.path) AS count
FROM
ds-open-datasets.orcid.summaries_2024,
UNNEST(person.keywords.keywords) AS
keyword
GROUP BY keyword
ORDER BY count DESC
```

There are 21,046,010 ORCID Records

Most Linked Profile: LinkedIn (238,827)

Growth of new ORCID records over time



Top 20 ORCID Profile Keywords